Current Topics

Indirect Readout of tRNA for Aminoacylation[†]

John J. Perona*[,‡] and Ya-Ming Hou[§]

Department of Chemistry and Biochemistry and Interdepartmental Program in Biomolecular Science and Engineering, University of California, Santa Barbara, California 93106-9510, and Department of Biochemistry and Molecular Biology, Thomas Jefferson University, 233 South 10th Street, BLSB 220, Philadelphia, Pennsylvania 19107

Received July 24, 2007; Revised Manuscript Received August 6, 2007

ABSTRACT: Aminoacylation of tRNA by aminoacyl-tRNA synthetases is the essential reaction that matches protein amino acids with the trinucleotide sequences specified in mRNA. Direct electrostatic interactions made by tRNA synthetases with discriminating functional groups on the tRNA bases have long been known to determine aminoacylation specificity. However, structural and biochemical studies have revealed a second "indirect readout" mechanism that makes an important contribution as well. In indirect readout, the sequence-dependent conformations of tRNA are recognized through protein contacts with the sugar−phosphate backbone and with nonspecific portions of the bases. This mechanism appears to function in single-stranded regions, in canonical A-type duplex segments, and in the complex tertiary core portion of the tRNA. Operation of the indirect mechanism is not exclusive of the direct mechanism, and both are further mediated by induced-fit rearrangements, in which enzyme and tRNA undergo precise conformational changes after formation of an initial encounter complex. The examples of indirect readout in tRNA synthetase complexes extend the concept beyond its traditional application to DNA duplexes and serve as models for the operation of this mechanism in more complex systems such as the ribosome.

Protein−nucleic acid complexes are central to all cellular processes involving genome maintenance, replication, and information transfer. To select among nucleic acid sequences, proteins may directly read the information content by forming specific hydrogen bonds with unique functional groups on the nucleic acid bases (*1*). However, proteins also discriminate among nucleic acids by making contacts with the sugar−phosphate backbone or with nonspecific portions of the bases (*2*, *3*). For duplex DNA, this "indirect readout" is validated by extensive experimental and theoretical investigations, which describe correlations between the sequence of the bases and the detailed conformation of the sugar−phosphate backbone (reviewed in ref *4*). The sequence-dependent backbone effects arise from differences in thermodynamic base stacking propensities among the individual bases and base pair steps, and from local differences in mechanical properties of the duplex due to steric overlap of bases in adjacent steps. Together, these forces give rise to two phenomena: (i) sequence-dependent differences in the distribution of equilibrium unliganded conformations of the nucleic acid and (ii) differences in the energetic cost of deforming the equilibrium conformation into the particular structure that is needed for productive protein binding and chemical transformation.

While both of these phenomena are crucial in determining the quantitative measures of selectivity, the latter is of particular importance because most protein−nucleic acid complexes form by induced-fit rearrangements of the unbound conformations of each partner (*5*). Because of this, indirect readout can operate to modulate the first-order rearrangements that occur after formation of an initial encounter complex, thus modulating the overall association kinetics and through this the binding affinity. Further, interactions of a protein with a noncognate nucleic acid may generate conformational changes different from those produced in the specific complex (*6*). This provides a mechanism by which induced fit can contribute to catalytic specificity, by driving the noncognate protein−nucleic acid complex into a low-energy ground state complex in which the active site is not fully assembled (*7*). The rate of the catalytic step in such a distorted noncognate complex is thereby slowed.

In this review, we shall employ the term indirect readout broadly, to encompass all forms of sequence discrimination among similarly structured nucleic acids that do not involve direct contacts of the protein with specific moieties of the nucleotide bases. The concept is thus extended from its traditional application to DNA duplexes to include protein−RNA complexes in which recognition occurs in double-stranded A-helices, in single-stranded loops and segments, and in more complex globular regions. Examples are drawn

[†] Supported by National Institutes of Health Grants GM63713 (to J.J.P.) and GM52666 (to Y.-M.H.).

* To whom correspondence should be addressed. Telephone: (805) 893-7389. Fax: (805) 893-4120. E-mail: perona@chem.ucsb.edu.

[‡] University of California.

[§] Thomas Jefferson University.

10.1021/bi7014647 CCC: $37.00
Published on Web 08/24/2007

from the aminoacyl-tRNA[1] synthetases and closely related proteins, especially in systems where detailed X-ray structures of the enzyme−tRNA complexes are available, because these represent a set of well-studied model systems allowing for a common conceptual framework. It is important to recognize that all the tRNA synthetases read their cognate tRNA sequences by direct readout as well, often at portions of the tRNA molecule directly proximal to where indirect effects are proposed to operate. The mechanisms are not mutually exclusive but operate together to ensure the very high levels of specificity required to maintain fidelity in protein synthesis.

Indirect Readout of DNA Duplexes: Model Systems for Binding and Catalysis

The concept of indirect readout first emerged from crystallographic analyses of repressor−operator complexes. It was discovered that when the *trp* repressor binds to its operator DNA sequence, the protein makes no contacts with base-specific moieties on the heterocyclic rings (*2*). Extensive interactions with the sugar−phosphate backbone of the nucleic acid were observed instead of the direct contacts. Further, the site-specific binding affinities of 434 repressor and Cro proteins were shown to depend on portions of their binding site sequences not directly contacted by the proteins (*3*). Because genetic and biochemical studies had established significant DNA binding sequence specificity for each of these proteins, it was necessary to postulate that the sequence of the nucleic acid allowed the duplex to adopt a particular conformation and that this conformation represented an indirect means by which the base sequence could be distinguished. The structure of the *trp* repressor complex further suggested that water-mediated contacts bridging the protein and nucleic acid contribute to the indirect readout mechanism (*2*). The role of indirect readout in transcription regulation, the major area in which this concept has so far been applied, has been recently reviewed (*4*).

Type II restriction endonucleases also provide model protein−DNA systems in which the underlying thermodynamic basis for indirect readout is particularly clear. For example, EcoRV endonuclease cleaves the DNA duplex 5′-GATATC-3′ at the center TA step, leaving blunt ends with 5′-phosphates. In the *Eco*RV−DNA complex, direct protein readout at the center TA step is limited to hydrophobic contacts with the exocyclic 5-methyl groups of the thymine bases (*8*). Using base−analogue studies, it was shown that the contribution of indirect readout to the overall enzyme discrimination against a noncognate CG pair at the center step is between 6 and 10 kcal/mol, greater than the contribution of 5 kcal/mol made by direct readout (*9*). The basis for the discrimination by indirect readout lies in the observation that the bases at the center TA step are partially unstacked by a 50° roll into the major groove, to permit formation of a productive Michaelis complex. Thermodynamic studies show that a TA step is the most easily unstacked dinucleotide step in duplex DNA (*4*); thus, the greater energetic cost associated with unstacking CG provides a basis for discrimination at the binding step. Greater discrimination occurs at the catalytic step, because even the small disruptions observed in the active site structures of *Eco*RV−DNA base−analogue complexes are sufficient to disrupt formation of the divalent metal binding sites required for catalysis (*9*).

Comparative analysis of the homologous *Hinc*II−DNA complex further illustrates how underlying thermodynamic differences in base pair stacking are exploited as a basis for achieving selectivity by indirect readout (*10*). *Hinc*II performs blunt-ended cleavage of the degenerate 5′-GTYRAC-3′ sequences at the center YR step. However, the center YR step opens by roll into the major groove by only 25° rather than the value of 50° observed in the *Eco*RV complex, and this opening is accompanied by an unusual deformation involving cross-strand stacking of the center step purines. The basis for the degenerate cleavage at all center step YR sequences, as compared with the exclusive TA specificity of *Eco*RV, could be explained in terms of compensating thermodynamic properties of the bases. A CG step incurs the greatest penalty in unstacking among the four YR steps, while guanine most favors stacking among the individual bases (*11*). Hence, it appears that the greater free energy cost of unstacking CG center steps in the *Hinc*II recognition site can be compensated by the improved cross-strand stacking of the guanines, consistent with the observed degeneracy.

These correlations, of course, are *post hoc* rationalizations from the observed X-ray structures that are consistent with experimental and theoretical determinations of base stacking energies. That such correlations are possible, however, suggests that differences among base stacking energies indeed do contribute to indirect readout in protein−nucleic acid complexes and that the thermodynamic scales of base stacking propensities derived from experimental and theoretical considerations may be more or less accurate. It has been suggested instead that calculations of base stacking energies are of little use in quantifying indirect readout in any particular system, because the energies involved are relatively small and because of stereochemical constraints imposed by the sugar−phosphate backbone (*4*). However, the experimental observations in the type II restriction endonucleases suggest that careful structure−function analyses of modified DNAs bound to protein, and systematic comparisons among homologous systems, may provide approaches to bridging the gap between the extensive thermodynamic and computational studies of isolated DNA duplexes and the properties of particular protein−DNA complexes.

Aminoacyl-tRNA Synthetase−tRNA Complexes

We consider now the operation of these fundamental principles in the context of the aminoacyl-tRNA synthetase family. These enzymes perform the essential decoding step in translation by matching amino acids with tRNA molecules possessing the appropriate anticodon triplet sequences (reviewed in ref *12*). The 22 known aminoacyl-tRNA synthetases [including enzymes that catalyze phosphoserylation (Sep) of $tRNA^{Cys}$ and pyrrolysylation (Pyl) of an unusual UAG-decoding tRNA in certain archaebacteria] perform the aminoacylation reaction in a two-step Mg^{2+}-dependent pathway:

[1] Abbreviations: aaRS, aminoacyl-tRNA synthetases; tRNA, transfer RNA; aa, amino acid; aa-tRNA, aminoacyl-tRNA; ATP, adenosine triphosphate; AMP, adenosine monophosphate; DHFR, dihydrofolate reductase; Ψ, pseudouridine; m^2A, 2-methyladenine.

$$\text{aaRS} + \text{aa} + \text{ATP} \leftrightarrow \text{aaRS}\cdot\text{aa}-\text{AMP} + \text{PP}_i$$

$$\text{aaRS}\cdot\text{aa}-\text{AMP} + \text{tRNA} \leftrightarrow \text{aaRS}\cdot\text{aa}-\text{tRNA} + \text{AMP}$$

ATP is used to activate the amino acid for transfer to either the 2′-OH or 3′-OH group at the conserved 3′-terminal A76 nucleotide of the tRNA. The ester linkage of the aa-tRNA then provides much of the energy required for peptide bond formation on the ribosome. Very strong specificity is manifested for both amino acid and tRNA: k_{cat}/K_m for noncognate aminoacylation is generally decreased by $\geq 10^5$-fold, with specificity manifested at both the binding and catalytic steps (*12*−*14*).

The tRNA synthetases are divided into two classes based on the structures of their active site domains. All class I enzymes (specific for Arg, Cys, Ile, Leu, Met, Val, Glu, Gln, Lys-I, Tyr, and Trp) possess a Rossmann dinucleotide fold that binds ATP, the amino acid, and the acceptor end of the tRNA and provides the site of catalysis for the two-step reaction. In class II enzymes (specific for Ala, Asp, Asn, Gly, His, Lys-II, Phe, Pro, Ser, Thr, Pyl, and Sep), these functions are provided by a mixed α/β-fold featuring an antiparallel β-sheet. All 22 enzymes possess idiosyncratic additional domains that provide specificity in tRNA recognition. Crystal structures are available for the 20 canonical enzymes (*12*, *15*−*18*) as well as for the noncanonical PylRS (*19*) and SepRS (*20*, *21*), with a majority also determined as complexes bound to cognate tRNA.

It is observed generally that the L-shaped tRNA molecule binds across multiple domains of its cognate synthetase such that the single-stranded 3′-acceptor end of the molecule is properly positioned in the active site. The horizontal arm of the compact L-shaped structure of tRNA is formed by a coaxial stack of the acceptor and T stems, while the perpendicular vertical arm is formed from a coaxial stack of the dihydrouridine (D) and anticodon stems. The two helical arms are joined by the D, T, and V (variable) loops to form the highly compact and globular core (*22*, *23*) (Figure 1a,b). The overall orientation of tRNA binding is generally a class-specific feature. Most class I enzymes are monomeric, and the tRNA binds such that the minor groove of the acceptor stem interacts with the active site domain of the synthetase (Figure 1c). In contrast, class II tRNA synthetases are primarily dimeric and approach the tRNA acceptor stem from the major groove side (Figure 1d). However, in the dimeric class I enzymes (TyrRS and TrpRS), the tRNA binds across the subunits with its acceptor stem oriented in a manner resembling that of class II complexes.

Indirect Readout in Double-Helical tRNA Stems

Most tRNAs contain 20 or 21 canonical Watson−Crick and G-U base pairs distributed among the acceptor, D, anticodon, and T stems (Figure 1a). However, the D and T stems form part of the complex tertiary core of the tRNA, and their detailed helical parameters are likely to be significantly influenced by surrounding portions of the chain (see below). In contrast, the acceptor and anticodon stems exist as isolated, canonical A-helical duplexes. With a few exceptions, these regions do not possess asymmetric internal loops or mismatched base pairs. Nearly all cocrystal structures of aminoacyl-tRNA synthetase−tRNA complexes reveal protein interactions with the acceptor stem, and many also feature anticodon stem contacts. Mutagenesis studies involving replacements of tRNA base pairs in the acceptor stem have shown that tRNA synthetases often discriminate among tRNAs by contacting this portion of the molecule (*13*). However, interactions with the anticodon stem are only rarely used as a means of discriminating among tRNAs.

As for DNA−protein complexes, indirect readout of the information in tRNA acceptor stems may be most readily inferred when crystal structures of a tRNA synthetase complex do not reveal direct hydrogen bonding contacts at particular base pairs, yet mutagenesis studies indicate that the identity of the base pairs is important to aminoacylation efficiency. Clear examples of this phenomenon, however, are not common. One case in which indirect readout does appear by these criteria is the aspartyl-tRNA synthetase (AspRS) system in *Escherichia coli* (*24*). The dimeric *E. coli* AspRS binds two tRNAs, with each $tRNA^{Asp}$ confined to interactions with one monomeric enzyme subunit. The acceptor stem of the tRNA interacts with the enzyme primarily through water-mediated interactions: the well-resolved structure reveals more than 20 water-mediated hydrogen bonds in this region, although only three of these represent direct protein−water−RNA contacts (*24*). In contrast, the crystal structure of the yeast AspRS−$tRNA^{Asp}$ complex does not reveal a large number of acceptor stem contacts: only the extremity of the acceptor arm from nucleotide A72 to A76 is in contact with the tRNA synthetase (*25*). In accordance with these findings, mutational analysis shows that the most important recognition nucleotides for both *E. coli* and yeast AspRS are located in the anticodon loop and at the discriminator nucleotide (position 73), rather than in the acceptor stem (*26*, *27*).

The evidence for indirect readout of the acceptor stem sequence by *E. coli* AspRS arises from studies of the heterologous complex formed between the dimeric *E. coli* AspRS and two yeast $tRNA^{Asp}$ molecules (*28*). The crystal structure of this heterologous complex showed that one yeast tRNA was bound to the *E. coli* enzyme only through anticodon loop contacts, while the second yeast $tRNA^{Asp}$ adopted an orientation similar to that of the cognate *E. coli* tRNA, maintaining a significant number of water-mediated contacts across the acceptor stem. Most importantly, *E. coli* AspRS aminoacylates yeast $tRNA^{Asp}$ 4000-fold less efficiently than the cognate *E. coli* $tRNA^{Asp}$, as judged by relative k_{cat}/K_m values. However, replacement of the U1-A72 and C2-G71 base pairs of yeast $tRNA^{Asp}$ with the G1-C72 and G2-C71 base pairs found in the *E. coli* tRNA, respectively, improved aminoacylation by *E. coli* AspRS by 150-fold (*28*). Thus, the base pairs at the extreme end of the acceptor arm are crucial to tRNA recognition by *E. coli* AspRS even though they are mainly contacted via water-mediated interactions with the sugar−phosphate backbone in both the native bacterial and heterologous tRNA complexes. The structural information for both homologous and heterologous complexes, together with the mutational data showing partial reconstitution of activity, provides very strong evidence demonstrating acceptor stem recognition by indirect readout. It is interesting to note that *E. coli* $tRNA^{Asp}$ possesses a G4-U69 wobble base pair, while the yeast $tRNA^{Asp}$ possesses a similar wobble pair, but at a different position (U5-G68). Because the G-U wobble pair introduces a distinct stereochemical signature into the backbone of the

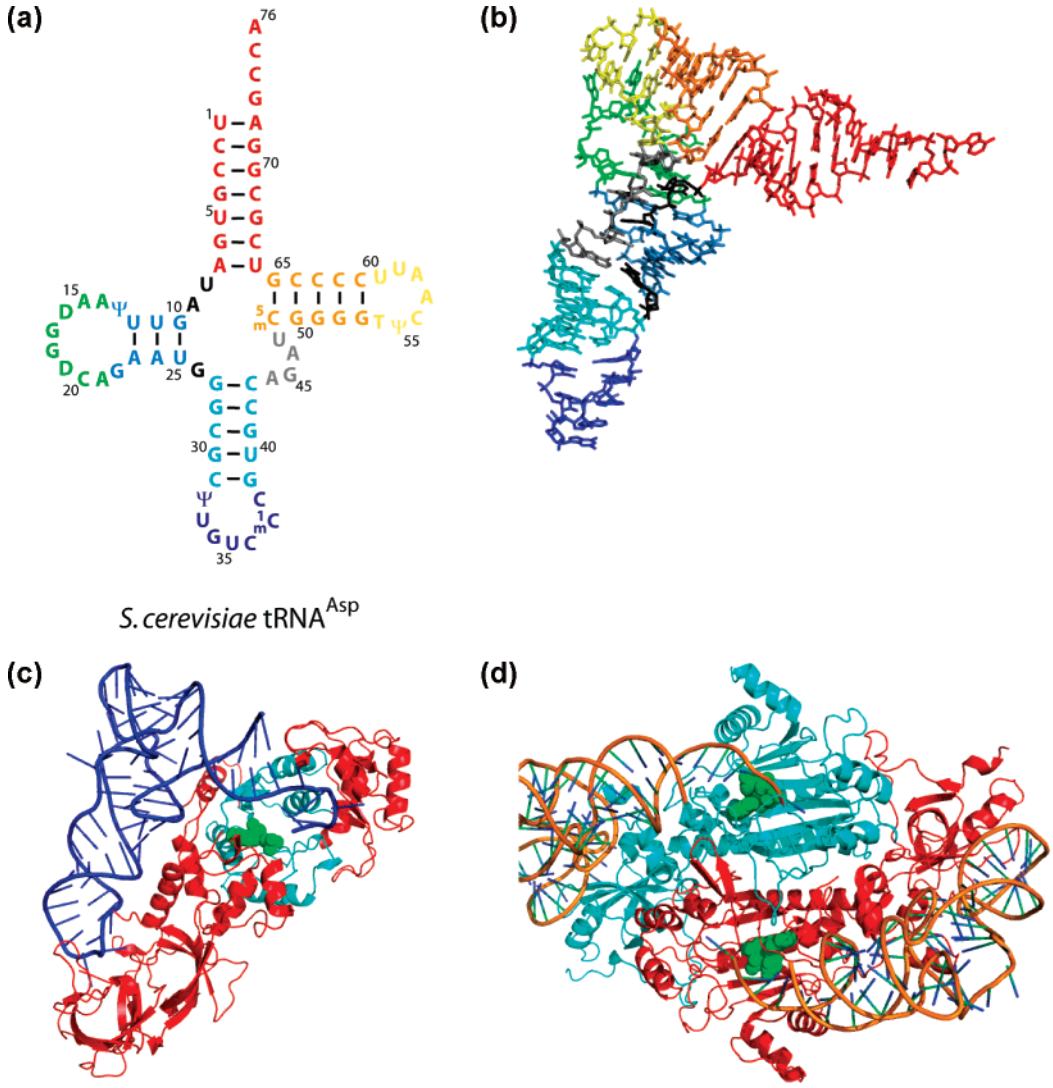

FIGURE 1: (a) Secondary and (b) tertiary structure of yeast $tRNA^{Asp}$ (PDB entry 2TRA). The elements of the structure are color-coded, as follows: red, acceptor stem (nucleotides 1−7 and 66−72) and single-stranded acceptor end (nucleotides 73−76); orange, T stem (nucleotides 49−53 and 61−65); yellow, T loop (nucleotides 54−60); slate, D stem (nucleotides 10−13 and 22−25); green, D loop (nucleotides 14−21); gray, variable loop (nucleotides 44−48); black, connector segments (nucleotides 8, 9, and 26); cyan, anticodon stem (nucleotides 27−31 and 39−43); and dark blue, anticodon loop (nucleotides 32−38). (c) Depiction of the structure of the *E. coli* glutaminyl-tRNA synthetase−$tRNA^{Gln}$−AMP ternary complex (PDB entry 1GTS), an example of a class I tRNA synthetase. The catalytic Rossmann fold of the enzyme is colored cyan, and the AMP ligand is shown in a space-filling representation (green). (d) Depiction of the structure of the dimeric yeast aspartyl-tRNA synthetase−$tRNA^{Asp}$−ATP ternary complex (PDB entry 1ASZ), an example of a class II tRNA synthetase. The ATP in each subunit is shown in a space-filling representation (green).

acceptor stem (*29*), it seems possible that replacement of sequences at these additional positions might reconstitute activity more fully. Although these experiments have not yet been conducted, they offer the potential to provide a clearer demonstration of the operation of indirect readout in this system (Figure 2).

The *E. coli* aspartylation system also offers perhaps the best-described example of how indirect readout in the acceptor stem is mediated by means of induced fit transitions. Binding of the $tRNA^{Asp}$ acceptor stem through the network of water molecules mentioned above results in closing of the tRNA hinge angle from 110° to 95° (*25*), and this tertiary change is accompanied by a set of global domain reorientations in the protein (*30*). The structural changes converge in the active site, where they trigger the movement of a "flipping loop" which adopts a conformation allowing productive binding of the acceptor end of cognate but not noncognate tRNA substrates (Figure 2a,b).

A second well-studied example of acceptor stem recognition is the alanylation system, where extensive studies over many years have established that the G3-U70 base pair in $tRNA^{Ala}$ is a major determinant for aminoacylation by *E. coli* AlaRS (*31*−*33*). This single base pair has been implicated in both direct and indirect readout. The NMR structure of the G3-U70 base pair in the acceptor helix of $tRNA^{Ala}$ shows that the pair adopts the expected wobble geometry (*29*), in which the G base shifts toward the minor groove while the U base shifts toward the major groove. The proposal for direct readout of the G3-U70 base pair by AlaRS is based on extensive studies of base−analogue substitutions at moieties in the major and minor grooves, as studied in a chemically synthesized acceptor stem of $tRNA^{Ala}$. These base−analogue studies show that the exocyclic 2-amino group of G3 in the minor groove of the base pair is essential for aminoacylation (*34*, *35*). Elimination of this functional group or committing it to an intrahelical hydrogen bond

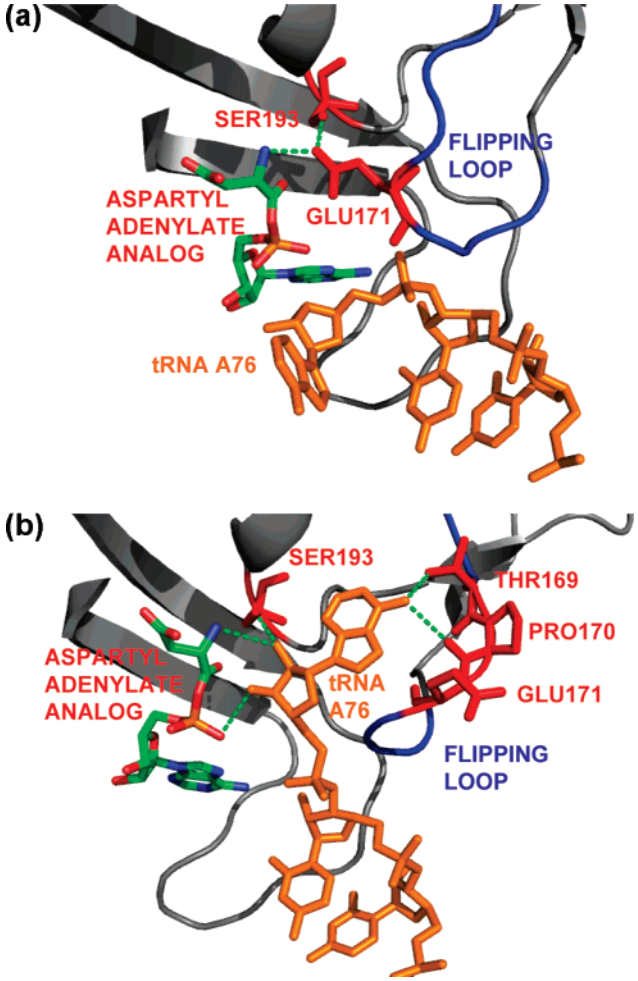

FIGURE 2: Active site structure of *E. coli* AspRS. (a) Crystal structure of *E. coli* AspRS bound to yeast $tRNA^{Asp}$ in a heterologous complex. The active site is in an unproductive conformation. Hydrogen bonds are depicted with green dots. (b) Crystal structure of *E. coli* AspRS bound to the cognate *E. coli* $tRNA^{Asp}$. Note that the flipping loop (blue) adopts a closed conformation in panel a, as tRNA A76 (orange) points away from the aspartyl adenylate substrate (left). In contrast, the flipping loop adopts an open conformation in panel b, and tRNA A76 (orange) appears properly juxtaposed with aspartyl adenylate. The following coordinates were used: noncognate *E. coli* AspRS−yeast $tRNA^{Asp}$ (PDB entry 1IL2) and cognate *E. coli* AspRS−*E. coli* $tRNA^{Asp}$ (PDB entry 1C0A). Adapted from ref *28*.

inactivates aminoacylation. The exocyclic amino group conferred alanylation activity regardless of the nature of the base pair in which it resided, suggesting a direct readout mechanism.

In contrast, genetic studies using amber suppressors of *E. coli* $tRNA^{Ala}$ show that certain substitutions of the G3-U70 base pair that do not retain the exocyclic 2-amino group in the minor groove are recognized by AlaRS in vivo. Examples of functional substitutions include the G-A, A-C, C-A, and C-C base pairs, none of which exhibits a minor groove configuration resembling that of the G-U base pair (*36*, *37*). A NMR analysis of the $tRNA^{Ala}$ acceptor helix containing the active C3-C70 base pair revealed structural irregularities, with some features shared by the wild-type G3-U70 pair (*37*). In this indirect readout model, AlaRS optimally aminoacylates acceptor stems of a particular conformation, which it recognizes by means of contacts to the sugar−phosphate backbone. Support for the backbone recognition hypothesis also arises from the study of deoxy sugar analogues, which shows that a number of 2′-OH groups in and around the G3-U70 base pair individually contribute between 1 and 2 kcal/mol to the stabilization of the transition state for aminoacylation (*38*). Together, the genetic and deoxy sugar analogue studies suggest that indirect readout also makes an important contribution to conferring recognition of the G3-U70 pair in the acceptor stem of $tRNA^{Ala}$ by *E. coli* AlaRS.

Work on the AlaRS system, although extensive, nonetheless reflects experimental difficulties in resolving the contribution of indirect readout that are prevalent in studies of all tRNA synthetase complexes. First, the use of in vivo genetic studies, while representing a powerful approach to appreciating the physiological significance of particular interactions, is inherently limited because in some cases a substantially impaired aminoacylation activity due to a mutated tRNA or synthetase may still sustain cellular viability. While these studies may allow determination of the fraction of tRNA that is aminoacylated in vivo or the degree of incorporation of an amino acid at a particular position in a target protein such as DHFR (*31*, *39*), such information is at best only suggestive of a particular recognition mechanism. Second, many of the chemical modification studies in the AlaRS system were necessarily (at the time) performed using short "minihelix" substrates comprising only the acceptor−T stem portion of the tRNA; these molecules are much poorer substrates than intact tRNA, and their use may lead to overestimation of the contributions of individual recognition determinants. For example, while the G3-U70 base pair of $tRNA^{Ala}$ confers alanylation properties upon noncognate minihelix substrates, such transplantation studies have not been performed in the context of full-length tRNAs. Finally, the AlaRS system lacks a high-resolution X-ray structure of the enzyme−tRNA complex to ground the extensive functional studies within a stereochemical context. Even in the many systems where high-resolution structures of the enzyme−tRNA complex are available (such as the aspartylation system described above), systematic structural studies of mutated complexes exhibiting altered recognition properties have not been performed. X-ray structures of modified complexes possessing altered specificity are essential, however, in elucidating how direct and indirect readout can jointly operate to allow proper relative juxtaposition of the reactive moieties of ATP, amino acid, and tRNA substrates in the pre-transition state complex on the enzyme. The *Eco*RV restriction endonuclease system provides a well-characterized example demonstrating how altered sugar−phosphate binding interactions in modified complexes can be correlated to the changes in binding affinity and single-turnover kinetics induced by the modifications (*9*).

Indirect Readout in Single-Stranded Regions

The secondary structure of tRNA includes seven distinct segments that are not present in canonical duplex stems: the T, D, variable, and anticodon loops; the four-nucleotide unpaired 3′-end terminating with the conserved CCA sequence; and the very short connectors bridging the acceptor−D stems and the D−anticodon stems (Figure 1a,b). Most of these segments, however, form part of the complex tertiary core. Only the 3′-acceptor nucleotides and the anticodon loop, at opposite ends of the L-shaped molecule, possess structures

that are not interdependent with other parts of the tRNA. Both of these are used as recognition sites by most tRNA synthetases (*13*). tRNA selectivity by indirect readout at these positions was first recognized in the crystal structure of the *E. coli* glutaminyl-tRNA synthetase (GlnRS)−tRNA^{Gln} complex (*40*). Although the details differ, similar principles operate for anticodon loop and acceptor end binding in many of the other systems.

The crystal structure of the GlnRS−tRNA^{Gln} complex shows that the tRNA^{Gln} anticodon loop adopts a conformation in which all three anticodon bases are unstacked to bind in complementary pockets formed by the C-terminal β-barrels of the enzyme (*40*, *41*). The anticodon stem of tRNA^{Gln} bound to GlnRS is also extended by two non-Watson−Crick base pairs: 2′-O-methyl-U32:Ψ38 and m^{2}A37:U33. Only one direct hydrogen bond links each of the two pairs, but a network of ordered water molecules assists in stabilizing the base-pairing configuration. The two new base pairs exhibit A-form geometry and stack well with the anticodon stem. Comparison of the anticodon loop structure of bound tRNA^{Gln} with that of other unliganded tRNAs (*23*, *42*) strongly suggests that the loop has been deformed from its unbound structure in an induced fit transition concomitant with protein binding (see below).

The discriminating Watson−Crick functional groups of all three anticodon nucleotides (C34, U35, and G36) are specifically recognized by direct hydrogen bonds with protein groups (*41*). In contrast, discriminating moieties of the U32:Ψ38 and m^{2}A37:U33 pairs are directly recognized only by a pair of hydrogen bonds donated by the side chain amide of Asn370 to exocyclic O2 of Ψ38 and N1 of m^{2}A37. Further, the unusual U32:Ψ38 and m^{2}A37:U33 pairs participate in a water-mediated base−phosphate interaction network that constrains the conformation of the nucleic acid backbone for extensive contacts with the protein at residues m2A37 and Ψ38. These observations suggest that the U32:Ψ38 and m^{2}A37:U33 pairs promote tRNA^{Gln} anticodon discrimination primarily by facilitating the adoption of the required anticodon loop structure at a lower free energy cost than other nucleotides. This provides a basis for tRNA selection by indirect readout (Figure 3a).

In the tRNA acceptor stem, operation of indirect readout is distinguished by the fact that the terminus of the tRNA acceptor stem is a site with increased flexibility and distortability upon protein interaction. These properties are evident in the GlnRS−tRNA^{Gln} complex and may be exploited to provide selective recognition by indirect readout. At the 3′-acceptor end of tRNA^{Gln}, the crystal structure shows that the U1-A72 base pair at the far end of the acceptor stem is completely disrupted and that the 3′-GCCA sequence forms a hairpin loop that positions the ribose sugar of A76 in the active site, adjacent to the α-phosphate of ATP (*40*). This conformation is stabilized by extensive protein interactions as well as by an intramolecular hydrogen bond donated by the exocyclic NH_2 group of G73 to the A72 phosphate (Figure 3b). The requirement for disruption of the 1-72 base pair provides discrimination against tRNAs possessing either a G1-C72 or a C1-G72 base pair, as the additional Watson−Crick hydrogen bond renders these pairs more difficult to break (*43*, *44*). Similarly, only G73 can provide the 2-NH_2 moiety for formation of the intramolecular hydrogen bond within the RNA hairpin, allowing selectivity against tRNAs

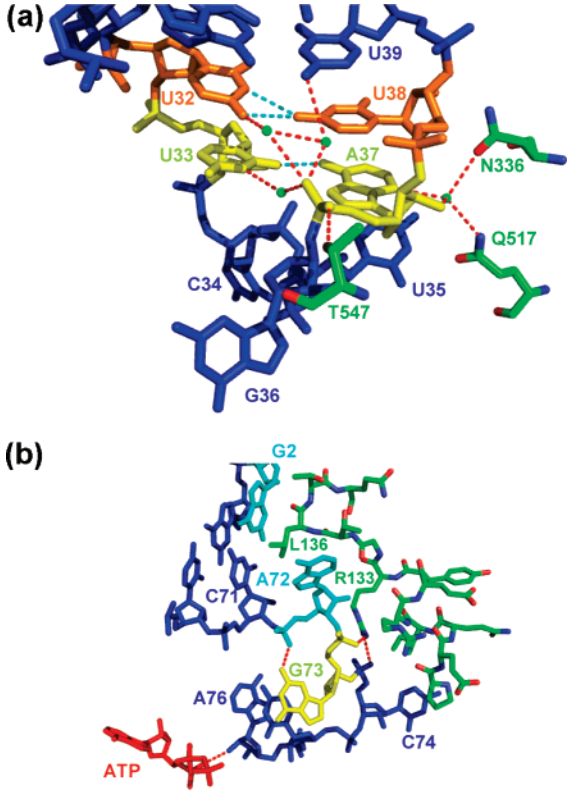

FIGURE 3: Indirect readout of single-stranded RNA by *E. coli* glutaminyl-tRNA synthetase (PDB entry 1GTR). (a) Interactions with the anticodon loop of tRNA^{Gln}. Green spheres depict key water molecules stabilizing the conformation of the anticodon loop and its interaction with protein. Hydrogen bonds between noncanonical U33-A37 and U32-U38 base pairs are shown as blue dashes; other hydrogen bonds are shown in red. (b) Hairpinning of the tRNA 3′-end into the active site. Intramolecular and intermolecular hydrogen bonds are depicted with red dashes. The ATP is colored red at bottom left. A hydrogen bond between the tRNA A76 2′-hydroxyl group and the α-phosphate of ATP assists in the mutual positioning of substrate moieties for catalysis. For clarity, only a small number of the interactions made are shown.

possessing any other base at this position. No specific protein interactions are made with U1, A72, or G73, yet biochemical and genetic studies show that mutations at these positions in the tRNA have significant effects on aminoacylation efficiency both in vivo and in vitro (*43*−*45*). Thus, these three nucleotides exert their effects on selectivity by facilitating adoption of the required 3′-hairpin structure at a lower free energy cost than other nucleotides (*40*).

The acceptor stem termini in tRNA are also the sites of 5′- and 3′-endonucleolytic processing and of the amino acid editing reaction. Site-specific cleavage by the ribozyme RNase P removes the 5′-extension of all precursor tRNAs to produce a 7 bp acceptor stem in nearly all mature species. However, an exception occurs in tRNA^{His}, where the mature tRNA possesses an additional guanosine residue (G^{-1}) at the 5′-terminus (*46*). In bacteria, archaea, and eukaryotic organelles, the G^{-1} nucleotide may form a base pair with C73, while in the eukaryotic cytoplasm, tRNA^{His} species possess A73 at the discriminator position, leading to a G^{-1}-A73 mismatch.

Although a cocrystal structure of HisRS bound to tRNA^{His} is not yet available, biochemical evidence in two systems shows that the G^{-1}-C73 pair is important to tRNA selectivity. In the *E. coli* histidine system, mutations of the G^{-1}-C73

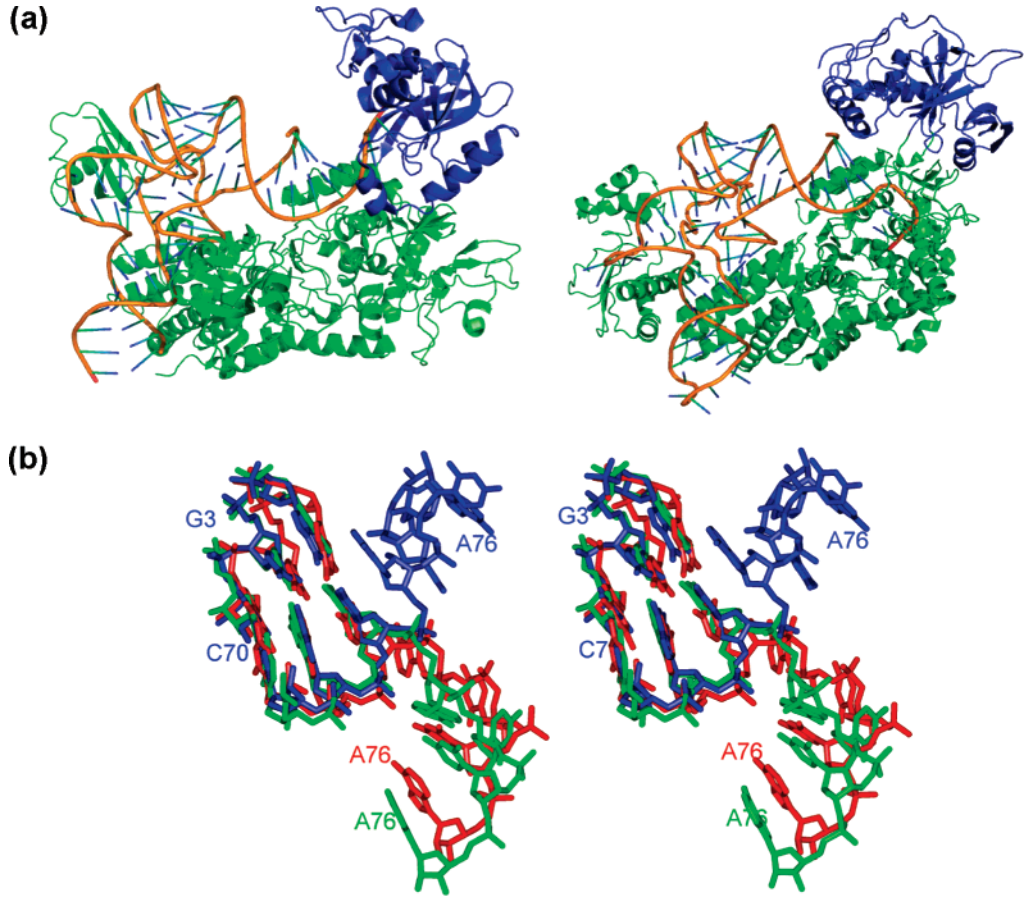

FIGURE 4: (a) Depiction of the cocrystal structure of *Thermus thermophilus* LeuRS bound to $tRNA^{Leu}$ in the post-transfer editing conformation (left) and depiction of the cocrystal structure of *Pyrococcus horikoshii* LeuRS bound to $tRNA^{Leu}$ in the aminoacylation conformation (right). In each drawing, the post-transfer editing domain is colored blue. (b) Detailed conformations of the acceptor 3′-end of $tRNA^{Leu}$ in the aminoacylation and post-transfer editing complexes: blue, trajectory of the 3′-end in the *T. thermophilus* post-transfer editing complex (PDB entry 2BYT); red, trajectory of the 3′-end in the *P. horikoshii* aminoacylation complex (PDB entry 1WZ2); and green, a second 3′-end trajectory in the *P. horikoshii* aminoacylation complex, corresponding to a conformation that may represent an intermediate along the pathway from the functionally relevant aminoacylation complex toward the post-transfer editing complex. Adapted from ref *16*.

pair decreased the steady state rate of histidylation (*47*), while short acceptor stem microhelix substrates derived from $tRNA^{Ala}$ and $tRNA^{Gly}$ were efficiently histidylated upon introduction of the G^{-1}-C73 pair (*48*, *49*). Further, the C73U mutation alone reduces the rate of the aminoacyl transfer step by ~1000-fold in single-turnover kinetics (*50*), suggesting the possibility that C73 might be directly read by HisRS. Evidence for indirect readout of the G^{-1}-C73 base pair comes from a study that investigated removal of the 5′-phosphate moiety from G^{-1}, which reduced the rate of histidylation by more than 500-fold (*51*). Similarly, *Saccharomyces cerevisiae* HisRS also recognizes the 5′-phosphate of $tRNA^{His}$; in this case, the rate of histidylation is reduced by 130-fold when the phosphate is removed (*52*). These comparative studies are of particular interest since they demonstrate that discrimination by indirect readout has been conserved across different domains in Nature.

Some tRNA synthetases possess editing mechanisms for removing incorrectly activated amino acids or misacylated tRNAs. In the latter event, known as post-transfer editing, the 3′-end of the misacylated tRNA dissociates from the synthetic active site and translocates across the surface of the enzyme to bind in a spatially separate editing domain that hydrolyzes the aminoacyl ester bond (*53*). In class I tRNA synthetases [such as IleRS (*54*, *55*), LeuRS (*16*, *56*), and ValRS (*57*)], the CCA sequence adopts a hairpin conformation to bind properly in the synthetic site but extends approximately along the acceptor stem helical axis to enter the editing site (Figure 4a). In near-mirror-image contrast, the CCA single strand in class II tRNA synthetases [such as ThrRS (*58*)] continues along the acceptor stem helical axis to enter the synthetic site but bends to enter the editing site. These differences, of course, reflect the distinct structural organization of class I versus class II enzymes.

The shuttling of the misacylated single-stranded tRNA 3′-end across the enzyme may be guided by interactions of specific amino acids with the sugar−phosphate backbone, the nucleotide bases, or both. Details of this complex conformational rearrangement are largely unknown, but some insight into the process has been recently provided by crystal structures of an archaeal class I LeuRS−$tRNA^{Leu}$ complex that reveal two conformations of the CCA terminus (*16*). In both conformations, the 3′-terminus of the tRNA is bound in the aminoacylation site (Figure 4b). Both conformations also reveal direct readout discrimination of the A73 nucleotide located immediately 5′ to the 3′-CCA terminus, consistent with biochemical studies that established this nucleotide as being crucial to tRNA identity in this system (*59*, *60*). However, the complexes differ in the precise acceptor end conformation, the global orientation of the editing domain, and the positioning of the upstream acceptor stem of the tRNA. Remarkably, the CCA end in the two

complexes appears to be marking a trajectory toward the editing domain, with a key conformational change altering the base stacking and phosphodiester backbone conformation at the junction between nucleotides A73 and C74 (Figure 4b).

The efficiency of the translocation step in LeuRS editing, then, appears likely to be influenced by conformational preferences of the CCA end as determined in part by stacking energies between the bases at the 3′-end of the molecule. Differences in the thermodynamic propensities of nucleotide bases to stack upon each other may thus provide the basis for variations in tRNA 3′-end translocation kinetics among the tRNA synthetases that possess editing domains (*4*). Some support for this notion arises from studies of *E. coli* ValRS, which is topologically similar to the archaeal LeuRS complex. Mutation of A73 to G73 in $tRNA^{Val}$ substantially reduced the level of overall editing (including both pre- and post-transfer steps), while mutations to C73 and U73, perhaps surprisingly, had little effect (*61*). In both of these systems (and very likely in editing reactions generally), the tRNA 3′-end translocation is also coupled to induced fit transitions in the enzymes: X-ray crystal structures show that tRNA synthetase editing domains undergo significant global reorientation upon tRNA binding and translocation (*54*−*58*, *62*). It is also relevant to note that thermodynamic and NMR studies of a model system consisting of the acceptor stem and the 3′-ACCA end of *E. coli* $tRNA^{Ala}$ showed that the single-stranded 3′-terminus significantly stabilizes the duplex, in a sequence-dependent manner (*63*). While the effect of the acceptor stem-proximal A73 nucleotide was greatest, all four 3′-dangling nucleotides were found to influence the flexibility and thermal melting transitions of the RNA. These studies further demonstrate the importance of base stacking interactions at the tRNA 3′-end.

Tertiary Core

The diversity of three-dimensional structure in RNA molecules arises from the ability of the polynucleotide chain to form complex, highly specific, and thermodynamically stable intramolecular interactions (*64*−*66*). Packing of small RNA secondary structural motifs often depends on the ability of divalent metal ions such as Mg^{2+} to bridge phosphate groups that are brought together in the tertiary core region, particularly when independent coaxially stacked helical domains are assembled together (*67*). However, stable tertiary RNA motifs that do not depend on divalent metal ions for stability are also known. Examples of these include double-pseudoknotted structures such as the hepatitis delta virus and *glmS* self-cleaving ribozymes (*68*, *69*). Transfer RNA falls into this latter category of stable globular RNAs that do not require precisely bound divalent metal ions within the core domain.

The two helical domains of tRNA are linked by the interactions of the D and T hairpin loops ("hinge" region) which together form a complex network of tertiary interactions with each other and with adjacent portions of the molecule (Figure 1; *22*, *23*). A set of highly conserved noncanonical base−base interactions forms an "augmented D stem", which stacks upon the D stem in the region closest to the hinge, extending the D stem−anticodon stem domain. For the majority of tRNAs possessing small variable loops of four or five nucleotides (the class I type), the variable loop also interacts in the major groove of the D and augmented D stems, forming base triple and other interactions. The innermost base pair of the augmented D stem forms the conserved "Levitt" pair interaction between the bases of nucleotide 15 in the D loop and nucleotide 48 in the variable loop. The Levitt pair, in turn, stacks beneath nucleotides 59 and 60 of the T loop, which are nestled underneath the perpendicular stack of the T stem. The number of bases in the T stem and T loop, and the conformation of this portion of the molecule, are nearly identical in all tRNAs (Figures 1 and 5).

Within this generally conserved structure of the tRNA hinge domain, sequence variations among species allow the sugar−phosphate backbones to adopt unique conformations. Part of the structural variation arises directly from small differences in the sizes of the variable and D loops. Even where these loops are identical in size, the pattern of base stacking interactions in the D and augmented D stems may vary because of sequence differences, generating significant diversity in the overall shape. For example, while $tRNA^{Gln}$ possesses a commonly found U8-A14-A21 base triple in the augmented D stem [as first observed in the structure of $tRNA^{Phe}$ (Figure 5a)], $tRNA^{Cys}$ possesses instead a U8-A14-A46 triple at the topologically equivalent position. Although the two tRNAs possess D loops of the same size, this difference likely arises because U21, which possesses the least propensity to stack among the four bases, flips out of the $tRNA^{Cys}$ core to interact with solvent.

While a number of examples of this type of diversity exist, the number of stacked bases, base pairs, and base triple layers in the vertical arm of class I tRNAs, from nucleotide 60 of the T loop to the 31-39 base pair at the end of the anticodon stem, remains constant at 14 (Figure 5a). Thus, it is meaningful to speak of topologically conserved backbone groups and hence to consider recognition of the core domain in class I tRNAs as an extension to the conceptual framework of indirect readout. In the different tRNAs, topologically equivalent backbone moieties occupy substantially different relative positions in space, separated by several angstroms or more, as may be easily seen by performing superpositions of distinct portions of the cores among species (*70*) (Figure 5b). Noncognate tRNAs with distinctly structured core domains may then be improperly bound on an enzyme surface as compared with cognate species, leading to mispositioning of the reactive 3′-end acceptor in the active site.

The compactness of the tertiary core and the large number of base stacking and electrostatic interactions that preserve its shape generally make it difficult to distinguish particular nucleotides as discrete identity determinants, because the function of any nucleotide in recognition usually depends on the tertiary structural context. A good example of this phenomenon is the unusual G15-G48 Levitt pair unique to $tRNA^{Cys}$ in *E. coli*, which is an important recognition element for CysRS (*71*−*74*). Introduction of a G15-G48 base pair into the tertiary core of *E. coli* $tRNA^{Gln}$ (which possesses the canonical G15-C48 base pair at this position), together with $tRNA^{Cys}$ identity determinants at the acceptor end and anticodon, produced a tRNA that was reduced in cysteinylation efficiency by 100-fold as compared with that of native $tRNA^{Cys}$ (*70*). The crystal structure of the G15-G48 $tRNA^{Gln}$ variant bound to GlnRS revealed that the base of G48 could

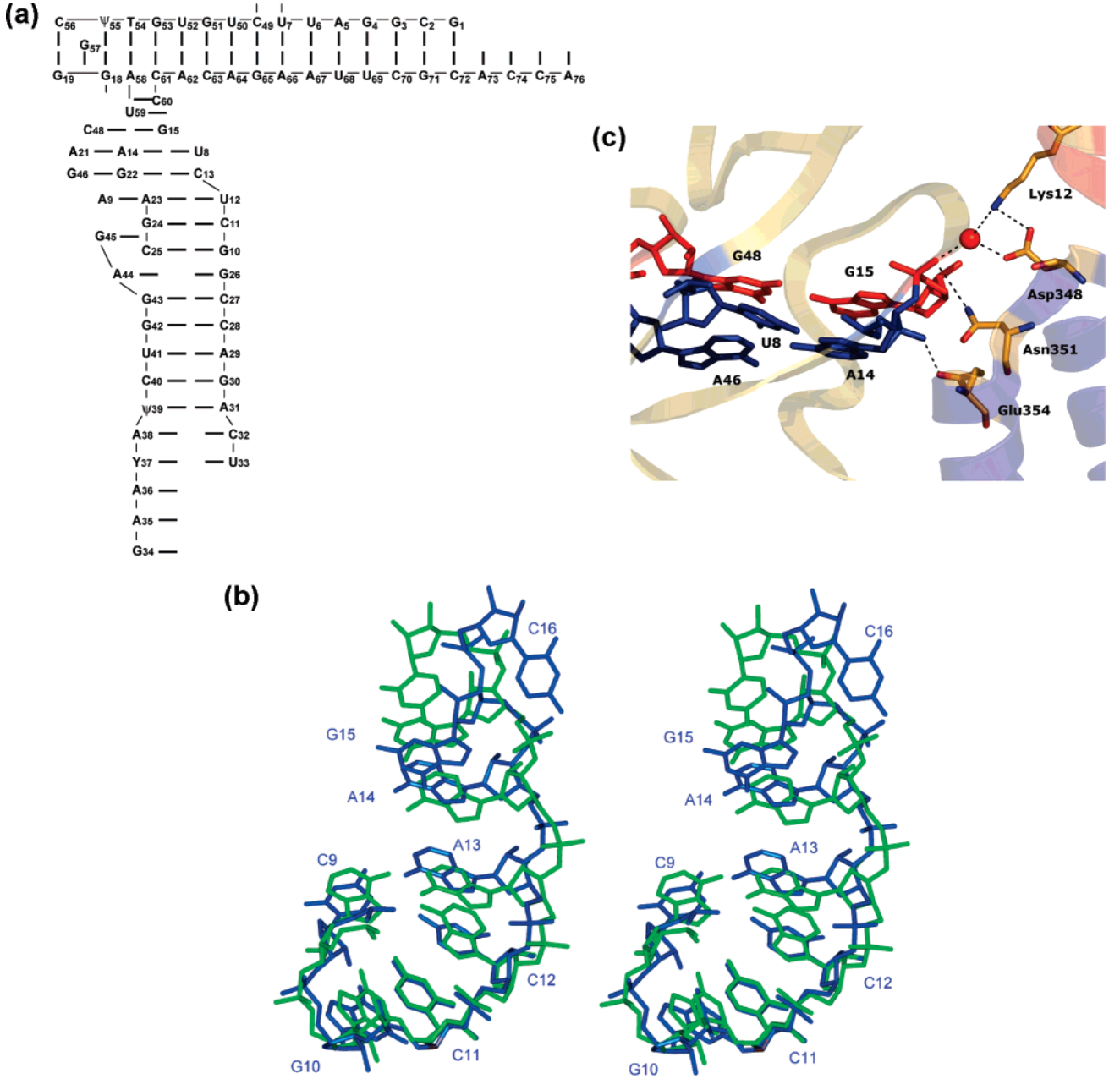

FIGURE 5: (a) Schematic drawing of the base stacking arrangement in yeast $tRNA^{Phe}$. (b) Superposition of a portion of the structures of *E. coli* $tRNA^{Cys}$ and *E. coli* $tRNA^{Gln}$, each bound to the cognate tRNA synthetase. The superposition is performed by aligning all backbone atoms in nucleotides 10−12 on the 5′-side of the D stem. The divergence in position of phosphate groups at nucleotides A14 and G15 is more than 2 Å, providing a basis for differential shape-dependent interactions with the tRNA synthetases. The following coordinates were used: GlnRS−$tRNA^{Gln}$ complex (PDB entry 1GTR) and CysRS−$tRNA^{Cys}$ complex (PDB entry 1U0B). (c) Indirect readout in the tertiary core of *E. coli* $tRNA^{Cys}$ by CysRS. Dashed lines denote hydrogen bonds between CysRS and tRNA backbone groups at nucleotides 14 and 15. The red sphere depicts a bridging water molecule. Adapted from refs *15* and *70*.

not be accommodated in the expected anti configuration within the $tRNA^{Gln}$ core but instead adopted the unusual syn orientation. The tertiary context provided by the surrounding nucleotides thus clearly determines the local conformation at positions 15 and 48. That the tertiary context is what is recognized is clear from the structures of the tRNA-bound GlnRS and CysRS complexes, both of which reveal enzyme interactions exclusively with the sugar−phosphate backbones in the core region (*15*, *40*).

E. coli CysRS recognizes the G15-G48 Levitt pair in *E. coli* $tRNA^{Cys}$ by indirect readout in its natural context. The structure of the CysRS−$tRNA^{Cys}$ complex showed that the enzyme makes a number of interactions at the backbone position of the Levitt pair. Hydrogen bonds are made at a nonbridging phosphate oxygen of G15 and at the 2′-OH group of A14 by the side chains of Asn351 and Glu354, respectively. In addition, a water-mediated hydrogen bond links the G15 phosphate with the side chains of Asp348 and Lys12, which form an adjacent salt bridge (Figure 5c) (*15*). To investigate the role of these interactions in indirect readout, the catalytic efficiencies of variant CysRS enzymes carrying mutations at position Asn351 or Glu354 were examined using both wild-type (G15-G48) and altered (G15-C48) tRNA substrates. Replacement of Glu354 with either alanine or glutamine reduces the catalytic efficiency by approximately 10-fold toward both wild-type and mutant

tRNAs, with the majority of the effects at the level of k_{cat}. Because the effect is identical toward both tRNA substrates, it appears that the Glu354 contact with the A14 ribose is not sensitive to any structural difference between these tRNA species. By contrast, CysRS mutants at Asn351 were altered in their relative preferences for wild-type versus mutant backbones. For example, the k_{cat}/K_m of the N351A mutant was reduced more than 30-fold compared to that of wild-type G15:G48 $tRNA^{Cys}$, but it retains catalytic efficiency similar to that of the wild-type enzyme with the G15:C48 variant. Thus, Asn351 mutants preferred G15:G48 $tRNA^{Cys}$ by only 6-fold, as compared with the 125-fold preference for the wild-type tRNA by native CysRS. It appears then that the Asn351-G15 phosphate contact mediates tRNA recognition by indirect readout, while the Glu354-A14 ribose contact does not. This combined mutational approach, in which both enzyme and tRNA are altered, provides a general methodology that can be used to explore which particular backbone contacts are important to substrate selectivity (*15*).

Use of the tRNA core domain as a basis for sequence discrimination is not limited to the tRNA synthetases. The tRNA-dependent amidotransferases also provide an example of tRNA selectivity via binding in this region. These enzymes play an important role in aminoacyl-tRNA synthesis in eukaryotic organelles, archaea, and most prokaryotes, all of which lack the canonical GlnRS and/or AsnRS enzymes (*75*). The synthesis of Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$ in these cases is a two-step process, in which the nondiscriminating glutamyl- and aspartyl-tRNA synthetases first catalyze misacylating reactions that generate Glu-$tRNA^{Gln}$ and Asp-$tRNA^{Asn}$, respectively. The tRNA-dependent amidotransferase then converts the carboxylate group of the acylated noncognate amino acid to the corresponding amide, using ammonia produced from hydrolysis of glutamine as the source of nitrogen. The amidotransferase enzymes must specifically recognize misacylated Glu-$tRNA^{Gln}$ and Asp-$tRNA^{Asn}$ while excluding the cognate Glu-$tRNA^{Glu}$ and Asp-$tRNA^{Asp}$ species that are also synthesized by the nondiscriminating tRNA synthetases.

The tRNA-dependent amidotransferases fall into two families. The GatCAB enzymes are able to convert both Glu-$tRNA^{Gln}$ and Asp-$tRNA^{Asn}$ to the correctly acylated products, while the GatDE subgroup is specific to the synthesis of Gln-$tRNA^{Gln}$. A recently determined crystal structure of the GatDE−$tRNA^{Gln}$ complex from *Methanothermobacter thermoautotrophicus* suggests that the enzyme discriminates between Glu-$tRNA^{Glu}$ and Glu-$tRNA^{Gln}$ by recognizing differences in the shape of the tRNA core (*76*). The GatE subunit of GatDE binds $tRNA^{Gln}$ by forming a concave surface complementary to the shape of the T stem−loop motif and D loop. Close steric complementarity to the D loop region of $tRNA^{Gln}$ appears to form the basis for exclusion of $tRNA^{Glu}$, which in this archaeon possesses a D loop that is three nucleotides larger. Enlargement of the $tRNA^{Gln}$ D loop to mimic that of $tRNA^{Glu}$ indeed eliminated tRNA binding, providing experimental confirmation of the importance of the overall shape (*76*). However, because the protein domain that binds the D loop is not well-ordered in the crystal structure, the extent to which recognition might also be based on a direct readout of functional groups is not yet known.

tRNA Discrimination by Shape Complementarity in tRNAs with a Long Variable Loop

As noted above, discrimination among class I tRNA core domains by tRNA synthetases and other enzymes can be considered in the conceptual frame of indirect readout: topologically equivalent backbone positions can be identified within the overall tertiary structure and their detailed relative orientations analyzed in terms of the sequence of the molecule. In contrast, tRNAs, including $tRNA^{Ser}$ and $tRNA^{Leu}$, as well as $tRNA^{Tyr}$ in some organisms, possess a much larger variable region consisting of a stem−loop motif that projects from the hinge (Figure 6a). The recognition of distinct globular shape in these class II tRNAs as a mechanism for selectivity is related to indirect readout, because usually most or all of the interactions are made with the sugar−phosphate backbone. However, it is distinct in the sense that the tertiary structures of the different tRNAs are not conserved. Crystal structures of all three class II tRNAs show that the orientation of the variable arm differs with respect to the global structure of the hinge region, depending on differences in the D and variable loop regions (*17*). The resulting large variations in overall shape give rise to recognition opportunities that are dependent more upon the relative orientations of structural motifs in space than upon the detailed sequence-dependent conformations of motifs conserved among class I tRNAs.

Class II tRNAs generally possess insertions of several nucleotides in the D loop as compared with class I species. One of these nucleotides interacts with the inner portion of the variable loop to stabilize a distinct conformation in each case (Figure 6b). The different orientations of the variable arm among the tRNAs thus depend on the identity of the interacting D loop nucleotide, the number of other inserted D loop nucleotides, and the number and identities of the unpaired variable loop nucleotides that flank the 5′- and 3′-sides of the stem−loop motif. For example, inserted D loop nucleotide G20b forms a platform on which the innermost A45-U48.1 base pair of the variable stem stacks in $tRNA^{Ser}$, while A20b of $tRNA^{Tyr}$ pairs with U48.2; the unpaired U48.1 serves as an inner platform in that tRNA (*17*, *77*, *78*).

Cocrystal structures of SerRS, TyrRS, and LeuRS that aminoacylate class II tRNAs show that the enzymes recognize the unique core domain shape arising from the large stem−loop variable region. In each case, structural elements of the enzyme bind between the long variable arm and the surrounding portion of the core domain (*17*, *77*, *78*). Because these enzymes recognize the junction between the variable arm and the core, they can be sensitive to differences in the orientations of the variable arm among the three class II tRNAs. Mutagenesis experiments have been performed that support the notion of shape recognition for each of the three enzymes. For example, studies of the *E. coli*, *S. cerevisiae*, and human SerRS enzymes have shown that mutation of variable stem−loop nucleotides in noncognate $tRNA^{Tyr}$, $tRNA^{Leu}$, or $tRNA^{Val}$ isoacceptors confers serylation activity (*79*−*82*). Similarly, alterations in the number of unpaired nucleotides at the 5′- and 3′-sides of the variable stem in $tRNA^{Tyr}$ conferred enhanced serylation activity, while decreasing tyrosylation activity (*79*). Finally, alteration of tertiary core nucleotides, including those within the variable stem−loop motif, allowed efficient aminoacylation of both

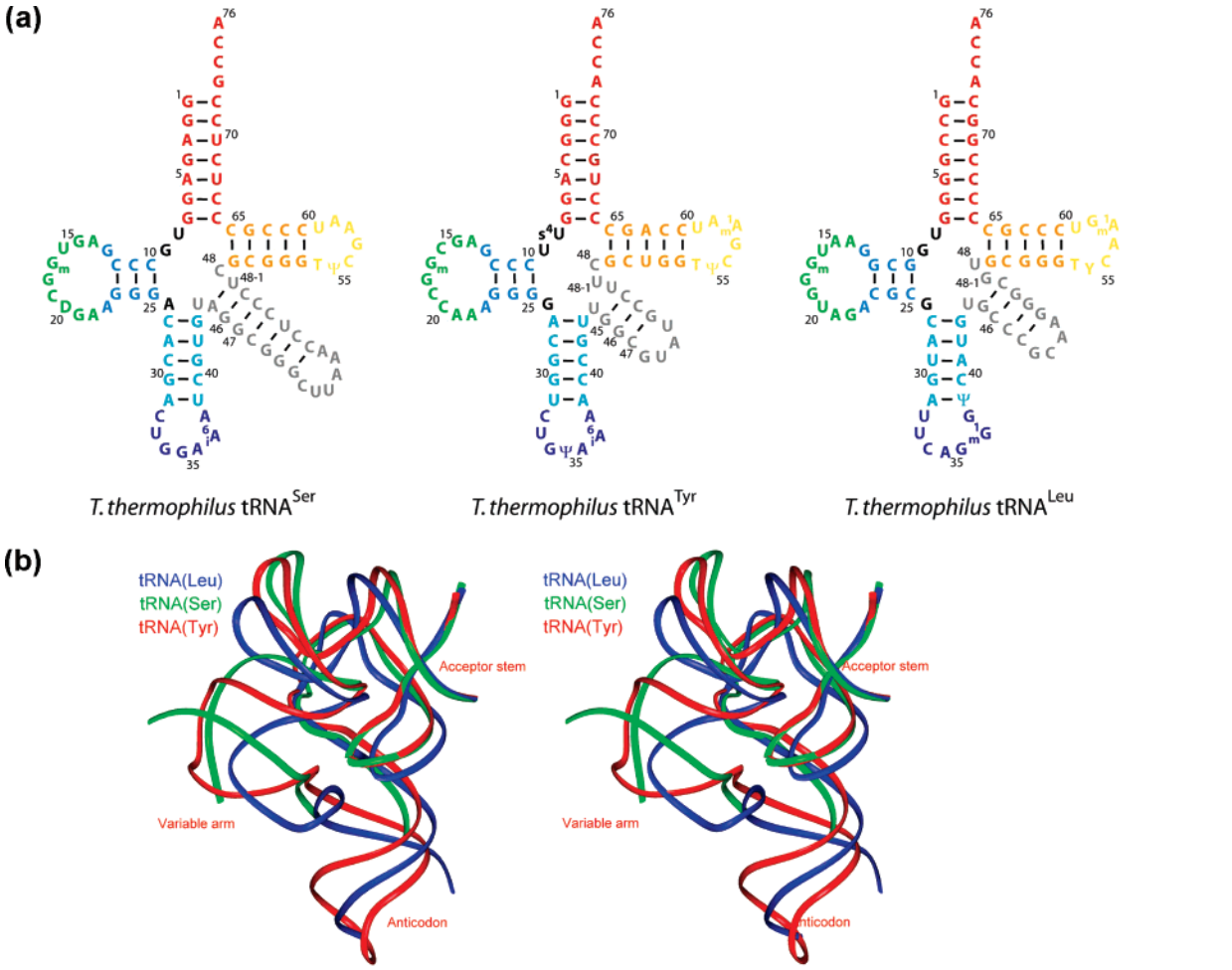

FIGURE 6: (a) Secondary structures of *T. thermophilus* $tRNA^{Ser}$, $tRNA^{Tyr}$, and $tRNA^{Leu}$, color-coded as in Figure 1. Nucleotides in the D and variable loops create the detailed architecture of the junction connecting the variable arm with the core of the tRNA, giving rise to distinct spatial orientations of the variable arm in each case. (b) Divergent stereodrawing depicting a superposition of the phosphate backbones of the three tRNAs depicted in panel a, using base pairs in the bottom portion of the acceptor stem for the superposition. The following coordinates were used: SerRS−$tRNA^{Ser}$ complex (PDB entry 1SER, TyrRS−$tRNA^{Tyr}$ complex (PDB entry 1H3E), and LeuRS−$tRNA^{Leu}$ complex (PDB entry 2BYT). Adapted from ref *17*.

$tRNA^{Ser}$ and $tRNA^{Tyr}$ isoacceptors by LeuRS (*59*, *60*). However, identifying which elements of the tertiary tRNA configuration are essential to synthetase selectivity and which are not is not straightforward (*17*). Experiments combining mutations in the enzyme and tRNA as described above for CysRS, as well as the study of backbone-modified substrates, should be informative with respect to further elucidating the shape selection problem.

Perspectives

Some distinct perspectives emerge from organizing the extensive literature addressing tRNA−aminoacyl-tRNA synthetase interactions, within the context of how the sequence information in the tRNA molecule is read by the enzymes. With the exception of the three isoaccepting sets that possess long variable arms, all tRNAs possess very similar tertiary structures. The numbers of base pair (and base triple) stacks in the vertical and horizontal arms are identical, while the number of single-stranded nucleotides varies minimally only in the D and small variable loops. The detailed tertiary conformational differences among the molecules then arise from sequence-dependent effects on the detailed positions of the sugar−phosphate backbone moieties. This very high degree of structural conservation permits an extension of the indirect readout concept from its common application in analyzing protein selectivity among double-helical nucleic acids, to include as well the discrimination among homologous single-stranded and globally structured portions of the molecules.

It is reasonable to identify indirect readout as a mechanism for selectivity when a reliable set of X-ray structures reveals local interactions with only the sugar−phosphate backbone (or nonspecific portions of the bases), while mutational studies show that the base sequence in this particular region of the tRNA−enzyme interface influences binding or catalysis. As described for the *E. coli* CysRS system, it also is possible to elucidate specific sugar−phosphate contacts that provide specificity via an analysis that combines mutations in both the enzyme and tRNA. Indirect readout may also (and probably does) function in localized structural contexts where the enzyme makes direct contact with specific moieties that distinguish the bases, although distinguishing the respective quantitative contributions of the two mechanisms is very challenging (*9*). Further, while we have included the well-studied examples of G-U wobble recognition in $tRNA^{Ala}$ and G^{-1}-C73 recognition in $tRNA^{His}$ as likely contexts where indirect readout functions, it is clear that taking these systems further will require X-ray structures of relevant functional

complexes to confirm the hypotheses based on functional data.

To date, there is no tRNA synthetase complex in which thorough and quantitative information describing how sugar−phosphate backbone contacts provide specificity is available. While a detailed catalog of the quantitative importance of specific protein−RNA backbone contacts to binding and catalysis is an important goal, understanding the role of indirect readout in its biological context requires further that the base sequence be mutated and the role of each specific contact assessed again on the variant tRNA. X-ray structures of the mutant complexes are also essential to understanding how the pattern of protein−RNA contacts might change. Carrying out such studies with the application of pre-steady state kinetic techniques will be important in distinguishing how indirect readout affects individual binding and catalytic steps (*75*−*78*). Such information is much more readily correlated with structural parameters than are the more commonly made measurements of steady state catalytic parameters.

The deeper understanding of tRNA synthetase−tRNA selectivity that will emerge from such studies is a satisfying and important acheivement in its own right. It might also provide a basis for protein engineering of the complexes to more practical ends and provide a model for studies on complex systems such as the ribosome. Another set of model protein−RNA systems that function by indirect readout consists of the EF-Tu−tRNA complexes, where protein discrimination across the cohelical tRNA acceptor and T stems is achieved primarily by contacts with the sugar−phosphate backbone (*83*). In these systems, a 10^3-fold variation in tRNA binding affinity, achieved in significant part by indirect readout, is compensated thermodynamically by an inverse set of affinities for the amino acid portions of the aminoacyl-tRNAs. On the ribosome, recent results suggest that the noncanonical base pair formed between nucleotides 32 and 38 in the anticodon loop may function to tune the affinity of tRNA to the ribosomal A site (*84*, *85*). If the different 32-38 pairs modulate affinity by conferring conformational preferences to the anticodon loop, which appears likely, this would demonstrate a role for indirect readout directly in decoding. The importance of tRNA conformational selection would appear even greater in other aspects of ribosomal function, such as translocation. As noted previously, it may be likely that no nucleotide within a tRNA is entirely random but rather has been selected, in combination with neighboring nucleotides, to fulfill one or more of the molecule's functions (*86*). Increasingly, it appears that sequence-dependent modulation of detailed backbone conformation is the underlying structural key that will fully unlock the biological role of transfer RNA.

ACKNOWLEDGMENT

We thank Dr. Marcel Dupasquier for assistance in the preparation of Figures 1, 5, and 6.

REFERENCES

1. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins, *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
2. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution, *Nature 335*, 321−329.
3. Koudelka, G. B., Harrison, S. C., and Ptashne, M. (1987) Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro, *Nature 326*, 886−888.
4. Koudelka, G. B., Mauro, S. A., and Ciubotaru, M. (2006) Indirect readout of DNA sequence by proteins: The roles of DNA sequence-dependent intrinsic and extrinsic forces, *Prog. Nucleic Acid Res. Mol. Biol. 81*, 143−177.
5. Williamson, J. R. (2000) Induced fit in RNA-protein recognition, *Nat. Struct. Biol. 7*, 834−837.
6. Post, C. B., and Ray, W. J., Jr. (1995) Reexamination of induced fit as a determinant of substrate specificity in enzymatic reactions, *Biochemistry 34*, 15881−15885.
7. Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state, *Biopolymers 44*, 153−180.
8. Kostrewa, D., and Winkler, F. K. (1995) Mg^{2+} binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Å resolution, *Biochemistry 34*, 683−696.
9. Martin, A. M., Sam, M. D., Reich, N. O., and Perona, J. J. (1999) Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease, *Nat. Struct. Biol. 6*, 269−277.
10. Horton, N. C., Dorner, L. F., and Perona, J. J. (2002) Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation, *Nat. Struct. Biol. 9*, 42−47.
11. Hobza, P., and Sponer, J. (1999) Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations, *Chem. Rev. 99*, 3247−3276.
12. Ibba, M., Francklyn, C., and Cusack, S. (2005) *The Aminoacyl-tRNA Synthetases*, Landes Bioscience, Georgetown, TX.
13. Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res. 26*, 5017−5035.
14. Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA Synthesis, *Annu. Rev. Biochem. 69*, 617−650.
15. Hauenstein, S., Zhang, C. M., Hou, Y. M., and Perona, J. J. (2004) Shape-selective RNA recognition by cysteinyl-tRNA synthetase, *Nat. Struct. Mol. Biol. 11*, 1134−1141.
16. Fukunaga, R., and Yokoyama, S. (2005) Aminoacylation complex structures of leucyl-tRNA synthetase and tRNA(Leu) reveal two modes of discriminator-base recognition, *Nat. Struct. Mol. Biol. 12*, 915−922.
17. 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, *Nat. Struct. Mol. Biol. 12*, 923−930.
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. Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Soll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation, *Proc. Natl. Acad. Sci. U.S.A. 104*, 11268−11273.
20. Kamtekar, S., Hohn, M. J., Park, H. S., Schnitzbauer, M., Sauerwald, A., Soll, D., and Steitz, T. A. (2007) Toward understanding phosphoseryl-tRNACys formation: The crystal structure of *Methanococcus maripaludis* phosphoseryl-tRNA synthetase, *Proc. Natl. Acad. Sci. U.S.A. 104*, 2620−2625.
21. Fukunaga, R., and Yokoyama, S. (2007) Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea, *Nat. Struct. Mol. Biol. 14*, 272−279.
22. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., and Rich, A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science 185*, 435−440.
23. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution, *Nature 250*, 546−551.
24. Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J. C., and Moras, D. (1999) Synthesis of aspartyl-tRNA(Asp) in *Escherichia coli*: A snapshot of the second step, *EMBO J. 18*, 6532−6541.
25. 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.
26. Putz, J., Puglisi, J. D., Florentz, C., and Giege, R. (1991) Identity elements for specific aminoacylation of yeast tRNA^{Asp} by cognate aspartyl-tRNA synthetase, *Science 252*, 1696−1699.
27. Nameki, N., Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992) *Escherichia coli* tRNA(Asp) recognition mechanism differing from that of the yeast system, *Biochem. Biophys. Res. Commun. 189*, 856−862.
28. Moulinier, L., Eiler, S., Eriani, G., Gangloff, J., Thierry, J. C., Gabriel, K., McClain, W. H., and Moras, D. (2001) The structure of an AspRS-tRNA^{Asp} complex reveals a tRNA-dependent control mechanism, *EMBO J. 20*, 5290−5301.
29. Ramos, A., and Varani, G. (1997) Structure of the acceptor stem of *Escherichia coli* tRNA Ala: Role of the G3•U70 base pair in synthetase recognition, *Nucleic Acids Res. 25*, 2083−2090.
30. Rees, B., Webster, G., Delarue, M., Boeglin, M., and Moras, D. (2000) Aspartyl tRNA-synthetase from *Escherichia coli*: Flexibility and adaptability to the substrates, *J. Mol. Biol. 299*, 1157−1164.
31. 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.
32. 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.
33. Francklyn, C., and Schimmel, P. (1989) Aminoacylation of RNA minihelices with alanine, *Nature 337*, 478−481.
34. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., and Schimmel, P. (1991) Specificity for aminoacylation of an RNA helix: An unpaired, exocyclic amino group in the minor groove, *Science 253*, 784−786.
35. Musier-Forsyth, K., Shi, J. P., Henderson, B., Bald, R., Furste, J. P., Erdmann, V. A., and Schimmel, P. (1995) Base-Analog-Induced Aminoacylation of an RNA Helix by a tRNA Synthetase, *J. Am. Chem. Soc. 117*, 7253−7254.
36. Gabriel, K., Schneider, J., and McClain, W. H. (1996) Functional evidence for indirect recognition of G•U in tRNA^{Ala} by alanyl-tRNA synthetase, *Science 271*, 195−197.
37. Chang, K. Y., Varani, G., Bhattacharya, S., Choi, H., and McClain, W. H. (1999) Correlation of deformability at a tRNA recognition site and aminoacylation specificity, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11764−11769.
38. Musier-Forsyth, K., and Schimmel, P. (1992) Functional contacts of a transfer RNA synthetase with 2′-hydroxyl groups in the RNA minor groove, *Nature 357*, 513−515.
39. McClain, W. H., and Foss, K. (1988) Changing the acceptor identity of a transfer RNA by altering nucleotides in a "variable pocket", *Science 241*, 1804−1807.
40. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA^{Gln} and ATP at 2.8 Å resolution, *Science 246*, 1135−1142.
41. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase, *Nature 352*, 213−218.
42. Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., and Giege, R. (1980) Crystal structure of yeast tRNA^{Asp}, *Nature 288*, 669−674.
43. Jahn, M., Rogers, M. J., and Soll, D. (1991) Anticodon and acceptor stem nucleotides in tRNA^{Gln} are major recognition elements for *E. coli* glutaminyl-tRNA synthetase, *Nature 352*, 258−260.
44. Ghysen, A., and Celis, J. E. (1974) Mischarging single and double mutants of *Escherichia coli* sup3 tyrosine transfer RNA, *J. Mol. Biol. 83*, 333−351.
45. Smith, J. D., and Celis, J. E. (1973) Mutant tyrosine transfer RNA that can be charged with glutamine, *Nat. New Biol. 243*, 66−71.
46. Sprinzl, M., and Vassilenko, K. S. (2005) Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res. 33*, D139−D140.
47. Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K., Miura, K., and Shimizu, M. (1989) Role of the extra G-C pair at the end of the acceptor stem of tRNA^{His} in aminoacylation, *Nucleic Acids Res. 17*, 7855−7863.
48. Francklyn, C., and Schimmel, P. (1990) Enzymatic aminoacylation of an eight-base-pair microhelix with histidine, *Proc. Natl. Acad. Sci. U.S.A. 87*, 8655−8659.
49. Francklyn, C., Shi, J. P., and Schimmel, P. (1992) Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices, *Science 255*, 1121−1125.
50. Guth, E. C., and Francklyn, C. S. (2007) Kinetic discrimination of tRNA identity by the conserved motif 2 loop of a class II aminoacyl-tRNA synthetase, *Mol. Cell 25*, 531−542.
51. Rosen, A. E., and Musier-Forsyth, K. (2004) Recognition of G-1: C73 atomic groups by *Escherichia coli* histidyl-tRNA synthetase, *J. Am. Chem. Soc. 126*, 64−65.
52. Rosen, A. E., Brooks, B. S., Guth, E., Francklyn, C. S., and Musier-Forsyth, K. (2006) Evolutionary conservation of a functionally important backbone phosphate group critical for aminoacylation of histidine tRNAs, *RNA 12*, 1315−1322.
53. Fersht, A. R. (1998) Sieves in sequence, *Science 280*, 541.
54. 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.
55. Nureki, O., Vassylyev, D. 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.
56. Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue, *EMBO J. 19*, 2351−2361.
57. 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.
58. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C., and Moras, D. (2000) Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem, *Cell 103*, 877−884.
59. Asahara, H., Himeno, H., Tamura, K., Hasegawa, T., Watanabe, K., and Shimizu, M. (1993) Recognition nucleotides of *Escherichia coli* tRNA^{Leu} and its elements facilitating discrimination from tRNASer and tRNA^{Tyr}, *J. Mol. Biol. 231*, 219−229.
60. Tocchini-Valentini, G., Saks, M. E., and Abelson, J. (2000) tRNA leucine identity and recognition sets, *J. Mol. Biol. 298*, 779−793.
61. Tardif, K. D., and Horowitz, J. (2002) Transfer RNA determinants for translational editing by *Escherichia coli* valyl-tRNA synthetase, *Nucleic Acids Res. 30*, 2538−2545.
62. Fukunaga, R., and Yokoyama, S. (2005) Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation, *J. Mol. Biol. 346*, 57−71.
63. Limmer, S., Hofmann, H. P., Ott, G., and Sprinzl, M. (1993) The 3′-terminal end (NCCA) of tRNA determines the structure and stability of the aminoacyl acceptor stem, *Proc. Natl. Acad. Sci. U.S.A. 90*, 6199−6202.
64. Recht, M. I., and Williamson, J. R. (2004) RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex, *J. Mol. Biol. 344*, 395−407.
65. Hendrix, D. K., Brenner, S. E., and Holbrook, S. R. (2005) RNA structural motifs: Building blocks of a modular biomolecule, *Q. Rev. Biophys. 38*, 221−243.
66. Stefl, R., Skrisovska, L., and Allain, F. H. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle, *EMBO Rep. 6*, 33−38.
67. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons, *Nature 430*, 45−50.
68. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme, *Nature 395*, 567−574.
69. Klein, D. J., and Ferre-D'Amare, A. R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate, *Science 313*, 1752−1756.
70. Sherlin, L. D., Bullock, T. L., Newberry, K. J., Lipman, R. S., Hou, Y. M., Beijer, B., Sproat, B. S., and Perona, J. J. (2000) Influence of transfer RNA tertiary structure on aminoacylation efficiency by glutaminyl and cysteinyl-tRNA synthetases, *J. Mol. Biol. 299*, 431−446.
71. Hou, Y. M., Westhof, E., and Giege, R. (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA, *Proc. Natl. Acad. Sci. U.S.A. 90*, 6776−6780.

72. Hou, Y. M. (1994) Structural elements that contribute to an unusual tertiary interaction in a transfer RNA, *Biochemistry 33*, 4677−4681.
73. Hamann, C. S., and Hou, Y. M. (1997) A strategy of tRNA recognition that includes determinants of RNA structure, *Bioorg. Med. Chem. 5*, 1011−1019.
74. Hamann, C. S., and Hou, Y. M. (2000) Probing a tRNA core that contributes to aminoacylation, *J. Mol. Biol. 295*, 777−789.
75. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L., and Soll, D. (2000) The adaptor hypothesis revisited, *Trends Biochem. Sci. 25*, 311−316.
76. Oshikane, H., Sheppard, K., Fukai, S., Nakamura, Y., Ishitani, R., Numata, T., Sherrer, R. L., Feng, L., Schmitt, E., Panvert, M., Blanquet, S., Mechulam, Y., Soll, D., and Nureki, O. (2006) Structural basis of RNA-dependent recruitment of glutamine to the genetic code, *Science 312*, 1950−1954.
77. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$, *Science 263*, 1404−1410.
78. Yaremchuk, A., Kriklivyi, I., Tukalo, M., and Cusack, S. (2002) Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition, *EMBO J. 21*, 3829−3840.
79. Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K., and Shimizu, M. (1990) Conversion of aminoacylation specificity from $tRNA^{Tyr}$ to $tRNA^{Ser}$ in vitro, *Nucleic Acids Res. 18*, 6815−6819.
80. Achsel, T., and Gross, H. J. (1993) Identity determinants of human tRNA(Ser): Sequence elements necessary for serylation and maturation of a tRNA with a long extra arm, *EMBO J. 12*, 3333−3338.
81. Himeno, H., Yoshida, S., Soma, A., and Nishikawa, K. (1997) Only one nucleotide insertion to the long variable arm confers an efficient serine acceptor activity upon *Saccharomyces cerevisiae* $tRNA^{Leu}$ in vitro, *J. Mol. Biol. 268*, 704−711.
82. Asahara, H., Himeno, H., Tamura, K., Nameki, N., Hasegawa, T., and Shimizu, M. (1994) *Escherichia coli* seryl-tRNA synthetase recognizes $tRNA^{Ser}$ by its characteristic tertiary structure, *J. Mol. Biol. 236*, 738−748.
83. Asahara, H., and Uhlenbeck, O. C. (2005) Predicting the binding affinities of misacylated tRNAs for *Thermus thermophilus* EF-Tu·GTP, *Biochemistry 44*, 11254−11261.
84. Olejniczak, M., Dale, T., Fahlman, R. P., and Uhlenbeck, O. C. (2005) Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding, *Nat. Struct. Mol. Biol. 12*, 788−793.
85. Olejniczak, M., and Uhlenbeck, O. C. (2006) tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition, *Biochimie 88*, 943−950.
86. Frugier, M., Helm, M., Felden, B., Giege, R., and Florentz, C. (1998) Sequences outside recognition sets are not neutral for tRNA aminoacylation. Evidence for nonpermissive combinations of nucleotides in the acceptor stem of yeast tRNAPhe, *J. Biol. Chem. 273*, 11605−11610.

BI7014647