Y. AMINOACYL tRNA SYNTHETASES

The Genetic Code

The Genetic Code is the product of an immense effort in the 1960's to determine how mRNA sequence is translated into protein sequence by the ribosome, the catalyst for peptide synthesis in the cell. The key observation is that nucleic acid sequence is translated in three nucleotide segments called **codons**. The code is redundant. Most proteins are formed from only twenty amino acids.[1] A three base codon has sixty-four (4^3) different possible sequences (See Figure 1). Even including an additional codon that directs termination of translation, the stop codon, there are more possible codons than amino acids to be encoded. This redundancy becomes important in the mechanism by which amino acids are incorporated into the growing polypeptide chain.

Second Base in Codon

First Base in Codon	U	C	A	G	Third Base in Codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figure 1. The Genetic Code.

---

[1] In some proteins, a twenty-first amino acid is added during translation. Selenocysteine has the same structure as cysteine, but replaces the sulfur with selenium. Typically, the UGA stop codon is used. In the absence of selenium, translation is terminated. There are additional recognition elements that allow the ribosome to distinguish a SeCys encoding UGA from a "regular" stop UGA.

Adaptor hypothesis in translation

In 1961, Francis Crick proposed the "Adaptor Hypothesis", noting that a single molecule that could be used to link a particular mRNA sequence to a particular amino acid could be used in translation. Quoting from Crick's book, *What Mad Pursuit*[2]

> "The main idea was that it was very difficult to consider how DNA or RNA, in any conceivable form, could provide a direct template for the side-chains of the twenty standard amino acids. What any structure was likely to have was a specific pattern of atomic groups that could form hydrogen bonds. I therefore proposed a theory in which there were twenty adaptors (one for each amino acid), together with twenty special enzymes. Each enzyme would join one particular amino acid to its own special adaptor. This combination would then diffuse to the RNA template. An adaptor molecule could fit in only those places on the nucleic acid template where it could form the necessary hydrogen bonds to hold it in place. Sitting there, it would have carried its amino acid to just the right place where it was needed."

In 1965 Robert Holley identified and sequenced a transfer RNA, or tRNA, molecule from yeast (Figure 2) and provided Crick with the necessary adaptor molecule. All tRNA molecules share two essential features with regards to the current discussion: (1) The anticodon loop contains a three base sequence that complements the sequence of the codon and (2) each tRNA has a four nucleotide overhang at the 3' end to which the amino acid in covalently linked.
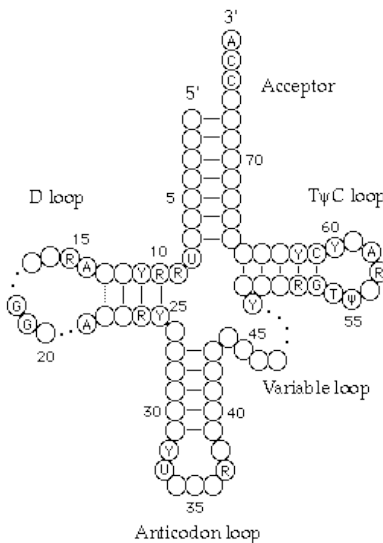

Figure 2. A schematic of a generalized tRNA molecule. Note especially the anticodon loop, which binds to the mRNA, and the acceptor stem, which is covalently linked to the cognate amino acid.

---

[2] F. C. Crick (1988) *What Mad Pursuit*

In *E. coli*, there are 86 genes that encode tRNA molecules with anticodons that are complementary to 39 of the 64 possible codons. The smallest genome sequenced so far, from *Mycoplasma genitalium* has only 36 tRNA genes that complement only 33 different codons.[3] As it turns out, one anticodon can complement more than one codon sequence through imaginative base pairing schemes. This was again an understanding that Crick recognized early on and labeled the "Wobble Hypothesis." We'll get to that later. The role of each tRNA can be provided by adding a superscript (for example, $tRNA^{Ala}$ is a tRNA molecule that is designed to carry alanine) and the anticodon can be specified (i.e. $tRNA^{Ala,GGC}$).

At the other end of the folded tRNA molecule, the covalent linkage to the amino acid is made via the 3' hydroxyl of the terminal adenosine (Figure 3). As predicted by Crick, that linkage is formed via the catalytic action of a specific enzyme that shows the necessary specificity to connect a specific amino acid with the appropriate tRNA molecule. In *E. coli* there are 21 of these enzymes, called aminoacyl tRNA synthetases (aaRS's), with one for each amino acid except lysine, which has two.[4]
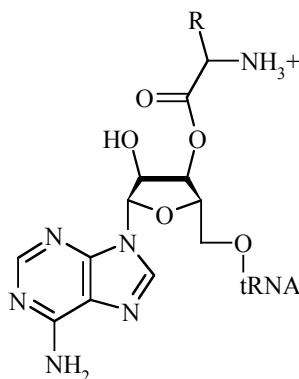

Figure 3. Covalent linkage between an amino acid and 3' terminal adenosine of a tRNA molecule.

Aminoacyl tRNA Synthetases Catalyze a Two Step Reaction

The general reaction catalyzed by aminoacyl-tRNA synthetases involves the formation of an ester between the α-carboxylate of an amino acid and the 3' hydroxyl of its cognate tRNA molecule, with the concomitant hydrolysis of ATP to produce AMP and inorganic pyrophosphate (PP_i).

$$AA + ATP + tRNA \rightarrow AA\text{-}tRNA + AMP + PP_i$$

The reaction proceeds with an intermediate that can be isolated in the absence of tRNA. The amino acid reacts with ATP to form an aminoacyl-AMP mixed acid anhydride. This intermediate can be added to a mixture of the aaRS and the cognate tRNA to form the aminoacyl-tRNA ester. The two-step mechanism is shown below (Figure 4).

---

[3] See the Genomic tRNA Database maintained by the Eddy group at Washington University: http://rna.wustl.edu/GtRDB/

[4] In some organisms glutamine and asparagine aaRS's are absent. For example, it is possible to covalently link Glu to $tRNA^{Gln}$ and then modify the amino acid side chain to glutamine by forming the amide.
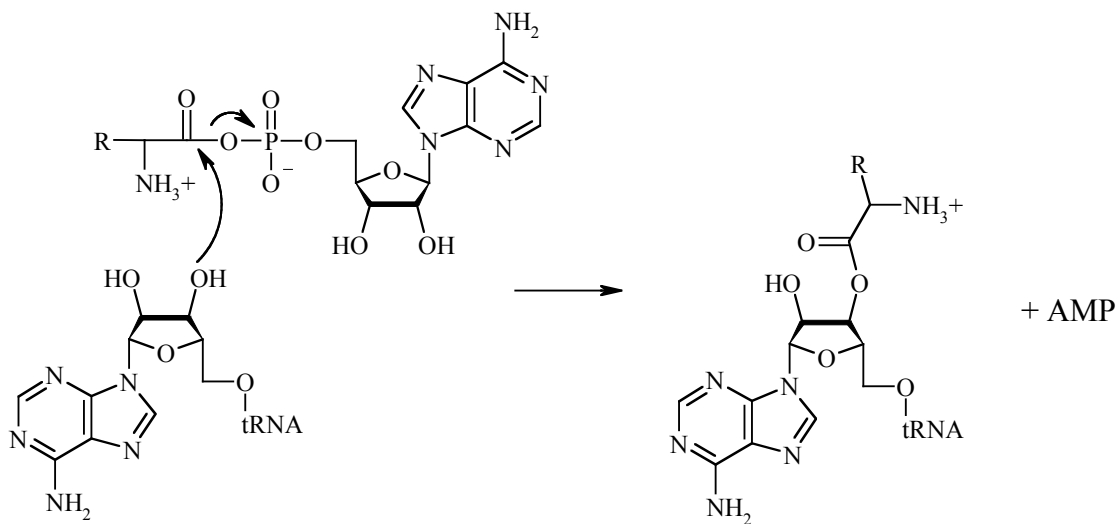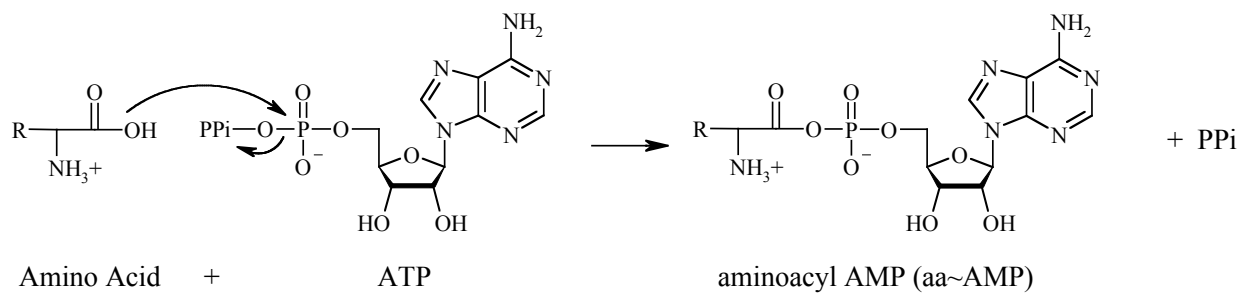
Figure 4. Two step chemical transformation catalyzed by aminoacyl tRNA synthetases.

The fidelity of translation is highly dependent upon the specificity of the above reactions. Should an enzyme misacylate a given tRNA molecule, there is no further mechanism by which the translation machinery will detect the error. Therefore, it is critical that a given aaRS show high specificity for a particular amino acid.

Thermodynamic Issues faced by aaRS's

An aaRS is charged with the task of preparing an ester (the aminoacyl-tRNA) from an acid (the amino acid) and an alcohol (the tRNA 3' OH group). That is a thermodynamically unfavorable process at pH 7.

$$aa + tRNA \rightarrow aa{\sim}tRNA \qquad\qquad \Delta G° = +7.3\ ^{kcal}/_{mol}$$
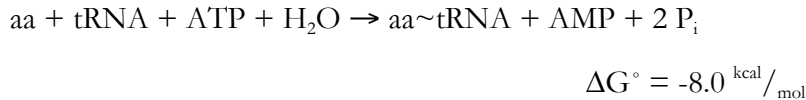
Thus there needs to an investment in energy to see that it happens. That is why ATP gets involved. However, the energy invested in the hydrolysis on the phosphate anhydride bond is insufficient to give the thermodynamic driving force need for the first step

$$ATP + H_2O \rightarrow AMP + PP_i \qquad\qquad \Delta G^\circ = -7.3 \, ^{kcal}/_{mol}$$
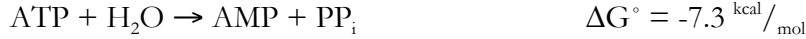
A second step of hydrolysis is needed – the breakdown of pyrophosphate (PP_i) to two phosphate ions (P_i):

$$PP_i + H_2O \rightarrow 2\,P_i \qquad\qquad \Delta G^\circ = -8.0 \, ^{kcal}/_{mol}$$

This makes the overall process favorable:

$$aa + tRNA + ATP + H_2O \rightarrow aa{\sim}tRNA + AMP + 2\,P_i$$

$$\Delta G^\circ = -8.0 \, ^{kcal}/_{mol}$$

Kinetic Issues faced by aaRS's

A given aaRS must show high selectivity for a single amino acid and a very few tRNA molecules. From a kinetic perspective, the Scheme 1 applies:
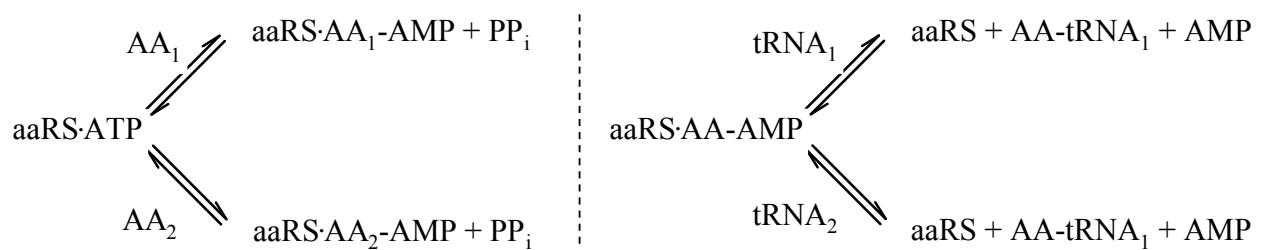
Scheme 1. Kinetic mechanism for selectivity in aaRS's.

The challenge presented to the aaRS is to show high selectivity towards the cognate amino acid and towards the cognate tRNA molecule in these two steps. The kinetic challenge is to have a rate of activation and transfer for cognates that wildly exceed the rates of reactions involving non-cognate amino acids and transfer RNA's. Returning to Michaelis-Menten kinetics, we expect to see an enhancement of the apparent second order rate constants of the cognate (A) vs. the non-cognate (B) as in Equation 1:

$$\frac{v_A}{v_B} = \frac{\left(k_{cat}/K_m\right)_A [E][A]}{\left(k_{cat}/K_m\right)_B [E][B]} > 40000 \qquad\qquad \text{(Eq. 1)}$$

As shown above, the minimum ratio encountered in these enzymes is 40,000:1, reflecting a difference in free energy of activation of about 6.5 $^{kcal}/_{mol}$ at 25°C (assuming that the relevant amino acids are at nearly equivalent concentrations). In some instances, this discrimination is easily

achieved. An enzyme that selectively binds alanine can easily discriminate against tryptophan, but other pairings are not so clear (see Table 1). What remains is for us to explain how these selectivities are achieved.

Table 1. Challenges to specificity in amino acyl tRNA synthetases.

Cognate Amino Acid	Potentially Difficult Non-Cognates
Alanine	Glycine
Cysteine	Serine, Alanine
Phenylalanine	Tyrosine, Leucine, Tryptophan
Histidine	Glutamine
Isoleucine	Valine
Proline	Alanine, Cysteine
Serine	Alanine, Cysteine
Threonine	Serine, Valine
Valine	Isoleucine/Threonine
Tyrosine	Phenylalanine

Typically, where steric exclusion or charge conflicts are possible, discrimination should not be problematic. However, in many instances, the 40,000-fold selectivity between two amino acids is not so easy to resolve.

Tyrosyl-tRNA Synthetase[5]

Alan Fersht performed the seminal studies on this enzyme from *Bacillus stearothermophilus* in the 1980's, using TyrRS as an archetypal enzyme for probing the energetics of catalysis and the bases of substrate specificity. In one study, Fersht and coworkers explored the relative specificity of TyrRS for the two alternate substrates, Tyr and Phe, and found that the level of specificity reflected in k_{cat}/K_m favors activation of tyrosine by ATP by a factor of 150,000. How is it achieved? The crystal structure of TyrRS complexed to tyrosyl-AMP shows two amino acids interacting with the hydroxyl group of tyrosine's phenol group. Asp176 acts as a proton acceptor for the phenolic proton while Tyr34 presumably acts as a proton donor (Figure 5).

---

[5] An extensive discussion of this enzyme is given in Fersht (1999) "Structure and Mechanism in Protein Science", WH Freeman, New York, pp. 422-450.
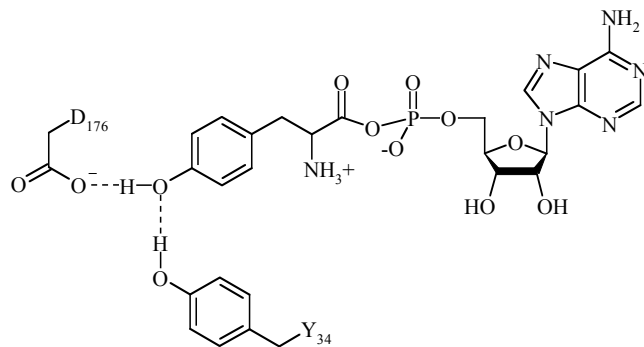
Figure 5. Active site geometry of TyrRS. Hydrogen bonds by Tyr34 and Asp176 contribute to specificity for tyrosine~AMP over phenylalanine~AMP.

Fersht and co-workers probed the importance of these two interactions by examining the kinetic profile of the wild-type enzyme to a site-specific mutant, Y34F, where tyrosine 34 has been replaced by a phenylalanine residue. The kinetic results of this study are presented in Table 2.

Table 2. Kinetic data for TyrRS.

	Tyr			Phe	Selectivity	$\Delta\Delta G^{\ddagger}$
	k_{cat}	K_m	k_{cat}/K_m	k_{cat}/K_m		$\left(^{kcal}/_{mol}\right)$
Wild-Type	5.4 s^{-1}	2.2 μM	2.5 x 10^6 M^{-1}s^{-1}	17 M^{-1}s^{-1}	150,000	7.2
Y34F	4.4 s^{-1}	4.4 μM	1.0 x 10^6 M^{-1}s^{-1}	100 M^{-1}s^{-1}	10,000	5.5
			$\Delta\Delta G^{\ddagger}$=.54 $^{kcal}/$mol	$\Delta\Delta G^{\ddagger}$=-1.0 $^{kcal}/$mol		

Interestingly, the loss of the hydroxyl group has negligible impact on the catalytic step itself (very little change in k_{cat}), but rather appears important in binding the substrate amino acid (K_m goes up two-fold), which in turn has an impact on the overall reaction rate (k_{cat}/K_m). The difference in energetics between wild type and mutant enzymes is only 0.5 $^{kcal}/_{mol}$, reflecting the loss of a relatively weak hydrogen bond between Y34 and the substrate tyrosine. However, when one compares the selectivity between wild type and mutant TyrRS with respect to phenylalanine as a substrate, one sees a 15-fold loss of substrate selectivity, as the loss of hydrogen bonding to Y34 has no negative impact on a substrate without an H-bond acceptor. In fact, Phe is accepted more readily in the mutant because it does not remove H-bonding opportunities for residue 34 upon binding (see Figure 6).
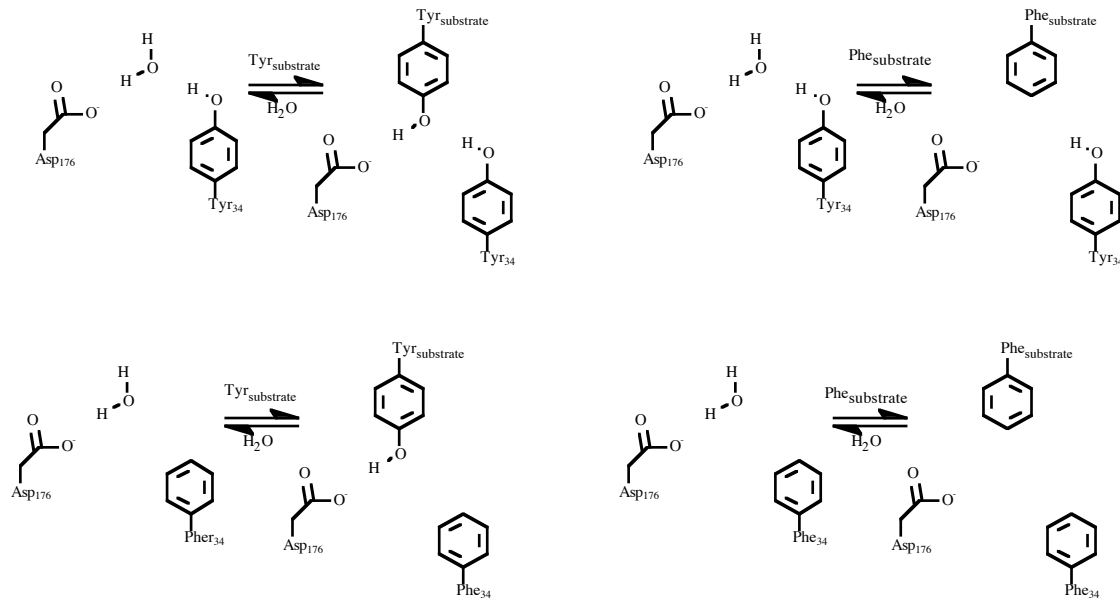
Figure 6. Comparison of phenylalanine and tyrosine binding to two variants of TyrRS, one with the wild-type Y34 and the other, Y34F, with a phenylalanyl residue in place of Y34.

So, what of the importance of Asp176? Tyrosine is still favored 10000-fold over phenylalanine as a substrate for Y34F. Presumably, a great deal of this specificity originates with the hydrogen bond between Asp176 and the substrate side-chain. In fact, that difference in reaction rates reflects an energy of interaction of 5.5 $^{kcal}/_{mol}$. While that difference in k_{cat}/K_m may be due to differences in stabilization of the bound substrate and in relative stabilization of the transition state, it does show the potential value of a single hydrogen bond in determining substrate specificity. It points to means by which aaRS's can easily distinguish non-cognate amino acid substrates lacking an H-bonding group. For example, SerRS should have no difficulty discriminating against alanine as a substrate, and AspRS should be able to exclude Asn. But what of cases where the structural differences are more subtle?

Cysteinyl-tRNA Synthetase

Cysteinyl-tRNA synthetase has a slightly more challenging task in discriminating against non-cognate amino acids. The thiol group does not provide hydrogen-bonding opportunities of the same magnitude as the hydroxyl group. Furthermore, CysRS must discriminate between these two similar functional groups, in loading cysteine onto its cognate tRNAs instead of serine. Fersht and Dingwall investigated the kinetics of CysRS from *Escherichia coli* in 1979, and found that the enzyme is nevertheless capable of selectively activating only cysteine.[6] Kinetic data for the reaction of several amino acids with ATP are given in Table 3.

---

[6]A. R. Fersht and C. Dingwall (1979) "Cysteinyl-tRNA Synthetase from *Escherichia coli* Does Not Need an Editing Mechanism to Reject Serine and Alanine. High Binding Energy of Small Groups in Specific Molecular Interactions" *Biochemistry* **18**, 1245-1249.

Table 3. Kinetic data from CysRS

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1}s^{-1})	$\Delta\Delta G^\dagger$ (kcal/mol)
Cysteine	22	0.05	440,000	0
Serine			0.004	> 11
α-Aminobutyrate*	>0.2	>180	1.29	≥ 7.6
Alanine	>0.03	>365	0.88	≥ 9.1

*α-aminobutyrate is a non-natural amino acid with an ethyl group side chain

From the data above, it can be seen that alanine, rather than serine, is the most problematic non-cognate, naturally occurring amino acid for CysRS, but even so it is activated at a rate 5×10^7 slower than cysteine. α-Aminobutyrate, an unnatural amino acid with an ethyl group side chain, is slightly better, but still not a significant challenge to CysRS's specificity. Clearly, the thiol provides adequate opportunities for specific interactions.

Comparison of the specificity constants for Cys vs. Ala show a 9 $^{kcal}/_{mol}$ lower free energy of activation for cysteine, indicating that the thiol provides substantial intermolecular interactions in both binding (compare K_m's – there is a 5 $^{kcal}/_{mol}$ difference in binding) and in transition state stabilization (comparing k_{cat} values). In comparing Cys to aminobutyrate, we see that a large fraction of that stabilization has yet to be recovered. Fersht and Dingwall argued that the selectivity is a result of the enhanced van der Waals interactions that are achieved between the sulfur and surrounding atom groups in the active site vs. a methylene group. Between the difference in atomic radii and the relative polarizabilities of the atom groups, CysRS has generated substantial specificity through both vdW interactions and perhaps through an H-bonding interaction that is sterically inaccessible to the hydroxyl group.

Table 4. Potential for vdW interactions of various atom groups.

Atom / Group	vdW Radius (Å)	Relative energy of interaction (kcal/mol) with CH$_2$ group
-O-	1.5	1
-CH$_2$-	1.8	1
-S-	2.0	2.5

However, Fersht and Dingwall did recognize two other possibilities: (1) that CysRS could use interactions with an active site Zn^{2+} to selectively bind a thiolate group of deprotonated cysteine, or (2) that CysRS could covalently bind cysteine in a disulfide between the enzyme active site and the substrate. As of 2002, we have the answer. It turns out that CysRS has an essential Zn^{2+} ion in its active site. Kate Newberry, working with John Perona at UCSB, solved the structure of CysRS with and without the cysteine substrate and showed that the active site zinc accommodates the thiolate of the substrate's side chain by expanding its coordination geometry from a distorted tetrahedron to

trigonal bipyramidal.[7] Fersht[7] had cited earlier work indicating a 10^5-fold preference for a thiolate complex with zinc over an alcohol complex. The lower pK_a of cysteine means that it is more likely to bind as an anion than the alcohol. Indeed, in a follow-up study to the crystallographic work, a Co^{2+}-bound form of CysRS was found to bind cysteine with a K_d of 42 μM, while serine was bound with a much higher K_d of 980 mM. The cobalt form was used because Co^{2+} is spectroscopically active in the visible range, permitting direct observation of the enzyme-substrate complex and a titratable spectroscopic signal.

Isoleucyl-tRNA Synthetase

So far we've examined enzymes where the intermolecular interactions are sufficient to drive high specificity for cognate amino acids, but what about situations in which the energy to be derived from those interactions is too small to provide adequate selectivity? The prime example of such difficulties is isoleucyl-tRNA synthetase (IleRS), which must discriminate against valine, an amino acid that differs by a single methylene group. In 1957 Linus Pauling noted that IleRS could only distinguish between Ile and Val on the basis of interactions with one methylene group. He calculated this difference to be worth ~1 kcal/mol or about a five-fold enhanced specificity for isoleucine. It has been observed that the actual selectivity is much higher, at about roughly 40000:1 (a difference of 6.4 kcal/mol in specificity).
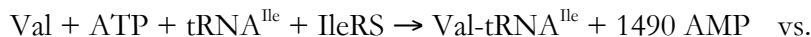
How is it that IleRS effectively excludes Val? Consider the reverse, ValRS is effective at excluding Ile from activation and transfer to tRNA^{Val}, despite a similar difference in energy of interaction. Why? Presumably through steric exclusion of the larger Ile side chain. This option is not available to IleRS, though. Or is it?

In the *absence of tRNA^{Ile}*, Fersht observed that the reverse of the activation step:

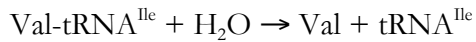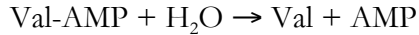$$AA\text{-}AMP + PP_I \;\rightarrow\; AA + ATP$$

as measured by ratio in k_{cat}/K_m is 250 in favor of Ile, reflecting a difference of 3.3 kcal/mol – substantially larger than that predicted by Pauling, but well below that observed (Freist, 1985).

In the presence of tRNA^{Ile}, however, the efficiency of Val-tRNA^{Ile} synthesis is dramatically less than for the isoleucyl-tRNA^{Ile}, and is accompanied by the massive production of AMP:

$$Val + ATP + tRNA^{Ile} + IleRS \rightarrow Val\text{-}tRNA^{Ile} + 1490\ AMP \quad \text{vs.}$$

$$Ile + ATP + tRNA^{Ile} + IleRS \rightarrow Val\text{-}tRNA^{Ile} + 1\ AMP$$

That is to say, that in E. coli, for every erroneous molecule of Val-tRNA^{Ile} synthesized, 1490 molecules of AMP are produced, where as joining Ile to its cognate tRNA^{Ile} leads only to the stoichiometric production of one molecule of AMP. Why the difference? There appears to be a hydrolytic reaction that takes place somewhere along the reaction pathway, in a **proofreading** step. There are two possible activities catalyzed by IleRS:

---

[7] KJ Newberry et al. (2002) "Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase" *EMBO J.* **21**, 2778-2787.

$$\text{Val-AMP} + H_2O \rightarrow \text{Val} + \text{AMP}$$

$$\text{Val-tRNA}^{\text{Ile}} + H_2O \rightarrow \text{Val} + \text{tRNA}^{\text{Ile}}$$

Cramer and colleagues[8] have performed an experiment to distinguish between these two possibilities. His lab prepared an analog of tRNA$^{\text{Ile}}$ in which the 3' hydroxyl group of the terminal adenosine was replaced by an amino group. When the 3' $-NH_2$ gets aminoacylated, a relatively stable and unreactive amide will form (Figure 7).
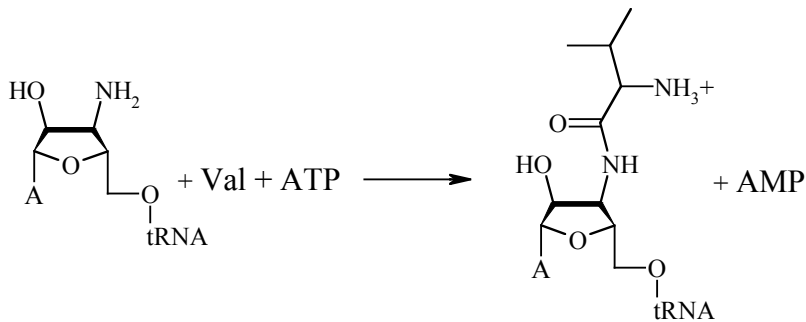

Figure 7. 3'-amino analog of tRNA reacts to make inert Val~tRNA analog.

Note that this substrate won't block the hydrolysis of Val-AMP by IleRS, but it will block hydrolysis of Val-tRNA$^{\text{Ile}}_{\text{NH2}}$.

When this substrate is used, AMP produced per Val-tRNA$^{\text{Ile}}$ decreases to 24, instead of the 1490 observed previously. Thus the adenylation reaction, where valine is mistakenly allowed to react with ATP in the active site of IleRS, has the possibility of being reversed by hydrolysis:

$$\text{Val} + \text{ATP} \rightarrow \text{Val-AMP} + H_2O \rightarrow \text{Val} + \text{AMP}$$

This hydrolysis provides 24-fold selectivity in the presence of tRNA$^{\text{Val}}$. However, by division (1490/24), we see that the hydrolysis of Val-tRNA$^{\text{Ile}}$ provides 62 fold selectivity. That proofreading reaction has been lost, since Val-tRNA$^{\text{Ile}}_{\text{NH2}}$ is not hydrolyzed by IleRS. Thus we see two results. In the absence of tRNA$^{\text{Ile}}$, the enzyme shows a 250-fold preference for isoleucine over the non-cognate valine. However, when tRNA$^{\text{Ile}}$ is present, as in Cramer's study, a hydrolysis reaction is introduced at each step, which dramatically increases selectivity.

How does the enzyme achieve this and why is tRNA$^{\text{Ile}}$ required for hydrolysis? IleRS is a huge protein of roughly 120 kDa (~1000 residues) with a multi-domain structure. In 1998, Schimmel showed that a single mutation, G56A, in *E. coli* IleRS could dramatically affect activation of Ile vs. Val. Both activation reactions are severely impeded, but the resulting mutant now shows no selectivity towards isoleucine (see Table below). However, the mutation has no effect on hydrolysis of the incorrect formation of Val-AMP in the presence of tRNA$^{\text{Ile}}$; the rate constant for Val-AMP

---

[8] W. Freist, H. Sternbach and F. Cramer (1987) "Isoleucyl-tRNA synthetase from baker's yeast and from *Escherichia coli* MRE 600" *Eur. J. Biochem.* **169**, 33-39.

hydrolysis is constant between WT and G56A IleRS. The interpretation of this result is that there are *two* active sites in this enzyme – one for activation and transfer to the tRNA molecule and a second for the proofreading hydrolysis reaction.

Table 5. Kinetic data for IleRS

	Amino Acid Activation			k_{cat} (s^{-1})	
	k_{cat}/K_m (Ile)	k_{cat}/K_m (Val)	ratio	hydrolysis of Val-AMP	hydrolysis of Val-tRNA
WT	6.9×10^6	3.8×10^4	180	2.7	0.07
G56A	690	660	1	2.7	0.13

The crystal structure[9] of IleRS revealed one binding pocket for Ile, but when Val added is added to the protein, two binding sites are evident, one that is identical to the site occupied by Ile, the other in an entirely different domain. The isoleucine binding pocket is surrounded by residues D85 an Q554 which H-bond to ammonium and carboxylate of isoleucine respectively. The non-polar side chain of the substrate is packed in by P46, W518 and W558. Isoleucine and valine both fit in this site, but leucine would be excluded due to gamma branching (i.e. two carbons coming off the γ carbon). Valine does not interact as thoroughly with surface area as isoleucine, but presumably the loss of vdW contacts is what leads to the 250-fold selection for isoleucine in the absence of a bound tRNA molecule.

A dramatic mutation, deletion of the entire domain that contains the second, novel valine binding site, (Δ219-265), the following kinetics were observed:

Table 6. Data related to the proofreading site of IleRS.

	k_{cat}	K_m	k_{cat}/K_m
WT + Ile ATP + tRNA	0.69 s^{-1}	45 μM	15300 s^{-1}M^{-1}
Δ + Ile ATP + tRNA	0.78 s^{-1}	91 μM	8571 s^{-1}M^{-1}
Δ + Val ATP + tRNA	0.21 s^{-1}	86 μM	2441 s^{-1}M^{-1}

Note that the ratio of specificity constants (k_{cat}/K_m) between isoleucine and valine without a hydrolytic reaction is about 3.5 fold, roughly 0.75 kcal/mol, as Pauling predicted. In a separate structural study with IleRS bound to its tRNA molecule, Steitz's lab uncovered the reason that

---

[9] O. Nureki *et al.* (1998) "Enzyme Structure with Two Catalytic Sites for Double-Sieve Selection of Substrate" *Science* **280**, 578-582.

tRNA stimulates hydrolysis of Val-tRNA^{Ile}. The editing domain only swings into place upon binding tRNA^{Ile}.[10]

The Second Half of the Problem: AlaRS

Selectivity for an amino acid is only part of the goal of an aminoacyl tRNA synthetase. The other objective is to selectively bind only those tRNA molecule(s) that possess anticodons necessary to deliver alanine to the correct codon on an mRNA transcript. In principle, this means that the aaRS only needs to specifically bind the three base anticodon region of the tRNA molecule to obtain specificity. However, it isn't that simple. For example, there are many mutant tRNA's that contain an "amber suppressor" anticodon, CUA, that allows them to complement the amber stop codon UAG. These suppressors, when charged with an amino acid, are able to load the amino acid into a growing peptide *despite* the presence of a stop codon at that position. While there are many problems that arise from this suppression, it's an interesting phenomenon that shows aaRS's are capable of recognizing and binding tRNA molecules without exerting specificity for the anticodon.

In 1988, Hou and Schimmel[11] presented the first chemical evidence for how a single synthetase, Alanyl tRNA synthetase (AlaRS) from *E. coli* is able to selectively bind tRNA^{Ala} vs. all other tRNA molecules in the cell, *without specifying the identity of the anticodon.* They chose an experimental system that used an alanine accepting tRNA molecule that contained an amber suppressing anticodon: tRNA^{Ala,CUA}. With only that difference, AlaRS still recognizes the tRNA molecule and adds alanine to the 3' OH group and then alanine will be added to growing peptide chains when an amber codon appears in the mRNA transcript (not always a good thing for the cell). To test that function *in vivo* (in the cell), they used a gene for tryptophan synthase (*trpA*) that contained a premature amber stop codon at position 234. Unless alanine is placed at position 234, tryptophan synthase is inactive. An inactive tryptophan synthase means that the cell can't make its own Trp. To make sure that alanine is placed at the amber position, the *E. coli* was grown on medium without tryptophan. Thus cells would not grow unless a functional tryptophan synthase was produced. This became their test for determining whether AlaRS could recognize mutations to tRNA^{Ala,CUA}. If they make a mutation to tRNA^{Ala,CUA} that does not affect AlaRS activity, then the cells survive. A mutation to the tRNA that inhibits AlaRS activity will lead to cell death.

Hou and Schimmel created 28 mutants of the tRNA^{Ala,CUA} (Figure 8). Of those 28 mutants: **20** of them retained activity in suppressing amber mutation (sup$^+$ phenotype), **5** mutants were inconclusive (no mutant tRNA observed in cell) and **3** mutants are lethal to the cell.

The three interesting mutations, then, are those that kill the cell, because those mutations abolish recognition of the tRNA^{Ala,CUA} by AlaRS. The three all occur at one particular base pair in the acceptor stem, between G3 and U70. Obviously, this is not a Watson-Crick base pair, but when it is replaced by a WC pairing, either by the G3→A mutation or the U70→C mutation, the tRNA is no

---

[10] L. F. Silvian, J. Wang and T. A. Steitz (1999) "Insights into Editing from an Ile-tRNA Synthetase Structure with tRNA^{Ile} and Mupirocin" *Science* **285**, 1074-1077.

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

longer recognized by AlaRS (the third mutation was actually a multiple mutant that included U70→C, so really it's just the same mutation twice.)
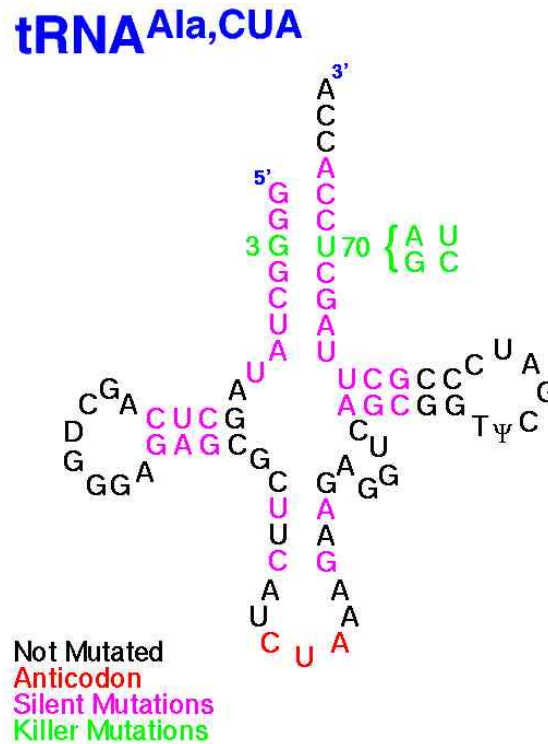

Figure 8. The structure of tRNA$^{\text{Ala,CUA}}$ used in the Hou and Schimmel study. Note that the anticodon reads CUA. Only mutations to the G3:U70 base pair in the acceptor stem abolished AlaRS activity with this substrate.

Could it be that only one base pair is necessary for the recognition of tRNA$^{\text{Ala}}$ by the *E. coli* AlaRS? Of the 86 tRNA genes in *E. coli*, the only ones to possess a G3•U70 base pair are those that accept alanine! As further confirmation of this finding, Hou and Schimmel appropriated the cysteine accepting tRNA and mutated it to contain an amber anticodon. The resulting tRNA molecule, tRNA$^{\text{Cys,CUA}}$, is able to add cysteine to position 234 of tryptophan synthase, but the enzyme is inactive with cysteine at that position. However, if a G3•U70 base pair is placed in the tRNA molecule, yielding (G3•U70) tRNA$^{\text{Cys,CUA}}$, then the tRNA molecule becomes a substrate for AlaRS and it succeeds in promoting the production of an active tryptophan synthase in the cell. The conclusion of the study is that a *single base pair* is the recognition determinant for AlaRS.

The obvious follow-up experiment is to see how AlaRS interacts with the G3•U70 base pair in order to confer specificity and activity. Schimmel's lab undertook a biochemical study outside the cell

using a small 35 base pair stem loop as a substrate for AlaRS.[12] The so-called "mini-helix" was designed to possess an stem analogous to the acceptor stem of tRNA[Ala] (Figure 9).
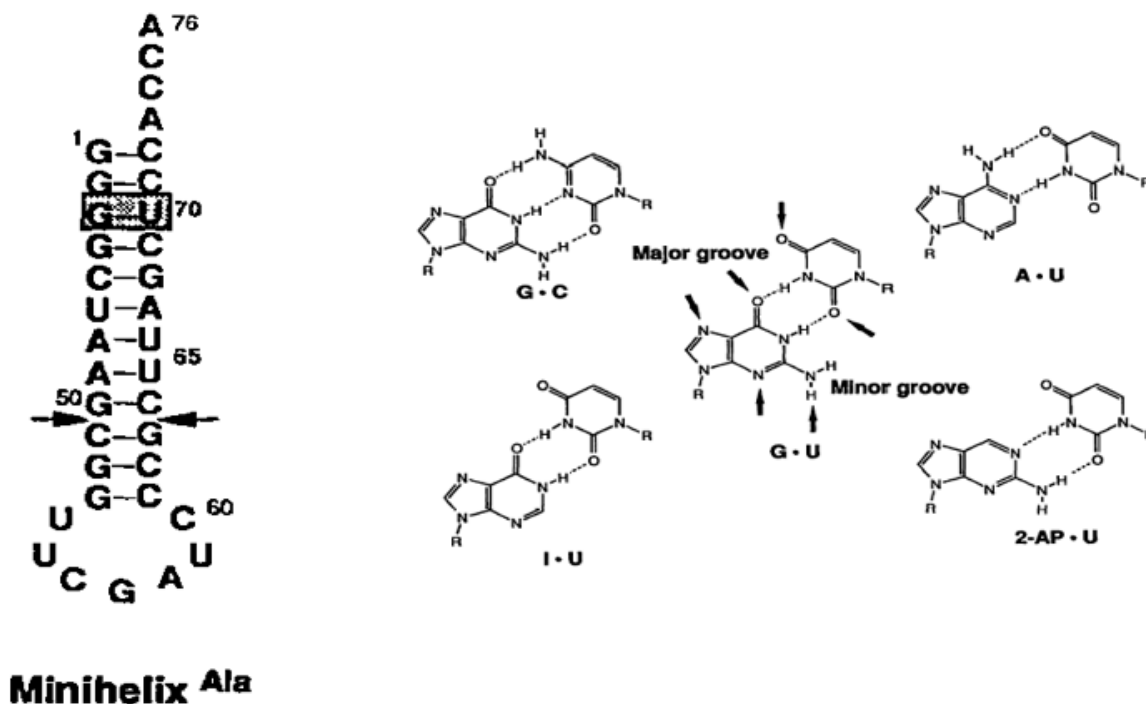

Figure 9. Structure of the minihelix and pictures of base pairs designed to replace the G3•U70 base pair. [13]

The minihelix presents the opportunity to replace specific bases through chemical synthesis. The authors investigated the role of the G3:U70 bp by creating synthetic RNAs with mutations to G:C, A:U, I:U, and 2-AP:U (see Figure X). None of the mutations was successfully aminoacylated by AlaRS, though the minihelix containing the G3•U70 base pair was a substrate. Of particular interest is the I3•U70 base pair, which differs from the "wild type" base pair only n the absence of an exocyclic amino group at C2. That small change presumably does not affect the conformation of the base pair, but only removes an atom group that can be used for binding. In response, the k_{cat}/K_m for aminoacylation of the IU helix is at least 600 fold lower than for GU, and the K_m value for the IU helix was judged to be at least 50 fold higher than for the GU helix. In conclusion, Schimmel's group argued that AlaRS obtains specificity through a single functional group on G3 that is only available when base paired to a uracil at position 70.

---

[12] Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). Specificity for Aminoacylation of an RNA Helix: An Unpaired, Exocyclic Amino Group in the Minor Groove. *Science* **253**, 784-786.

[13] Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) "Structure of *E. coli* Glutaminyl-tRNA Synthetase Complexed with tRNAGln and ATP at 2.8 Å Resolution" *Science* 246, 1135-1142.

A More Complicated Interaction: GlnRS

The GlnRS•tRNA^{Gln} was the first crystal structure of an aminoacyl tRNA synthetase bound to tRNA. The 547 residue enzyme forms extensive interactions between the protein and its cognate tRNA substrate, with contacts between GlnRS and the acceptor stem of tRNA^{Gln} and with the anticodon loop (Figure 10).
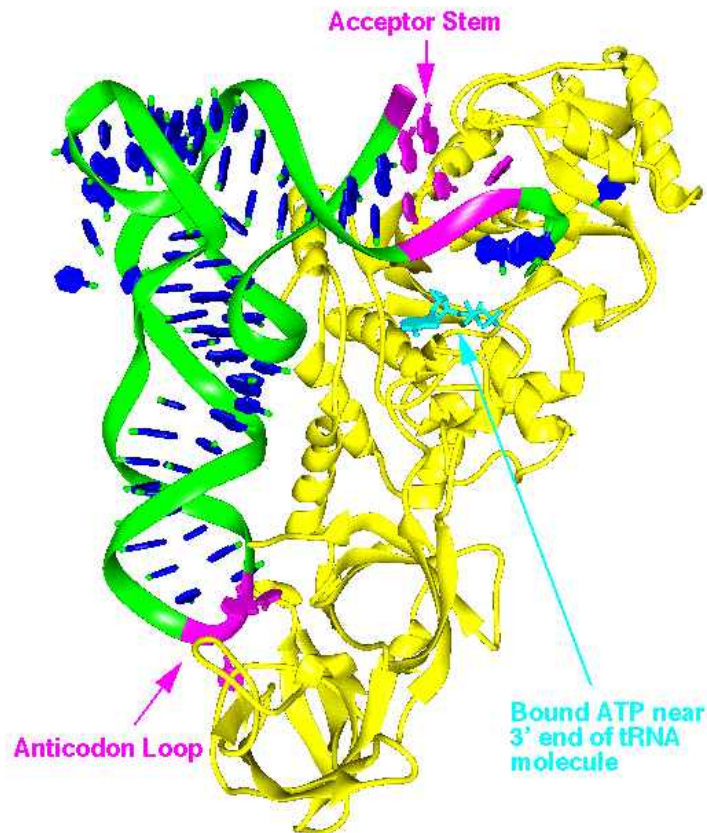
Glutaminyl–tRNA Synthetase

Figure 10. Structure of the GlnRS•tRNA^{Gln} complex.

Unlike AlaRS, GlnRS will not accept any sequence in the anticodon loop, but it is not generally accepting of mutations in the acceptor stem either. Thus, rather than focusing on a single structural determinant for specificity, GlnRS requires many structural features. For example, GlnRS forms contacts with each base of the anticodon (Figure 11), which has the sequence $C_{34}U_{35}G_{36}$. The most extensive interactions are made the central uracil, U35, with hydrogen bonds to each of the three donors and acceptors of the base. This position tends to be the most important in conferring specificity of the codon-anticodon pairing on the ribosome, so there is a good rationale for placing emphasis on specificity for this position. Note that the 3' position of the anticodon, G36, is also strongly specified by a direct interaction with Arg437, the 5' position of the anticodon, also known as the wobble position, receives two hydrogen bonds, both to the O2 oxygen of C34, thus assuring that a pyrimidine will be specified in the anticodon. Since two codons code for Gln, CAG and

CAA, it stands to reason that the enzyme would accept either CUG or UUG as anticodons. The interactions allow both tRNA's to bind.
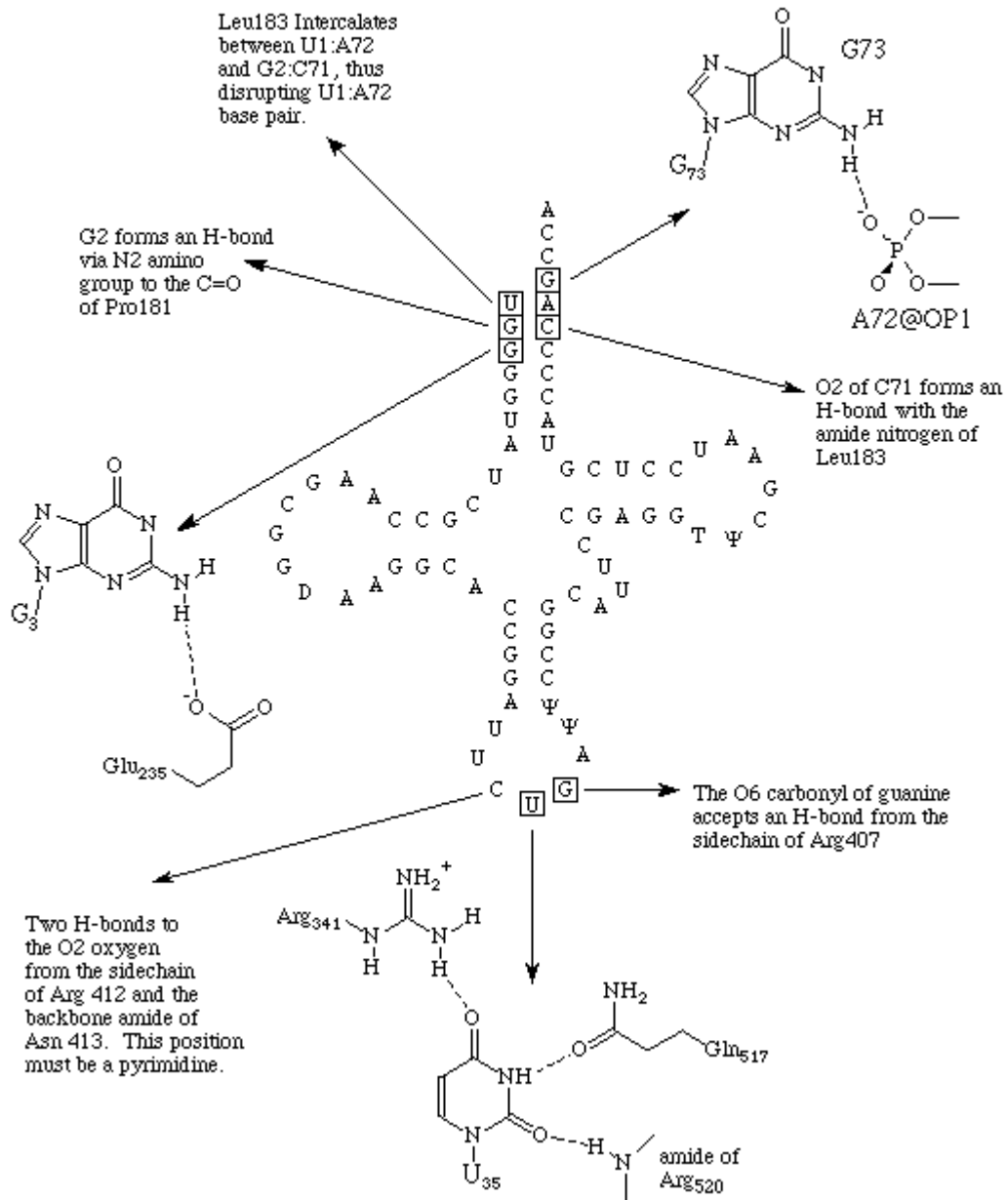

Figure 11. Summary of contacts made between GlnRS and tRNA[Gln].

From a kinetic perspective, we see how important each of these interactions in the anticodon are (Table 7). Compared to mutations made in the acceptor stem, non-cognate mutations in the anti-codon loop are extremely serious, resulting in 1000 fold losses in k_{cat} and 40 fold increases in K_m. Clearly the mutation of U35 to cytosine is the most damaging, as would be expected by the close complementarity of the protein to uracil at that position. Also, we see that the "wobble position", while being permissive to either pyrimidine at position 34, does not accept a purine, namely A34. And finally, the contact of Arg437 with G36 discriminates against adenine by 3.3 $^{kcal}/_{mol}$.

Table 7. Kinetic constants for the aminoacylation of tRNA^{Gln} possessing substitutions to its nucleotide sequence.[14]

tRNA Variant	k_{cat} (s^{-1})	K_m (µM)	k_{cat}/K_m (M^{-1}s^{-1})	$\Delta\Delta G^{\ddagger}$ (kcal/mol)
Wild type	0.2	0.15	1.3×10^6	
Discriminator Base				
A73	0.14	0.2	7.0×10^5	+0.4
U73	0.0068	8.0	850	+4.4
Acceptor Stem				
G1:A72	0.92	0.66	1.4×10^6	
G1:C72	0.17	1.3	1.3×10^5	+1.4
A2:U71	0.011	10	1.1×10^3	
A3:U70	0.046	3.3	1.4×10^4	+2.7
Anticodon Loop				
A34	0.00065	2.5	260	+5.1
C35	**0.00034**	**6.7**	**50**	**+6.1**
A36	0.036	6.6	5500	+3.3

In the acceptor stem, there are several important interactions, but the most significant is related to guanine at position 73 – the position known as the discriminator base. The name arises from the observation that the identity of the nucleotide at position 73 of tRNA's is frequently critical to the specificity of the aminoacyl tRNA synthetase. For example, if one makes one change in tRNA^{Leu}, A73 to G, then the resulting tRNA molecule is charged only with serine in the cell. Interestingly, the specificity for G73 is achieved through *indirect readout*. Rather than making direct contact with the base, GlnRS appears to specifically recognize a conformation of the acceptor stem that is only possible through an H-bond between the C2 amino group of guanine and the phosphate of A72,

---

[14] This data comes from a study by Dieter Söll's group: M. Jahn, M. J. Rogers and D. Söll (1991) "Anticodon and acceptor stem nucleotides in tRNA^{Gln} are major recognition elements for *E. coli* glutaminyl-tRNA synthetase" *Nature* **352**, 258-260.

which causes the single stranded portion of the acceptor stem to bend back at a sharp angle (Figure 12). Another example of indirect readout can be seen in the disruption of the A1:U72 base pair by Lue183. Replacement of A1:U72 with a G1:C72 pairing causes a ten fold decrease in k_{cat}/K_m. On the other hand, placing guanine at position 72, opposite the adenine, utterly disrupts the base pairing at that position and has no effect on the specificity constant of GlnRS. There is one point of direct contact between GlnRS and the acceptor stem bases. Glutamate 235 forms a hydrogen bond in the minor groove to the C2 amino group of G3. Replacement of the G3:C70 base pair by A3:U70 leads to a 100 fold decrease in the specificity constant. (Parenthetically, the G3:U70 base pair favored by AlaRS leads to a three-fold loss of activity in GlnRS).
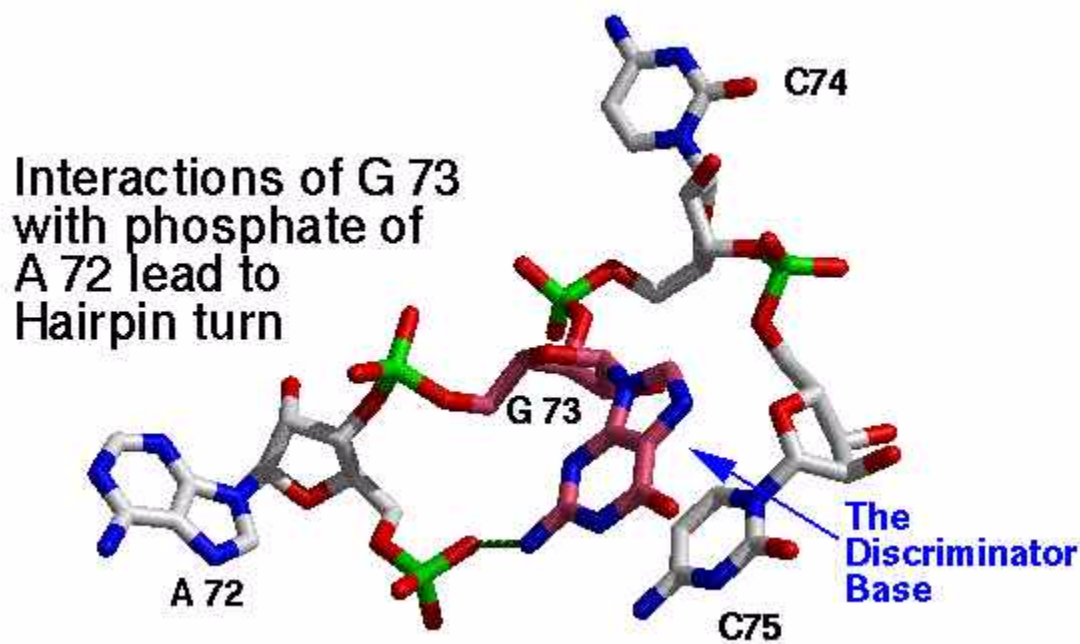

Figure 12. Conformation promoted by G73's interaction with the phosphate of A72.

Thus, we see in these two examples, that the means by which an aminoacyl tRNA synthetase recognizes its tRNA substrate can be as simple as interactions with a single atom group or be a complex set of interactions involving a significant number of bases in the tRNA molecule, using both direct and indirect readout.