The EMBO Journal Vol.17 No.18 pp.5449–5457, 1998

Functional analysis of peptide motif for RNA microhelix binding suggests new family of RNA-binding domains

Lluís Ribas de Pouplana, Douglas Buechter[1], Niranjan Y.Sardesai and Paul Schimmel[2]

The Skaggs Institute for Chemical Biology and Departments of Molecular Biology and Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

[1]Present address: The US Surgical Corporation, 195 McDermott Road, North Haven, CT 06473, USA

[2]Corresponding author

RNA microhelices that recreate the acceptor stems of transfer RNAs are charged with specific amino acids. Here we identify a two-helix pair in alanyl-tRNA synthetase that is required for RNA microhelix binding. A single point mutation at an absolutely conserved residue in this motif selectively disrupts RNA binding without perturbation of the catalytic site. These results, and findings of similar motifs in the proximity of the active sites of other tRNA synthetases, suggest that two-helix pairs are widespread and provide a structural framework important for contacts with bound RNA substrates.

Keywords: peptide motifs/photo-crosslinking/RNA–protein interactions

Introduction

An RNA-binding motif that has not been identified previously (Burd and Dreyfuss, 1994; Arnez and Cavarelli, 1997) provides a conceptual framework for understanding how the acceptor stems of certain tRNAs are recognized by tRNA synthetases. In particular, this analysis relates to the long-standing question of how the acceptor stems of alanine tRNA are identified (Hou and Schimmel, 1988; McClain and Foss, 1988).

Class II tRNA synthetases are defined by a common active site that is based on a seven-stranded anti-parallel structure with three α-helices (Cusack *et al*., 1990; Ruff *et al*., 1991; Arnez *et al*., 1995; Logan *et al*., 1995; Mosyak *et al*., 1995; Åberg *et al*., 1997). Three sequence motifs, motifs 1, 2 and 3, are common to these enzymes and form a helix–loop–strand, strand–loop–strand, and strand–helix, respectively (Eriani *et al*., 1990; Moras, 1992). The solved crystal structures of members of this family reveal that acceptor-stem interactions are achieved through contacts with insertion regions that branch out of the conserved active site (Ruff *et al*., 1991; Biou *et al*., 1994). In particular, the variable loop of motif 2 is prominent in the acceptor-stem contacts made by aspartyl- and seryl-tRNA synthetases (Ruff *et al*., 1991; Biou *et al*., 1994).

Alanyl-tRNA synthetase is a member of the class II tRNA synthetases whose modular organization has been recognized for some time (Jasin *et al*., 1983, Buechter and Schimmel, 1993; Sardesai and Schimmel, 1998). Its different functions can be assigned to the different domains that are organized in a linear way along its sequence. While the structure is unsolved, functional analysis, sequence alignments and homology modeling have located the class-defining active site to the first 250 amino acids (Eriani *et al*., 1990; Ribas de Pouplana *et al*., 1993). This structural unit contains the three conserved motifs that define class II enzymes, and specific residues within it have been shown to be important for catalytic activity (Davis *et al*., 1994; Lu and Hill, 1994; Shi *et al*., 1994; Ribas de Pouplana and Schimmel, 1997) (Figure 1).

While the wild-type *Escherichia coli* enzyme is a tetramer comprised of identical chains of 875 amino acids, a monomeric N-terminal fragment of 461 residues (N461) is also active (Jasin *et al*., 1983; Ho *et al*., 1985). In particular, this fragment catalyzes aminoacylation of microhelix substrates that are based on the acceptor stem of $tRNA^{Ala}$ (Francklyn and Schimmel, 1989; Buechter and Schimmel, 1993) (Figure 1). Like the full tRNA, this aminoacylation is dependent on a single G:U bp that is located at the third position from the end of the acceptor stem (Buechter and Schimmel, 1993). Throughout evolution from bacteria to humans, the G3:U70 bp marks a tRNA for charging with alanine (Hou and Schimmel, 1989; Ripmaster *et al*., 1995; Shiba *et al*., 1995). The challenge is to find the determinants within the protein that are required for this recognition.

Extensive mutagenesis within the N250 domain identified residues involved in functions expected for this region of the protein, but failed to identify residues affecting solely the tRNA-dependent step of aminoacylation (Davis *et al*., 1994; Lu and Hill, 1994; Shi *et al*., 1994; Ribas de Pouplana and Schimmel, 1997). In particular, mutations in the loop of motif 2 do not suggest that this loop plays an essential role in acceptor–helix recognition (Davis *et al*., 1994; Lu and Hill, 1994). Thus, determinants for recognition of the acceptor stem of $tRNA^{Ala}$ lie elsewhere in the structure of N461 (Buechter and Schimmel, 1995; Sardesai and Schimmel, 1998). To identify these determinants, extensive alanine-scanning mutagenesis was done at highly conserved residues that have functional group side chains. These experiments were followed by biochemical analyses and further mutagenesis, combined with computational analysis and homology modeling. The complete body of results was used to build a structural model of a highly conserved two-helix pair. This motif has a positively charged surface where the two critical residues are located. Similar folds have been found in crystal structures of two other class II

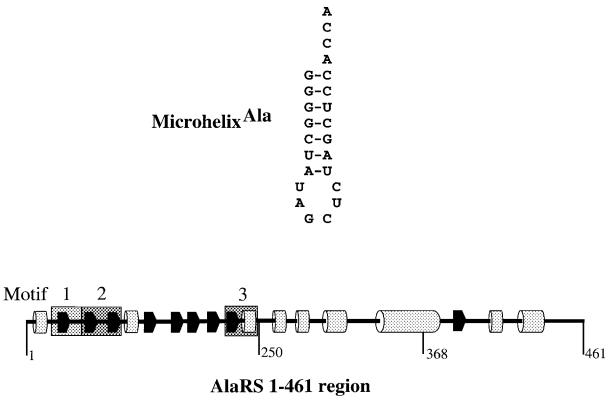

Fig. 1. Microhelix structure and domain distribution in AlaRS. Distribution of secondary structure elements in the N461 fragment of AlaRS was based on the secondary structure prediction of PHD (Rost and Sander, 1994). Helices are represented as cylinders, β-strands as arrows and other elements as lines.

synthetases (Delarue *et al.*, 1994; Åberg *et al.*, 1997), where two-helix pairs are ideally placed to interact with the acceptor stem of a bound tRNA.

Results

Initial mutagenesis experiments and purification of non-complementing mutants

The tRNA synthetases typically carry out aminoacylation in a two-step reaction:

(Eqn 1) $E + AA + ATP \leftrightarrow E(AA\text{-}AMP) + PPi$

(Eqn 2) $E(AA\text{-}AMP) + tRNA \leftrightarrow AA\text{-}tRNA + AMP + E$

In the first reaction (adenylate synthesis), the amino acid AA is activated to an enzyme-bound aminoacyl adenylate (AA-AMP), while in the second reaction, the aminoacyl moiety is transferred to the 3′-end of the tRNA to give the aminoacyl-tRNA (AA-tRNA) (Schimmel, 1990). We used assays of adenylate synthesis (Eqn 1) and of aminoacylation of microhelices (Eqn 1 and Eqn 2) to test the functional effects of our mutations. Mutations affecting the RNA binding capacity should result only in a decrease of the rate of the aminoacyl transfer step (Eqn 2), but not of the rate of adenylate synthesis (Eqn 1).

The purpose of our initial experiments was to identify residues between V250 and L461 that are important for acceptor-stem recognition. Because this region is not part of the active site architecture (Ribas de Pouplana *et al.*, 1993; Buechter and Schimmel, 1995), we reasoned that the substitutions of critical residues for acceptor-stem recognition might result in mutant enzymes with reduced capacity to charge a microhelix, but with full activity for adenylate synthesis. Thirty-seven residues were selected for mutagenesis based on their level of conservation among sequences of alanyl-tRNA synthetases and the potential of their side chains to be involved in hydrogen-bond interactions (Figure 2). Each of these residues was individually substituted with an alanine.

The single-substitution mutants were constructed as described below, and their capacity to support growth of an *alaS* null strain was tested. This strain has a deletion of the gene *alaS* from the chromosome and cells are maintained by an *alaS*-encoding plasmid that has a temperature-sensitive replicon. Because the plasmid-borne replicon is defective at 42°C, no growth occurs at this temperature unless a second plasmid encoding an active alanyl-tRNA synthetase is introduced. Thus, this system provides a way to test whether mutant enzymes have an activity that is sufficient to sustain cell growth and rescue the temperature-sensitive phenotype.

Point mutations at only two of the 37 positions, D285 and R314, resulted in the non-complementation phenotype. We verified that these mutant proteins were stable and accumulated in the cell, so that the reason for their non-complementation phenotype was not that they were unstable and subject to degradation within the cell (data not shown). The two mutants were then purified by Ni-NTA-affinity chromatography. The purification of both mutants was relatively inefficient, particularly for D285A, for which only small amounts of purified protein were obtainable. The poor purification yields were mainly due to the detrimental effect that the expression of these two mutant enzymes had on the growth of the *E.coli* strain W3110 used for their purification.

Combinatorial mutagenesis of residues D285 and R314

The small number of positions (i.e. two) that were sensitive to mutation, and the results of the modeling experiments described below, suggest that D285 and R314 may be functionally and/or structurally linked. In order to test whether these two positions are involved in a reciprocal interaction (such as a salt bridge), we constructed a combinatorial library of substitutions for the two residues. In this library several combinations of sequences at positions D285 and R314 were constructed in order to test whether these two residues could be interchanged, and to test their sensitivity to conservative substitutions that should preserve a salt bridge-type interaction. The results (Table I) demonstrated that the two positions are virtually immutable. Even conservative changes such as D285E affect *in vivo* complementation by the mutant enzymes. If the two residues are reversed, as in the D285R/R314D double mutant, the resulting enzyme entirely loses its complementation capacity (we again verified that this mutant accumulates *in vivo*; data not shown). These results indicate that, even if D285 and R314 are in close spatial proximity, the loss of activity associated with mutation of either D285 or R314 is not simply due to the loss of an ionic interaction between them.

Kinetic analysis

The amino acid activation and aminoacylation activities of the purified enzymes were analyzed as described below. Each of the two mutant enzymes is dramatically impaired for aminoacylation (Figure 3). By increasing the concentration of the mutant enzymes in the aminoacylation reactions we estimated that their activities with microhelix substrates are each ~700-fold less than that of the wild-type enzyme. Very similar effects were found when the aminoacylation activity of the mutant enzymes was tested with purified $tRNA^{Ala}$ (data not shown). However, both mutant proteins retain wild-type levels of activity for adenylate synthesis (Figure 3). This kinetic behavior is consistent with a role in tRNA recognition for D285 and R314. The results also

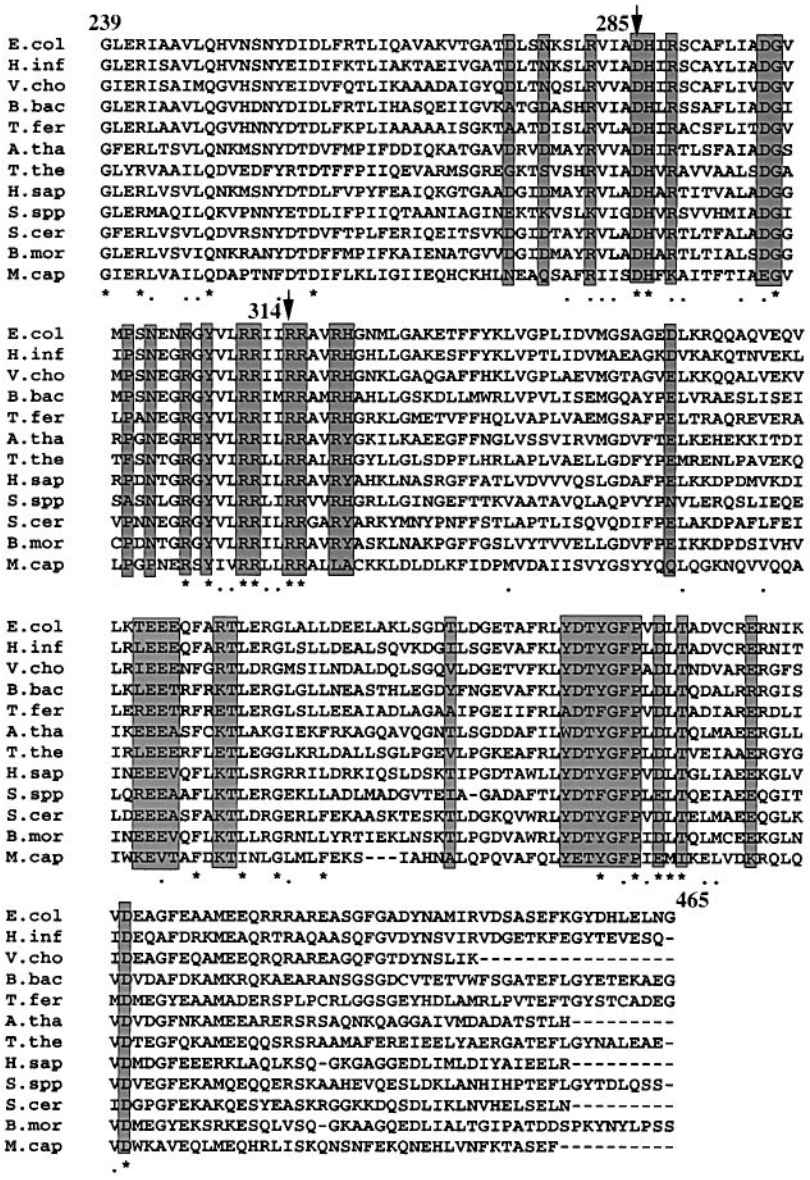

Fig. 2. Alignment of sequences of alanyl-tRNA synthetases in the region from G239 to G465 (numbered according to the *E.coli* enzyme that is shown as the top sequence). The boxed positions represent the residues that were mutated to alanine in this work. The two positions found to be sensitive to mutation (D285 and R314) are numbered and marked by arrows. (Sequence code: E.col, *Escherichia coli*; H.inf, *Haemophilus influenzae*; V.col, *Vibrio cholera*; B.bac, *Bacillus bacteroides*; T.fer, *Thiobacillus ferrooxidans*; A.tha, *Arabidopsis thaliana*; T.the, *Thermus thermophilus*; H.sap, *Homo sapiens*; S.spp, *Synechocystis spp*; S.cer, *Saccharomyces cerevisiae*; B.mor, *Bombyx mori*; M.cap, *Mycobacterium capricolum*).

Table I. *In vivo* results of combinatorial mutagenesis at positions 285 and 314 of *E.coli* AlaRS

Residues at positions: 285	314	Complementation of alaS null strain[a]
Asp (wt)	Arg (wt)	+
Ala	Arg (wt)	–
Asp (wt)	Ala	–
Arg	Arg (wt)	–
Asp (wt)	Asp	–
Arg	Asp	–
Glu	Arg (wt)	±
Asp (wt)	Lys	±

[a]Complementation assay was as described in Materials and methods. ±, mutations that allowed for weak growth of colonies when assayed.

suggest that their location in the structure is removed from the active-site pocket.

tRNA-binding assays

In order to assess the potential effect of the mutation R314A on binding to $tRNA^{Ala}$, we used a nitrocellulose filter-binding assay (Yarus and Berg, 1967). Because of the difficulty of obtaining adequate amounts of the D285A protein, only the R314A enzyme was studied. For this assay we used the full-length tRNA rather than the microhelix, to take advantage of some of the tRNA–protein contacts that occur outside of the acceptor stem. Because they enhance binding affinity, we reasoned that these interactions would add more sensitivity to the filter-binding assay and, in addition, would test whether a single point mutation in the protein would prevent binding of the full tRNA.

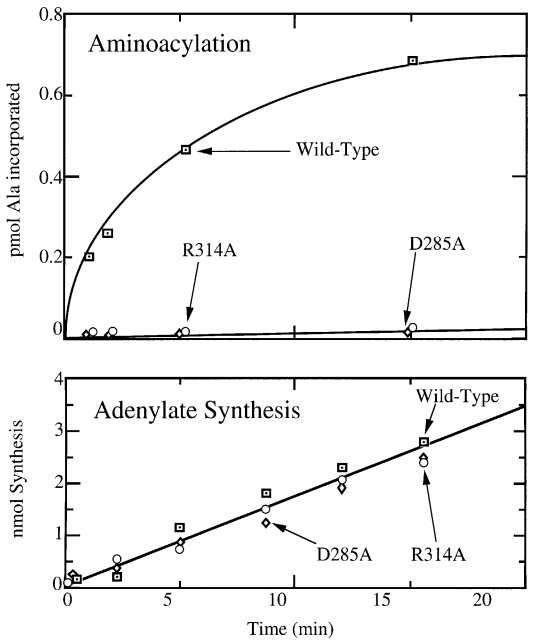

Fig. 3. Aminoacyl adenylate synthesis and charging activities of wild-type and D285A and R314A mutant enzymes. The two mutant enzymes have wild-type levels of adenylate formation activity, but are ~700-fold reduced in microhelix charging activity with respect to the wild-type enzyme.

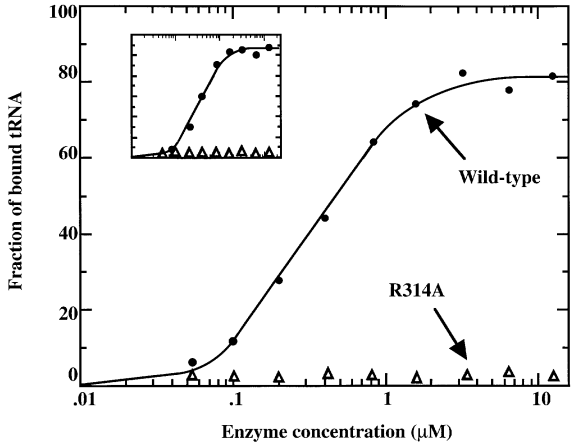

Fig. 4. Binding assays for wild-type and R314A AlaRS with $tRNA^{Ala}$ at pH 7.5. The R314A mutation decreases $tRNA^{Ala}$ binding by AlaRS down to that of the background (non-specific binding of $tRNA^{Lys}$ by wild-type AlaRS) which has been subtracted from the curves. This effect is seen at both pH 7.5 and 6 (see inset).

We found that the R314A mutant enzyme had a severe defect in binding to $tRNA^{Ala}$, and was indistinguishable from background (the non-specific binding of $tRNA^{Lys}$) (Figure 4). This result demonstrated a role for R314 in binding of $tRNA^{Ala}$. Because the enzyme has full activity levels for adenylate synthesis we concluded that the R314A substitution did not result in a structural perturbation, but rather in an ablation of an enzyme–tRNA contact.

Photo-crosslinking to an acceptor-stem duplex

We further probed the effect of the R314A mutation on the protein–acceptor helix interaction by testing the capacity of wild-type and mutant AlaRS to crosslink to azidophenacyl-modified RNA duplex substrates based on

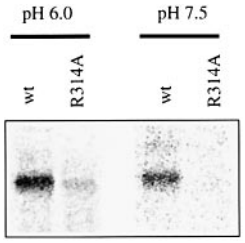

Fig. 5. Crosslinking of wild-type and AlaRS R314A with modified RNA $duplex^{Ala}$. The same amount of total RNA (free and bound) was shown to be loaded in each lane by exposing the phosphorimager screen to the wet gel prior to transferring the proteins to a PVDF membrane. Amido black staining of the PVDF membrane also allowed a visual estimate that the same amount of protein was loaded in each lane.

the acceptor stem of $tRNA^{Ala}$. Azide-substituted photoactive probes offer the advantage of rapid generation of short-lived intermediates and thereby eliminate long exposure to UV light which can be damaging to proteins. Azidophenacyl bromide alkylation of the single phosphorothioate linkage between dC69 and U70 to form 13-AP RNA proceeds with high yields (>80%) in 3 h. Upon annealing 13-AP with the complementary 9-mer RNA, the modified 9+13–AP duplex is a competent substrate for aminoacylation by AlaRS (Sardesai and Schimmel, 1998). The alanine acceptance levels are comparable with unmodified phosphodiester or phosphorothioate duplex substrates (data not shown).

Irradiation of 9+13–AP in the presence of wild-type AlaRS generated an RNA–protein crosslinked species (Figure 5), as ascertained by the incorporation of ^{32}P label into the protein band on an SDS–polyacrylamide gel. This protein–RNA crosslinked band migrated slower than the free protein (detected by transferring the protein products to a PVDF membrane and staining with amido black) and correlated well with the expected molecular weight of protein (97 kDa)+13–AP–RNA (4.2 kDa) (data not shown). In control experiments, we determined that no radioactive band was obtained when the RNA was mixed with AlaRS in the absence of irradiation or when irradiation was carried out in the presence of a random protein (maltose-binding protein).

In contrast, no crosslinked product is observed with the AlaRS-R314A mutant enzyme. Even at a lower pH (pH 6.0), where the overall binding affinity for the protein–RNA interaction is increased and discrimination between cognate and non-cognate systems is decreased (Schimmel and Soll, 1979; Park *et al*., 1989), only a weak crosslink is observed with the mutant protein (Figure 5). These data confirm that the effect of the R314A mutation is on acceptor–helix recognition.

Computational analysis of the L280–G320 region of AlaRS

Once we had identified two residues involved in tRNA binding and recognition we attempted to use molecular modeling techniques to gain further insight into the structure of this region of AlaRS. The secondary structure prediction for the L275–A325 region of AlaRS strongly indicated the presence of two helices (of ~20 and 15 residues, respectively) separated by a loop region of 15 residues. The first helix contains D285, while the second

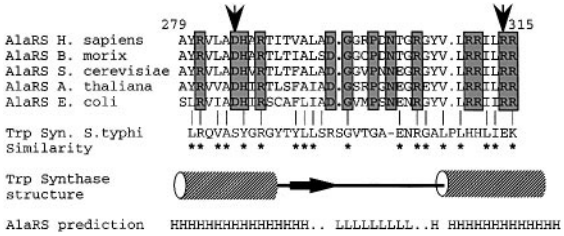

Fig. 6. Computational analysis of the S279–R315 region of AlaRS. The AlaRS sequences on top show the high level of conservation in this region and highlight the positions mutated in this work (see also Figure 2). The group of AlaRS sequences was then aligned to the sequence of *Salmonella typhimurium* tryptophan synthase (the vertical lines indicate identical residues between tryptophan synthase and sequences within the AlaRS alignment). The distribution of secondary structure elements (cylinders for helices, arrows for β-strands, black line for loops) in the crystal structure of tryptophan synthase is depicted underneath its sequence (Hyde *et al*., 1988). For comparison, the secondary structure prediction for *E.coli* AlaRS (H, helical; L, loop; dots, poorly predicted positions) calculated by the program PHD (Rost and Sander, 1994) is shown at the bottom.

contains R314. Both predicted helices display high levels of amphiphilicity, with a total number of six to eight arginine residues in their hydrophilic surface.

In a sequence similarity search through the Protein Database (Brookhaven, NY) we detected a high (~70%) level of sequence identity between the L280–N320 region of AlaRS and a region of the biosynthetic enzyme tryptophan synthase (Figure 6). This region of tryptophan synthase folds into two adjacent α-helices linked by a loop–strand–loop structure that forms one of the building elements of a β-barrel (Hyde *et al*., 1988).

Further analysis of the sequence of the same L280–N320 region was done using fold recognition programs based on two different sequence threading methods and hidden Markov statistical analysis (Stultz *et al*., 1993; Alexandrov *et al*., 1996; Rost and Sander, 1994). The three methods used for fold recognition are based on independent prediction approaches, and use different structural parameters to construct their predictions. Despite these differences all the predictions strongly favored a helix hairpin arrangement (either parallel or anti-parallel) for the L280–N320 region of AlaRS, exposing a common, positively charged surface to the solvent (data not shown). No other protein fold was predicted by these three methods as a potential structure for the region.

Using homology-based techniques (Bajorath *et al*., 1993), backed by the computational results with other fold recognition methods (Stultz *et al*., 1993; Alexandrov *et al*., 1996; Rost and Sander, 1994), we modeled the sequence between L280 and N320 into a two-helix pair (Figure 7). In this arrangement, D285 and R314 fall in close proximity of each other. More importantly, the whole domain is shown to display a positively charged surface, largely reminiscent of other RNA-binding domains (Burd and Dreyfuss, 1994).

Discussion

The high level of conservation of the $tRNA^{Ala}$ identity element (the G3:U70 bp) in evolution correlates well with the two-helix pair described here being among the most conserved sequences in an alignment of alanyl-tRNA synthetases. This region, however, shows no homology to any region in other class II tRNA synthetases based on sequence comparisons. Thus, the two-helix pair may have been an early addition to the class II core structure, perhaps following an ancient duplication that produced the ancestral AlaRS.

Other two-helix pairs with structures similar to the one we propose for AlaRS are found in two other synthetases. The crystal structures of *Thermus thermophilus* aspartyl- and histidinyl-tRNA synthetases (AspRS and HisRS) show the presence of two-helix pairs with positively charged surfaces in close proximity to the hypothetical position of the bound tRNA acceptor stem (Delarue *et al*., 1994; Åberg *et al*., 1997). The two-helix pair in the AspRS structure is part of a large 140 amino acid insertion located between motifs 2 and 3 of that enzyme's active site. This insertion is not found in the *Saccharomyces cerevisiae* AspRS crystal structure (Ruff *et al*., 1991), and sequence searches suggested that it was idiosyncratic to bacterial AspRS. The two helices are long (14 and 18 residues each), pack against each other in anti-parallel fashion, and are separated in sequence by other secondary structure elements that are part of the insertion domain. Approximately thirty percent of all residues in the two helices and the intervening loop regions are arginines or lysines that face the solvent-exposed side of the helices, towards the enzyme's active site.

In the case of the class II *T.thermophilus* HisRS, the two-helix pair is part of a 64 amino acid insertion that is also located between motifs 2 and 3. The helices are shorter in this case (seven and six residues, respectively) and are directly connected by a four residue turn. A 19-residue segment that covers the two-helix pair contains four arginines or lysines. Three of these basic residues (R197, R204 and K209; *T.thermophilus* HisRS numbering) are completely conserved among bacterial organisms, together with another four residues in the same region (N201, P202, L206 and D207).

Docking analyses of $tRNA^{Asp}$ from the yeast AspRS-$tRNA^{Asp}$ co-crystal (Ruff *et al*., 1991) with the crystal structures of AspRS and HisRS from *T.thermophilus* were carried out by Delarue *et al*. (1994) and Åberg *et al*. (1997), respectively. In both models the two-helix pairs came in close proximity to the first base pair of the acceptor stem. Given that the interaction mechanism between aaRSs and their cognate tRNAs tends to be highly conserved between evolutionarily related systems (Ruff *et al*., 1991; Biou *et al*., 1994), it is reasonable to assume that these models are close to the structures of the real complexes. The striking similarity in the positions of the two-helix pairs found in AspRS and HisRS from *T.thermophilus* suggests a common function in acceptor-stem binding for these helices, despite the clear differences in the structural environments that surround them. Whether this two-helix pair is directly or indirectly important for specific contacts with tRNA by AspRS or HisRS is not known. However, the experiments presented here demonstrate its functional significance for binding of the related AlaRS to the acceptor helix of $tRNA^{Ala}$.

In the crystal structure of the class I *T.thermophilus* GluRS, another two-helix pair (with similar structural- and charge-distribution characteristics) is located in the

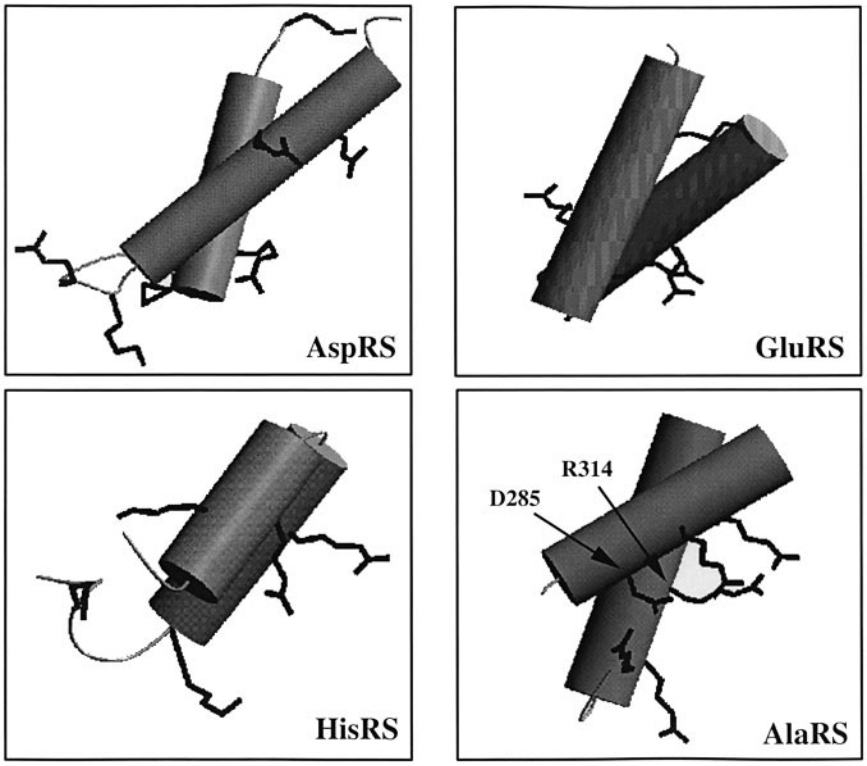

Fig. 7. Structures of proposed RNA-binding two-helix pairs. Tubular representations of the two-helix pairs of *T.thermophilus* AspRS (Delarue *et al*., 1994), HisRS (Åberg *et al*., 1997), GluRS (Nureki *et al*., 1995) and the model of the S279–G320 region of *E.coli* AlaRS. The helices in these motifs are shown as gray tubes, and the side chains of lysines and arginines are depicted as thin black lines.

anticodon-binding domain (Nureki *et al*., 1995). The two helices extend for 13 residues each (P392 to E405 and G455 to A468, in the *T.thermophilus* enzyme), and contain a total of eight arginines and lysines (~30% of the total length). The pattern of conserved residues in this region of GluRS is similar to that found in the two-helix pair of HisRS. In particular, the sequence pattern around the most conserved residues are almost identical (i.e. PIRVA in *Staphylococcus aureus* HisRS, and PLRVL in *Salmonella typhimurium* GluRS). If the two-helix hairpin in GluRS is indeed involved in tRNA binding, then this fold can be adapted to bind different regions of a tRNA molecule, because its position in the GluRS structure would not be close to the acceptor stem of $tRNA^{Glu}$.

As far as we know, this is the first example of a structural motif that has been incorporated into both classes of tRNA synthetases. Despite their clear structural similarities (Figure 7), the length, sequence and connectivity of the two-helix pair motifs from AspRS, HisRS, GluRS, and the predicted one in AlaRS, vary from one case to another. This variability suggests that the two-helix pair may either be an example of functional convergence from different initial structures, or an example of wide divergences from an ancestral motif that was incorporated into many proteins. In either case, its frequent occurrence may arise from a particular suitability of this kind of domain for providing a framework for RNA binding. Whether the actual RNA contacts are made directly by the motifs, or whether residues appended to the motifs make RNA contacts, may vary from case to case and remains to be determined.

A possibly related example of a motif involved in RNA binding is offered by the protein rop, which is involved in the regulation of the replication of plasmid ColE1 (Banner *et al*., 1987). The crystal structure of rop shows that this protein forms four helix bundles through the dimerization of two identical two-helix hairpins of 63 residues each. The distribution of charges in this structure is built around the amphiphilic nature of the helices, which also present positively charged surfaces likely to be involved in RNA binding. Despite these similarities, placing rop in the same structural family with the two-helix pairs described here does not seem justified. No four-helix bundles have been found in the tRNA binding regions of synthetases, and the two-helix pairs in synthetases show large differences in helix packing and connectivity with respect to rop.

In an earlier study we identified the region between Arg368 and Asp461 as important for recognition of the G3:U70 bp that marks a tRNA for aminoacylation with alanine (Buechter and Schimmel, 1995). With this observation in mind we tested the low activity (~700-fold reduced, see above) of the R314A mutant enzyme for its sensitivity to substitutions at the 3:70 position. Within the limitations imposed by the low activity, we found that the R314A protein was still sensitive to the nature of the bp at the 3:70 position. Thus, the two-helix pair studied here may provide a general platform for acceptor-stem binding, around which the addition of specificity-determining elements are assembled. For

example, the recognition of the 3:70 bp may be achieved through interactions that require the presence of both the 250–368 and 368–461 domains in order to recognize specifically the G3:U70 bp of the $tRNA^{Ala}$ acceptor stem.

Materials and methods

Sequence and structure analysis

All class II tRNA synthetase sequences were obtained from the Swissprot database (Benson *et al*., 1994). Initial database searches were performed with BLAST (Altschul *et al*., 1990). Sequence alignments were done using CLUSTALW (Thompson *et al*., 1994). Secondary structure predictions were done with PHD (Rost and Sander, 1994). Fold recognition experiments by threading analysis were done using methods based on secondary structure predictions and neural networks [with TOPITS (Rost and Sander, 1994)] and on contact capacity potentials (with 123D; Alexandrov *et al*., 1996). Another fold recognition method based on statistical analysis with Hidden Markov Models was also used (PSA; Stultz *et al*., 1993). The results from these three independent methods were very similar. The three-dimensional coordinates used were from the Protein Database at Brookhaven (Bernstein *et al*., 1977). Structure-based alignments and three-dimensional models were built manually using QUANTA (Molecular Simulations, Waltham, MA), and were based on the results of the sequence alignments, secondary structure predictions, sequence threading analysis, information from class II synthetase crystal structures and the biochemical results presented here.

Mutagenesis and in vivo complementation assays

The gene *alaS* coding for *E.coli* alanyl-tRNA synthetase (harbored by plasmid $pBSKS^+$_*alaS* (Davis *et al*., 1994) was modified by site-directed mutagenesis using the uracil-incorporation method (Kunkel, 1985). The region around the mutation site was sequenced using dideoxy sequencing methods (Barrell *et al*., 1980) and subcloned back into the original plasmid to ensure that no other mutations were incorporated into the gene.

The complementation assays were carried out as described (Jasin *et al*., 1983) using the *alaS* null strain W3110 (*lacI*^{q} *rec*AΔ1 Kan^Γ *alaS*Δ2) maintained by plasmid pMJ901 (Tet^Γ marker) that expresses full-length AlaRS and that has a temperature-sensitive replicon (Jasin *et al*., 1985). At the elevated temperature of 42°C, plasmid pMJ901 is lost and the cells cease to grow. When an additional plasmid is introduced that encodes a mutant AlaRS, growth is rescued at 42°C only if the mutation does not significantly disrupt enzyme activity.

Protein expression and purification

Protein expression and purification was carried out using the *E.coli alaS* null strain W3110 (maintained by plasmid pT461) (Ribas de Pouplana and Schimmel, 1997) transformed with plasmid pQE-*alaS*-6H, as described.

The genomic copy of *alaS* of this strain is disrupted with a Kan^Γ marker, and the cells contain a copy of the *lac*I^q gene that ensures tight control of the lac promoter of pQE-70-based plasmids in the absence of IPTG (Jasin and Schimmel, 1984). Growth of the W3110 *alaS* null strain is maintained by plasmid pT461 (Tet^Γmarker) which encodes the active N-terminal 461 amino acids fragment of AlaRS (N461) (Regan, 1986).

The plasmid pQE-*alaS*-6H (based on pQE-70; Qiagen, Chatsworth, CA) is compatible with plasmid pT461, contains an Amp^Γ marker, and encodes full length AlaRS (wild-type or mutant) fused to a coding sequence for a C-terminal 6-histidine extension. The full length (tetrameric) protein produced by pQE-*alaS*-6H does not interact with the N-terminal 461-mer fragment (monomeric) of AlaRS encoded by pT461 (Jasin *et al*., 1983; Ho *et al*., 1985). Because only the full-length enzymes contain a 6-histidine tail, the two molecules can be separated by affinity chromatography on a Ni-NTA column.

The expression product of pQE-*alaS*-6-His was designated as AlaRS-6H. Mutant AlaRS-6H were purified from the W3110 *alaS* null strain (Ribas de Pouplana and Schimmel, 1997) that had been transformed with mutant versions of plasmids pQE-*alaS*-6H. The purification protocol was essentially the same for both wild-type and mutated proteins, and was carried out as described previously (Ribas de Pouplana and Schimmel, 1997).

Kinetic analysis

Alanyl adenylate synthesis was measured at 25°C in a thermostatted waterbath in 100 mM Tris–HCl (pH 8.0), 2 mM ATP, 2 mM pyrophosphate, 10 mM KF, 2 mM alanine, 10 mM β-mercaptoethanol, and 5 mM $MgCl_2$ as previously described (Calendar and Berg, 1966). Aminoacylation activity was measured at 25°C in a thermostatted waterbath in 50 mM HEPES pH 7.5, 20 μM alanine, 4 mM ATP, 20 mM KCl, 10 mM $MgCl_2$, 20 mM β-mercaptoethanol, and 0.1 mg/ml bovine serum albumin as reported (Hill and Schimmel, 1989). Unmodified microhelix RNA substrates were synthesized chemically using procedures previously published (Francklyn *et al*., 1992). All phosphoramidites used were from Chemgenes (Waltham, USA). To ensure proper folding, RNA microhelix substrates were heated to 65°C in water, and cooled in the absence of magnesium prior to adding to the aminoacylation reaction at concentrations ranging from ~10–130 μM. Enzyme concentrations were typically 10 nM as determined by UV absorbance at 280 nm and Bradford assays (Hill and Schimmel, 1989). RNA concentrations of the single strands were determined by summation of extinction coefficients for monoribonucleotides (Puglisi and Tinoco, 1989) such that ε_{260} (13-mer) = 135 850 $M^{-1}cm^{-1}$ and ε_{260} (9-mer) = 110 700 $M^{-1}cm^{-1}$. Wild-type enzyme was used at a concentration of 5 nM and mutant enzymes were assayed at concentrations (10–50 nM) that gave aminoacylation rates >10-fold over background values.

tRNA binding assays

Escherichia coli $tRNA^{Ala}$ (Subriden, Rolling Bay, WA) and $tRNA^{Lys}$ (Sigma Chemical Co., St Louis, MO) were 5′-labeled and purified as previously described (Regan *et al*., 1987). The *in vitro* binding of both tRNAs to purified wild-type and mutant AlaRS was measured as previously reported (Regan *et al*., 1987). Increasing concentrations of purified mutant and wild-type AlaRS (typically 0.1–100 μM) were incubated with a fixed concentration of tRNA. Typically 60 nM tRNA (containing a constant amount of ^{32}P-labeled tRNA) was incubated for 20 min at 37°C with increasing concentrations of enzyme. The incubations were carried at either pH 6 or 7.5, in a buffer containing 50 mM $NaPO_3$, 10 mM Mg_2Cl and 1 mM dithiothreitol. After 20 min of incubation, 10 μl of the solution were applied to nitrocellulose filters, previously equilibrated with incubation buffer. The filters were washed with 1 ml of incubation buffer in a vacuum manifold. The radioactivity bound to the filter was counted in a liquid scintillation counter and converted to pmols of bound tRNA using the specific activities of the labeled substrates.

Synthesis of azidophenacyl-modified RNA substrate

RNA oligonucleotides (9-mer and 13-mer) were synthesized on a Pharmacia synthesizer (model Gene Assembler Special) using standard phosphoramidite chemistry. A unique phosphorothioate linkage was introduced between dC69 and U70 of the 13-mer oligonucleotide (see below) by using Beaucage Reagent (3H-1, 2-benzodithiole-3-one 1, 1-dioxide) (Glen Research, Sterling, VA) instead of a 0.02 M iodine solution in pyridine/water/THF at the oxidation step during synthesis (Iyer *et al*., 1990). The single deoxyribonucleotide was introduced on the 3′-side of the phosphorothioate to prevent probe elimination (Musier-Forsyth and Schimmel, 1994). Azidophenacyl bromide (Aldrich, Milwaukee, WI) was reacted with the phosphorothioate-containing RNA oligomers following a protocol similar to that reported previously (Conway and McLaughlin, 1991; Yang and Nash, 1994). The reaction mixture was extracted three times with 2-butanol to remove excess reagent (organic layer) and the azidophenyl-modified 13-mer (13-AP RNA) was ethanol precipitated from the aqueous layer. The 13-AP RNA pellet was resuspended in 0.1 M triethylammonium acetate (TEAA) pH 7.0 and further purified to homogeneity by reverse phase HPLC using a C-18 column and a 0.1 M TEAA/acetonitrile gradient (0–40% in 40 min). Alkylation yield was estimated to be 80% by comparing peaks corresponding to the recovered unmodified RNA and the 13-AP RNA in the HPLC elution profile. The two diastereomers of 13-AP RNA were collected together and used without further diastereomeric resolution.

The 13-AP RNA (200 pmol) was 5′-end labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase, and purified by PAGE in a 16% gel with 8 M urea (Sambrook *et al*., 1989). The 5′-^{32}P-labeled RNA was recovered from the gel by electro-elution followed by ethanol precipitation (Sambrook *et al*., 1989). The single-stranded oligonucleotides (9-mer and 13-AP RNA) at the appropriate concentrations were mixed and annealed by heating to 70°C followed by slow cooling to room temperature (rt).

Protein–RNA crosslinking

Samples (20 μl) containing protein (2 μM) and 9+13–AP duplex (5 μM; 30 000 c.p.m.) were prepared in reaction buffer (50 mM HEPES pH 7.5 or 50 mM Na-acetate pH 6.0, 10 mM $MgCl_2$, 20 mM KCl, 20 mM

β-mercaptoethanol and 20 μM alanine), incubated at rt for 5 min and irradiated at 300 nm for 2 min in a Rayonet mini-photochemical reactor (model RMR-500, The Southern New England Ultraviolet Co., Hamden, CT). The samples were then mixed with 10× SDS loading buffer (3 μl) and denatured at 37°C for 10 min before electrophoresis on a 10% SDS–polyacrylamide gel for 1 h. The proteins were transferred to a PVDF membrane (Millipore, Immobilon-P, Stanford, CA) by electro-blotting and visualized by staining with amido black (free protein) or by phosphorimaging of the [^{32}P]13-AP RNA–protein complex (Sardesai and Schimmel, 1998). A modified protocol for active-site titrations using the adenylate burst assay (Fersht *et al*., 1975) was used to determine enzyme concentration for the crosslinking experiments (the release of labeled inorganic phosphate in the supernatant was measured over time instead of the decrease of [γ-^{32}P]ATP bound to charcoal).

Acknowledgements

We thank Dr Luc Moulinier and Professor Dino Moras for helpful comments and sharing unpublished data with us. This work was supported by grants GM15539 and GM23562 from the National Institutes of Health.

References

Åberg,A., Yaremchuk,A., Tukalo,M., Rasmussen,B. and Cusack,S. (1997) Crystal structure analysis of the activation of histidine by *Thermus thermophilus* histidyl-tRNA synthetase. *Biochemistry*, **36**, 3084–3094.

Alexandrov,N.N. and Fischer,D. (1996) Analysis of topological and nontopological structural similarities in the PDB: new examples with old structures. *Proteins*, **25**, 354–365.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Arnez,J.G. and Cavarelli,J. (1997) Structures of RNA-binding proteins. *Q. Rev. Biophys.*, **30**, 195–240.

Arnez,J.G., Harris,D.C., Mitschler,A., Rees,B., Francklyn,C.S. and Moras,D. (1995) Crystal structure of histidyl-tRNA synthetase from *Escherichia coli* complexed with histidyl-adenylate. *EMBO J.*, **14**, 4143–4155.

Bajorath,J., Stenkamp,R. and Aruffo,A. (1993) Knowledge-based model building of proteins: concepts and examples. *Protein Sci.*, **2**, 1798–1810.

Banner,D.W., Kokkinidis,M. and Tsernoglou,D. (1987) Structure of the ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.*, **196**, 657–675.

Barrell,B.G. *et al*. (1980) Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc. Natl Acad. Sci. USA*, **77**, 3164–3166.

Benson,D.A., Boguski,M., Lipman,D.J. and Ostell,J. (1994) GenBank. *Nucleic Acids Res.*, **22**, 3441–3444.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Biou,V., Yaremchuk,A., Tukalo,M. and Cusack,S. (1994) The 2.9 A crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA (Ser). *Science*, **263**, 1404–1410.

Buechter,D.D. and Schimmel,P. (1993) Dissection of a class II tRNA synthetase: determinants for minihelix recognition are tightly associated with domain for amino acid activation. *Biochemistry*, **32**, 5267–5272.

Buechter,D.D. and Schimmel,P. (1995) Minor groove recognition of the critical acceptor helix base pair by an appended module of a class II tRNA synthetase (published erratum appears in *Biochemistry*, **34**, 16532). *Biochemistry*, **34**, 6014–6019.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Calendar,R. and Berg,P. (1966) The catalytic properties of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry*, **5**, 1690–1695.

Conway,N.E. and McLaughlin,L.W. (1991) The covalent attachment of multiple fluorophores to DNA containing phosphorothioate diesters results in highly sensitive detection of single-stranded DNA. *Bioconjug. Chem.*, **2**, 452–457.

Cusack,S., Berthet-Colominas,C., Hartlein,M., Nassar,N. and Leberman,R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 A. *Nature*, **347**, 249–255.

Davis,M.W., Buechter,D.D. and Schimmel,P. (1994) Functional dissection of a predicted class-defining motif in a class II tRNA synthetase of unknown structure. *Biochemistry*, **33**, 9904–9911.

Delarue,M., Poterszman,A., Nikonov,S., Garber,M., Moras,D. and Thierry,J.C. (1994) Crystal structure of a prokaryotic aspartyl tRNA-synthetase. *EMBO J.*, **13**, 3219–3229.

Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.

Fersht,A.R., Ashford,J.S., Bruton,C.J., Jakes,R., Koch,G.L. and Hartley,B.S. (1975) Active site titration and aminoacyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry*, **14**, 1–4.

Francklyn,C. and Schimmel,P. (1989) Aminoacylation of RNA minihelices with alanine. *Nature*, **337**, 478–481.

Francklyn,C., Shi,J.P. and Schimmel,P. (1992) Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science*, **255**, 1121–1125.

Hill,K. and Schimmel,P. (1989) Evidence that the 3′ end of a tRNA binds to a site in the adenylate synthesis domain of an aminoacyl-tRNA synthetase. *Biochemistry*, **28**, 2577–2586.

Ho,C., Jasin,M. and Schimmel,P. (1985) Amino acid replacements that compensate for a large polypeptide deletion in an enzyme. *Science*, **229**, 389–393.

Hou,Y.M. and Schimmel,P. (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*, **333**, 140–145.

Hou,Y.M. and Schimmel,P. (1989) Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. *Biochemistry*, **28**, 6800–6804.

Hyde,C.C., Ahmed,S.A., Padlan,E.A., Miles,E.W. and Davies,D.R. (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.*, **263**, 17857–17871.

Iyer,R.P., Egan,W., Regan,J.B. and Beaucage,S.L. (1990) 3H-1,2-Benzodithiole-3-one 1,1-dioxide as an improved sulfurizing reagent in the solid-phase synthesis of oligodeoxyribonucleoside phosphorothioates. *J. Am. Chem. Soc.*, **112**, 1253–1254.

Jasin,M. and Schimmel,P. (1984) Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. *J. Bacteriol.*, **159**, 783–786.

Jasin,M., Regan,L. and Schimmel,P. (1983) Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. *Nature*, **306**, 441–447.

Jasin,M., Regan,L. and Schimmel,P. (1985) Two mutations in the dispensable part of alanine tRNA synthetase which affect the catalytic activity. *J. Biol. Chem.*, **260**, 2226–2230.

Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.

Logan,D.T., Mazauric,M.H., Kern,D. and Moras,D. (1995) Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J.*, **14**, 4156–4167.

Lu,Y. and Hill,K.A. (1994) The invariant arginine in motif 2 of *Escherichia coli* alanyl-tRNA synthetase is important for catalysis but not for substrate binding. *J. Biol. Chem.*, **269**, 12137–12141.

McClain,W.H. and Foss,K. (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end. *Science*, **240**, 793–796.

Moras,D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.*, **17**, 159–164.

Mosyak,L., Reshetnikova,L., Goldgur,Y., Delarue,M. and Safro,M.G. (1995) Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*. *Nature Struct. Biol.*, **2**, 537–547.

Musier-Forsyth,K. and Schimmel,P. (1994) Acceptor helix interactions in a class II tRNA synthetase: photoaffinity crosslinking of an RNA miniduplex substrate. *Biochemistry*, **33**, 773–779.

Nureki,O., Vassylyev,D.G., Katayanagi,K., Shimizu,T., Sekine,S., Kigawa,T., Miyazawa,T., Yokoyama,S. and Morikawa,K. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase (published erratum appears in *Science*, **268,** 625). *Science*, **267**, 1958–1965.

Park,S.J., Hou,Y.M. and Schimmel,P. (1989) A single base pair affects binding and catalytic parameters in the molecular recognition of a transfer RNA. *Biochemistry*, **28**, 2740–2746.

Puglisi,J.D. and Tinoco,I.,Jr (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, **180**, 304–325.

Regan,L. (1986) The functional organization of alanine-tRNA synthetase: analysis of deletion and point mutations. PhD Thesis. Massachusetts Institute of Technology, Cambridge.

Regan,L., Bowie,J. and Schimmel,P. (1987) Polypeptide sequences essential for RNA recognition by an enzyme. *Science*, **235**, 1651–1653.

Ribas de Pouplana,L. and Schimmel,P. (1997) Reconstruction of quaternary structures of class II tRNA synthetases by rational mutagenensis of a conserved domain. *Biochemistry*, **36**, 15041–15048.

Ribas de Pouplana,L., Buechter,D.D., Davis,M.W. and Schimmel,P. (1993) Idiographic representation of conserved domain of a class II tRNA synthetase of unknown structure. *Protein Sci.*, **2**, 2259–2262.

Ripmaster,T.L., Shiba,K. and Schimmel,P. (1995) Wide cross-species aminoacyl-tRNA synthetase replacement *in vivo*: yeast cytoplasmic alanine enzyme replaced by human polymyositis serum antigen. *Proc. Natl Acad. Sci. USA*, **92**, 4932–4936.

Rost,B. and Sander,C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19**, 55–72.

Rost,B., Schneider,R. and Sander,C. (1997) Protein fold recognition by prediction-based threading. *J. Mol. Biol.*, **270**, 471–480.

Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.C. and Moras,D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA (Asp). *Science*, **252**, 1682–1689.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sardesai,N.Y. and Schimmel,P. (1998) Noncovalent assembly of microhelix recognition by a class II tRNA synthetase. *J. Am. Chem. Soc*., **120**, 3269–3270.

Schimmel,P. (1990) Alanine transfer RNA synthetase: structure–function relationships and molecular recognition of transfer RNA. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **63**, 233–270.

Schimmel,P.R. and Soll,D. (1979) Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. *Annu. Rev. Biochem.*, **48**, 601–648.

Shi,J.P., Musier-Forsyth,K. and Schimmel,P. (1994) Region of a conserved sequence motif in a class II tRNA synthetase needed for transfer of an activated amino acid to an RNA substrate. *Biochemistry*, **33**, 5312–5318.

Shiba,K., Ripmaster,T., Suzuki,N., Nichols,R., Plotz,P., Noda,T. and Schimmel,P. (1995) Human alanyl-tRNA synthetase: conservation in evolution of catalytic core and microhelix recognition. *Biochemistry*, **34**, 10340–10349.

Stultz,C.M., White,J.V. and Smith,T.F. (1993) Structural analysis based on state-space modeling. *Protein Sci*., **2**, 305–314.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.

Yang,S.W. and Nash,H.A. (1994) Specific photo-crosslinking of DNA–protein complexes: identification of contacts between integration host factor and its target DNA. *Proc. Natl Acad. Sci. USA*, **91**, 12183–12187.

Yarus,M. and Berg,P. (1967) Recognition of tRNA by aminoacyl tRNA synthetases. *J. Mol. Biol*., **28**, 479–490.

Received June 5, 1998; revised July 14, 1998; accepted July 15, 1998