Structure of *E. coli* Glutaminyl-tRNA Synthetase Complexed with $tRNA^{Gln}$ and ATP at 2.8 Å Resolution

MARK A. ROULD, JOHN J. PERONA, DIETER SÖLL, THOMAS A. STEITZ

The crystal structure of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS) complexed with its cognate glutaminyl transfer RNA ($tRNA^{Gln}$) and adenosine triphosphate (ATP) has been derived from a 2.8 angstrom resolution electron density map and the known protein and tRNA sequences. The 63.4-kilodalton monomeric enzyme consists of four domains arranged to give an elongated molecule with an axial ratio greater than 3 to 1. Its interactions with the tRNA extend from the anticodon to the acceptor stem along the entire inside of the L of the tRNA. The complexed tRNA retains the overall conformation of the yeast phenylalanine tRNA ($tRNA^{Phe}$) with two major differences: the 3′ acceptor strand of $tRNA^{Gln}$ makes a hairpin turn toward the inside of the L, with the disruption of the final base pair of the acceptor stem, and the anticodon loop adopts a conformation not seen in any of the previously determined tRNA structures. Specific recognition elements identified so far include (i) enzyme contacts with the 2-amino groups of guanine via the tRNA minor groove in the acceptor stem at G2 and G3; (ii) interactions between the enzyme and the anticodon nucleotides; and (iii) the ability of the nucleotides G73 and U1 · A72 of the cognate tRNA to assume a conformation stabilized by the protein at a lower free energy cost than noncognate sequences. The central domain of this synthetase binds ATP, glutamine, and the acceptor end of the tRNA as well as making specific interactions with the acceptor stem. It is structurally similar to the dinucleotide binding motifs of the tyrosyl- and methionyl-tRNA synthetases, suggesting that all synthetases may have evolved from a common domain that can recognize the acceptor stem of the cognate tRNA.

AMINOACYL–TRANSFER RNA SYNTHETASES ARE THE ENzymes responsible for translating the genetic code. Each synthetase must recognize its cognate tRNA's, discriminating against all others, and attach its corresponding amino acid to the acceptor end of the tRNA in a reaction driven by the hydrolysis of ATP. Although the three bases of the anticodon completely embody the essential recognition element at the time of protein synthesis, recognition of the tRNA during aminoacylation frequently involves other tRNA features; in fact, some synthetases may not recognize the anticodon at all. The desire to understand more completely these recognition-discrimination features by direct structural methods gave impetus to the x-ray crystallographic study presented in this article.

Molecular, genetic, and biochemical studies of the specificity of charging of tRNA's have shown that a relatively small number of nucleotides, termed identity elements, can comprise fully the basis of proper tRNA selection by synthetases (*1–3*). Both by the study of mischarging of mutant suppressor tRNA's in vivo and the charging of T7 tRNA transcripts in vitro, some of the identity elements in tRNA's have become clear. By these methods, U35 in the anticodon (*4–6*), G73 (*7–9*) and base pair 1 · 72 (*8, 9*) in the acceptor stem have been implicated as identity elements in *Escherichia coli* glutaminyl-tRNA ($tRNA^{Gln}$). We address here the structural basis for recognition of these identity elements in the tRNA by GlnRS and propose other potential recognition elements implied by the crystal structure of the synthetase-tRNA complex.

The crystal structures of two aminoacyl-tRNA synthetases, those of *Bacillus stearothermophilus* tyrosyl-tRNA synthetase (TyrRS) (*10, 11*) and *E. coli* methionyl-tRNA synthetase (MetRS) (*12, 13*), have been determined as complexes with either adenosine triphosphate (ATP) or amino acid substrates but lacking their cognate tRNA's. Comparison of the structures of these two enzymes (*14*) revealed a similar five-stranded parallel β sheet of a dinucleotide binding motif that had previously been observed in virtually all dehydrogenase and kinase structures (*15*). A β-α-β motif that is responsible for many of the interactions with ATP has a nearly identical structure in these two enzymes.

The crystal structures of $tRNA^{Phe}$ (*16, 17*), $tRNA^{Asp}$ (*18*), and $tRNA^{Met}$ (*19, 20*) exhibit a very similar L-shaped structure. While some differences exist in the conformation of the acceptor end, the anticodon, and in the overall bend between the two halves, the overall similarity of these structures suggests that all tRNA's have approximately the same structure in solution, as their functioning in protein synthesis requires. This makes the question of how a

The authors are in the Department of Molecular Biophysics and Biochemistry, Department of Chemistry, and Howard Hughes Medical Institute, Yale University, New Haven, CT 06511.

synthetase discriminates between cognate and noncognate species an interesting challenge.

The *E. coli* GlnRS is a 63,400-dalton monomer of 553 amino acids whose gene has been cloned, sequenced, and overexpressed by Söll and co-workers (*21, 22*) as has the $tRNA^{Gln}$ isoacceptor 2 (*23*). The ability to make large quantities of both the GlnRS and its cognate tRNA have made possible their cocrystallization with ATP (*23*).

We now present the crystal structure of this ternary complex derived from a 2.8 Å resolution electron density map. We find differences in the structure of bound $tRNA^{Gln}$ from its presumed solution structure, specific protein-RNA interactions that might account for tRNA selectivity and similarities in the structure of the GlnRS enzyme to those of the TyrRS and MetRS enzymes. In general, it appears that recognition of the tRNA arises from direct hydrogen-bonding interactions between the protein and bases of the tRNA and the requirement for tRNA conformational changes whose free energy costs are sequence-dependent.

Structure determination. Crystals of the complex were obtained as described (*23*) except that 2.0*M* ammonium sulfate was used as precipitant instead of sodium citrate. Crystals are in the orthorhombic space group $C222_1$, with cell dimensions $a = 242.8$, $b = 93.6$, and $c = 115.7$ Å. The asymmetric unit contains a single monomeric complex resulting in a solvent content of 70 percent by volume. X-ray diffraction intensities for the native crystal were collected on a Xuong-Hamlin multiwire area detector with which we used the University of California at San Diego data collection and reduction software. An overall *R* factor of 8.1 percent on intensity was calculated for merging 152,900 measurements to 31,768 unique reflections in the resolution range of 18 to 2.8 Å, giving an *R* factor between final Friedel mates of 4.4 percent.

Four heavy atom derivatives (Table 1) provided multiple isomorphous replacement (MIR) phases suitable for identification of the tRNA and the molecular boundaries. Three mercurial derivatives were prepared by crystallizing the GlnRS-tRNA complex previously reacted with stoichiometric amounts of the mercurial. Mercury derivatives used in the structure determination contained a 3 to 1 molar ratio of ethylmercury phosphate to complex, 5 to 1 parachloromercuribenzoate, and 3 to 1 dimercury acetate. Lower molar ratios resulted in smaller isomorphous differences; higher molar ratios prevented crystallization. The more common method of soaking previously formed native crystals in low concentrations of various mercurials cracks the crystals, presumably by reaction with one or more of the weaker affinity sites (there are ten cysteine residues in the GlnRS), which subsequently disrupts the protein structure or crystal packing.

Since the heavy atom derivatives share most of the heavy atom binding sites, but with varying relative occupancies of those sites, it was necessary to decouple refinement of the heavy atom parameters from the parent MIR phases generated from the major heavy atom sites. This was accomplished by solvent flattening (*24*) the electron density map generated with these phases, extracting the new "solvent flattened" phases, and refining the heavy atom parameters against these improved phases without updating the phases during refinement (*25*). New MIR phases were calculated and the process was repeated. Three rounds of solvent flattening and parameter refinement resulted in convergence of the heavy atom parameters and identification of the minor heavy atom sites. The final mean figure of merit for the MIR phases was 0.58 to 2.8 Å. At this point we were able to unambiguously fit synthetase residues 6 to 353, 364 to 391, and 456 to 547, the tRNA, and the ATP into the solvent flattened electron density map (Fig. 1, A and B). The course of the polypeptide backbone was clearly defined by the electron density for most of the remaining residues, comprising part of the distal β barrel; but their side chains were too disordered to build with confidence. The crystallographic *R* factor for this initial model, lacking residues 347 to 455, was 38.7 percent for all data from 18 to 2.8 Å resolution. Preliminary refinement of the complex with X-PLOR (*26*), including molecular dynamics simulated annealing, allowed approximate fitting of the main chain for most of the remaining residues, and reduced the R factor to 27.9 percent for all data between 7.0 and 2.8 Å, with good stereochemistry. Further refinement should provide better definition of the interactions at the anticodon, reliable values for specific interaction distances, and clarification of the involvement of water and magnesium ions.

Structure of $tRNA^{Gln}$ differs from that of uncomplexed yeast $tRNA^{Phe}$. A comparison of the phosphodiester backbone structure of $tRNA^{Gln}$ in this complex with that of uncomplexed yeast $tRNA^{Phe}$ shows several major differences, the most dramatic of which is in the acceptor stem and 3′ end (Fig. 2). Although there is no direct evidence, we assume that the overall structures of *E. coli* $tRNA^{Gln}$ and yeast $tRNA^{Phe}$ are very similar in solution and therefore that the major differences between these two structures arise as a consequence of the formation of a complex with the enzyme. The most striking difference in the $tRNA^{Gln}$ structure is that the base pair between nucleotides 1 and 72 is disrupted and the 3′ CCA end of the tRNA forms a hairpin in the direction of the anticodon rather than continuing on from the acceptor stem in a helical fashion. The bases of A76, C75, and G73 are stacked on each other and the nucleotide C74 is looped out with its base interacting with a complementary pocket in the protein (Fig. 1A). This alternative structure of the 3′ end is stabilized in part by a hydrogen bond between the 2-amino group of G73 and the phosphate group of residue A72 and in part by numerous interactions of the sugar-phosphate backbone and the base of nucleotide C74 with the protein. This results in the 3′ terminus plunging into a deep protein pocket that contains the binding sites for the other two substrates, ATP and glutamine.

A second major difference between these two tRNA molecules involves the conformation of the anticodon loop and stem. In

Table 1. Heavy atom derivative statistics. The constellation of heavy atom binding sites for the four derivatives, their percent occupancies normalized to the most highly substituted site, and their phasing power (f_h/E), where f_h is the root-mean-square (rms) heavy atom structure factor and *E* is the rms lack of closure error. DiHgAc, $EtHgPO_4$, and PCMB are dimercuryacetate, ethylmercury phosphate, and parachloromercury benzoate, respectively. Two conformations for most of the metal-liganded cysteines were seen, primarily as rotomers about chi. Cys^{48} and Cys^{310} appear to be involved in a disulfide linkage in the electron density map, but must be reduced to a large extent in order for the heavy metals to bind. The low- and high-resolution $EtHgPO_4$ and PCMB data sets were collected at different times and kept as separate data sets during refinement; each high-resolution data set required several crystals.

Derivative	Resolution (Å)	f_h/E	Occupancies					
			Cys^{206}	Cys^{48} Cys^{310}	Cys^{525}	Cys^{229}	Cys^{330}	Cys^{61}
K_2AuCl_4	18–3.5	0.93	67 64	19 17	43	8		
DiHgAc	18–3.0	1.65	100	85 7	30 12	34	10 27	
$EtHgPO_4$ No. 1	18–3.5	2.31	73	41 41	26	8	21	26
No. 2	8.5–2.8	1.51	61	24 24	22	6	18	20
PCMB No. 1	18–3.5	2.02	99	42	70	24	88	
No. 2	8.5–2.8	2.16	94	45	51	20	84	

$tRNA^{Gln}$ the base of U35 is stacked underneath that of A37 and the bases of C34 and G36 are unstacked and project outward to interact with groups on the protein. These stabilizing contacts with the protein presumably compensate for the free energy cost of unstacking the bases.

Structure and ligand binding of the GlnRS enzyme. The structure of GlnRS consists of four domains, one of which contains the active site and makes sequence-specific interactions with the acceptor stem of the tRNA while the other three appear to have roles in specific recognition of $tRNA^{Gln}$ and discrimination against noncognate tRNA's (Fig. 3B). The structure is unusually elongated, having an axial ratio greater than 3 to 1. The protein is about 100 Å long, a dimension that is significantly larger than any in the tRNA molecule. This elongated protein is seen to interact continuously with one side of the L-shaped $tRNA^{Gln}$ molecule from the anticodon loop to the acceptor end (Figs. 3A and 4). The region of interaction is similar to that hypothesized by Rich and Schimmel (*27*) to be characteristic of many synthetase-tRNA interactions. The active site of the enzyme lies at the bottom of a very deep pocket in which the ATP, glutamine, and the acceptor end of the tRNA bind.

Dinucleotide fold domain binds ATP, glutamine, and tRNA. The domain that is perhaps most pivotal to the functioning of this enzyme is at the amino terminus and consists of a five-stranded parallel β-pleated sheet folded in the manner known as the dinucleotide (or Rossmann) fold (*15*). This common structural motif was first observed in dehydrogenases and is found in kinases, synthetases, and many other proteins that utilize ATP, GTP (guanosine triphosphate), or other high-energy nucleotide analogues. As in the case of other dinucleotide folds, the domain in the GlnRS is folded into two symmetrically related halves. The first half consists of three β strands and two intervening α helices and provides many of the major contacts for the ATP substrate. The second half of the domain consists of two β strands and two α helices and appears to form important contacts with the glutamine substrate as well as to make a sequence-specific interaction with the acceptor stem of tRNA. The acceptor binding domain is inserted between the first and second halves of this dinucleotide fold (Fig. 3B).

The overall structure of this dinucleotide fold domain is similar to that found in the TyrRS (*10, 11*) and the MetRS (*12, 13*) enzymes as is its interaction with the amino acid and ATP substrates. Of the 114 α-carbons in the dinucleotide fold of GlnRS, as many as 89 can be superimposed on analogous α-carbons in the TyrRS with an rms difference of 1.76 Å. The length of the polypeptide inserted between these two half-domains is considerably smaller in the case of the TyrRS. Insertion of a domain between two halves of a dinucleotide fold has been found in the structure of yeast hexokinase (*28*) and is predicted to exist in the structure of the isoleucyl-tRNA synthetase (*29*).

ATP binds to a sequence in the first half of the dinucleotide fold that is highly conserved among several synthetases. The ATP is near the imidazole side chains of the conserved His^{40}-Ile^{41}-Gly^{42}-His^{43} sequence in a manner similar to that observed in crystals of TyrRS complexed with tyrosinyl adenylate. The specificity of the enzyme for ATP may be attributed to hydrogen bonding interactions observed between the N1 and N6 of ATP and the peptide backbone amide NH and carbonyl oxygen of residue 261. Similar interactions were not seen when the adenine moiety of tyrosinyl adenylate was bound to the TyrRS (*11*). The α phosphate is within hydrogen bonding distance of the 2′ hydroxyl group of A76 while the β and γ phosphates are adjacent to the imidazole groups of His^{43} and His^{40}, respectively. Another sequence that is conserved among a group of synthetases, Met-Ser-Lys, corresponds to residues 268 to 270 in GlnRS; Lys^{270} makes interactions with the phosphates of ATP. This

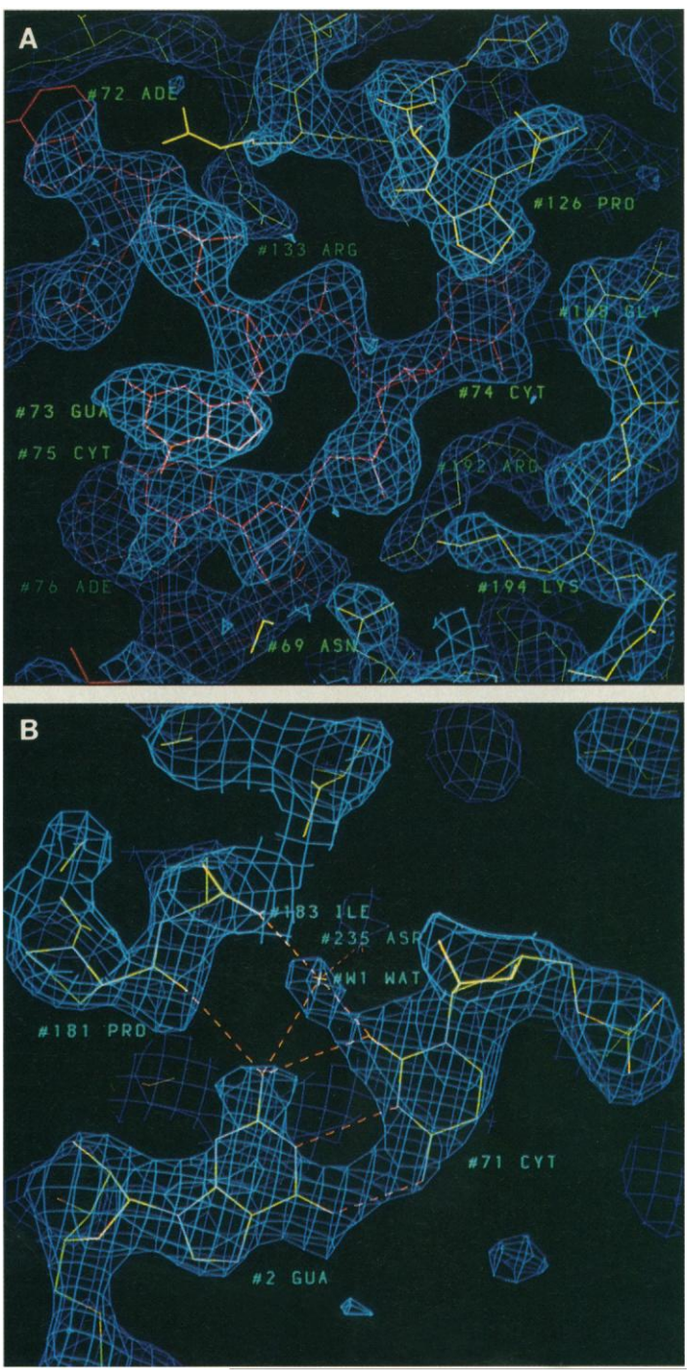


Fig. 1. (**A**) A section through the 2.8 Å resolution solvent flattened, MIR electron density map in the region of the hairpin acceptor strand, contoured at 2 σ. The protein model fitted to the map is yellow and the tRNA model is orange. The base of C74, with its exocyclic atoms clearly visible is looped out and fits snugly into a binding pocket formed by the enzyme. The bases of G73, C75, and A76 are seen to be stacked on each other with the 2-amino group of G73 well within bonding distance to the phosphate group of A72. (**B**) Electron density from the solvent flattened, MIR map around base pair $G_2 \cdot C_{71}$, contoured at 2 σ. A well-ordered buried water molecule (W1 WAT) is bound to the protein backbone amide of residue 183 and the carboxylate side chain of Asp^{235} (in background). The remaining hydrogen bond donor and acceptor from this water molecule together with the backbone carbonyl oxygen of Pro^{181} form a hydrogen-bonding surface that is specifically complementary to the G2 · C71 base pair in the minor groove.

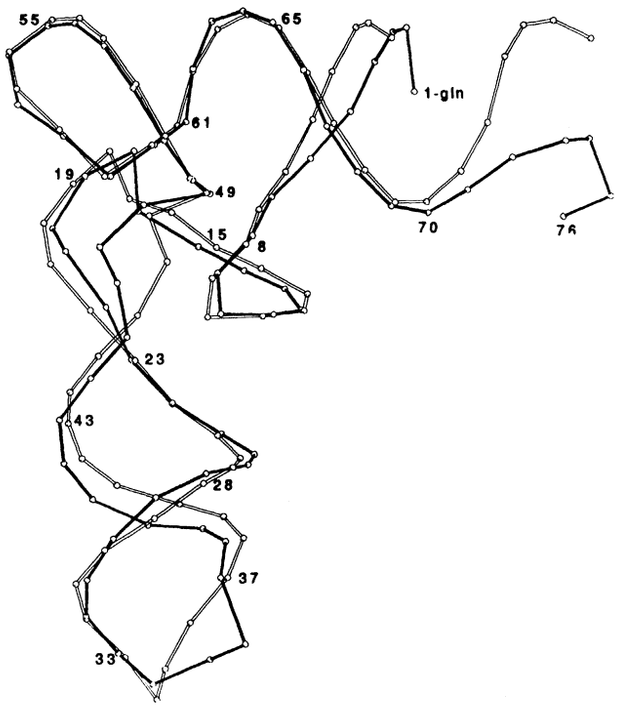


Fig. 2. Superposition of the phosphate backbone of uncomplexed yeast $tRNA^{Phe}$ on that of the $tRNA^{Gln}$ complexed with the synthetase. The corresponding phosphorus atoms of nucleotides 6 to 12, 23 to 31, 39 to 43, and 49 to 70 were used to determine the least-squares best superposition of the two molecules, giving an rms deviation of 1.68 Å for these 43 phosphorus atoms, and an rms deviation of 5.49 Å for all 75 of the corresponding phosphorus atoms. Major differences are seen in the acceptor strand conformation, the anticodon loop, and the width of the grooves of the acceptor stem and anticodon stem.

sequence and the His-Ile-Gly-His are simultaneously present in most of the synthetases having either one, which may point to a common essential role of these residues in catalysis (*30*). Indeed, Lys^{230} in the *B. stearothermophilus* TyrRS enzyme, hypothesized to be involved in transition state binding (*31*), occupies a position in a mobile loop of TyrRS that may allow it to play the analogous role to Lys^{270} of GlnRS. There is no interaction of the Met-Ser-Lys sequence with the 3′ end of the tRNA.

Although the mode of interaction of glutamine with the enzyme has not been delineated in detail, a difference electron density map of complex crystallized in the presence of glutamine and adenosine monophosphate (AMP) shows glutamine bound in a pocket formed in part by the second half of the dinucleotide fold.

Other GlnRS domains are involved in tRNA discrimination. A domain of 110 residues (Fig. 3B) is inserted between the two halves of the dinucleotide fold domain and consists of a five-stranded antiparallel β sheet flanked by three helices. Its primary role is to distort the acceptor end of the tRNA. This is achieved by binding an alternative conformation of the acceptor end that consists of a denatured terminal base pair of the stem and a hairpin curvature of the single-stranded 3′ end in the direction of the anticodon. An antiparallel β loop of protein inserts between the 3′ and 5′ strands of tRNA and prevents the formation of the U1 · A72 base pair. This domain also provides a binding pocket for the looped out base of C74. Pro^{126} at the end of α helix E is stacked on top of C74 and the aliphatic portions of Arg^{192} and Cys^{171} are packed underneath the ring. Two backbone carbonyl groups are hydrogen-bonded to the N4 of the cytosine.

The two protein domains at the carboxyl-terminal half of the molecule fold into six-stranded antiparallel β-barrel–like structures that interact with the anticodon loop of the tRNA. One barrel is directly adjacent to the dinucleotide fold domain. A single strand of this proximal barrel follows directly in amino acid sequence a long α

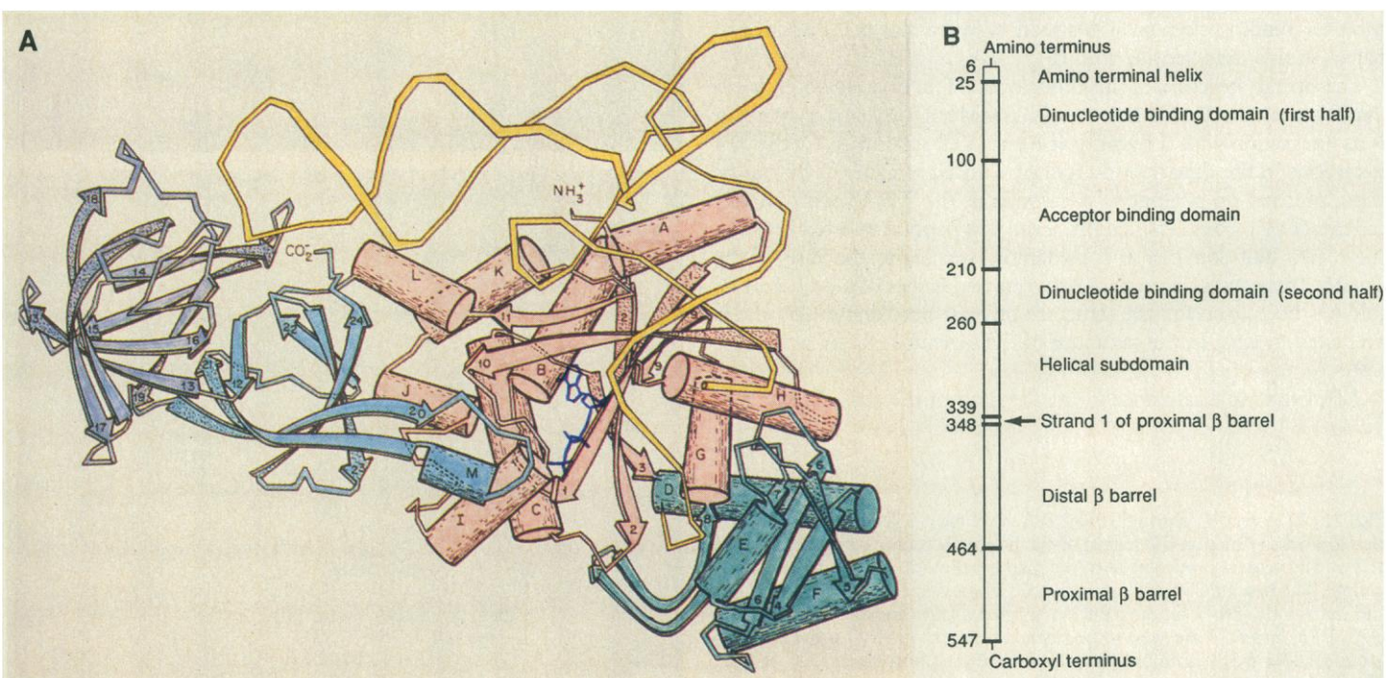


Fig. 3. (**A**) GlnRS complexed with $tRNA^{Gln}$ and ATP. For the protein, α helices are represented as tubes sequentially lettered and β strands as arrows sequentially numbered, both from the amino terminus. The dinucleotide fold domain includes residues from the β-strand 1 through β-strand 3 plus α-helix G through β-strand 10. The helical subdomain extends from α-helix I to α-helix L. The acceptor end binding domain includes the polypeptide chain between the amino end of α-helix D to the carboxyl end of β-strand 8. The two β-barrel anticodon binding domains consist of β-strand 13 to β-strand 19 for the distal domain and β-strand 20 to the carboxyl terminus plus β-strand 12 for the proximal domain. (**B**) The positions of the four protein domains and some other structural features mapped along the amino acid sequence.

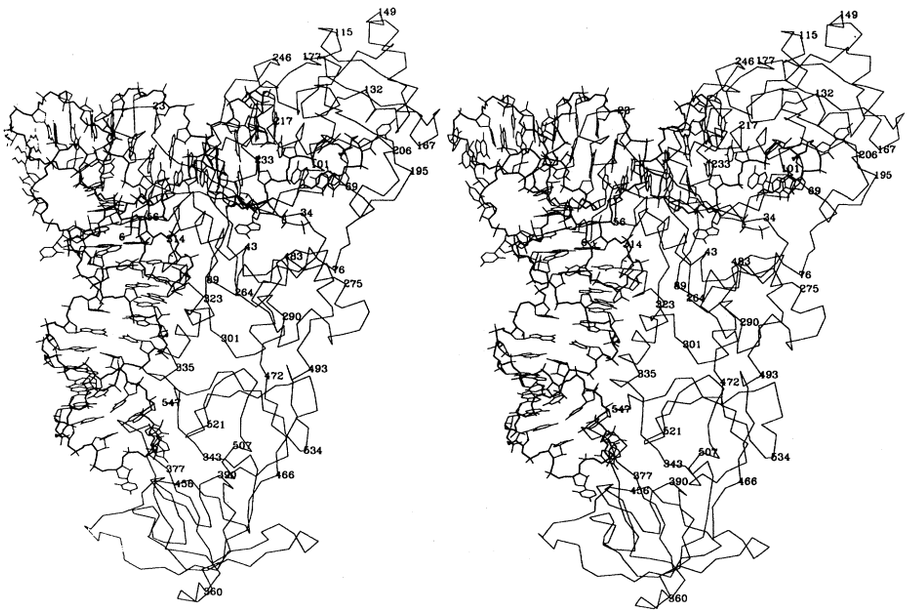


Fig. 4. Stereo drawing of the α-carbon backbone of GlnRS and all atoms of $tRNA^{Gln}$ and ATP. The first five residues of the amino terminus and the last six of the carboxyl terminus are disordered and not shown. While the secondary structural elements of one-half of the distal β-barrel domain (residues 392 to 455) are clear in the electron density map, their connectivity is not certain.

helix that runs alongside the D and anticodon stems of the tRNA and connects the dinucleotide fold to the carboxyl-terminal domains. The other five strands of the proximal barrel comprise the final 90 amino acids of the polypeptide chain. Inserted between the strands of the proximal barrel are sequences that make up the distal β barrel. This domain is the most distant from the acceptor end and is highly mobile in this crystal form.

tRNA discrimination. Of the very large number of interactions between the tRNA and GlnRS that are observed in this complex, only a small subset are likely to play an important role in discriminating among different tRNA's (Fig. 5). We find direct interactions of GlnRS with bases in single-stranded loops and in the minor groove of the duplex acceptor stem that are likely to be base-specific as well as nucleotides whose role in tRNA specificity is achieved through facilitation of an RNA conformation imposed by the protein (Fig. 6). Both in vivo and in vitro molecular genetic experiments, some from the early 1970's, have identified several important recognition elements in $tRNA^{Gln}$ both in the anticodon (U35) and at the acceptor end (G73, U1 · A72) of the molecule (*1–3*) that are entirely consistent with the structure of the complex. In addition to these previously identified elements, we observe that GlnRS is penetrating the minor groove of the acceptor stem and presenting a structural surface (including an important buried water molecule) that is complementary to base pairs 3 · 70 and 2 · 71. Other interactions that may (or may not) contribute to tRNA discrimination are seen between the N2 of G10 and the carboxylate of Glu^{323} as well as between C16 and Gln^{13}.

Discriminator base. Nucleotides at position 73 were termed "discriminator bases" by Crothers *et al.* (*32*), who found a correlation between the identity of the nucleotide at this position and the chemical nature of the amino acid encoded by the tRNA; for example, tRNA's with an A at position 73 tended to code for hydrophobic amino acids while G at this position correlated with polar amino acids. They suggested that the identity of the base at position 73 might be significant both in the evolution of tRNA specificity and in current recognition. It has been shown that a mutant of *E. coli supF* $tRNA^{Tyr}$ carrying an A73 to G73 mutation can insert glutamine in vivo (*7, 8, 33*). More recently, Seong, Lee, and RajBhandary have found in vitro that mutating a suppressor initiator tRNA from one containing an A at position 73 to one containing G73 results in an approximately five times higher

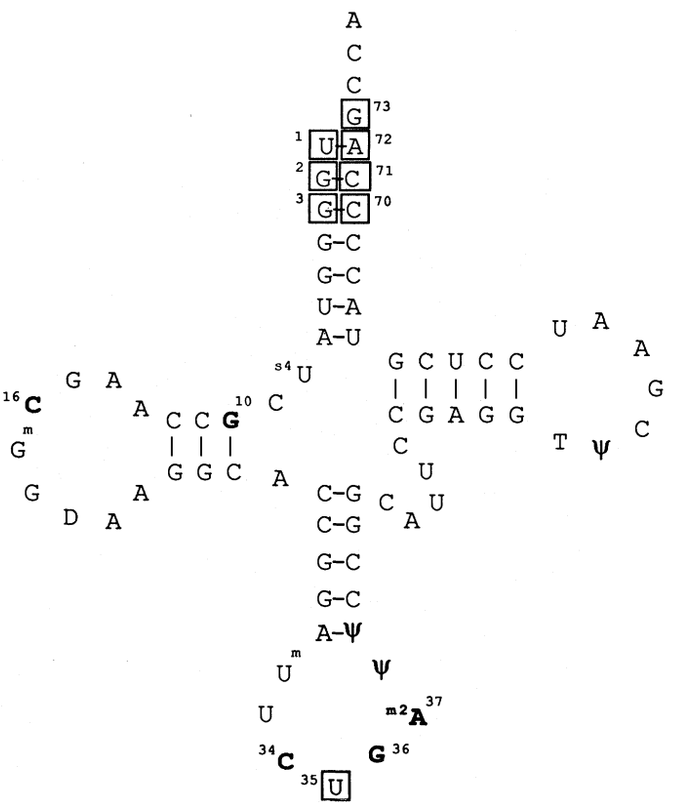


Fig. 5. Secondary-structure diagram of $tRNA_2^{Gln}$. Nucleotides clearly playing a role in recognition by GlnRS, as shown by the structure of the complex, are boxed. Nucleotides whose bases are interacting with the synthetase but whose role in recognition is less clear are in boldface.

V_{max}/K_m^{app} (V_{max} is the maximum velocity and K_m^{app} is the apparent Michaelis constant) for its charging with glutamine by GlnRS (*9*).

The structural explanation for the apparent preference of GlnRS for G at position 73 is that only guanine at position 73 can make the observed hydrogen bond between its 2-amino group and the phosphate oxygen of the previous nucleotide (Fig. 7). This hydrogen bonding interaction serves to stabilize the hairpin 3′ end of the

RNA. This base thus appears to be an important recognition element by virtue of an RNA-RNA interaction rather than an RNA-protein interaction. In this case, the free energy cost for the tRNA to assume the appropriate conformation required by the enzyme is sequence-dependent, a phenomenon also known to be important in sequence-dependent DNA recognition (*34–37*).

Acceptor stem base pairs. Sequence discrimination in duplex RNA presents an interesting problem since the major groove is both deep and narrow, and thus generally inaccessible to protein secondary structure elements such as the α helix. Moreover, there are fewer features presented by base pairs in the minor groove (as compared to the major groove) that allow discrimination among the two base pairs and their two orientations (*38*). The hydrogen bond acceptors (N3 on guanine and adenine and O2 on cytosine and uracil) occur in nearly the same position in the minor groove for all four bases. Only the exocyclic N2 of guanine distinguishes G · C from A · U base pairs (and perhaps from C · G, U · G, and G · U pairs). In the case of the GlnRS-tRNA^{Gln} complex the interaction of GlnRS with the acceptor stem is achieved via three secondary structure elements that enter the minor groove: two β turns from the acceptor stem binding domain and the amino end of α-helix H (Fig. 3) from the dinucleotide fold domain (Fig. 8). It is the propensity of the U1 · A72 base pair to be melted and hydrogen bonding with the exocyclic 2-amino groups of G2 and G3 that form specific recognition elements in the acceptor stem of tRNA^{Gln}.

Both in vivo and in vitro experiments have implicated nucleotides U1 · A72 as a significant recognition element for tRNA^{Gln} charging. Genetic selection for mutants of the *supF* tRNA^{Tyr} that could be charged with glutamine resulted in changes in base pair G1 · C72 to the mismatch A1 · C72 or to A1 · U72 (*8*). Using purified GlnRS, Seong *et al.* (*9*) have measured V_{max}/K_m^{app} for glutamine charging of three mutant suppressor initiator tRNA's in which base pair 1 · 72

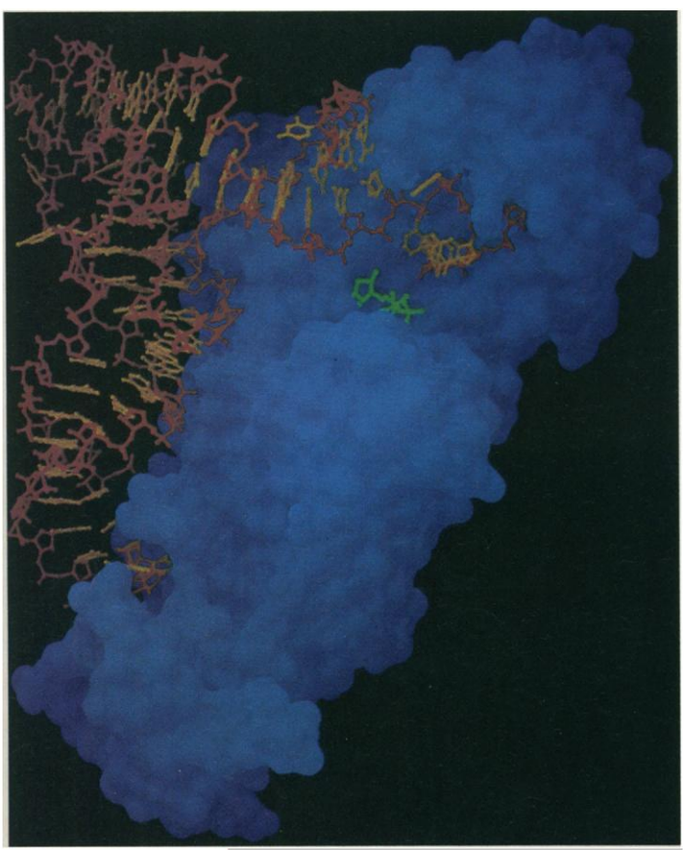


Fig. 6. Solvent accessible surface representation of the GlnRS enzyme complexed with tRNA^{Gln} and ATP. The region of contact between tRNA and protein extends across one side of the entire enzyme surface and includes interactions from all four protein domains. The acceptor end of the tRNA and the ATP are seen in the bottom of a deep cleft. Protein is inserted between the 5′ and 3′ ends of the tRNA and disrupts the expected base pair between U1 and A72.

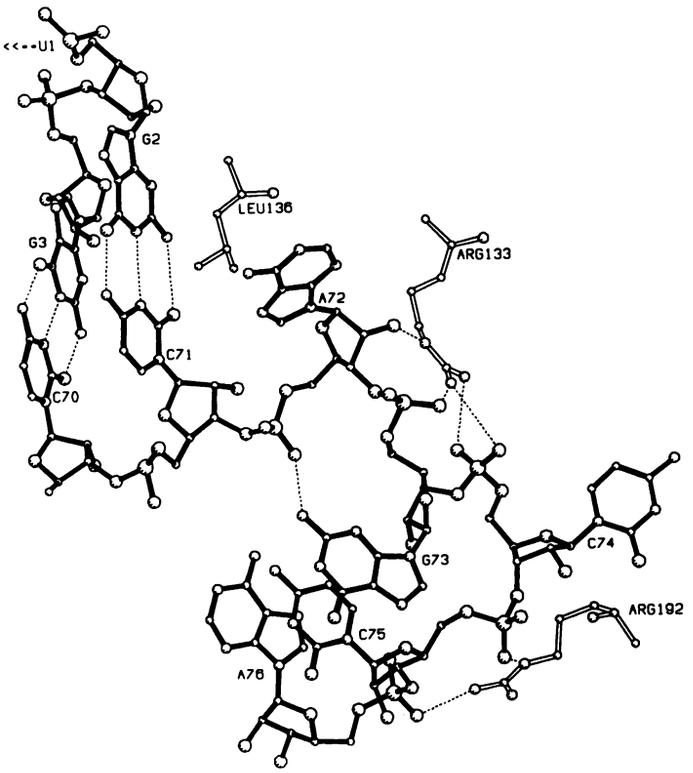


Fig. 7. The acceptor strand of the tRNA as seen in the complex. The side chain of Leu^{136} extends from a β turn and wedges between the bases of nucleotides A72 and G2, disrupting the last base pair of the acceptor stem, U1 · A72. The enzyme stabilizes the hairpin conformation via the interaction of several basic side chains with the sugar-phosphate backbone. An intramolecular hydrogen bond between the 2-amino group of G73 and the phosphate group of A72 further stabilizes this conformation.

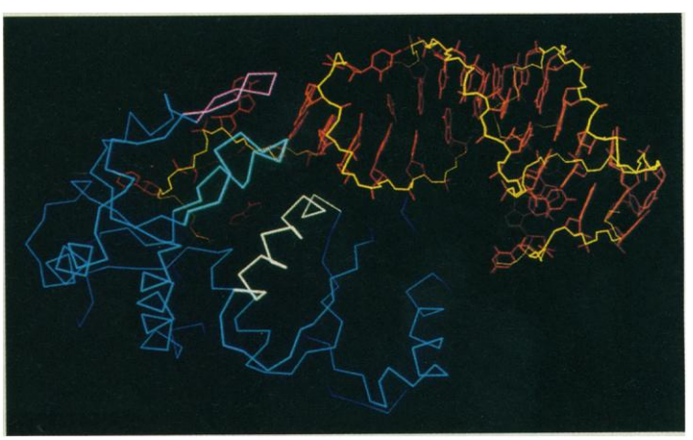


Fig. 8. A view from the acceptor stem and "down" the anticodon stem of the tRNA that shows three protein "fingers" interacting in the minor groove. This view is from the right of Fig. 3A. The β hairpin in pink contains Leu^{136}. The β loop in green and the amino terminus of α-helix H in yellow together with a buried water molecule form a recognition surface for base pairs G2 · C71 and G3 · C70 in the minor groove.

was the wild type C · A, U · A, or C · G. Initiator methionyl-tRNA ($tRNA_i^{Met}$) containing U · A at this position was charged nine times better by GlnRS than the same species containing C · G, while the tRNA possessing the wild-type mismatch C · A pair was improved by about threefold over one with U · A.

These data were interpreted to mean that the specific identity of the terminal base pair in glutamine tRNA's (conserved as a U · A pair in all known isoacceptors) was less important than its ability to be denatured when interacting with GlnRS, thus explaining why initiator tRNA possessing C · A at position 1 · 72 is better charged with glutamine than that possessing U · A. The structural data support this interpretation. There are no sequence-specific contacts between GlnRS and base pair U1 · A72; instead, the side chain of Leu^{136} emanating from a β turn of the acceptor binding domain is wedged between the base of A72 and base pair G2 · C71 (Fig. 7) in a manner very analogous to that observed with the binding of single-stranded DNA to the Klenow fragment (*39*). In both cases a leucine side chain is packed between two nonparallel bases. The peptide bond at the end of the β turn is stacked against the bottom of base pair G2 · C71. The specificity for U · A at base pair 1 · 72 therefore derives from the smaller free energy cost of its denaturation relative to that of a G · C pair at this position.

Although there are no definitive molecular genetic data on the roles of base pairs G2 · C71 and G3 · C70, the structure of the complex suggests that they are important recognition elements (Fig. 1B). Pro^{181} forms part of a β turn between two β strands of the acceptor binding domain and is inserted into the minor groove of the tRNA stem. The backbone carboxyl oxygen of Pro^{181} is hydrogen bonded to the N2 of G2. The backbone amide of Ile^{183} is hydrogen bonded to a buried water molecule that in turn hydrogen bonds to the O2 of C71 and the N2 of G2. A fourth hydrogen bond is made from this buried water to a carboxylate oxygen of Asp^{235}. The Asp^{235}, which emanates from the amino end of α-helix H in the dinucleotide fold, is also making hydrogen bonds with the N2 group of G3. Thus, these three protein groups and the buried water molecule together form a hydrogen-bonding surface that is complementary to base pairs 3 · 70 and 2 · 71 in the minor groove and which undoubtedly discriminates against other possible base pairs at these positions. Mutation of Asp^{235} results in an enzyme that mischarges certain tRNA's as discussed by Perona *et al.* (*40*). In the Trp repressor · DNA complex there are buried water molecules bound between the DNA bases and the protein that are proposed to play a role in DNA recognition via the major groove (*41*).

Anticodon recognition. There is extensive interaction between the anticodon nucleotides 34 to 36 of $tRNA^{Gln}$ and the protein. U35 has been implicated by many studies, both in vivo (*4*) and in vitro (*5*), as being critical to the identity of $tRNA^{Gln}$. It is seen to be buried in a cleft formed by the juxtaposition of the two carboxyl-terminal β-barrel domains in a fashion reminiscent of the antigen binding sites of antibodies. In addition, the bases of the other two anticodon nucleotides are also seen to be interacting with binding pockets formed by the two β-barrel domains. The exact details of the specific interactions between the protein and the functional groups on the bases of these nucleotides that might readily account for their role in $tRNA^{Gln}$ identity remain to be determined.

A large loop that connects two long antiparallel β strands of the proximal β barrel is packed against the front of the ATP binding site and may provide structural connectivity between recognition of the anticodon and events at the catalytic site (Fig. 3A). This structural link could account for the large effect (*5*) that changes in the anticodon have on the k_{cat} (where k_{cat} is the catalytic rate constant) of acylation with glutamine.

None of the other tRNA's containing a U at position 35 (Glu, His, Tyr, Asn, Lys, and Asp) have all of the recognition elements in the acceptor stem proposed to be critical for specific recognition by GlnRS. Since each recognition element may contribute a factor of 10 to 100 in k_{cat}/K_m to the charging of tRNA, noncognate species lacking several important elements will not be significantly acylated by GlnRS.

Implications for synthetase evolution. It appears plausible to suggest that all early aminoacyl-tRNA synthetases consisted of a dinucleotide fold domain that could carry out the enzymatic reaction and correctly recognize the amino acid and the tRNA. Because of the size and the nature of this domain, the only portion of the present-day tRNA molecule that could have been easily recognized is the acceptor stem and acceptor end region. In the case of the glutamine enzyme this recognition involves base pairs 3 · 70, nucleotide G73, and, to a lesser extent, possibly base pair 1 · 72 although it interacts primarily with an inserted domain. Although an identifiable sequence similarity does not exist among all of the different synthetases for the amino acids in the half of the dinucleotide binding domain that interacts with the ATP, it remains possible that all synthetases possess a dinucleotide binding motif performing a similar function and that all 20 synthetases evolved from a common precursor dinucleotide binding domain. Since the early synthetases would presumably only have been capable of interacting with the acceptor stem of the present-day tRNA, recognition elements responsible for distinguishing among tRNA's would have had to reside only in the acceptor stem. This expectation is consistent with the ability of the AlaRS to specifically charge a "mini-helix" consisting of only the acceptor stem (*42*) and the apparent importance of base-pair 3 · 70 in this system (*43, 44*).

Since all other domains in the GlnRS are different from the additional domains in both the TyrRS and the MetRS, the synthetases for the 20 amino acids may have diverged idiosyncratically. Both the MetRS and the TyrRS show a high α-helical content in the additional domains, whereas the GlnRS enzyme consists mostly of β-sheet structure. The tRNA elements recognized outside the acceptor stem and the mechanisms by which this recognition is achieved are most likely to be widely variant among the synthetases.

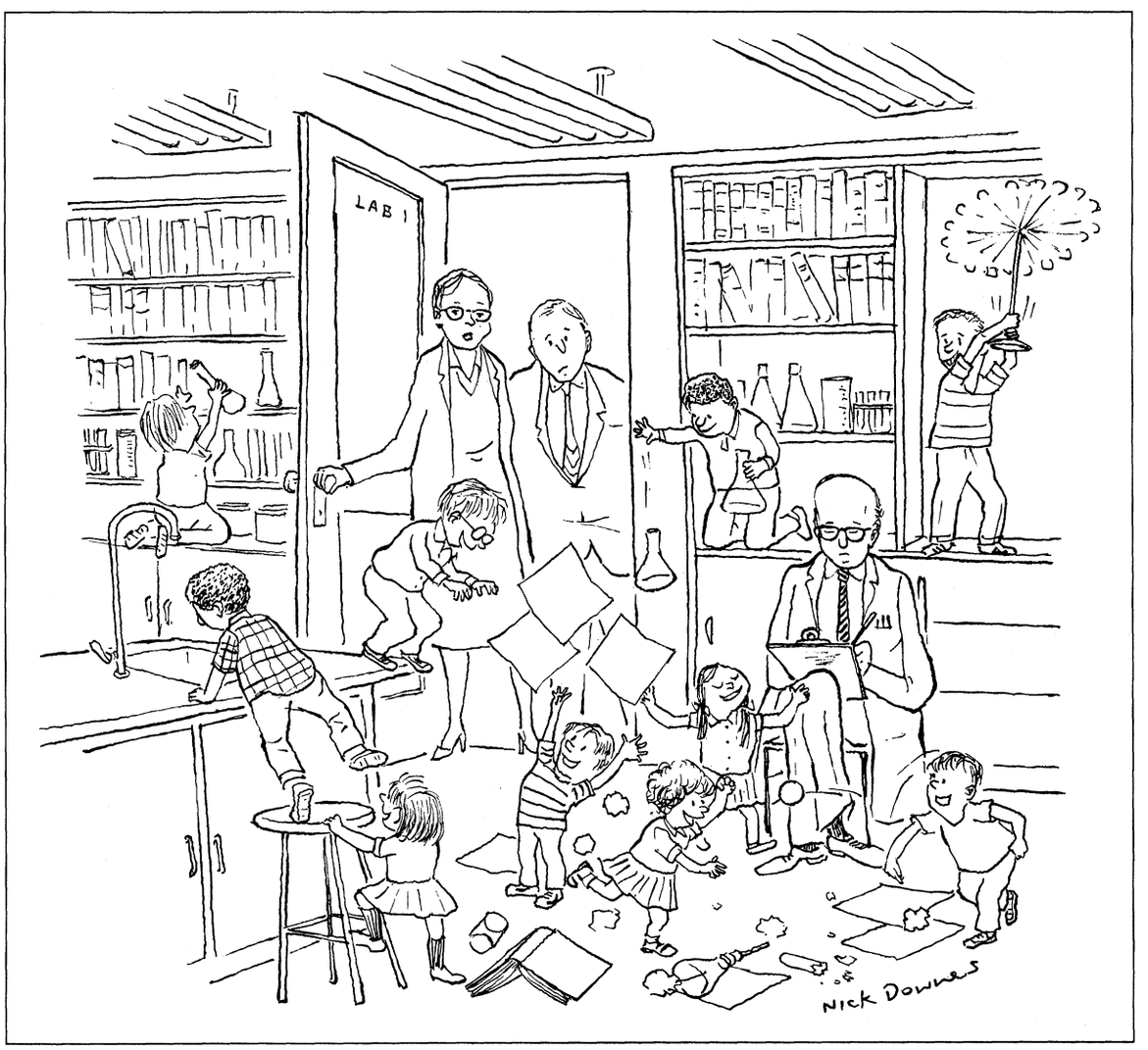
How a particular amino acid became associated with a specific tRNA early in evolution remains an interesting and unanswered question. However, the structure of the GlnRS · $tRNA^{Gln}$ complex is not consistent with some existing hypotheses. There is no direct interaction observed between the amino acid and the tRNA in the GlnRS complex; rather, both tRNA and amino acid recognition are mediated by protein interactions. Thus, the suggestion of a recognition code that arose through specific tRNA interactions with specific amino acids (*45*) appears unlikely in view of this crystal structure. Furthermore, the specific interactions between the protein and the tRNA do not constitute a code in any conventional sense, but a complex set of requirements for achieving structural complementarity between these two macromolecules.

REFERENCES AND NOTES

1. M. Yarus, *Cell* **55**, 739 (1988).
2. L. H. Schulman and J. Abelson, *Science* **240**, 1591 (1988).
3. J. Normanly and J. Abelson, *Annu. Rev. Biochem.*, in press.
4. M. Yaniv, W. Folk, P. Berg, L. Soll, *J. Mol. Biol.* **86**, 245 (1974).
5. L. H. Schulman and H. Pelka, *Biochemistry* **24**, 7309 (1985).
6. M. Yarus, R. Knowlton, L. Soll, in *Nucleic Acid–Protein Recognition*, H. J. Vogel, Ed. (Academic Press, New York, 1977), pp. 391–409.
7. Y. Shimura *et al.*, *FEBS Lett.* **22**, 144 (1972).
8. M. L. Hooper, R. L. Russell, J. D. Smith, *ibid.*, p. 149.
9. B. L. Seong, C-P. Lee, U. L. RajBhandary, *J. Biol. Chem.* **246**, 6504 (1989).
10. T. N. Bhat, D. M. Blow, P. Brick, *J. Mol. Biol.* **158**, 699 (1982).
11. P. Brick, T. N. Bhat, D. M. Blow, *ibid.* **208**, 83 (1989).
12. S. Brunie *et al.*, *J. Mol. Graphics* **5**, 18 (1987).
13. C. Zelwer, J. L. Risler, S. Brunie, *J. Mol. Biol.* **155**, 63 (1982).
14. D. M. Blow *et al.*, *ibid.* **171**, 571 (1983).
15. M. G. Rossmann, A. Liljas, C. I. Branden, L. J. Banaszak, in *The Enzymes*, P. D.

Boyer, Ed. (Academic Press, New York, 1976), pp. 62–102.
16. J. D. Robertus *et al.*, *Nature* **250**, 546 (1974).
17. S. H. Kim, *et al.*, *Science* **185**, 435 (1974).
18. D. Moras *et al.*, *Nature* **288**, 669 (1980).
19. N. H. Woo, B. A. Roe, A. Rich, *ibid.*, **286**, 346 (1980).
20. R. W. Schevitz *et al.*, *ibid.* **278**, 188 (1979).
21. F. Yamao, H. Inokuchi, A. Cheung, H. Ozeki, D. Söll, *J. Biol. Chem.* **257**, 11639 (1982).
22. H. Uemura, J. Conley, F. Yamao, J. Rogers, D. Söll, *Protein Sequence Data Anal.* **1**, 479 (1988).
23. J. J. Perona, R. Swanson, T. A. Steitz, D. Söll, *J. Mol. Biol.* **202**, 121 (1988).
24. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
25. Program PHARE of the MRC-CCP program suite, written by G. Bricogne, allows heavy atom parameter refinement against externally supplied phases.
26. A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988) (Program X-PLOR, Yale University, 1989).
27. A. Rich and P. R. Schimmel, *Nucleic Acids Res.* **4**, 1649 (1977).
28. T. A. Steitz, R. J. Fletterick, W. F. Anderson, C. M. Anderson, *J. Mol. Biol.* **104**, 197 (1976).
29. R. M. Starzyk, T. A. Webster, P. Schimmel, *Science* **237**, 1614 (1987).
30. C. Hountondji, P. Dessen, S. Blanquet, *Biochimie* **68**, 1071 (1986).
31. A. R. Fersht, *Biochemistry* **26**, 8031 (1987).
32. D. M. Crothers, T. Seno, D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3063 (1972).
33. A. Ghysen and J. E. Celis, *J. Mol. Biol.* **83**, 333 (1974).
34. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988).
35. M. R. Gartenberg and D. M. Crothers, *Nature* **333**, 824 (1988).
36. Z. Otwinowski *et al.*, *ibid.* **335**, 321 (1988).
37. J. Warwicker, B. P. Engelman, T. A. Steitz, *Proteins Struct. Funct. Genet.* **2**, 283 (1987).
38. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
39. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *ibid.* **85**, 8924 (1988).
40. J. J. Perona, R. M. Swanson, M. A. Rould, T. A. Steitz, D. Söll, *Science* **246**, 1152 (1989).
41. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
42. Y.-M. Hou, C. Francklyn, P. Schimmel, *Trends Biochem. Sci.* **14**, 233 (1989).
43. W. H. McClain and K. Foss, *Science* **240**, 793 (1988).
44. Y.-M. Hou and P. Schimmel, *Nature* **333**, 140 (1988).
45. G. de Duve, *ibid.*, p. 117.
46. Supported by NIH grants (T.A.S. and D.S.) and an NSF grant (D.S.). The protein Cα and tRNA phosphorus coordinates will be deposited in the Brookhaven Protein Structure Data Bank.

4 August 1989; accepted 20 October 1989



" Dr. Hardwick is studying the mathematics of chaos."