Recognition of $tRNA^{Cys}$ by the *E. coli* cysteinyl-tRNA synthetase: *in vivo* and *in vitro* studies.

Thesis by

George A. Komatsoulis

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Submitted February 16, 1993)

Acknowledgment:

The path leading to a Ph.D. degree is always crooked, and so I wish to acknowledge some of the people who helped to direct the turns. I would like to thank Joseph J. Kroto and Chris Pouler of the Prince George's County Public Schools who aimed me away from computer science and towards microbiology. I also wish to thank my undergraduate advisor, Bill Ghiorse, and the members of his lab, especially Lee Adams and Nancy Schenck, who tolerated my bumbling around as I got something of a feel for working in the laboratory.

In the Abelson Laboratory, I wish to acknowledge the invaluable assistance of Jennifer Normanly, N. Kyle Tanner, Michael W. Clark, Phil Green, Shawn Westaway, and Giuseppe Tocchini-Valentini. Special mention must go to Peggy Saks and Jeff Sampson, who I am sure consider me their personal nemesis by now, and to Christine O'Day and David Horowitz, my friends in and out of the lab (in David's case, despite a bumpy start involving some phenol/chloroform and parafilm).

On the more personal side I wish to thank my parents Joan and Tony Komatsoulis for supporting what I chose to do all of these years. I also wish to thank my Aunt and Uncle, Diane and Jack Peci, whose San Diego home provided many needed escapes. There are several friends and neighbors; Kurt Eakle, Anna Bruhn, Heidi Skidmore, Gabrielle Szechenyi, Robin Burks, Debra Hori, Linda Evans, Larry Fuller, Karin Fuller, and Dylan and Mike O'Day who made my time in California something special. Finally I wish to express my deep gratitude to my wife, Ann Miller, who has been with me through thick and thin for the last 6 years.

And last (but by no means least) I want to thank my advisor John Abelson, for giving me the freedom to pursue my own ideas. He truly made all of this possible.

Abstract:

A study of the recognition of $tRNA^{Cys}$ by *E. coli* cysteinyl-tRNA synthetase using *in vivo* and *in vitro* methods was performed. All three anticodon nucleotides, the discriminator U73, and some element(s) within the tertiary domain (the D stem/loop, the TΨC stem/loop and extra loop) are important for recognition; the anticodon stem and acceptor stem appear to contain no essential elements. A T7 RNA polymerase transcribed $tRNA^{Cys}$ is a 5.5-fold worse substrate than native $tRNA^{Cys}$ (in terms of the selectivity constant, k_{cat}/K_m) mainly due to an increase in K_m. This may reflect recognition of modified nucleotides or subtle effects on the folding of the tRNA. The greatest loss of specificity caused by mutation of a single nucleotide occurs when the discriminator U73 is changed; k_{cat}/K_m declines 3 to 4 orders of magnitude depending on the substitution. Mutations in the wobble nucleotide of the anticodon also cause reductions in the selectivity constant of 3 orders of magnitude, while mutations in the other anticodon nucleotides caused lesser effects. Interestingly, a C35A mutation had no effect on aminoacylation by the cysteinyl-tRNA synthetase. Several amber suppressor tRNAs were constructed whose *in vivo* identity did not correlate with their *in vitro* specificity, indicating the need for both types of experiments to understand the factor(s) which maintain tRNA specificity. Future *in vitro* experiments will attempt to explain the *in vivo* discrimination between the glycine, phenylalanine, and cysteine tRNAs by the cysteinyl-tRNA synthetase. Finally, these results suggest that the notion that a small set of isoacceptor specific elements define tRNA identity (the so-called "second genetic code") is incorrect. A better model is based on competition between synthetases for tRNA substrates which contain differing amounts of partially overlapping identity determinants.

Contents:

Chapter I

tRNA discrimination in *E. coli*

General Considerations:

Importance of tRNA recognition:

The accurate transformation of genes into their physical manifestation as the protein products of the cell is critically dependent upon the fidelity of translation. To maintain high levels of translational fidelity, aminoacyl-tRNA synthetases must accurately discriminate between cognate and non-cognate tRNA species. Two characteristics of the system of tRNAs increase the difficulty of this task. First, all tRNAs share a similar three-dimensional structure, presumably imposed by the requirement that they interact with the ribosome, mRNA, and elongation or initiation factors. Second, although there are twenty aminoacyl-tRNA synthetases in *E. coli*, one for each normal amino acid, many synthetases charge more than one tRNA (Komine et al., 1990, Sprinzl et al., 1991). These groups of tRNAs which are aminoacylated with the same amino acid are known as isoacceptors; the members of an isoaccepting group frequently exhibit great sequence or structural variations (Komine et al., 1990, Sprinzl et al., 1991).

tRNA structure:

The *in vivo* function of tRNAs dictate that they share a certain degree of structural similarity at both the secondary and tertiary levels. Since this structural conservation defines one of the problems of tRNA discrimination, and dictates the features which are accessible for interactions with synthetases some knowledge of tRNA structure is useful. The most common model of tRNA structure is yeast $tRNA^{Phe}$, so this will be the reference for this section. Where there is data to suggest variations in other tRNAs, the observed differences will be described in the text.

Figure 1 shows the cloverleaf representation of the secondary structure of yeast $tRNA^{Phe}$. For convenience, the molecule can be divided into several

secondary structural domains: the acceptor stem, D stem and loop, anticodon stem and loop, variable loop, and TΨC stem and loop. The acceptor stem, site of amino acid attachment, contains 7 base pairs and four single stranded nucleotides in all but one normal *E. coli* elongator tRNA. (There are eight base pairs and only three single stranded nucleotides in $tRNA^{His}$.) Three of the single stranded nucleotides, the terminal CCA, are invariant in all tRNAs from all species. The amino acid is attached to the terminal adenosine residue on either the 2' or 3' hydroxyl group depending on the aminoacyl-tRNA synthetase (for review see Moras, 1992). The other single stranded base is nucleotide 73, the discriminator (Crothers et al., 1972). The D stem and loop derive their names from the one or more dihydrouridine residues present in this loop in all tRNAs. The D loop contains two absolutely conserved G residues (G18G19), involved in tertiary interactions with the TΨC loop (see below). The number or residues in the D loop, and the relative position of the invariant G19G20 sequence varies from tRNA to tRNA, as does the size of the stem (between 3 and 4 Watson-Crick base pairs depending on the tRNA). The anticodon stem and loop obviously derive their name from the presence of the anticodon, nucleotides 34-36. The variable loop is as the name suggests; in *E. coli* tRNAs it ranges in size from 4 nucleotides up to 21 nucleotides. The size of this loop is often used to classify tRNAs; type I tRNAs have short variable loops (4-5 nucleotides) while type II tRNAs have long (≥13 nucleotides) variable loops. The TΨC stem and loop, named for the three invariant residues in the loop are of constant size in all *E. coli* tRNAs, 5 base pairs in the stem, 7 bases in the loop. The secondary structure of a tRNA is often described in an abbreviated manner by the number of Watson-Crick base pairs in the D-stem and the number of bases in the variable loop. Thus a tRNA with 4 base pairs in the D stem and 5 bases in the variable loop is of type D4V5.

Crystal structures of two uncomplexed tRNAs, yeast $tRNA^{Phe}$ (Kim et al., 1974) and $tRNA^{Asp}$ (Westhof et al., 1985) have been solved, providing models of tertiary structure for D4V5 (Phe) and D3V4 (Asp) tRNAs. These structures are thought to be representative of most type I tRNAs since many of the interactions which produce the tertiary structure involve invariant nucleotides. The crystal structure of both tRNAs shows an "L" shaped molecule with extensive tertiary interactions between the secondary structural elements. Figure 2A shows the tertiary structure of $tRNA^{Phe}$ with the various secondary structural domains labeled. The tertiary interactions which produce the structure in Figure 2A are shown on a cloverleaf representation of $tRNA^{Phe}$ in Figure 2B. The structure shows extensive contacts between the D stem and loop and the variable and TΨC loops. Several of the tertiary interactions deserve special mention. The base pair between U8 and A14 is usually called the reverse-Hoogsteen base pair. The base pair between G15 in the D loop and C48 in the variable loop is called the Levitt pair (Levitt, 1969). All *E coli* tRNAs with the exception of $tRNA^{Cys}$ (which is G15:G48) have a G15:C48 or an A15:U48 Levitt pair (Komine et al., 1990, Sprinzl et al., 1991). Nucleotides 26 and 44 are frequently referred to as the "propeller twist" nucleotides, because their base pairing interaction is not planar, but twisted like the blades of a propeller. The tertiary structure of the tRNA produces a feature known as the variable pocket (Ladner, 1975), a group of 5 single stranded nucleotides which are placed in close proximity in an accessible depression near the bend of the "L". The variable pocket nucleotides are 16, 17, 20, 59, and 60; the location of these nucleotides is indicated on both Figure 2A and 2B.

The crystal structure of $tRNA^{Asp}$ (Westhof et al., 1985) also shows an "L" shaped molecule, but there are some minor differences in the tertiary structure. There is no tertiary interaction between the invariant G19 and C56, as seen in

$tRNA^{Phe}$. Also the nature of the interaction between nucleotide 21 and the reverse-Hoogsteen base pair (U8:A14) is slightly different. In yeast $tRNA^{Phe}$ nucleotide 21 is only hydrogen bonding to the phosphate backbone of U8, whereas in $tRNA^{Asp}$ the base appears to hydrogen bond with both the phosphate backbone of U8 and the base of nucleotide 14. In addition, there is functional evidence involving aminoacylation by the yeast aspartyl- and phenylalanyl-tRNA synthetases which support the theory that the conformation of $tRNA^{Asp}$ is subtly different than of yeast $tRNA^{Phe}$.(see Giegé et al., 1990, Perret et al., 1992)

Recognition and Identity:

Discrimination by aminoacyl-tRNA synthetases between structurally similar cognate and non-cognate tRNAs involves two elements, recognition and identity (for review see Schulman, 1991). Recognition is the term applied to the interaction between a tRNA and its cognate synthetase, and the elements on a tRNA which are required for this interaction are called recognition elements. Recognition elements can be specific nucleotides which are recognized by a synthetase, or structural features which correctly orient other elements which are directly recognized by the synthetase. All recognition elements are by definition positive elements, that is, the presence of the element aids in the creation of a productive interaction between tRNA and synthetase. Removal of a recognition element therefore causes a reduction in the catalytic efficiency of aminoacylation by the cognate synthetase. Identity is the term used to describe the amino-acid-acceptance of a tRNA *in vivo*, and the sequence and structural elements which are required to confer a specific amino-acid-acceptance are called identity elements. Unlike recognition elements, identity elements can act in either a positive or a negative fashion. A positive identity element is a recognition element, that is, the presence of the element aids in the formation of

a productive interaction with the cognate synthetase. Negative elements on the other hand, work by preventing productive interactions with non-cognate synthetases. From a phenotypic standpoint, removal of either type of identity element should have the same effect, a loss of *in vivo* identity.

***In vivo* versus *in vitro* experiments:**

There are two obvious ways to study the interaction of tRNA and aminoacyl-tRNA synthetase; both involve changing the sequence and/or structure of tRNAs and seeing what effect the mutations have on aminoacylation of the tRNA. The difference comes from the way that the effects are determined; either a tRNA may be put into a cell, and its amino-acid-acceptance determined, or the tRNA may aminoacylated *in vitro* and the kinetic constants for the new tRNA compared against those of a wild-type tRNA. The different environments in which the two types of experiments occur define the aspects of the tRNA:synthetase interaction which each type of experiment can study. Since *in vitro* experiments generally involve one, or at most two aminoacyl-tRNA synthetases, they study tRNA recognition, not tRNA identity. *In vivo* experiments on the other hand take place in a milieu where all twenty aminoacyl-tRNA synthetases are present, hence they study tRNA identity rather than recognition. Ultimately, both types of experiments are required to understand discrimination, since (as indicated above) the *in vivo* effect of removing an identity element is the same regardless of whether the element is positive or negative. A distinction can be made *in vitro* between a positive and a negative identity element because removal of a purely negative element will have little effect on aminoacylation with the cognate synthetase, but only an *in vivo* experiment could suggest that the negative element was there in the first place.

Practical considerations also dictate what can and cannot be studied by the two types of experiments. The experimental details of *in vitro* studies are

fairly straightforward; mutant tRNAs are prepared and then aminoacylated at different RNA concentrations to estimate the value of the coefficients of the Michaelis-Menten equation (k_{cat} and K_m). The only variations involve the source of the RNA and the way that the mutants are generated. Initially, tRNAs had to be isolated *in vivo* and then mutated by chemical or enzymatic means, but this was later supplanted by preparing RNA from cloned tRNA genes with mutations introduced at the DNA level. A simpler way to isolate tRNAs was pioneered by Sampson and Uhlenbeck (1988) who used T7 RNA polymerase to produce an unmodified yeast $tRNA^{Phe}$ which was aminoacylated with nearly the same kinetics as the native tRNA. This method has the advantage that it can be used to rapidly prepare and test the effects of any mutation at any position in the tRNA; the limitation of this method, of course, is that it cannot study the effects of modified nucleotides, which have been shown to be important for recognition by several aminoacyl-tRNA synthetases including the isoleucyl- (Muramatsu et al., 1988) and lysyl- (Tamura et al., 1992) tRNA synthetases.

To understand the practical strengths and limitations of *in vivo* studies, it is important to understand the methods used in such experiments. Most *in vivo* studies attempt to determine the identity elements for an isoaccepting group by performing an identity swap; that is, mutations are made in a noncognate tRNA until it is converted into a cognate tRNA. This approach was first used by Normanly et al. (1986a) who converted a leucine accepting tRNA into a serine accepting tRNA. The most common type of *in vivo* study uses an amber or opal suppressor tRNA as the experimental substrate. The *in vivo* identity of the tRNA can then be determined by studying the effect of the suppressor tRNA on a reporter gene with a nonsense mutation, either by testing for an amino acid specific phenotype or by directly sequencing the protein product. The limitation of this method, of course is that the anticodon must be either CUA or UCA, due

to the requirement that the tRNA be a nonsense suppressor. Since many aminoacyl-tRNA synthetases use the anticodon as recognition and/or identity elements, this is a serious problem (see below and Schulman, 1991). In addition, the amber anticodon contains residues which are important for the recognition of the lysyl-tRNA synthetase (U35, Tamura et al., 1992) and glutaminyl-tRNA synthetase (C34 and U35, Jahn et al., 1991). Thus identity elements for these two tRNAs are introduced into all identity swap experiments which use amber suppressor tRNAs. A new method (Chattapadhyay et al., 1990) allows the role of the anticodon to be investigated by incorporating the anticodon of interest into an *E. coli* initiator tRNA ($tRNA^{fmet}$), and then sequencing protein produced from a reporter gene in which the corresponding codon replaces the normal AUG initiation sequence. Although useful, this method has weaknesses also, principally that the only tRNA body which can be used is that of $tRNA^{fmet}$.

Competition as the driving force behind identity:

In vivo, there are two factors which define the identity of a tRNA; the total catalytic efficiency of the reaction with the cognate synthetase, and the sum of the catalytic efficiencies with the other twenty synthetases. To retain a specific *in vivo* identity, a tRNA must contain enough positive and negative identity elements so that the cognate synthetase can successfully compete against all other synthetases for the substrate. If the competitive model is true then altering the factors which influence the total catalytic efficiency of aminoacylation (namely k_{cat} and K_m) should affect identity. Modeling studies using tRNAs with dual identities suggest that the primary factor which would affect discrimination *in vivo* is k_{cat} (Hou and Schimmel, 1989a). While k_{cat} is fixed its manifestation V_{max} can be changed by increasing synthetase concentration. Thus, if competition is important to recognition, changing the balance between the

concentrations of the tRNA and the synthetase should affect the fidelity of aminoacylation and the identity of certain tRNAs.

Several workers have dramatically demonstrated the principle that the balance between synthetase and tRNA can in fact, affect identity. Swanson et al. (1988) observed that in the presence of excess glutaminyl-tRNA synthetase an amber suppressor tRNA which was normally aminoacylated with tyrosine was also aminoacylated with glutamine. Increasing the concentration of $tRNA^{Gln2}$ could compensate for the effect of the increased levels of synthetase because the cognate $tRNA^{Gln2}$ could out compete the amber suppressor for the glutaminyl-tRNA synthetase. The complementary effect was observed by Hou and Schimmel (1989a). They reported that a mutant amber suppressor $tRNA^{Tyr}$ which contained recognition elements for the alanyl-tRNA synthetase, was aminoacylated with tyrosine (and about 5% glutamine) but no alanine *in vivo*. However, the addition of a plasmid which overexpressed the alanyl-tRNA synthetase caused the tRNA's identity to change so that it was aminoacylated with a mixture of tyrosine and alanine. Recently Sherman et al. (1992) was able to directly demonstrate both *in vivo* and *in vitro* that competition can affect identity. The *supF* amber suppressor tRNA (derived from $tRNA^{Tyr}$) is aminoacylated *in vivo* with both tyrosine and glutamine. Aminoacylation by the glutaminyl-tRNA synthetase is prevented when the competing tyrosyl-tRNA synthetase is overproduced in the cell. Similarly, in an *in vitro* aminoacylation assay the fraction of *supF* tRNA aminoacylated with glutamine was reduced 60% when tyrosyl-tRNA synthetase was added. This effect is not specific to this tRNA, the addition of glutamate-tRNA synthetase reduced the amount of misacylation of $tRNA^{Glu}$ with glutamine in proportion to the amount of GluRS added.

Features used by synthetases to identify tRNAs:

General:

The tRNA sequence elements which are used to distinguish different isoacceptors vary from group to group. Nevertheless, these elements tend to fall in only five major regions of the tRNA. These regions are the anticodon, the acceptor stem (excluding the discriminator), the discriminator, the variable pocket, and the extra loop (in type II tRNAs). Although this may appear on the surface to be a fairly random grouping, these elements constitute the most accessible elements on the tRNA. The anticodon, discriminator, variable pocket and many extra loop nucleotides are single stranded and present no physical impediments to interactions (i.e., they are not blocked by other residues), while the acceptor stem and the remaining variable loop nucleotides are helical RNA, whose minor groove is accessible by proteins. The next section details which synthetases use elements in these regions as part of their recognition or identity sets.

Anticodon:

The obvious solution to identifying tRNAs is for the synthetase to contact and "read" the anticodon. While logical from an informational perspective, there were two problems from a mechanistic standpoint. First is the problem of isoacceptors; namely how can a protein recognize several different nucleotides at one position while excluding non-cognate tRNAs which also have one or more anticodon nucleotides in common? This is a particular problem for the seryl-tRNA synthetase, which has no absolutely conserved anticodon nucleotides (Komine et al., 1990, Sprinzl et al., 1991). The second problem is the question of distance; how can a protein transfer the identity information gained at the anticodon to the active site at the end of the acceptor stem? Although allosteric enzymes were available as models for enzymes which can change substrate K_m's by interaction at a second, distal site, there were no

models in which the modulation of activity was at the level of k_{cat} (at least until the cocrystal structure of *E. coli* $tRNA^{Gln}$:glutaminyl-tRNA synthetase was solved, Rould et al., 1989).

There was however, growing evidence (see for example, Saneyoshi and Nishimura, 1971, Squires and Carbon, 1971, and Kern and LaPointe, 1979) that the anticodon was involved in tRNA recognition and identity. Schulman and Pelka (1983) are credited with the first definitive proof that the anticodon is recognized *in vitro* by aminoacyl-tRNA synthetases. They showed that the anticodon was important for recognition of the *E. coli* initiator $tRNA^{fmet}$ by the methionyl-tRNA synthetase by demonstrating that substitutions of the anticodon nucleotides reduced the rate of aminoacylation *in vitro*. Direct proof of the importance of the anticodon in identity was somewhat slower in coming, although the misacylation with glutamine and lysine of several amber suppressor tRNAs (anticodons switched to CUA) suggested that the anticodon was important for identity in some cases (Normanly et al., 1990). Although comparison of the effects of amber and opal suppressor mutants proved the importance of some anticodon residues in some systems *in vivo* (for instance McClain et al., 1990), direct testing of the effects of anticodon nucleotides was not available until the development of the *E. coli* initiation test system (Chattapadhyay et al., 1990, and above).

Application of *in vivo* and *in vitro* methods have shown that the anticodon is involved in discrimination in 17 isoaccepting groups in *E. coli* (see Table I). Only three *E. coli* aminoacyl-tRNA synthetases completely ignore all three anticodon nucleotides, those for alanine, leucine, and serine. Two of these synthetases (the leucine and serine enzymes) recognize type II tRNAs with the distinctive long variable loop.

Acceptor stem (excluding discriminator):

The theoretical problems associated with anticodon recognition (especially the distance problem) are not associated with acceptor stem recognition since this region is in close proximity to the site of the aminoacylation reaction. Recognition in the acceptor stem has proven very amenable to study by both *in vivo* and *in vitro* methods using normal tRNAs, amber and opal suppressor tRNAs, and model substrates such as minihelices (a stem loop structure composed of the acceptor stem connected to the TΨC stem and loop), microhelices (the acceptor stem closed by an RNA loop), and linear RNA duplexes with as few as 4 base pairs (Schulman, 1991, Musier-Forsyth et al., 1991a). Historically, the initial demonstration that acceptor stem elements were important in identity was performed by Normanly et al. (1986a) who found that serine acceptor stem nucleotides were required to switch a leucine accepting tRNA into a serine accepting tRNA. The first evidence of acceptor stem recognition was from Hou and Schimmel (1988) who demonstrated that a G3:U70 base pair was the prime determinant of alanine acceptor activity.

Acceptor stem discrimination (excluding the discriminator nucleotide 73) has been shown to be important in eight isoaccepting groups (Table II). In at least one isoaccepting group (alanine) acceptor stem recognition is the dominant component in both the recognition and identity processes (see references in Table II). To date only two aminoacyl-tRNA synthetases have been shown to ignore the acceptor stem (excluding the discriminator) in recognition: the phenylalanyl- and cysteinyl-enzymes (Table II).

Discriminator:

The discriminator was originally proposed as an important identity/recognition element by Crothers et al. (1972) based on the limited number of tRNA sequences then available. Although not a universal recognition

element the discriminator has been shown to be important in recognition and/or identity in 18 isoaccepting groups (Table III). Only the serine and threonine enzymes have been shown to not utilize the discriminator in either recognition and/or identity. There are currently two models for discriminator action; one is direct recognition of the single stranded nucleotide by the synthetase. The other (based on evidence from the $tRNA^{Gln}$/glutaminyl-tRNA synthetase cocrystal structure) is that the discriminator helps position the CCA terminus in the active site by intramolecular hydrogen bonding with the acceptor stem (Rould et al., 1989, Perona et al., 1989).

Variable pocket:

The variable pocket, first proposed as a defined region by Ladner et al. (1975), has also been shown to be involved in tRNA discrimination both *in vivo* and *in vitro*. The variable pocket is composed of single stranded nucleotides 16, 17, 20, 59 and 60 (numbering of Sprinzl et al., 1991), which are placed in close proximity in a depression near the bend in the "L" structure of the molecule (see above). Since all 5 nucleotides are single stranded, they would be particularly amenable to making sequence specific contacts with synthetases. The paradigm for variable pocket recognition is the yeast phenylalanyl-tRNA synthetase, which requires G20 for efficient aminoacylation (Sampson and Uhlenbeck, 1988, Sampson et al., 1989, Sampson et al., 1992). To date, five *E. coli* isoaccepting groups have been shown to utilize variable pocket elements as part of their identity and/or recognition mechanisms (Table IV).

Unusual elements:

Isoaccepting groups which are known to depend on other types of elements to define their identity are shown in Table V. These elements fall into the general categories of modified or nonstandard nucleotides, central core nucleotides (which includes the D-stem), and large variable loops. As seen in

Table V, there are two known cases of modified residues being involved in tRNA recognition: lysidine34 in $tRNA^{Ile}$ and mnm^5s^2U34 found in $tRNA^{Lys}$. The nonstandard G at the -1 position in $tRNA^{His}$ is known to be important for recognition by the histidyl-tRNA synthetase. Central core nucleotides are known to be important in recognition or identity by the phenylalanyl-, glutaminyl-, and seryl-tRNA synthetases. Finally, the large variable loops found in type II tRNAs are important for recognition and/or identity of the seryl- and tyrosyl-tRNA synthetases.

Integrating data from different sources:

Introduction:

Studying the experiments whose data are summarized in the tables it can be seen that the correspondence between the *in vitro* and *in vivo* columns is not always 100%. As indicated earlier, identity and recognition are not the same process. For this reason it is instructive to compare the results of identity and recognition studies. It is also instructive to compare the results of recognition and identity experiments with structural studies from cocrystals of tRNAs and aminoacyl-tRNA synthetases. Thus, the remainder of this chapter will focus on comparisons of data obtained from different types of experiments. Specifically, the relationship between the recognition and identity sets in two well studied *E. coli* isoaccepting groups, alanine and phenylalanine, will be investigated, and the results of the only tRNA:aminoacyl-tRNA synthetase cocrystal structure (glutamine) will be compared to discrimination data obtained from molecular genetic experiments.

Alanine:

The primary determinant of alanine recogntion and identity in *E. coli* (and many other organisms as well) is a single base pair G3:U70 in the acceptor stem (Figure 3, Hou and Schimmel, 1988, Hou and Schimmel, 1989b, Shi et al.,

1990). Several theories have been advanced to explain the importance of the wobble (G:U) pair (see for instance, McClain et al., 1988c) but a study with nucleotide analogs has conclusively demonstrated that the function of the G:U is to leave the exocyclic amine of G3 unpaired in the acceptor stem (Musier-Forsyth et al., 1991b).

Several other nucleotides (Figure 3) contribute to alanine recognition *in vitro*, these include the discriminator A73 (Shi et al., 1990, Shi and Schimmel, 1991, Tamura et al., 1991b), base pair G2:C71 (Francklyn et al., 1992) and G20, a variable pocket nucleotide in the D-loop (Tamura et al., 1991b). All of these elements are less important than the wobble pair at 3:70; that is changing these elements causes less dramatic effects than those caused by mutations in the wobble pair.

The identity set of $tRNA^{Ala}$ matches the recognition set fairly well. The G3:U70 base pair is a major determinant of alanine identity; transfer of this element into amber suppressors of $tRNA^{Lys}$ (McClain et al., 1988c), $tRNA^{Phe}$ or $tRNA^{Cys}$ (Hou and Schimmel, 1988) is sufficient to switch the identity of the tRNA to Ala (in Lys and Cys backgrounds) or at least add some alanine identity (in a Phe background). The discriminator has also been shown to be an identity element (McClain et al., 1991b), as was base pair 2:71 and the unpaired G20 (McClain et al., 1988c). Only two elements were found to be part of the identity set which were not previously defined as recognition elements, base pair 1:72 (McClain et al., 1991b), and C60 (McClain et al., 1988a) a variable pocket nucleotide in the TΨC loop. The effect of G1:C72 on aminoacylation with alanine is unknown, but C60 mutations have only minor effects (Tamura et al., 1991b), suggesting that C60 is a negative identity element. The fact that a C60U mutation in $tRNA^{Ala}$ causes some lysine to appear in an amber suppressor tRNA tends to support this view.

Phenylalanine:

The definitive study of recognition by the *E. coli* phenylalanyl-tRNA synthetase (FRS) is by Peterson and Uhlenbeck (1992); all recognition data reported herein are derived from this study (see Figure 4). They found that the important recognition elements for FRS found in *E. coli* $tRNA^{Phe}$ are the anticodon, U20 and U59 in the variable pocket, the propeller twist nucleotides, and three residues (10,25, and 45) which form a tertiary interaction in the central core of the molecule. Peterson and Uhlenbeck also list the discriminator as a recognition element, although mutations at this site cause reductions in k_{cat}/K_m of only 1.6- to 2-fold in an *E. coli* $tRNA^{Phe}$ background. In a yeast $tRNA^{Phe}$ background 6-fold effects are observed when the discriminator is mutated, hence the element was included in the recognition set.

The identity set for *E. coli* $tRNA^{Phe}$ was investigated by McClain and Foss (1988) in amber suppressors and by Chattapadhyay et al. (1990) using an initiation system. The identity set is composed of at least 11 nucleotides: A36 in the anticodon (Chattapadhyay et al., 1990) U20, U59, and U60 (and possibly U16 and U17) in the variable pocket, G27:C43 and G28:C42 in the anticodon stem, G44 and U45 in the variable loop, and the discriminator A73 (McClain and Foss, 1988). The effects on phenylalanine identity of the other two anticodon nucleotides, the D stem, 16 and 17 in the D loop, G1:C72, A26 in the propeller twist, and the remaining 3 nucleotides of the variable loop have not been examined.

Several correlations appear immediately. Anticodon nucleotide A36 is both an identity element and a recognition element, as are nucleotides U20 and U59 in the variable pocket, G44 and U45 in the variable loop, and the discriminator U73. Equally apparent is the fact that U16, U17, and U60 in the variable pocket, and G27:C43 and G28:C42 in the anticodon stem which

appeared to be important for identity are not important in the recognition process. Similarly, nucleotides 34 and 35 which seemed to be unimportant for identity were in fact quite important for recognition. No correlations between the *in vivo* and *in vitro* behavior of the remaining recognition nucleotides (10, 25, and 26) can be ascertained because they have not been tested in both systems.

Although at first approximation a jumble, this data can be adequately integrated. The elements contained in both the identity and recognition sets are obviously positive identity elements, while the nucleotides which are found to be important for identity but not important for recognition (U16, C17, U60 in the variable loop and G27:C43 and G28:C42 in the D stem) are negative identity elements, whose sole function is to prevent productive interactions with non-cognate synthetases. Since the variable pocket nucleotides are used as recognition elements by many aminoacyl-tRNA synthetases (see above and Schulman, 1991) U16, C17 and U60 are logical negative elements. The anticodon stem elements are less understandable, since no known aminoacyl-tRNA synthetase uses the anticodon stem as a recognition element. However C34 and U35 in the amber anticodon are known to be recognition elements for the glutaminyl-tRNA synthetase (Jahn et al., 1991). Perhaps these base pairs can change the orientation of the anticodon with regard to their binding pocket on the glutaminyl-tRNA synthetase. There is circumstantial evidence to support this theory; Komatsoulis and Abelson (1993) found that changes in the anticodon stem can affect the level of glutamine identity of amber suppressor tRNAs *in vivo*. To integrate the rest of the data it is necessary to presuppose that $tRNA^{Phe}$ is overspecified. If $tRNA^{Phe}$ is overspecified then the loss of one or two recognition elements (i.e., G34 and A35) may not affect identity significantly. The fact that at least five isoaccepting groups which recognize anticodon

elements retain *in vivo* identity as amber suppressors suggests that this is not an unreasonable postulate.

Glutamine:

A cocrystal structure of the glutaminyl-tRNA synthetase and $tRNA^{Gln}$ has been solved to 2.5Å resolution (Rould et al., 1989, Perona et al., 1989, Rould et al., 1991), and concurrent studies conducted of the recognition of the tRNA by the synthetase (Jahn et al., 1991, Hayase et al., 1992). The result of this work is a detailed, coherent view of the interaction of a tRNA and an aminoacyl-tRNA synthetase.

Figure 5 shows the nucleotides which have been shown to be involved in specific interactions with the synthetase. Unsurprisingly, the synthetase makes extensive stacking and hydrogen bonding contacts with the CCA terminus, particularly C74 and A76, the site of amino acid attachment. Specific contacts are seen between the synthetase and the exocyclic amines of G2 and G3 (which are found in the minor groove of the acceptor helix) partially mediated by a water molecule. Specific contacts are also seen between the exocyclic amine of G10 and the synthetase, and between C16 and Gln13 of the synthetase. All three anticodon nucleotides in the anticodon (34-36) and two nucleotides in the anticodon loop (37 and 38) are also recognized by the synthetase in a sequence specific manner. Although not recognized directly U32 and U33 are required to base pair with nucleotides 37 and 38 for proper interaction of the anticodon with the synthetase. Base pair 1:72 and nucleotide 73 are also not involved in nucleotide specific contacts with the synthetase, but the cocrystal indicates that they form structures which are important for the tRNA to adopt the proper configuration for aminoacylation. The first base pair is melted by the enzyme; thus the interaction is specific for a base pair which is disrupted easily. The exocyclic amine of the discriminator, G73, forms an intramolecular

hydrogen bond with the phosphate of residue A72 (which is not base paired after interaction with the synthetase), to maintain a hairpin turn that positions the CCA terminus in the active site.

All of these residues have been studied for their effects on *in vitro* aminoacylation with purified glutaminyl-tRNA synthetase (Jahn et al., 1991, Hayase et al., 1992). The conclusions which can be drawn from this data is that mutations in all but one of the contact residues tested (G2, G3, G10, C34, U35, G36, A37, and U38) cause reductions in the total catalytic efficiency of the aminoacylation reaction. The only exception to this was C16; a C16U mutation caused only a 2-fold reduction in total catalytic efficiency of aminoacylation. Furthermore, mutations which affect the helix at the G2 or G3 base pair (i.e., by introducing a helical irregularity with a G:U wobble pair) also cause reductions in the total catalytic efficiency. The cocrystal data indicated that the interactions with G2, G3 and G10 were mediated through the exocyclic amino group of the guanosines. This was directly tested by making mutant tRNAs in which inosine residues individually replaced each of the G residues. Since inosine lacks the minor groove exocyclic amine, these replacements would be expected to affect k_{cat}/K_m, and in fact this is what is observed. Mutations at the discriminator position (which helps hold the CCA terminus in place) also cause reductions in the catalytic efficiency of aminoacylation. The cocrystal structure indicates that the only important characteristic of the U1:A72 base pair is that it is easily denatured, and experiments bear this out. A G1:A72 mismatch has the same catalytic efficiency as wild-type U1:A72 $tRNA^{Gln}$, although the wild-type tRNA has a K_m approximately four times lower than the G1:A72 mutant. Furthermore, a tRNA with a deletion of nucleotide 1 (i.e., the transcript starts at G2) is only a 2-fold worse substrate in terms of the selectivity constant than the U1:A72 wild-

type tRNA (although K_m is increased 6-fold in the deletion mutant). In summary, there is excellent fit between the results of genetic and physical studies.

Concluding remarks:

There is a great deal of order emerging from the chaos which until recently described the results of studies on tRNA recognition and identity. What had previously seemed to be a set of idiosyncratic rules governing the interaction between synthetases are becoming clearer as more study indicates that tRNA:synthetase interactions generally occur in five well defined regions of the tRNA: the anticodon, the acceptor stem, the discriminator, the variable pocket, and the extra arm (in type II tRNAs). Additionally, a new paradigm for *in vivo* identity has developed out of the relatively recent understanding of the importance of competition between synthetases in maintaining fidelity (see Söll, 1990). The new paradigm states that identity is determined by competition among synthetases for substrate molecules which contain different numbers of partially overlapping identity elements. No doubt, as work in this field continues, the gaps in Tables 1 to 4 will be filled, and our understanding of tRNA discrimination will be more complete.

References:

Asahara, H., Himeno, H., Tamura, K., Hasegawa, T., Watanabe, K., Shimizu, M. (1993) in preparation.

Chattapadhyay, R., Pelka, H., & Schulman, L. H. (1990) *Biochemistry 29*, 4263-4268.

Crothers, D. M., Seno, T., and Söll, D. (1972) *Proc. Natl. Acad. Sci. USA 69*, 3063-3067.

Francklyn, C., & Schimmel, P. (1989) *Nature 337*, 478-481.

Francklyn, C., & Schimmel, P. (1990) *Proc. Natl. Acad. Sci. USA 87*, 8655-8659.

Francklyn, C., Shi, J. -P., & Schimmel, P. (1992) *Science 255*, 1121-1125.

Giegé, R., Florentz, C., Garcia, A., Grosjean, H., Perret, V., Puglisi, J., Théobald-Dietrich, A., & Ebel, J.P. (1990) *Biochimie 72*, 453-461.

Hasegawa, T., Himeno, H., Ishikura, H., & Shimizu, M. (1989) *Biochem. Biophys. Res. Comm. 163*, 1534-1538.

Hayase, Y., Jahn, M., Rogers, M. J., Sylvers, L. A., Koizumi, M., Inoue, H., Ohtsuka, E., & Söll, D. (1992) *EMBO J. 11*, 4159-4165.

Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K., Miura, K. -i., Shimizu, M. (1989) *Nucleic Acids Res. 17*, 7855-7863.

Himeno, H., Hasegawa, T., Ueda., T., Watanabe, K., & Shimizu, M., (1990) *Nucleic Acids Res. 18*, 6815-6819

Himeno, H., Hasegawa, T., Asahara, H., Tamura, K., Shimizu, M. (1991) *Nucleic Acids. Res. 19*, 6379-6382.

Hou, Y. -M., & Schimmel, P. (1988) *Nature 333*, 140-145.

Hou, Y. -M., & Schimmel, P. (1989a) *Biochemistry 28*, 4942-4947.

Hou, Y. -M. , & Schimmel, P. (1989b) *Biochemistry 28*, 6800-6804.

Jahn, M., Rogers, M. J., & Söll, D. (1991) *Nature 352*, 258-260.

Kern, D., & LaPointe, J. (1979) *Biochemistry 18*, 5819-5826.

Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Proc. Natl. Acad. Sci USA 71*, 4970-4974.

Komatsoulis, G. A., & Abelson J. N. (1993) submitted to *Biochemistry*.

Komine, Y., Adachi, T., Inokuchi, H., & H. Ozeki (1990) *J. Mol. Biol. 211*, 579-598.

Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C., and Klug, A. (1975) *Proc. Natl. Acad. Sci. USA 72*, 4414-4418.

Levitt, M. (1969) *Nature 224*, 759-763.

Maramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T., Yokoyama, S. (1988) *Nature 336*, 179-181.

McClain, W. H., & Foss, K. (1988a) *Science 240*, 793-796.

McClain, W. H., & Foss, K. (1988b) *Science 241*, 1804-1807.

McClain, W. H., Chen, Y. -MM., Foss, K., & Schneider, J. (1988c) *Science 242*, 1681-1684.

McClain, W. H., & Foss, K. (1988d) *J. Mol. Biol. 202*, 697-709.

McClain, W. H., Foss, K., Jenkins, R. A., & Schneider, J. (1990) *Proc. Natl. Acad. Sci. USA 87*, 9260-9264.

McClain, W. H., Foss, K., Jenkins, R. A., & Schneider, J. (1991a) *Proc. Natl. Acad. Sci. USA 88*, 6147-6151.

McClain, W. H., Foss, K., Jenkins, R. A., & Schneider, J. (1991b) *Proc. Natl. Acad. Sci. USA 88*, 9272-9276.

Moras, D. (1992) *TIBS 17*, 159-164.

Musier-Forsyth, K., Scaringe, S, Usman, N. , & Schimmel, P. (1991a) *Proc. Natl. Acad. Sci USA 88*, 209-213.

Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., & Schimmel, P. (1991b) *Science 253*, 784-786.

Normanly, J., Ogden, R. C., Horvath, S. J., & Abelson, J. (1986a) *Nature 321*, 213-219.

Normanly, J., Masson., J. -M., Kleina, L. G., Abelson, J., & Miller, J. H. (1986b) *Proc. Natl. Acad. Sci. USA 83*, 6548-6552.

Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J., & Miller, J. H. (1990) *J. Mol. Biol. 213*, 719-726.

Normanly, J. Ollick, T., & Abelson, J. (1992) *Proc. natl. Acad. Sci USA 89*, 5680-5684.

Pak, M., Pallanck, L., & Schulman, L. H. (1992) *Biochemistry 31*, 3303-3309.

Pallanck, L., & Schulman, L. H. (1991) *Proc. Natl. Acad. Sci. USA 88*, 3872-3876.

Pallanck, L., Li., S., & Schulman, L. H. (1992) *J. Biol. Chem. 267*, 7221-7223.

Park, S. J., & Schimmel, P., (1988) *J. Biol. Chem. 263*, 16527-16530.

Perona, M. J., Swanson, R. N., Rould, M. A., Steitz, T. A., & Söll, D. (1989) *Science 246*, 1135-1142.

Perret, V, Florentz, C., Puglisi, J. D., & Giegé, R. (1992) *J. Mol Biol. 226*, 323-333.

Peterson, E. T., & Uhlenbeck, O. C. (1992) *Biochemistry 31*, 10380-10389.

Rogers, M. J., & Söll, D. (1988) *Proc. Natl. Acad. Sci. USA 85*, 6627-6631.

Rogers, M. J., Adachi, T., Inokuchi, H., & Söll, D. (1992) *Proc. Natl. Acad. Sci. USA 89*, 3463-3467.

Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135-1142.

Rould, M. A., Perona, J. J., & Steitz, T. A. (1991) *Nature 353*, 213-218.

Sampson, J. R., & Uhlenbeck, O. C. (1988) *Proc. Natl. Acad. Sci USA 85*, 1033-1037.

Sampson, J. R., DiRenzo, A. B., Behlen, K. S., & Uhlenbeck, O. C. (1989) *Science 243*, 1363-1366.

Sampson, J. R., Behlen, L. S., DiRenzo, A. B., & Uhlenbeck, O. C. (1992) *Biochemistry 31*, 4161-4167.

Saneyoshi, M. & Nishimura, S. (1971) *Biochem Biophys Acta 246*, 123-131.

Schulman, L. H., & Pelka, H (1983) *Proc. Natl. Acad. Sci. USA 80*, 6755-6759.

Schulman, L. H., & Pelka, H (1985) *Biochemistry 24*, 7309-7314.

Schulman, L. H., & Pelka, H (1988) *Science 242*, 765-768.

Schulman, L. H., & Pelka, H (1989) *Science 246*, 1595-1597.

Schulman, L. H., & Pelka, H (1990) *Nucleic Acids Res. 18*, 285-289.

Schulman, L. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 23-87.

Sherman, J. M., Rogers, M. J., & Söll, D. (1992) *Nucleic Acids Res. 20*, 2847-2852.

Shi, J. -P., Francklyn, C., Hill, K., & Schimmel, P. (1990) *Biochemistry 29*, 3621-3626.

Shi, J. -P., & Schimmel, P. (1991) *J Biol. Chem. 266*, 2705-2708.

Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T., & Himeno, H. (1992) *J. Mol. Evol. 35*, 436-443.

Söll, D. (1990) *Experientia 46*, 1089-1096.

Sprinzl, M., Dank, N., Nock, S., & A. Schön (1991) *Nucleic Acids Res. 19*, 2127-2171.

Squires, C. & Carbon, J. (1971) *Nature New Biol. 233*, 274-277.

Swanson, R., Hoben, P., Sumner-Smith, M., Uemura, H., Watson, L., & Söll, D. (1988) *Science 242*, 1548-1551.

Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., & Shimizu, M. (1991a) *Biochem Biophys. Res. Comm. 177*, 619-623.

Tamura, K., Asahara, H. Himeno, H., Hasegawa, T., Shimizu, M. (1991b) *ISAS Research Note 489*, 1-13.

Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., & Shimizu, M. (1992) *Nucleic Acids Res. 20*, 2325-2339.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119-145.

Table I

Anticodon discrimination by *E. coli* aminoacyl-tRNA synthetases

Synthetase	*in vitro*	*in vivo*	References:
Alanyl-	No	?*	Park & Schimmel, 1988, Hou & Schimmel, 1988, Tamura et al., 1991b
Cysteinyl-	Yes	Yes *	Pallanck et al., 1992, Shimizu et al., 1992, Komatsoulis & Abelson, 1993
Aspartic-	Yes	Yes	Normanly et al., 1990, Shimizu et al., 1992
Glutamic-	Yes	Yes	Saneyoshi & Nishimura, 1971, Kern & LaPointe, 1979, Normanly et al., 1990
Phenylalanyl-	Yes	Yes *	Peterson & Uhlenbeck, 1992, Chattapadhyay et al., 1990, Pallanck & Schulman, 1991
Glycyl-	Yes	Yes	Squires & Carbon, 1971, McClain et al., 1991a
Histidyl-	Yes	?*	Shimizu et al., 1992
Isoleucyl-	Yes	Yes	Maramatsu et al., 1988, Pallanck & Schulman, 1990
Lysyl-	Yes	Yes*	McClain et al., 1990, Normanly et al., 1990, Tamura et al., 1992
Leucyl-	No	?*	Shimizu et al., 1992
Methionyl-	Yes	Yes	Schulman & Pelka, 1983, Schulman & Pelka, 1985, Schulman & Pelka, 1988, Schulman & Pelka, 1990, Chattapadhyay et al., 1990, Pallanck & Schulman, 1991
Asparaginyl-	Yes	?‡	Shimizu et al., 1992
Prolyl-	Yes	?*	Shimizu et al., 1992
Glutaminyl	Yes	Yes *	Schulman & Pelka, 1985, Jahn et al., 1991, Rould et al., 1989, Rould et al., 1991, Normanly et al., 1990.
Arginyl-	Yes	Yes	Schulman & Pelka, 1989, McClain et al., 1990, Tamura et al., 1992
Seryl-	No	?*	J. Sampson & M. Saks, pers. comm., Himeno et al., 1990
Threonyl-	Yes	Yes	Schulman & Pelka, 1990
Valyl-	Yes	Yes	Schulman & Pelka, 1988, Pallanck & Schulman, 1991, Tamura et al., 1991a
Tryptophanyl-	Yes	Yes	Himeno et al., 1991, Pak et al., 1992, Rogers et al., 1992
Tyrosyl-	Yes	?*	Hou & Schimmel, 1989a, Himeno et al., 1990,

* Amber suppressor retains its *in vivo* identity-data taken from Normanly et al., 1990, except for Leu & Ser (Normanly et al., 1986a) and Cys & Phe (Normanly et al., 1986b)

‡No functional amber suppressor $tRNA^{Asn}$ has ever been produced.

Table II
Acceptor stem discrimination by *E. coli* aminoacyl-tRNA synthetases (does not include discriminator nucleotide)

Synthetase	*in vitro*	*in vivo*	References
Alanyl-	Yes	Yes	Hou & Schimmel, 1988, McClain & Foss, 1988a, McClain et al., 1988c, Park & Schimmel, 1988, Francklyn & Schimmel, 1989, Hou & Schimmel, 1989a, Hou & Schimmel, 1989b, Shi et al., 1990, Musier-Forsyth et al., 1991b, Tamura et al., 1991b, Francklyn et al., 1992
Cysteinyl-	No	?	Komatsoulis & Abelson, 1993
Aspartic-	?	?	
Glutamic-	?	?	
Phenylalanyl-	No	No	McClain & Foss, 1988d, Peterson & Uhlenbeck, 1992
Glycyl-	Yes	Yes	McClain & Foss, 1991, Francklyn et al., 1992
Histidyl-	Yes	?	Francklyn & Schimmel, 1989, Himeno et al., 1989, Francklyn et al., 1992 (see Table V)
Isoleucyl-	?	?	
Lysyl-	?	?	
Leucyl-	?	?	
Methionyl-	?	?	
Asparaginyl-	?	?	
Prolyl-	?	?	
Glutaminyl	Yes	Yes	Rogers & Söll, 1988, Perona et al., 1989, Rould et al., 1989, Jahn et al., 1991, Rould et al., 1991, Hayase et al., 1992
Arginyl-	?	?	
Seryl-	?	Yes	Normanly et al., 1986a, Normanly et al., 1992
Threonyl-	?	?	
Valyl-	Yes	?	Tamura et al., 1991a
Tryptophanyl-	Yes	No*	Himeno et al., 1991,
Tyrosyl-	?	?	

* Mutations of base pairs 1:72, 2:71, and 3:70 in an opal suppressor retained Trp identity, but suppression efficiency was reduced.

Table III
Discriminator recognition and identity in *E. coli* aminoacyl-tRNA synthetases

Synthetase	*in vitro*	*in vivo*	References
Alanyl-	Yes	Yes	Hou & Schimmel, 1988, McClain et al., 1991b, Shi et al., 1990, Tamura et al., 1991b, Shi & Schimmel, 1991,
Cysteinyl-	Yes	?	Komatsoulis & Abelson, 1993, Pallanck et al., 1992
Aspartic-	Yes	?	Hasegawa et al., 1989
Glutamic-	Yes	?	Shimizu et al., 1992
Phenylalanyl-	No‡	Yes	Peterson & Uhlenbeck, 1992, Pallanck, Li, & Schulman, pers. comm.
Glycyl-	Yes	Yes	McClain et al., 1991a, Francklyn et al., 1992
Histidyl-	Yes	?	Himeno et al., 1989, Francklyn & Schimmel, 1990, Francklyn et al., 1992
Isoleucyl-	Yes	Yes	Shimizu et al., 1992, Pallanck, Li, & Schulman, pers. comm.
Lysyl-	Yes	Yes	McClain et al., 1990, Tamura et al., 1992
Leucyl-	Yes	Yes	Asahara et al., 1993, Normanly et al., 1992
Methionyl-	Yes	?	Shimizu et al., 1992
Asparaginyl-	Yes	?	Shimizu et al., 1992
Prolyl-	Yes	?	Shimizu et al., 1992
Glutaminyl	Yes	?	Perona et al., 1989, Rould et al., 1989, Jahn et al., 1991, Rould et al., 1991,
Arginyl-	Yes	Yes §	Tamura et al., 1992, McClain et al., 1990
Seryl-	No	Yes	Himeno et al., 1990, Normanly et al., 1992, Shimizu et al., 1992
Threonyl-	No	?	Shimizu et al., 1992
Valyl-	Yes	Yes	Tamura et al., 1991a, Pallanck, Li, & Schulman, pers. comm.
Tryptophanyl-	Yes	Yes	Himeno et al., 1991, Pak et al., 1992.
Tyrosyl-	Yes	?	Himeno et al., 1990

‡ Based on results of assay with *E. coli* phenylalanyl-tRNA synthetase and *E coli* $tRNA^{Phe}$. Assays with the same synthetase but in a yeast $tRNA^{Phe}$ background suggest that the discriminator is a recognition element.
§ Based on reduction of suppression efficiency

Table IV
Variable pocket discrimination by *E. coli* aminoacyl-tRNA synthetases

Synthetase	*in vitro*	*in vivo*	References
Alanyl-	Yes	Yes	McClain & Foss, 1988a, McClain et al., 1991b, Tamura et al., 1991b
Cysteinyl-	?	?	
Aspartic-	?	?	
Glutamic-	?	?	
Phenylalanyl-	Yes	Yes	McClain & Foss, 1988a, McClain & Foss, 1988c, McClain & Foss, 1988d, Peterson & Uhlenbeck, 1992
Glycyl-	?	?	
Histidyl-	?	?	
Isoleucyl-	?	?	
Lysyl-	?	?	
Leucyl-	?	?	
Methionyl-	?	?	
Asparaginyl-	?	?	
Prolyl-	?	?	
Glutaminyl	Yes	?	Rould et al., 1991
Arginyl-	Yes	Yes	McClain & Foss, 1988b, Schulman & Pelka, 1989, McClain & Foss, 1990, Tamura et al., 1992,
Seryl-	Yes	?	Himeno et al., 1990
Threonyl-	?	?	
Valyl-	?	?	
Tryptophanyl-	?	?	
Tyrosyl-	?	?	

Table V
Unusual elements used in tRNA discrimination by *E. coli* aminoacyl-tRNA synthetases

Synthetase	Element	*in vitro*	*in vivo*	References
Phenylalanyl-	central core	Yes	?	Peterson & Uhlenbeck, 1992
Histidyl-	nucleotide -1	Yes	?	Himeno et al., 1989 Francklyn & Schimmel, 1990 Francklyn et al., 1992
Isoleucyl-	Lysidine 34	Yes	Yes	Maramatsu et al., 1988, Pallanck & Schulman, 1991
Lysyl-	mnm^5s^2U 34	Yes	?	Tamura et al., 1992
Glutaminyl-	G10 (D-stem)	Yes	?	Rould et al., 1989, Jahn et al., 1991, Hayase et al., 1992
Seryl-	C11:G24	?	Yes	Normanly et al., 1986a, Normanly et al., 1992
Seryl-	Variable loop	Yes	Yes	Himeno et al., 1990, Normanly et al., 1992
Tyrosyl	Variable loop	Yes	Yes	Himeno et al., 1990, Normanly et al., 1992

Figure legends:

Figure 1: Sequence of yeast $tRNA^{Phe}$ shown in the cloverleaf representation of the secondary structure of all tRNAs. Important secondary structural domains are labeled.

Figure 2: Tertiary structure of yeast $tRNA^{Phe}$ (Kim et al., 1974). A) Three-dimensional structure of yeast $tRNA^{Phe}$ showing the location of the secondary structural domains and the variable pocket (drawing courtesy of J. Sampson). B) Cloverleaf representation of yeast $tRNA^{Phe}$ showing tertiary interactions required to produce the three-dimensional structure shown in Figure 2A. Important interactions are indicated by name. The nucleotides which make up the variable pocket are circled. The tertiary interaction which involves A21 is with the phosphate backbone of U8.

Figure 3: Recognition and identity elements of $tRNA^{Ala}$. Circles with nucleotides labeled are recognition elements. Squares indicate identity elements. (*) Identity element G1:C72 has not been tested for effects on recognition *in vitro*.

Figure 4: Recognition and identity elements of *E. coli* $tRNA^{Phe}$. Circles with nucleotides indicated are recognition elements. Squares indicate identity elements. The propeller twist nucleotides 26:44, were always co-mutated. (*) Recognition elements G10, C25 and A26 have not been tested for *in vivo* effects; recognition elements G34 & A35 have not been tested in the *in vivo* initiation system (although they were mutated without effect in amber suppression system).

Figure 5: Recognition and identity elements of *E. coli* $tRNA^{Gln}$ and contact points for the glutaminyl-tRNA synthetase. Modified nucleotides are listed as the parent nucleotide. Circles, recognition or identity elements; squares, nucleotide specific contacts between the synthetase and tRNA. Dashed lines indicate base pairing interactions required for specific recognition. Base pair 1:72 and the discriminator do not have nucleotide specific contacts with the synthetase, but are required for the assumption of an optimal conformation for aminoacylation (see text). (*) U32 and U33 are believed to be involved in maintaining the structure of the anticodon region (see text) but have not been tested for effects on aminoacylation.

Figure 1

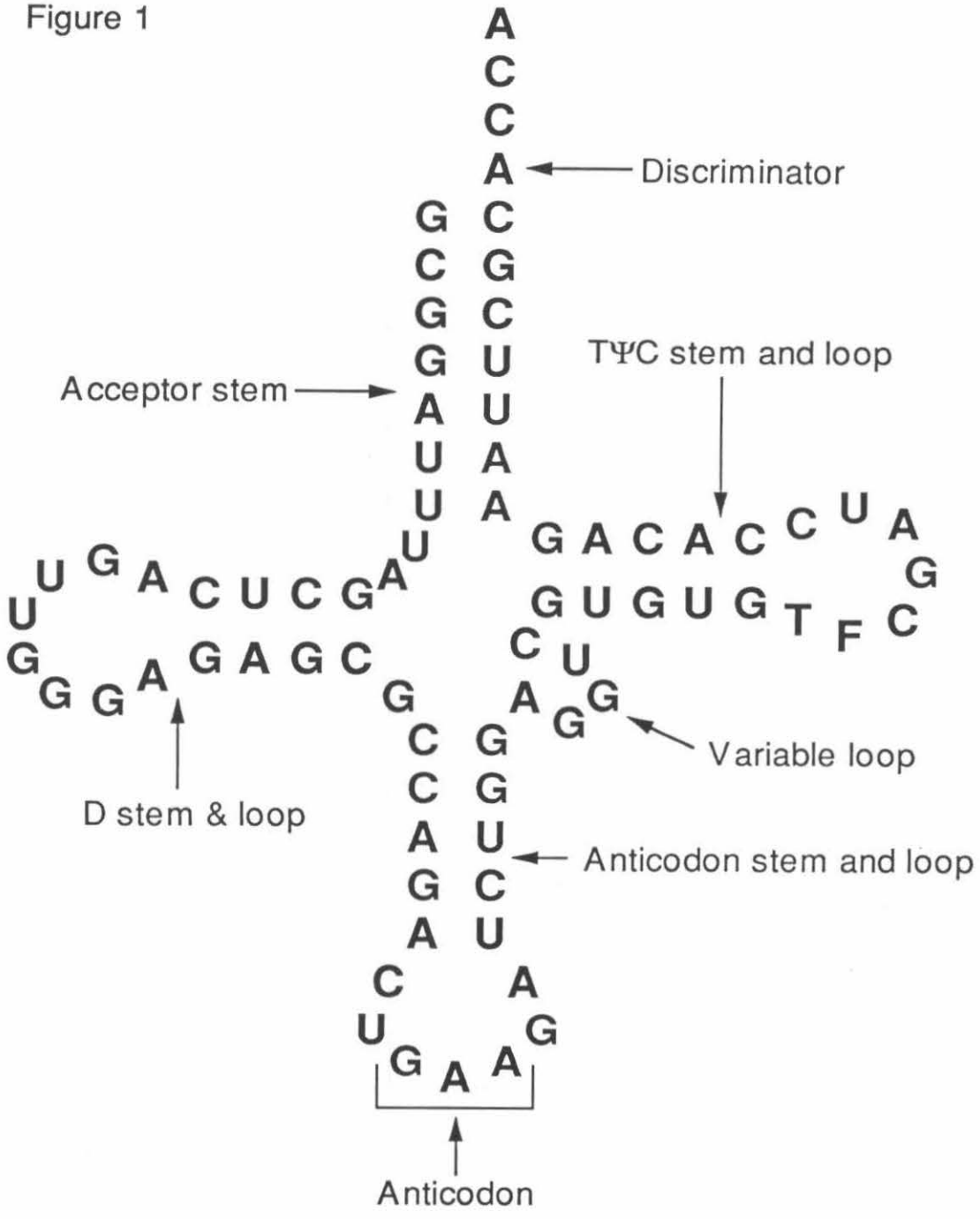

A
C
C
A
Discriminator
G C
C G
G C
G U
A U
U A
U A
Acceptor stem
TΨC stem and loop
G A C A C C U A
G U G U G T F C
U G A C U C G A U
U
G G G A G A G C
C U
A G G
Variable loop
D stem & loop
C G
C G
A U
Anticodon stem and loop
G C
A U
C A
U G
G A A
Anticodon

Figure 2A)

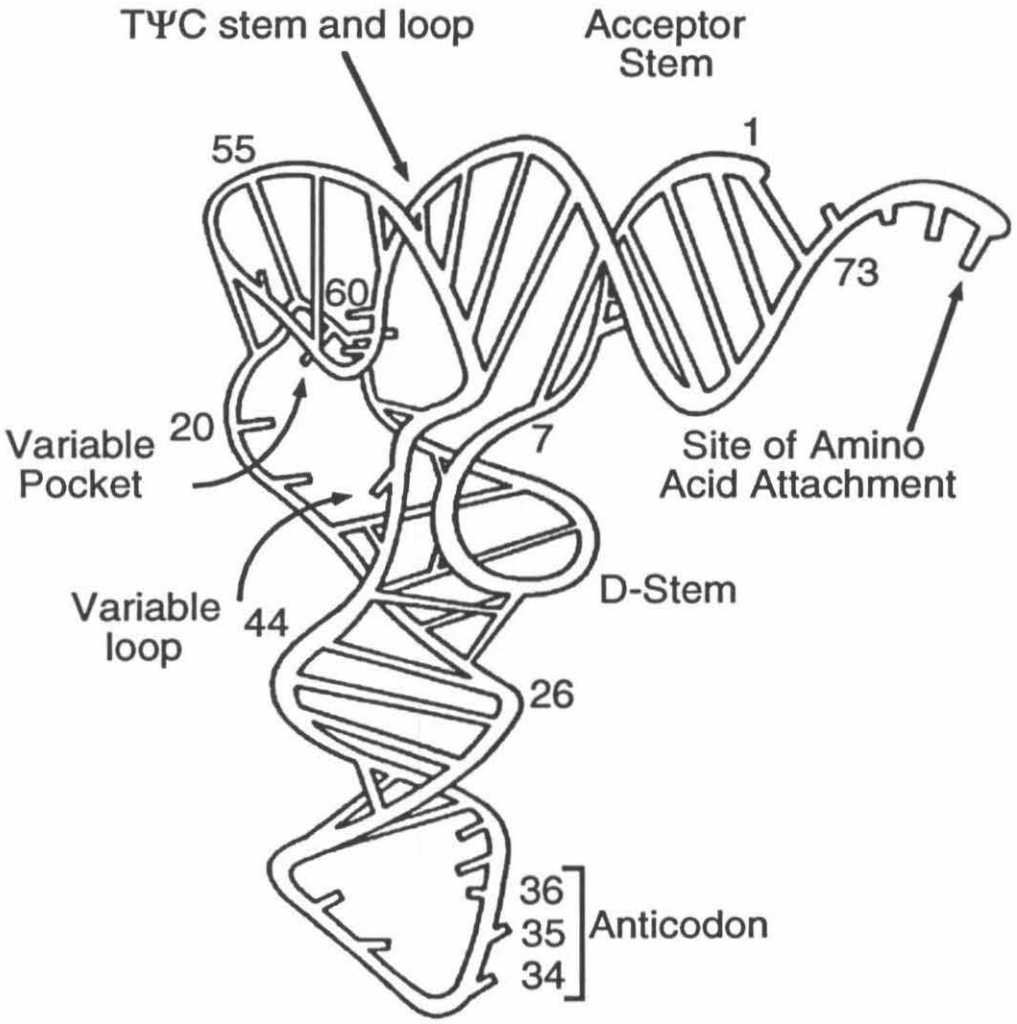

TΨC stem and loop
Acceptor Stem
55
1
60
73
Variable Pocket
20
7
Site of Amino Acid Attachment
Variable loop
44
D-Stem
26
36
35
34
Anticodon

Figure 2B

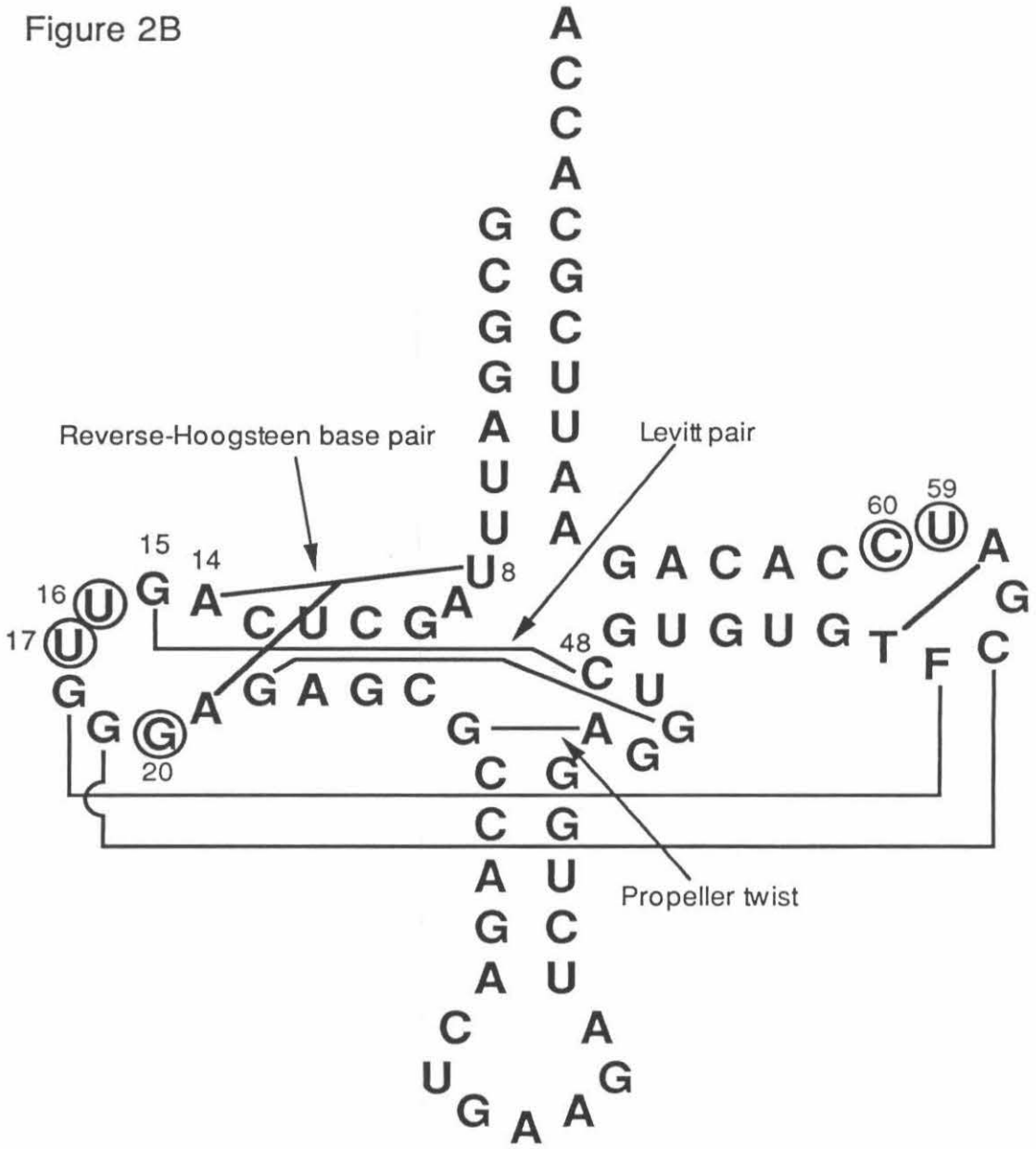

Reverse-Hoogsteen base pair
Levitt pair
Propeller twist
15
14
16
17
20
48
60
59

Figure 3

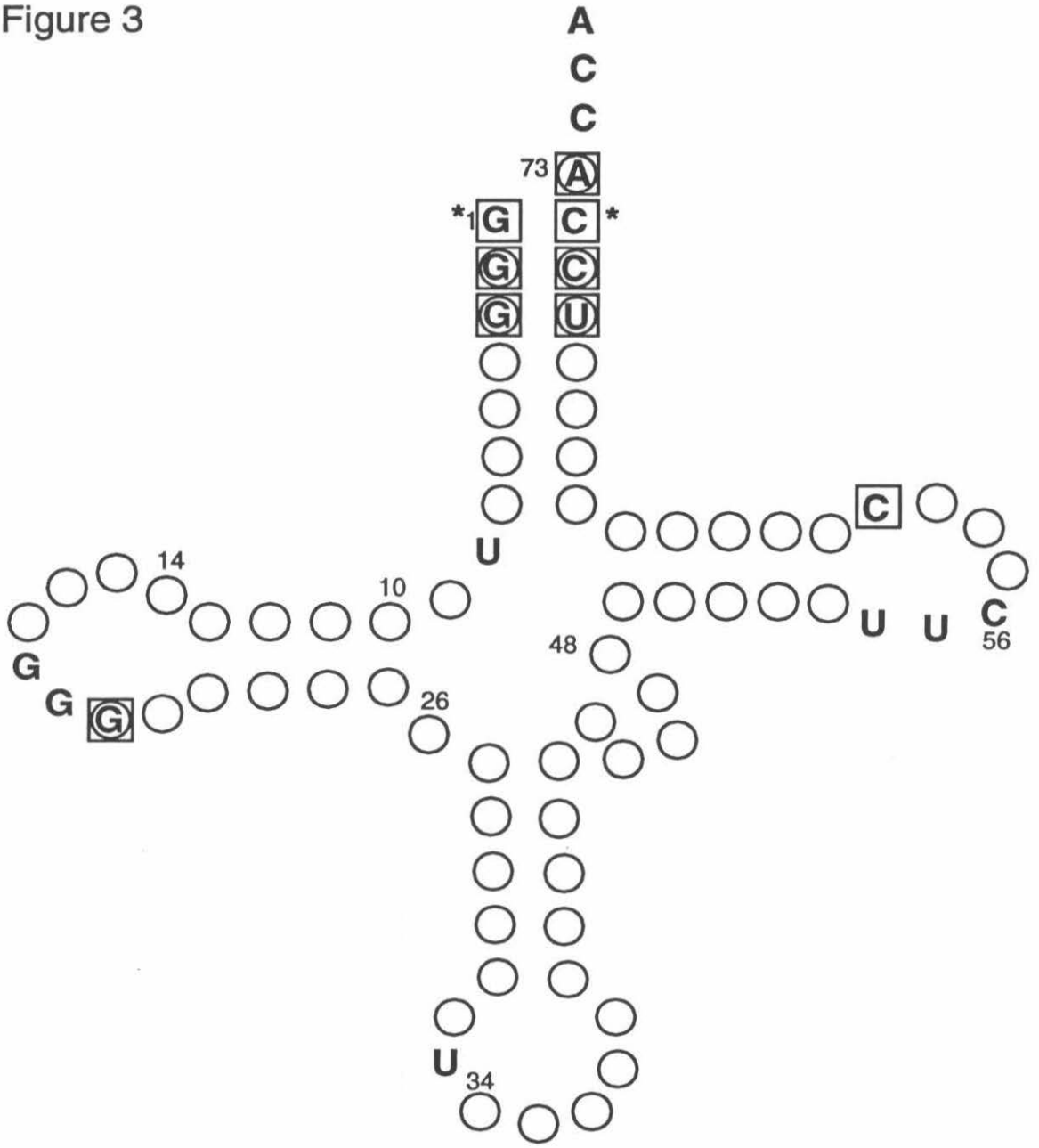


Figure 4

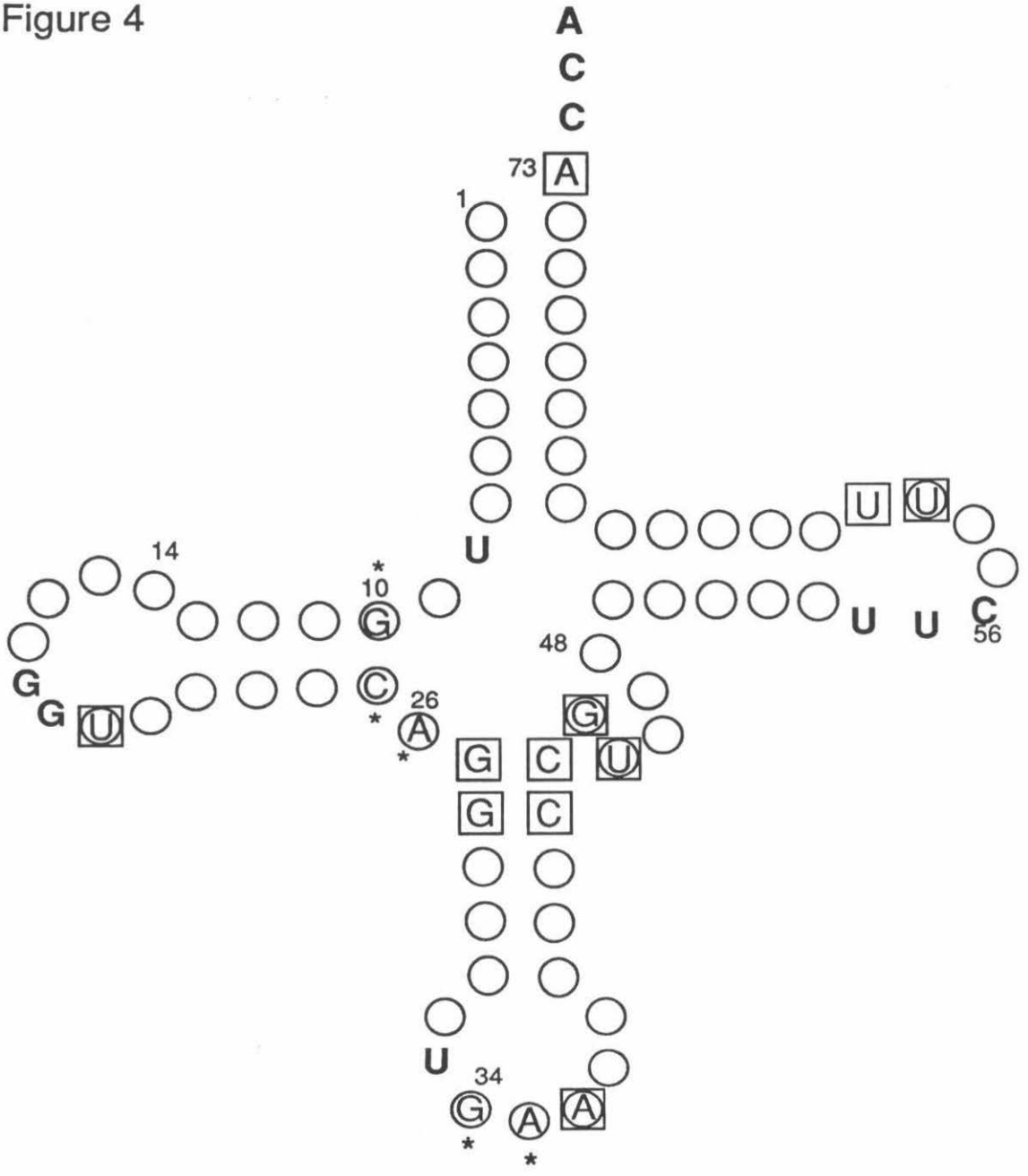
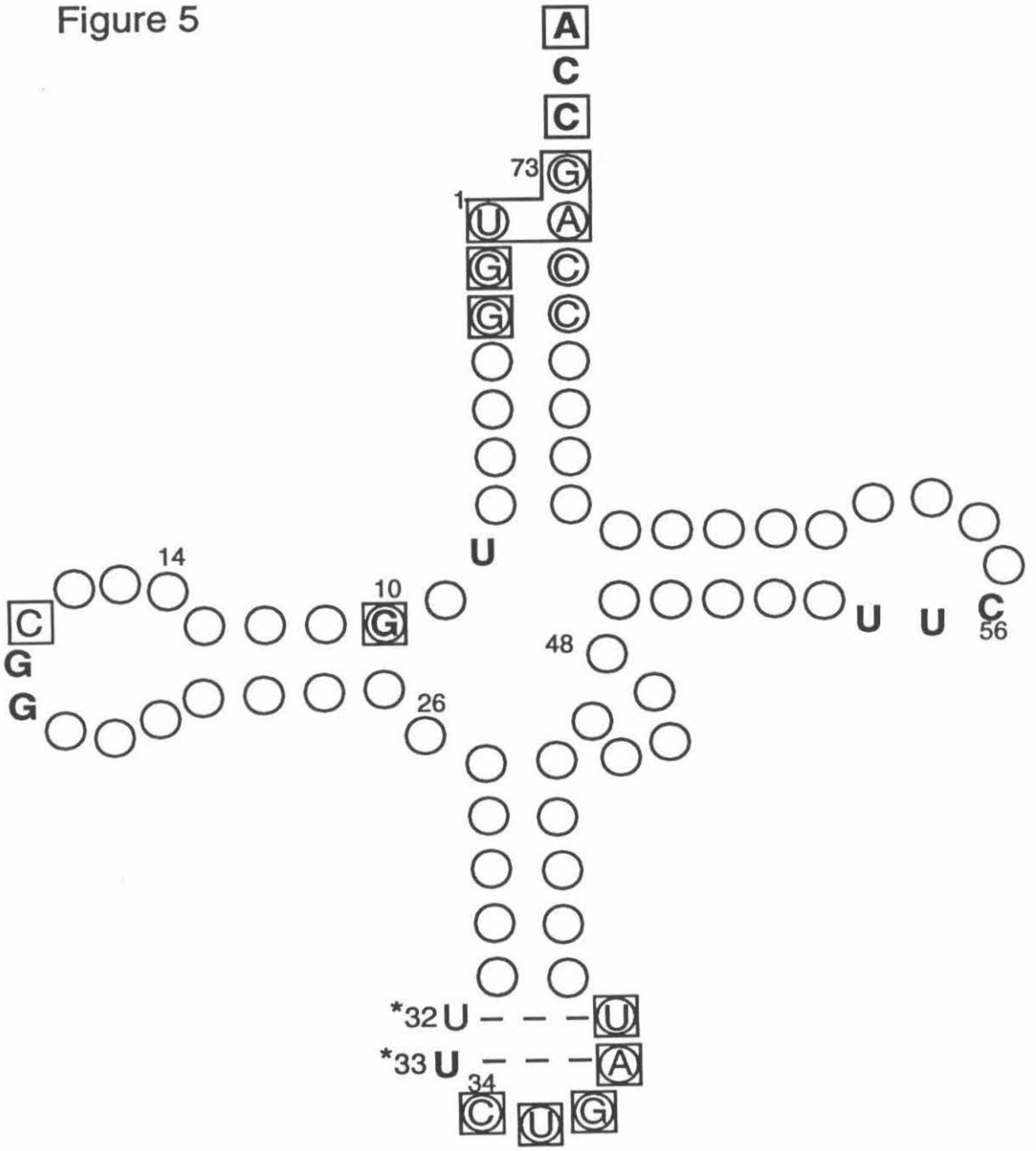


Figure 5



Chapter II

in vivo studies of the identity of $tRNA^{Cys}$ in *Escherichia coli*

Portions of the work described herein appeared in Komatsoulis, G. A., and Abelson, J. N. (1993) *Biochemistry* submitted

Abstract:

Eight amber suppressor tRNAs which were hybrids of $tRNA^{Cys}$ and $tRNA^{Ala}$ (UGC) were constructed in an effort to determine the nature and location of the elements required to confer cysteine identity *in vivo*. A wild-type amber suppressor $tRNA^{Cys}$ retains its cysteine identity *in vivo* (94% Cys, 6% Gln), but when three base pairs in the anticodon stem are mutated to the alanine sequence, the level of cysteine identity is reduced dramatically (37% Cys, 63% Gln). When the base pairs are mutated individually, cysteine identity is retained (89-93% Cys, 5 to 11% Gln).These results suggest that the wild-type amber suppressor $tRNA^{Cys}$ is already seriously impaired for aminoacylation by the cysteinyl-tRNA synthetase *in vivo*, and that the anticodon stem mutations are fine-tuning the competition between the cysteinyl- and glutaminyl- tRNA synthetases. *In vitro* aminoacylation experiments will be required to verify these conclusions and continue the study of the interaction of $tRNA^{Cys}$ and the cysteinyl-tRNA synthetase.

Introduction:

The scientific study of transfer RNA identity, the process of determining what sequence or structural elements within a tRNA are required to define a specific amino acid acceptance *in vivo*, has undergone a renaissance in recent years driven by advances in molecular biology. The application of the Merrifield technique to automated oligonucleotide synthesis has allowed the rapid and simple construction of any desired tRNA sequence, and new molecular biology techniques have advanced the methods used to determine the amino-acid-acceptance of a tRNA *in vivo*.

One way of determining the amino-acid-specificity of a tRNA involves amber or opal suppressors and reporter genes with nonsense mutations. Many proteins have positions which are specific for only one or two amino acids; other amino acids at these positions produce nonfunctional products. Thus, aminoacylation of a nonsense suppressor with a particular specificity can be tested by assaying for active protein produced from a gene with an amber or opal mutation at the amino acid restricted site. One example of such a genetic test involves the ß-lactamase gene; which has been subjected to extensive site directed mutatgenesis (Dalbadie-McFarland et al., 1982). Rescue of an amber mutation at the active site of this gene requires that the suppressor tRNA insert serine or cysteine. Other examples include suppression of an amber mutation at the active site of the thymidylate synthetase gene, which requires cysteine for reasonable activity (Dev et al., 1988, Michaels et al., 1990), and the *lacZ*$_{1000}$ amber mutation in strain BT235, which can be suppressed by glutamine (or several other amino acids) but not by serine. The approximate amino-acid-specificity for an amber suppressor tRNA can thus be easily determined by assaying for functional compensation of amber mutants sites specific for the twenty different amino acids.

An alternative approach to test the amino-acid-specificity of an amber suppressor tRNA is to simply sequence the protein product of the reporter gene. The relative amount of any amino acid found at the site of the amber mutation would then give an indication of the frequency with which the various aminoacyl-tRNA synthetases were charging the tRNA. The reporter gene which is most commonly (and originally) used is the *fol*amber10 gene from *E. coli,* (Normanly et al., 1986a, 1986b) which produces Dihydrofolate Reductase (DHFR). DHFR was chosen because it can be easily purified by methotrexate affinity chromatography and anion exchange chromatography. The approximate level of aminoacylation by any given synthetase can then be determined from the relative amount of the corresponding amino acid at position 10.

The first use of these new methods was by Normanly and coworkers (1986a), who successfully converted a naturally occurring leucine amber suppressor tRNA into a serine specific amber suppressor. Later that year Normanly et al. (1986b) created two amber suppressor tRNAs which did not exist in nature, one specific for phenylalanine and one specific for cysteine. Other systems were rapidly investigated using these methods (for review see Schimmel, 1987, Normanly and Abelson, 1989, and Schulman, 1991) These methods did not work with all tRNAs however; an attempt to produce a set of amber suppressor tRNAs for all twenty amino acids produced 11 functional suppressors which retained their identity, and 8 suppressor tRNAs which inserted significant amounts of glutamine or lysine, presumably because the anticodon was a significant determinant of *in vivo* identity in many tRNAs (Normanly et al., 1990, and Kleina et al., 1990). Recently, Chattapadhyay (1990) et al. described a new method based on initiation rather than amber suppression, which allows *in vivo* studies of tRNAs which use the anticodon as an identity element.

Despite the early experiments of Normanly et al. (1986b) showing that a wild-type cysteine amber suppressor retains cysteine identity *in vivo*, no further work had been reported when this study was begun. There is a single cysteine accepting tRNA in *E. coli* with anticodon GCA which recognizes two codons UGC and UGU (Komine et al., 1990, Sprinzl et al., 1991). Although this prevents using comparisons of isoacceptors to help guess identity nucleotides (see McClain and Nicholas Jr., 1987) there were several advantages to working with cysteine. A genetic screen specific for cysteine was available (Michaels et al., 1990, and C. Kim and J. Miller, pers. comm.) and the amber suppressor $tRNA^{Cys}$ (Normanly et al., 1986b) provided a good starting point for study. Furthermore, the cysteinyl-tRNA synthetase, originally characterized by Bohmann and Isaksson (1979) was cloned and sequenced shortly after these studies began (Eriani et al., 1991, Hou et al., 1991), making possible the future application of *in vitro* aminoacylation experiments. We therefore sought to determine the identity elements of *E. coli* $tRNA^{Cys}$.

Materials and Methods:

Strains and plasmids:

Strain XAC-1 is F' *lacI* $lacZ_{am}$ / $argE_{am}$ *gyrA rpoB Δ(lac pro) ara*. JM101 is F'$traD_{36}$ *proAB* $lacI^qZ_{M15}$ */Δ(lac pro) supE thi. E. coli* strain i-p+ is F' *lacpro/Δ(lacpro) thi*. Both were obtained from J. Normanly. XACΔTK is F⁻ */Δ(lac pro) ara gyrA rpoB* $argE_{am}$ *nal thi ΔthyA recA::Tn10* and was obtained from Choll W. Kim (UCLA). Plasmids pGFIB-1 and pDaYQ are described in Normanly et al. (1986b) and (1990) respectively. Plasmid pThyA is pACYC184 with $ThyA_{amber166}$ cloned into the HindIII site and was the gift of Choll W. Kim (UCLA).

Cloning of tRNA genes:

Genes for tRNAs were constructed from six oligonucleotides and cloned into the EcoRI-PstI sites in pGFIB-1 under the control of an lpp (lippoprotein) promoter and rrnC terminator (Normanly et al., 1986b). The lpp promoter is a strong constitutively on promoter; rrnC is a rho-independent terminator from the risbosomal RNA genes. Each tRNA gene included a T7 RNA polymerase promoter, the tRNA gene itself , and a BstNI site which included the terminal CCA of the tRNA (see Figure 1). Annealing of the oligonucleotides and cloning were carried out according to the methods of Normanly et al.(1986a). Once cloned and sequenced (Sanger et al., 1977) the tRNA was transformed into appropriate strains for characterization.

Genetic methods of studying *in vivo* identity:

The initial determination that the tRNA was a functional amber suppressor was the ability to suppress the amber mutations in the lacI/lacZ fusion and argE genes in strain XAC-1. The former was assayed by production of a blue color on indicator plates containing isopropyl-ß-D-galactopyranoside (IPTG) and X-gal; the latter by assaying for growth on minimal media lacking arginine (Miller, 1972). Suppression efficiency was assayed by determining ß-galactosidase levels produced from the episomal lacIlacZ$_{amber}$ fusion in strain XAC-1 by the cloned amber suppressor tRNA and then comparing this value with levels produced in a wild-type strain, *E. coli* I-p+. ß-galactosidase activity was determined using the method of Miller (1972). The cysteine specific amber suppression assay (Michaels et al., 1990 and C. W. Kim and J. Miller, pers. comm.) utilized a cysteine requiring amber mutation in the thymidylate synthetase gene (Dev et al., 1988). It was carried out by transforming a plasmid containing the suppressor tRNA of interest and pThyA (with the *thyA$_{amber}$* gene) into strain XACΔTK (which is Thy$^-$) and assaying for growth on minimal M9 medium (Miller, 1972) without thymidine. To verify that the results were

caused by the suppressor tRNA of interest in the correct strain, the proline auxotrophy and *argE*$_{amber}$ mutation were tested in parallel.

DHFR assays:

DHFR expression and purification was performed using a variation of the method described by Normanly et al. (1986a). DHFR was purified from XAC-1 cells using a compatible plasmid system. The two plasmids were pGFIB::suppressor-tRNA, and pDaYQ which contained *fol*amber10, the mutant DHFR gene.. Cells were grown until A_{600} reached 0.6 and then induced with 1mM IPTG and grown for another 8 hours. After recovery, cells were resuspended in 40mM Tris pH8.0 and then frozen in liquid nitrogen. Lysis was by lysozyme-Brij58 treatment, after which the extract was treated with DNAseI. A 50% ammonium sulfate cut was performed on the extract, followed by a 100% cut on the supernatant. The pellet was resuspended in dialysis buffer (50mM potassium phosphate pH6.0, 200mM KCl, 500μM EDTA, 1mM DTT), mixed with 3mL methotrexate sepharose (Pierce) and dialyzed overnight. The methotrexate sepharose was then batch washed or washed in column with 200mM potassium phosphate pH 6.0, 1M KCl, 1mM DTT. The DHFR was eluted with folic acid (200mM potassium borate pH9.0, 1M KCl, 1 mM DTT, 1mM EDTA, 2mM folic acid). Protein-containing fractions were combined and dialyzed against KP8 buffer (50mM potassium phosphate pH8.0, 1mM DTT). The DHFR was purified away from nucleic acids, other proteins, and folic acid by MonoQ™ FPLC using a linear gradient from 0 to 400mM KCl with a slope of 16mM KCl/min at a flow rate of 1 mL/min. Fractions containing DHFR were pooled, dialyzed against 1mM DTT, and then concentrated with a Centricon10 filter unit (Amicon). The protein was derivatized and sequenced at the Caltech Biopolymer Synthesis and Analysis Resource Center, Protein/Peptide Sequencing Laboratory. To more accurately quantitate cysteine, the DHFR was

first pyridylethylated and purified on reverse phase C18 HPLC (D. Teplow and T. Bauer, pers. comm.) before sequencing.

Results:

Construction and *in vivo* studies of amber suppressor mutants of $tRNA^{Cys}$:

To begin a study of cysteine identity and recognition a cysteine amber suppressor tRNA (CysCUA) was constructed in plasmid pGFIB-1. Such a gene had been constructed before (Normanly et al., 1986b) but it did not contain a T7 promoter (for later *in vitro* studies) and its identity had been assayed using ß-galactosidase (rather than DHFR) as the reporter gene. The suppression efficiency for the new amber suppressor tRNA was 11-24%, which is consistent with the results (17-50%) of Normanly et al., 1986b. The ability of tRNA CysCUA to suppress a mutant with an amber codon in the cysteine requiring active site of thymidylate synthetase indicates that it is charged at some level by cysteine although the assay does not indicate whether other aminoacyl-tRNA synthetases are also charging the tRNA. To determine the complete identity of CysCUA, a second plasmid, pDaYQ which contains the $fol_{amber10}$ gene under the control of a *tac* promoter, was transformed into the strain carrying pGFIB:CysCUA, and DHFR protein was purified and sequenced. The relative ability of the various synthetases to aminoacylate the tRNA is reflected as the amino acid composition of the protein at position 10; for CysCUA the results were 94% Cys, 6% Gln (Table 1). These results independently confirm the results of Normanly et al., (1986b) that CysCUA retains its cysteine identity *in vivo*.

Construction and *in vivo* studies of acm1 to acm4:

Since the amber suppressor $tRNA^{Cys}$ had retained its *in vivo* identity a series of tRNAs was constructed to attempt to define the minimal set of elements

required to specify a cysteine tRNA. These elements would be identified by their ability to confer cysteine identity on an amber suppressor $tRNA^{Ala}$, and so the tRNAs were labelled acm, Alanine to Cysteine Mutants. Alanine (see figure 3 for sequence) was chosen as the heterologous tRNA for several reasons; its main identity element, a G-U base pair at position 3:70 (all numbering of tRNAs follows Sprinzl et al., 1991) had been described, making the problem of eliminating alanine identity easier. Furthermore, when this G-U pair was introduced into an amber suppressor $tRNA^{Cys}$ the result was an alanine tRNA, suggesting that the overall structure of the two tRNAs were not incompatible. In order to ensure that all of the tRNAs were efficient suppressors, nucleotide 38 was mutated from C to A to better conform to the extended anticodon rule (Yarus 1982). This meant however that all of the acm tRNAs contained a cysteine element (A38) in the anticodon loop.

The first two tRNAs of this series (acm1 and acm2) were designed to rapidly determine the approximate size and location of the identity set elements. Acm1 (figure 4), has only a few residues in a single region, the acceptor stem changed to cysteine sequence ; while acm2 (figure 5), would have cysteine sequence in all but the anticodon stem. If both were charged only with cysteine future experiments could concentrate on the acceptor stem only; if acm2 was charged specifically with cysteine but acm1 was not then other regions of the tRNA would need to be studied as well.

The two tRNAs, acm1 and acm2, were cloned into pGFIB-1 and the initial genetic experiments performed. Both were functional suppressors, able to suppress the *lacIlacZ*$_{amber}$ and *argE*$_{amber}$ mutations of strain XAC-1. Their suppression efficiencies were calculated to be 0.6% and 11-24% respectively (Table I). Both were tested for suppression of the cysteine specific *thyA*$_{amber}$ mutant, but only acm2 was able to grow in the absence of thymidine. This did

not necessarily indicate that acm1 was being charged with no cysteine. The suppression efficiency of this tRNA was sufficiently low that several days were required for growth without arginine in strain XAC-1; the result with the $thyA_{amber}$ might simply be the combination of a bad suppressor and a multiple *in vivo* identity.

Two new tRNAs which incorporated more cysteine sequence were designed and cloned in an effort to correct whatever defects were causing the poor suppression efficiency of acm1. New tRNA acm3 (figure 6) has cysteine sequences in the acceptor stem, D stem and loop, and variable loop, and alanine sequences in the TΨC stem and loop, and anticodon stem. Acm4 (figure 7) has cysteine sequence in the acceptor stem and D loop, and alanine sequences in the D stem, TΨC stem and loop, variable loop, and anticodon stem. In both of these tRNAs base pairs 6:67 and 7:66 (at the bottom of the acceptor stem) retain the alanine sequence in order to preserve any stacking interactions with the alanine TΨC stem. These two tRNAs were extremely inefficient suppressors, incapable of suppressing the $argE_{amber}$ mutation in XAC-1. (Table I) Their calculated suppression efficiencies were 0.01% for acm3 and 0.001% for acm4. The cysteine specific thymidylate synthetase assay was never attempted.

In order to determine the full range of synthetases which could aminoacylate acm2, the tRNA was used to suppress the $fol_{amber10}$ mutation. The sequenced DHFR protein indicated that acm2 was aminoacylated with 37% cysteine and 63% glutamine (Table I). Thus, it appeared that one or more of the six anticodon stem nucleotides which differed between $tRNA^{Ala}$ and $tRNA^{Cys}$ might be identity elements.

Construction and analysis of anticodon stem mutations:

To determine which base pair(s) in the anticodon stem was the identity element(s), three mutant tRNAs were constructed. Each tRNA is a cysteine tRNA in which one of the three base pairs which are different between $tRNA^{Ala}$ and $tRNA^{Cys}$ changed to the alanine sequence. Mutant acm5 (figure 8) has cysteine base pair A31:Ψ39 changed to C31:G39; acm7 (figure 9) has cysteine base pair G29:C41 mutated to U29:A41, and acm8 (figure10) has G27:U41 mutated to C27:G41. All three tRNAs suppressed the $argE_{amber}$ and $lacIlacZ_{amber}$ fusion, with suppression efficiencies of 4%, 9%, and 20% respectively (Table I). All were aminoacylated to some extent with cysteine as determined by their ability to suppress the thymidylate synthetase amber mutation (Table I).

The results of the DHFR assay for these suppressor tRNAs are shown in Table I. All three suppressors inserted primarily cysteine with lesser amounts of glutamine. The aminoacylation pattern resembles that of CysCUA, the wild-type amber suppressor. All four tRNAs (CysCUA, acm5, acm7, and acm8) retain their identity, the variations were within the error of the assay.

Discussion:

The results of the *in vivo* studies presented here do not form a coherent data set on the surface. CysCUA, the wild-type amber suppressor $tRNA^{Cys}$ is a reasonably efficient suppressor tRNA which retains its identity *in vivo* (94% Cys, 6% Gln). When the anticodon stem of this tRNA is mutated to the sequence of an alanine tRNA its *in vivo* cysteine identity is partially lost and the new tRNA, acm2, is charged primarily with glutamine (63%) with a smaller amount of cysteine. This would seem to suggest that one (or more) of the three anticodon stem base pairs changed in acm2 was an identity element, and presumably the element(s) could be identified by changing each of the base pairs individually and assaying for cysteine identity. When this was done however, all of the

tRNAs (acm5, acm7, and acm8) retained their cysteine identities just like the wild-type amber suppressor tRNA.

This data only makes sense if the anticodon is an important identity element for the cysteinyl-tRNA synthetase, but can be mutated without changing the identity of the tRNA. This theory postulates that the amber supressor $tRNA^{Cys}$ is a much worse substrate for the cysteinyl-tRNA synthetase than a wild-type tRNA with the GCA anticodon, but retains cysteine identity because no other synthetase can compete successfully with the cognate synthetase (although the presence of 6% glutamine in the DHFR from CysCUA suggests that the glutaminyl-tRNA synthetase is beginning to compete successfully for the tRNA, see Sherman et al., 1992 for an interesting paper showing the effects of synthetase competition). Thus CysCUA is sitting on the knife edge, a relatively small change in the total catalytic efficiency with either synthetase is enough to produce large changes in identity. The sum of six mutations in acm2 might marginally increase the catalytic efficiency with the glutaminyl-tRNA synthetase, while mutations in each individual base pair (acm5, acm7, and acm8) cannot. Alternatively, the sum of the mutations might reduce the total catalytic efficiency of the reaction with the cysteinyl-tRNA synthetase, while the individual mutations cannot. One possible source for the changes in catalytic efficiency could be the effect of the anticodon stem on the position of the anticodon nucleotides with regard to the recognition site on the synthetases.

Regardless, the only way to ascertain which of these hypotheses is correct is to perform *in vitro* aminoacylation experiments on these tRNAs. In particular it will be necessary to evaluate the importance of the anticodon in recognition of $tRNA^{Cys}$.

References:

Bohmann, K. & Isaksson, L. (1979) *MGG, Mol. Gen. Genet 176*, 53-55

Chattapadhyay, R., Pelka, H., & Schulman, L. H. (1990) *Biochemistry 29*, 4263-4268

Dalbadie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, C., Itakura, K, & Richards, J. H. (1982) *Proc. Natl. Acad. Sci. 79*, 6409-6413.

Dev, I. K., Yates, B. B., Leong, J. D., Dallas, W. S. (1988) *Proc. Natl. Acad. Sci USA 85*, 1472-1476.

Eriani, G., Dirheimer, G., & Gangloff, J. (1991) *Nucleic Acids Res. 19*, 265 -269

Hou, Y.-M., Shiba, K., Mottes, C., & Schimmel, P. (1991) *Proc. Natl. Acad. Sci. USA 88*, 976-980.

Kleina, L. G., Masson, J. -M., Normanly, J., Abelson, J., & Miller, J. H. (1990) *J. Mol. Biol. 213*, 705-717.

Komine, Y., Adachi, T., Inokuchi, H., & H. Ozeki (1990) *J. Mol. Biol. 211*, 579-598.

McClain, W. H., & Nicholas Jr., H. B. (1987) *J. Mol. Biol. 194*, 635-642.

Michaels, M. L., Kim, C. W., Matthews, D., A., & Miller, J. H. (1990) *Proc. Natl. Acad. Sci. USA 87*, 3957-3961.

Miller, J. H. (1972) in *Experiments in Molecular Genetics* , Cold Spring Harbor Laboratory Press, New York.

Normanly, J., Ogden, R.C., Horvath, S. J., & Abelson, J. (1986a) *Nature 321*, 213-219.

Normanly, J., Masson, J.-M., Kleina, L., Abelson, J., & Miller, J. H. (1986b) *Proc. Natl. Acad. Sci. USA 83*, 6548-6552.

Normanly, J., & Abelson, J. (1989) *Annu. Rev. Biochem 58*, 1029-1049.

Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J., & Miller, J. H. (1990) *J. Mol. Biol. 213*, 719-726.

Normanly, J., Ollick, T., & Abelson, J. (1992) *Proc. Natl. Acad. Sci. USA 89*, 5680-5684.

Ozeki, H., Inokuchi, H., Yamao, F., Kodaira, M., Sakano, H., Ikemura, T. & Shimura, Y. (1980) in *Transfer RNA: Biological Aspects*, eds. Söll, D., Abelson, J. N., and Schimmel, P. R. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA 74*, 5463-5467.

Schulman, L. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 23-87.

Schimmel, P. (1987) *Annu. Rev. Biochem. 56*, 125-158.

Sherman, J. M., Rogers, M. J., & Söll, D. (1992) *Nucleic Acids Res. 20*, 2847-2852.

Sprinzl, M., Dank, N., Nock, S., & A. Schön (1991) *Nucleic Acids Res. 19*, 2127-2171.

Yarus, M. (1982) *Science 218*, 646-652.

tRNA	Suppression efficiency (% wild-type)*	Suppression of $argE_{amber}$ (nonspecific)	Suppression of $ThyA_{amber}$ (cys specific)	DHFR Results	
				% Cysteine	% Glutamine
CysCUA	11-24	(+)	(+)	94	6
acm1	0.5	(+)	(-)	nd	nd
acm2	5-12	(+)	(+)	37	63
acm3	0.001	(-)	(-)	nd	nd
acm4	0.01	(-)	(-)	nd	nd
acm5	4	(+)	(+)	90	10
acm7	9	(+)	(+)	93	5
acm8	20	(+)	(+)	89	11

Table I: Results of in vivo experiments with amber suppressor tRNAs. See text for details of experiments. Values for glutamine acceptance include glutamic acid (12% for acm2, 2% for all others) which is produced from glutamine during the Edman degradation.

Figure legends:

Figure 1: Graphic representation of a typical tRNA gene construction used in these studies.

Figure 2: RNA sequence of the amber suppressor $tRNA^{Cys}$. Modifications are inferred from the sequence of the wild-type tRNA (Sprinzl et al., 1991).

Figure 3: RNA sequence of $tRNA^{Ala}$, modified nucleotides shown as parent nucleotides

Figure 4: RNA sequence of mutant acm1, shown as changes from $tRNA^{Ala}$. Modified nucleotides shown as parent nucleotides.

Figure 5: RNA sequence of mutant acm2, shown as changes from $tRNA^{Ala}$. Modified nucleotides shown as parent nucleotides.

Figure 6: RNA sequence of mutant acm3, shown as changes from $tRNA^{Ala}$. Modified nucleotides shown as parent nucleotides.

Figure 7: RNA sequence of mutant acm4, shown as changes from $tRNA^{Ala}$. Modified nucleotides shown as parent nucleotides.

Figure 8: RNA sequence of mutant acm5, shown as changes from $tRNA^{Cys}$. Modified nucleotides shown as parent nucleotides.

Figure 9: RNA sequence of mutant acm7, shown as changes from $tRNA^{Cys}$. Modified nucleotides shown as parent nucleotides.

Figure 10: RNA sequence of mutant acm8, shown as changes from $tRNA^{Cys}$. Modified nucleotides shown as parent nucleotides.

Figure 1

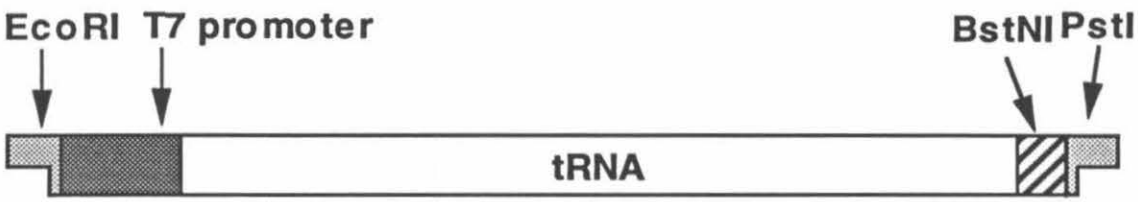


Figure 2

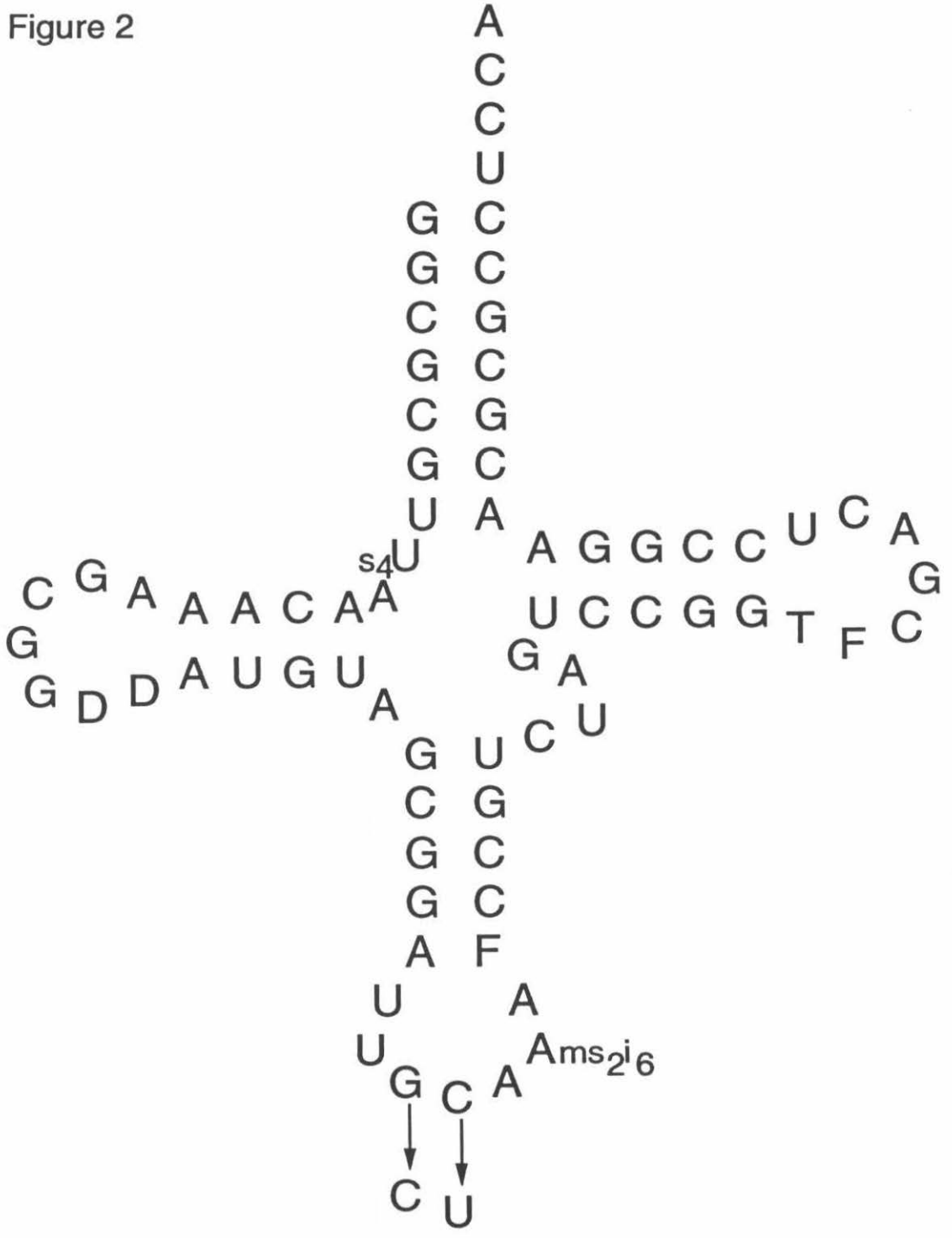

A
C
C
U
G C
G C
C G
G C
C G
G C
U A
s4U
A G G C C U C A
C G A A A C A A
U C C G G T F C
G
G D D A U G U
G A
A
C U
G U
C G
G C
G C
A F
U A
U
Ams2i6
G C A
C U

Figure 3

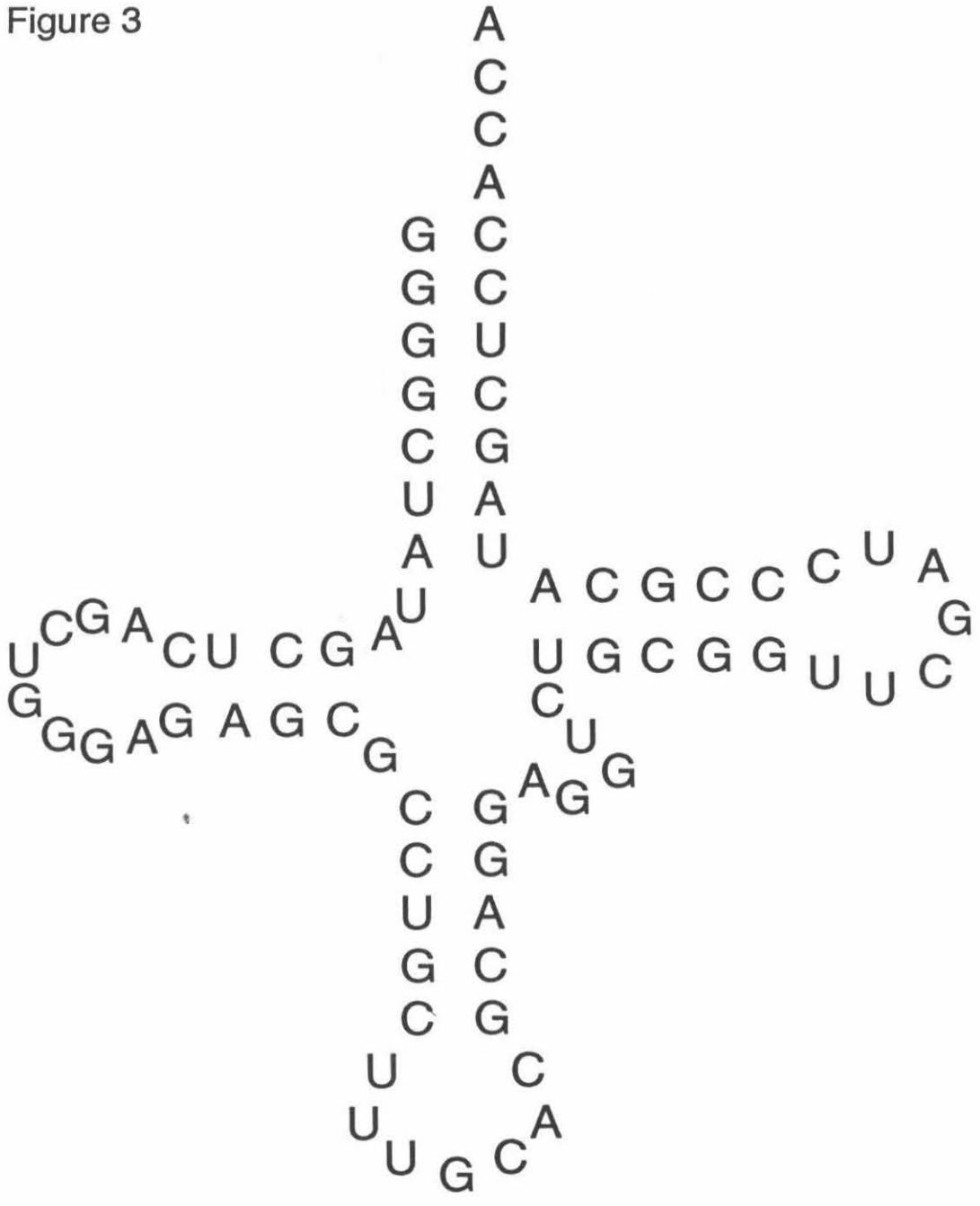

A
C
C
A
G C
G C
G U
G C
C G
U A
A U
A C G C C C U A G
U G C G G U U C
C U G
A G
C G
C G
U A
G C
C G
U C
U U G C A
U C G A C U C G A U
G G G A G A G C G

Figure 4

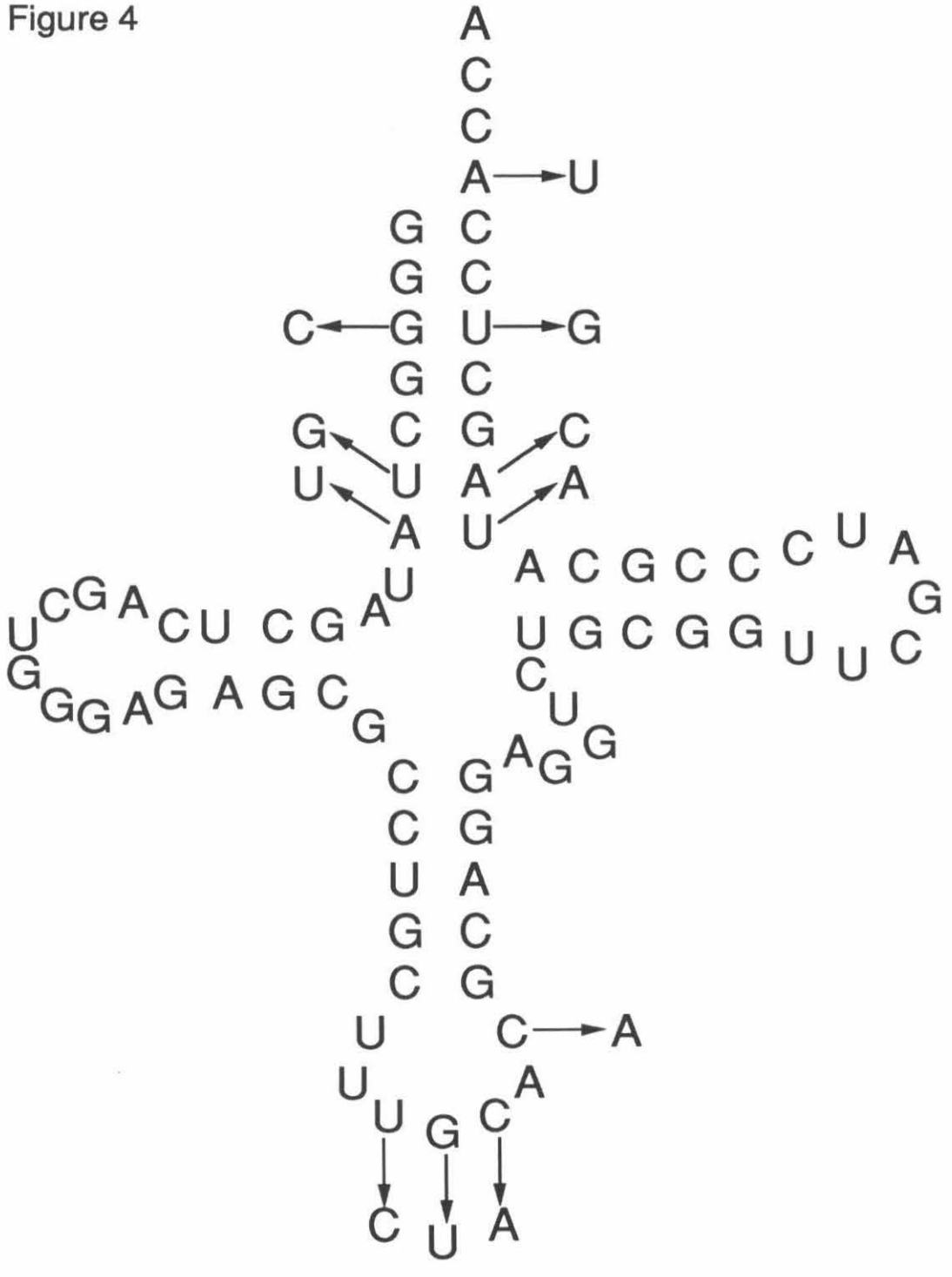

A
C
C
C
A→U
G C
G C
C←G U→G
G C
G C G C
U U A A
A U
A C G C C C U A G
U G C G G U U C
C U
A G G
C G
C G
U A
G C
C G
U C→A
U A
U G C
C U A

Figure 5

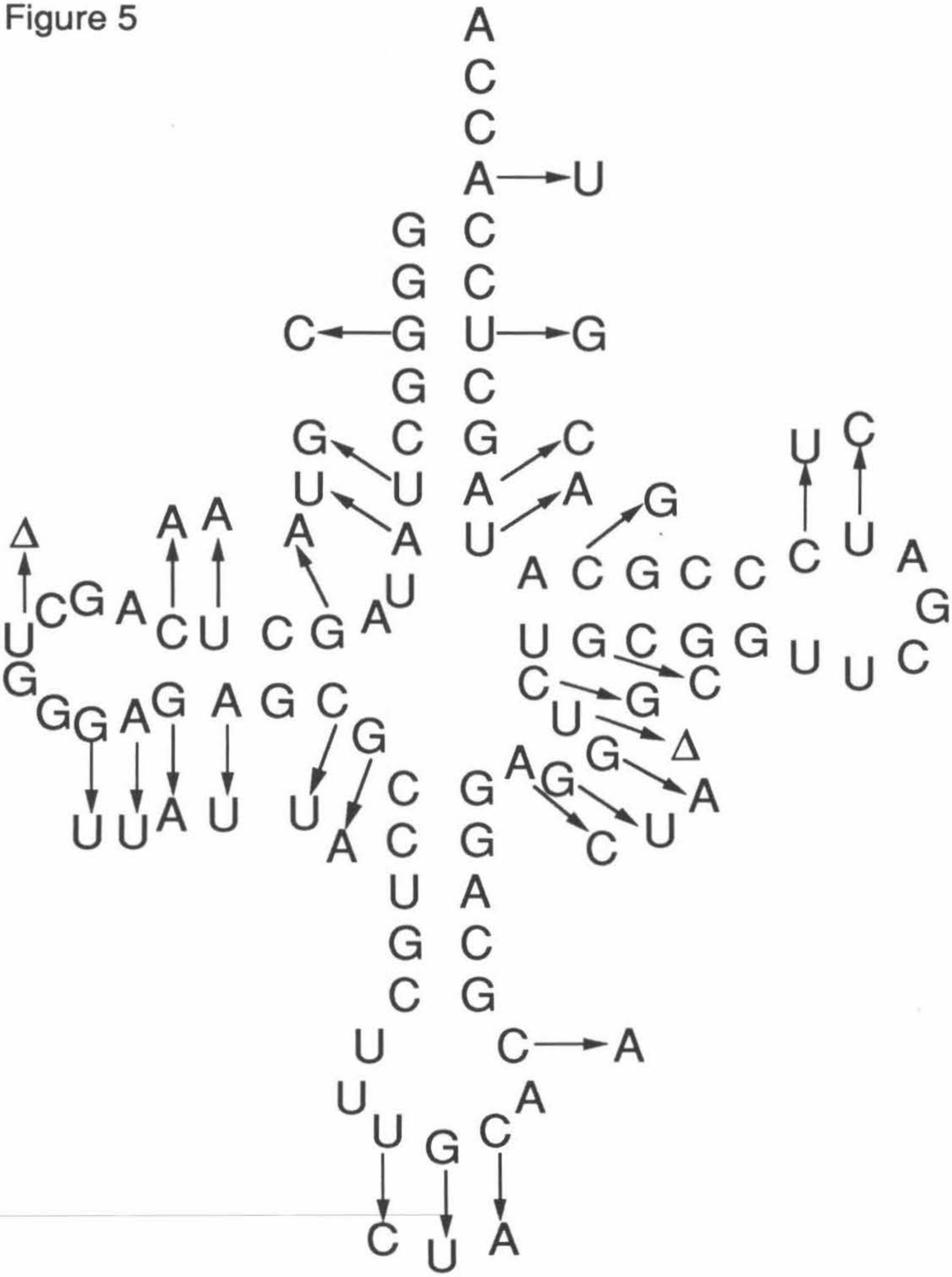


Figure 6

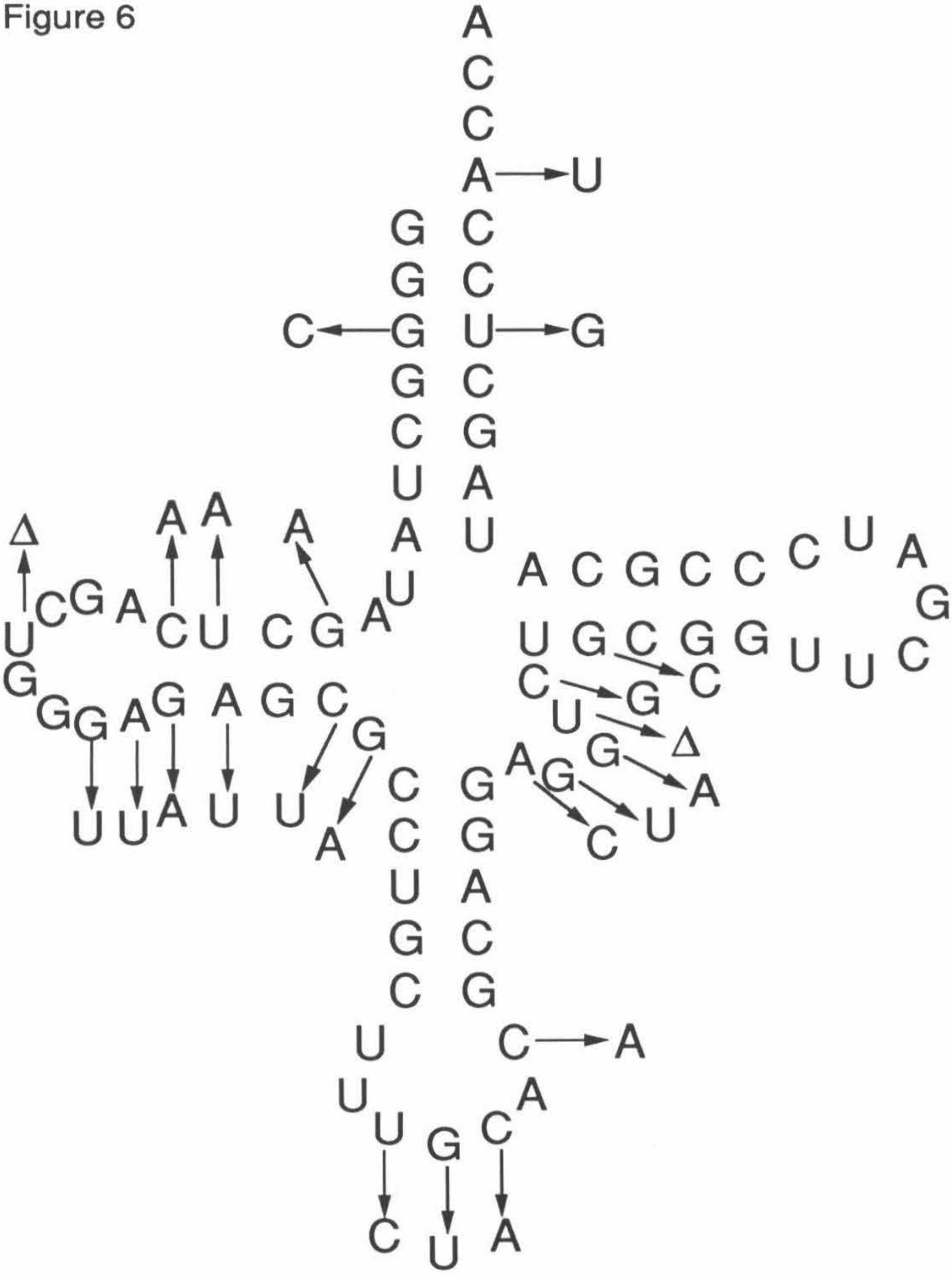

A
C
C
A → U
G C
G C
C ← G U → G
G C
C G
U A
Δ A A A A U
A C G C C C U A
U C G A C U C G A U
G
U G C G G U U C
C G C
G G A G A G C U G Δ
G A G A
U U A U U A C G C U
C G
U A
G C
C G
U C → A
U A
U G C
C U A

Figure 7

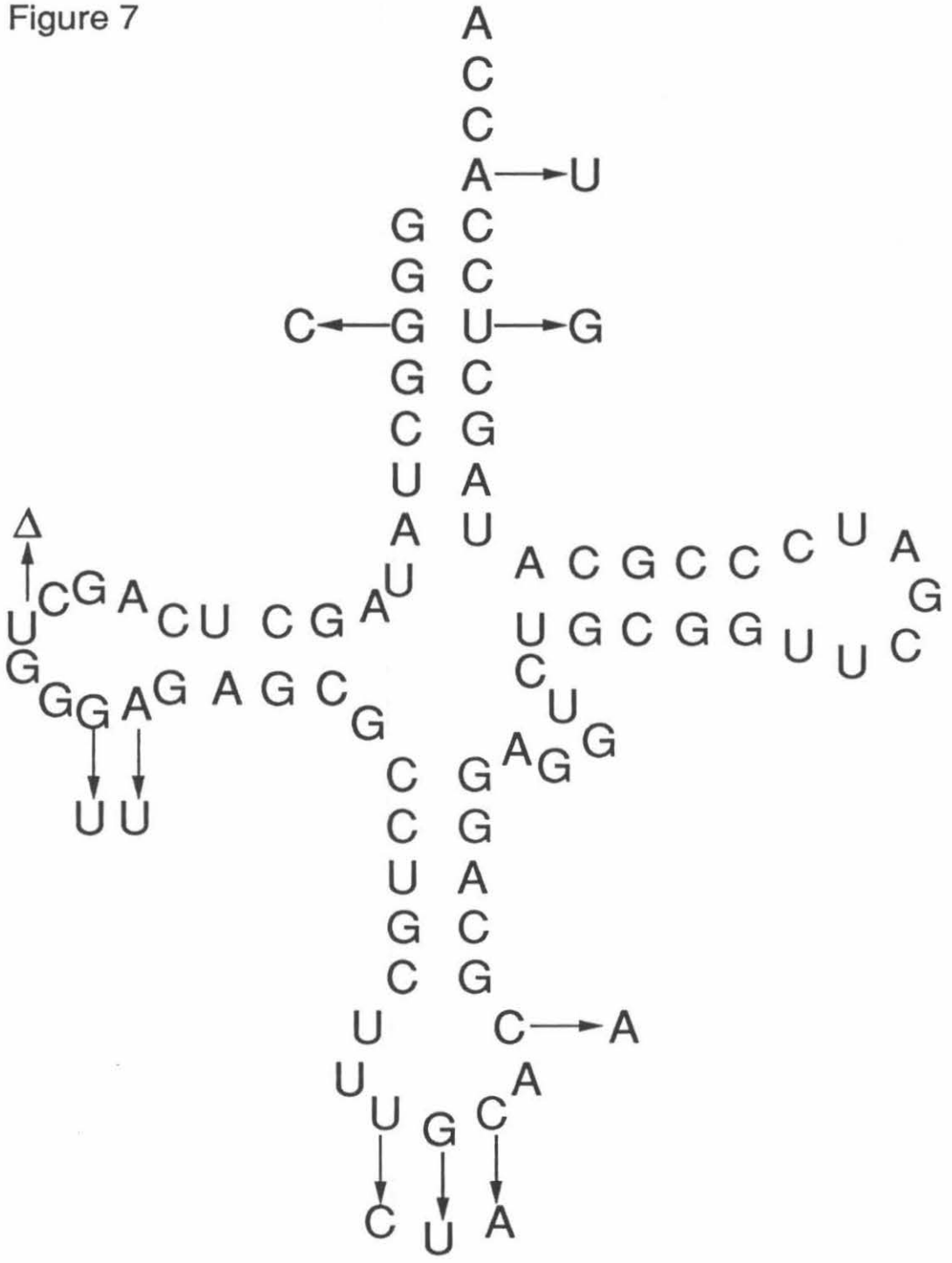

A
C
C
A→U
G C
G C
C←G U→G
G C
C G
U A
A U
A C G C C C U A G
U G C G G U U C
C U G
A G
C G
C G
U A
G C
C G
U C→A
U A
U G C
C U A

Figure 8

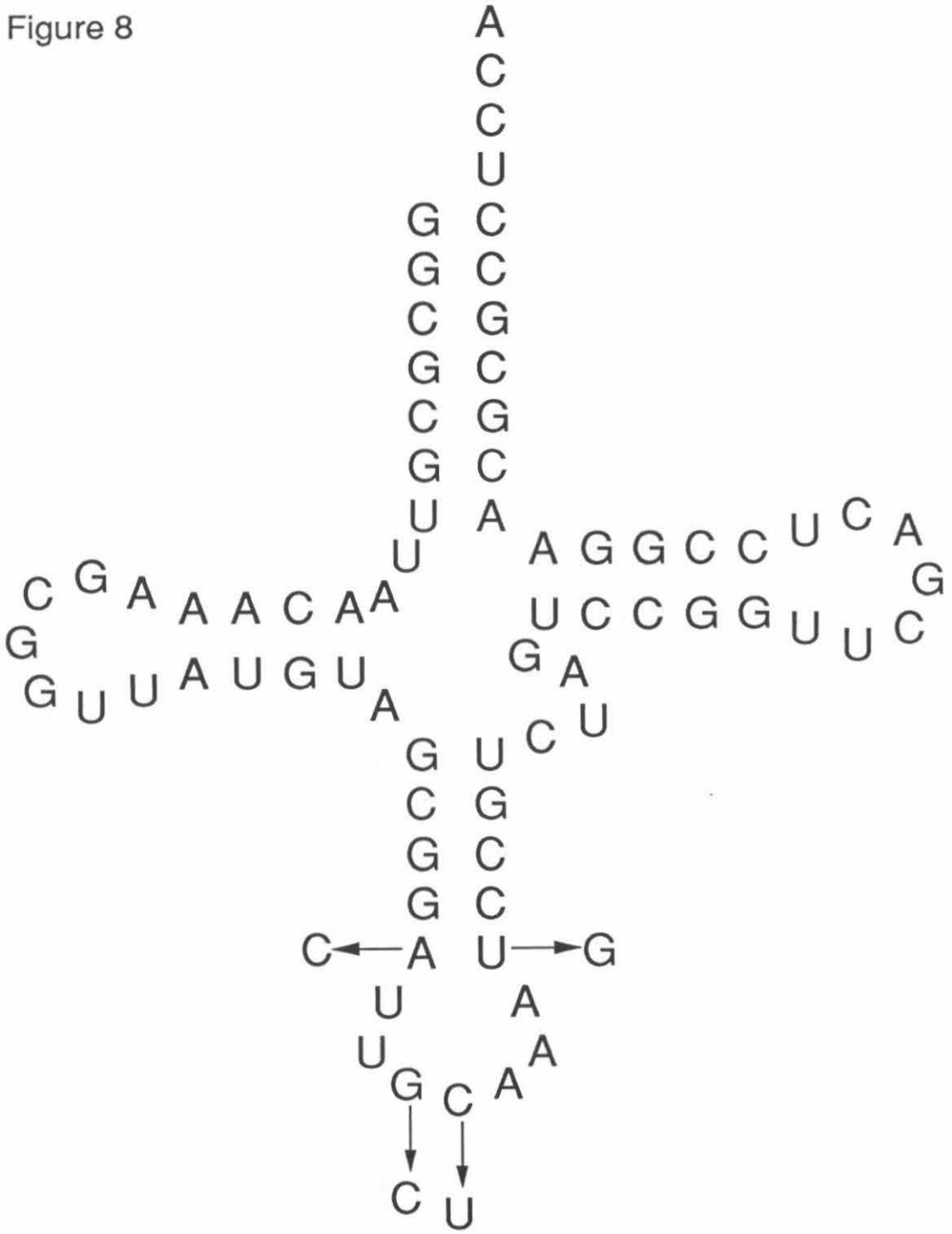

A
C
C
U
G C
G C
C G
G C
C G
G C
U A
U
A G G C C U C A
G
C G A A A C A A
U C C G G U U C
G
G A
G U U A U G U A
C U
G U
C G
G C
G C
C A U G
U A
U A
G C A
C U

Figure 9

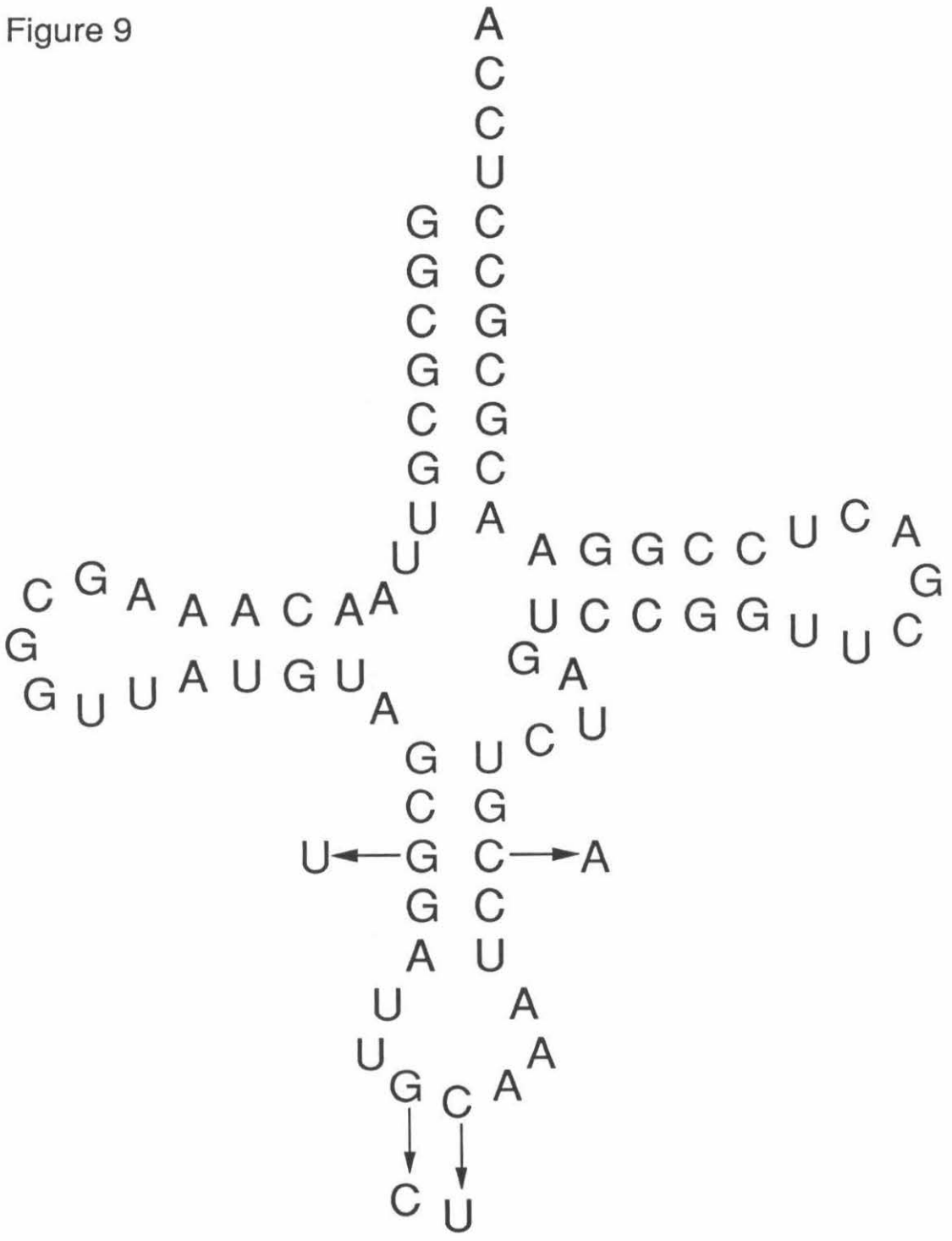
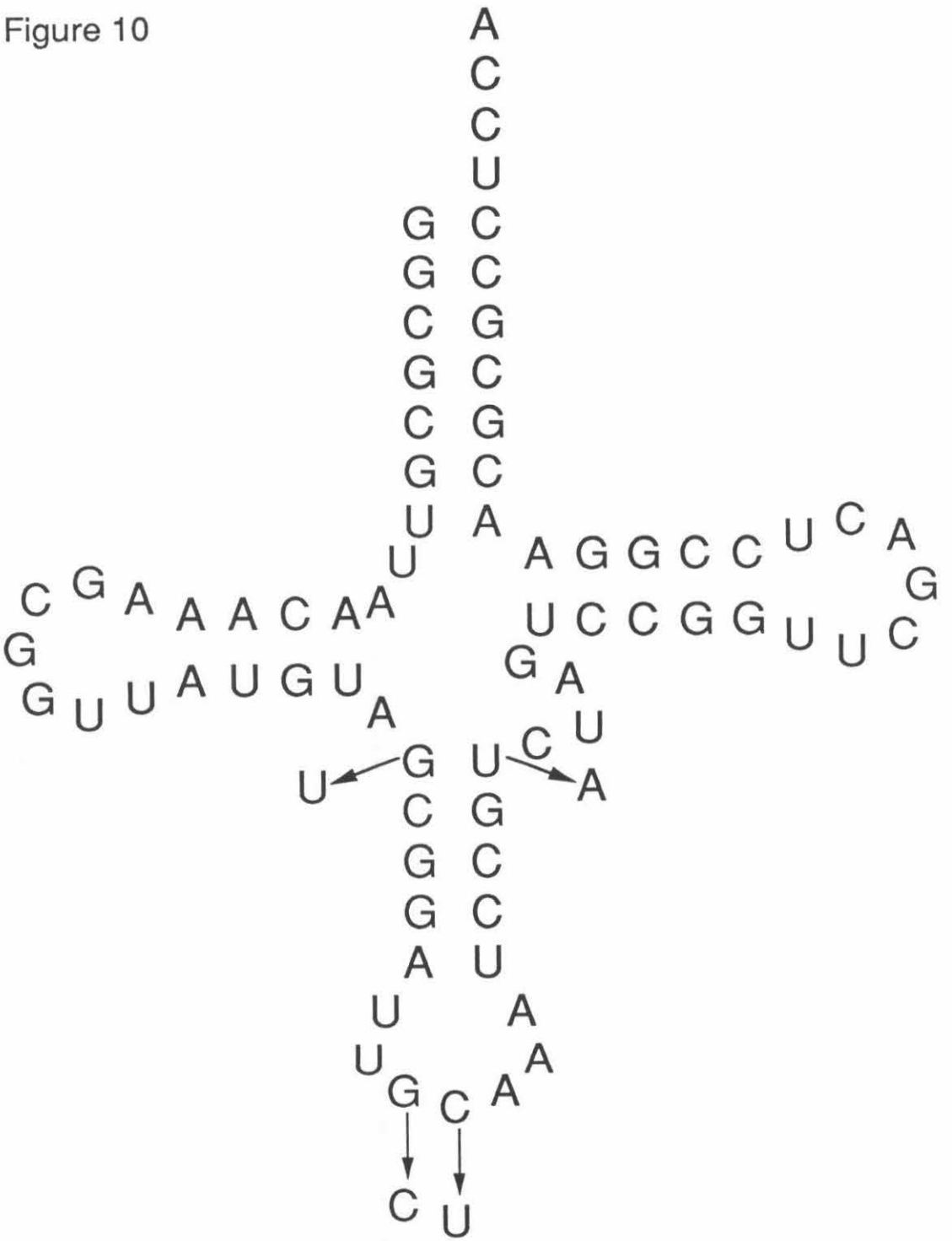

A
C
C
U
G C
G C
C G
G C
C G
G C
U A
U
A G G C C U C A G
C G A A A C A A
U C C G G U U C
G
G U U A U G U
G A
A
C U
G U
C G
U G C A
G C
A U
U A
U A
G C
C U

Figure 10


A
C
C
U
G C
G C
C G
G C
C G
G C
U A
U
A G G C C U C A
G
C G A A A C A A
U C C G G U U C
G
G A
G U U A U G U
A
C U
G U
U A
C G
G C
G C
A U
U A
U A
G C A
C U

Chapter III

In vitro studies of the recognition of $tRNA^{Cys}$ by the *E. coli* cysteinyl-tRNA synthetase

This work appeared in Komatsoulis, G. A., and Abelson, J. N., (1993) *Biochemistry*, submitted

Abstract:

A study of the recognition of $tRNA^{Cys}$ by *E. coli* cysteinyl-tRNA synthetase using *in vitro* methods was performed. All three anticodon nucleotides, the discriminator U73, and some element within the tertiary domain (the D stem/loop, the TΨC stem/loop and extra loop) are important for recognition; the anticodon stem and acceptor stem appear to contain no essential elements. A T7 RNA polymerase transcribed tRNA is a six-fold worse substrate (in terms of the selectivity constant, k_{cat}/K_m) mainly due to an increase in K_m. This may reflect recognition of modified nucleotides or subtle effects on the folding of the tRNA. The greatest loss of specificity caused by mutation of a single nucleotide occurs when the discriminator U73 is changed; k_{cat}/K_m declines 3 to 4 orders of magnitude depending on the substitution. Mutations in the wobble nucleotide of the anticodon also cause reductions in the selectivity constant of 3 orders of magnitude, while mutations in the other anticodon nucleotides caused lesser effects. Interestingly, a C35A mutation had no effect on aminoacylation by the cysteinyl-tRNA synthetase. Several amber suppressor tRNAs were constructed whose *in vivo* identity did not correlate with their *in vitro* specificity, indicating the need for both types of experiments to understand the factor(s) which maintain tRNA specificity.

Introduction:

The obvious alternative to *in vivo* experiments such as described in the previous chapter is to aminoacylate purified tRNAs *in vitro* with purified (or partially purified) aminoacyl-tRNA synthetases and then calculate the kinetic parameters for the reaction (for review see Schulman, 1991). These sorts of experiments can generate much useful information both in terms of how important an element is for recognition, and how that element mediates its effect. The former is usually measured by observing the effect of a mutation on the value of k_{cat}/K_m for the tRNA; this number is often called the selectivity constant and represents the total catalytic efficiency of the reaction. The individual components of the selectivity constant indicate whether an effect is mediated through binding of the substrate (K_m) or the formation of the transition state (k_{cat}).

Although *in vitro* aminoacylation experiments have been conducted for many years (see for instance Abelson et al., 1970, Squires and Carbon, 1971) their utility in tRNA recognition had been severely limited due to the difficulty of preparing mutant substrates. Early methods used chemical or enzymatic (Schulman and Pelka, 1983) methods to make mutations in pre-existing tRNAs, but this was time consuming and difficult to use in a site specific fashion. The advent of molecular cloning techniques allowed a researcher to make specific mutations in a tRNA and then overexpress that tRNA *in vivo*, but the purification was still clumsy and time consuming.

This difficulty was overcome by improvements in the understanding of promoters and reaction conditions for bacteriophage RNA polymerases. The RNA polymerase from phage T7 seemed particularly well suited to the purpose of transcribing mutant tRNA molecules, its promoter preferences were well characterized, and it initiates at the +1 site downstream from its promoter with

excellent accuracy (Milligan et al., 1987). Sampson and Uhlenbeck (1988) were the first to report on the aminoacylation of a yeast $tRNA^{Phe}$ produced by transcription with T7 RNA polymerase. The tRNA gene had been cloned with a T7 promoter just proximal to the 5' end of the tRNA molecule, and they converted the CCA terminus to a BstNI site by the addition of a GG dinucleotide. Thus, by cutting the plasmid with BstNI they generated a template from which an unmodified tRNA with the proper CCA terminus could be produced. At 15mM $MgCl_2$ the T7 transcribed tRNA behaved very similarly to the fully modified tRNA, with only a 4-fold increase in K_m and a slight reduction in k_{cat} giving a change in the selectivity constant k_{cat}/K_m, of only 5-fold. The thermal melting profile of the two tRNAs suggested that the difference was caused by a slightly looser structure on the part of the T7 transcript. They showed the utility of this method by rapidly preparing and analyzing a tRNA with a mutation in nucleotide 20, which had been shown previously to be important for recognition by the phenylalanyl-tRNA synthetase. This approach has since been extended in the Uhlenbeck laboratory, and the recognition elements of both yeast (Sampson et al., 1990, Sampson et al., 1992) and *E. coli* (Peterson et al., 1992) $tRNA^{Phe}$ have been studied extensively by this method. Since then many laboratories have adopted this approach for studies of tRNA recognition (review: Schulman, 1991)

A cloned synthetase is a useful tool in *in vitro* aminoacylation studies, and fortunately the *E. coli* cysteinyl-tRNA synthetase had been cloned and sequenced by two laboratories (Hou et al., 1991, Eriani et al., 1991) by complementation of a temperature sensitive mutant (Bohmann and Isaksson, 1979). The sequence indicates that the monomer unit is a peptide of 461 amino acids with a molecular weight of 52.3 kD and is a synthetase of type I (Hou et al., 1991, Eriani et al., 1991). Type I synthetases contain two conserved sequences

HIGH and KMSKS, which are indicative of the Rossman fold, a structural motif corresponding to an ATP binding domain (Eriani et al., 1990, reviewed in Moras, 1992).

The confusing *in vivo* results observed with mutant *E. coli* cysteine tRNAs suggested that *in vitro* aminoacylation experiments would be required to identify the elements important to define a tRNA as cysteine. Furthermore, if the anticodon was as important for recognition and identity as the *in vivo* experiments suggested, all future experiments would have to be performed *in vitro* since most *in vivo* methods rely on amber or opal suppressors. For this reason the cysteinyl-tRNA synthetase was partially purified from an overproducing strain and used to conduct a study of the recognition of $tRNA^{Cys}$ (Figure 1A), initially focusing on the amber suppressor tRNAs described in the previous section.

Materials and Methods:

Strains and plasmids:

Strain XAC-1 is F' *lacI*$_{373}$*lacZ*$_{u118am}$ *pro*B$^+$ / Δ(*lac pro*)$_{x111}$ *argE*$_{am}$ *gyrA rpoB ara*. JM101 is F'*traD*$_{36}$ *proAB lacI*^{q}*Z*$_{M15}$ /Δ(*lac pro*) *supE thi*. *E. coli* strain i-p+ is F' *lacpro*/Δ(*lacpro*)$_{x111}$ *thi*. Both were obtained from J. Normanly. Plasmids pGFIB-1 and pDaYQ are described in Normanly et al. (1986a) and (1990) respectively. Plasmid pCysS containing the cysteinyl-tRNA synthetase is the gift of G. Eriani and J. Gangloff (CNRS, Strasbourg, France).

Design and cloning of tRNA genes:

Genes for tRNAs were constructed from six oligonucleotides and cloned into the EcoRI-PstI sites in pGFIB-1 if the tRNA was to be used *in vivo*, or pUC18 if the tRNA was to be used solely *in vitro*. The exceptions to this rule are tRNAs Cys73A and CysGCA, which were cloned into pGFIB-1 and CysGAA which was cloned into pUC19. Each tRNA gene included a T7 RNA polymerase promoter, the tRNA gene itself , and a downstream BstNI site to produce the terminal CCA

of the tRNA (see figure 1B). Annealing and ligation of the oligonucleotides and cloning were carried out according to the methods of Normanly et al.(1986b). All tRNAs were sequenced using the dideoxy method of Sanger et al. (1977).

Purification of the cysteinyl-tRNA synthetase (CysRS):

Cysteinyl-tRNA synthetase was purified from an overproducing strain of JM101 which carried plasmid pCysS (Eriani et al., 1991), using a variation on the method of Hou et al. (1991). Two liters of cells were grown to an A_{600} of 2 in LB medium supplemented with 100μg/mL of ampicillin, harvested, and resuspended in 20 mL lysis buffer (50mM Tris pH7.6, 1mM EDTA, 1mM DTT, 500μM PMSF). The cells were washed twice in lysis buffer, sonicated and centrifuged at 100,000 times g (40,000 RPM in a Beckman 70.1Ti rotor) for one hour at 4°C. The S100 extract was loaded onto a Q-sepharose™ column which had been equilibrated with 50mM Tris pH7.6, $MgCl_2$ 1mM EDTA, 1mM DTT, 500μM PMSF and washed with 158mL of this buffer. CysRS was eluted with a linear gradient of 0-500mM KCl with a slope of 6.7mM KCl/min and a flow rate of 2 mL/min. Fifty 3mL fractions were collected; six active fractions were pooled and dialyzed against 50mM Tris pH7.6, 10mM $MgCl_2$, 10% glycerol, 1mM DTT, 500μM PMSF, concentrated with a Centricon10 filter unit (Amicon), increased to 50% glycerol, and stored at -20°C. The final protein concentration was 13μg/μL, and the activity was 590units/mg protein (unit definition: 1nmole aminoacylated native $tRNA^{Cys}$ per minute). The synthetase was approximately 25 to 50% pure, and was judged free of RNAse activity by incubating it under aminoacylation conditions (15mM $MgCl_2$ rather than the usual 7mM) with 5μM native $tRNA^{Cys}$ (Subriden RNA) for 1 hour at 37°C, and then checking for degradation products on a 10% polyacrylamide (29:1 crosslink ratio), 8M urea, 1X TBE gel.

***In vitro* transcription of tRNAs:**

Plasmid DNA was prepared using CsCl gradients, or by anion exchange

chromatography (with Qiagen <plasmid> tip 500 columns). Plasmids were then digested with restriction endonuclease BstNI, extracted twice with phenol/chloroform/isoamyl alcohol (25/24/1), extracted once with chloroform, precipitated, and recovered in TE8 (10mM Tris pH8.0, 1mM EDTA). Transcriptions used the following conditions: 0.1 μg/ μL digested DNA, 40mM Tris pH8.3, 20mM $MgCl_2$, 1mM spermidine, 5mM DTT, 5mM each NTP (Pharmacia),10mM GMP (Pharmacia), 50μg/mL acetylated BSA (New England Biolabs), 7u/mL pyrophosphatase (Sigma), 0.8 u/μL RNAsin (preincubated 20 minutes on ice with 20mM DTT), and 26u/μL T7 RNA polymerase (prepared by M. Saks and J. Sampson). Reactions were incubated for 4 to 5 hours at 37° C. Transcription reactions were terminated by adding 1/20 volume 0.5M EDTA and 1/10 volume 3M NaOAc, and were then extracted with phenol equilibrated with 300mM NaOAc pH5.3. The RNA was precipitated overnight at -20° C after adding 3 volumes ice cold ethanol.

The products were recovered in TE7.5 (10mM Tris pH7.5, 1mM EDTA) and mixed with equal volumes of RNA loading buffer heated to 85-90° C for 3 minutes, and then loaded onto 40 cm long denaturing 8% polyacrylamide (19:1 crosslink ratio) gels (except for AS1 and AS2, which were purified on 10% polyacrylamide, 29:1 crosslink gels). Gels were run at 600-900V. RNA was visualized by UV shadowing, excised from the gel, and eluted twice with 200mM KOAc pH5.4, 10mM EDTA at 4° C. The RNA was precipitated and recovered in TE7.5 (10mM Tris pH7.5, 1mM EDTA) three times. A fourth precipitation was carried out using 2M NH_4OAc in place of 0.3M NaOAc. The RNA was dissolved in 10mM HEPES pH 7.4 and stored at -20° C. The RNA was quantitated by measuring the absorbance at 260nm of a nuclease digested sample (1μL of 0.9μg/μL RNAseA, 11u/μL RNAse T1, and 0.2u/μL RNAseT2 in 700μL tRNA sample, overnight at room temperature) at 260nM. Extinction coefficients were

calculated from the sequence of the tRNA.

Aminoacylation reactions:

Aminoacylation reactions were performed at 37°C in 30mM HEPES pH7.4, 2mM DTT, 15mM KCl, 7mM $MgCl_2$, 2mM ATP, 20μM cysteine, 0.4μCi/μL ^{35}S-cysteine (stabilized, SJ15232, Amersham), and cysteinyl-tRNA synthetase. Synthetase was used at a final dilution of 1:200 to 1:100,000. The final protein concentration in all reactions was adjusted to 0.1μg/μL with acetylated BSA. RNA and distilled H_2O were prepared at 2X final concentration, and were heated to 90°C and allowed to cool slowly before use. At appropriate time points 8μL samples were spotted onto 1 inch by 0.5 inch segments of Whatman 3MM paper. The papers had been soaked one to two days before in 10% TCA, 10mM DTT, 25 μM all amino acids except cysteine, 100μM cysteine, and then allowed to dry. Spotted papers were soaked in 10% TCA, washed twice in 10% TCA, three times in 5% TCA (all 20 minutes per wash), and once in 95% ethanol (15 minutes). The washed papers were dried, placed in vials with 4 ml fluor (15.3g PPO in 4L of Mallinkrodt ScintillAR toluene) and counted with a Beckman LS7800 scintillation counter (window 0-1000), two minutes per sample. Several control reactions were performed without tRNA in order to estimate background counts.

Data analysis:

Backgrounds were subtracted from the counting data and aminoacylation rates calculated by linear regression. Enzfitter (Biosoft), a non-linear regression program designed for the calculation of kinetic parameters, was used to generate Michaelis-Menten plots and Eadie-Hofstee plots, as well as the values for k_{cat}, K_m, and k_{cat}/K_m. ^{35}S-cysteine has a high rate of non-radiolytic cleavage, making exact determinations of specific activities difficult. Values for k_{cat} were converted from cpm·min^{-1} to nmole·min^{-1}·mg protein^{-1} in one of two ways: 1) An aminoacylation reaction with 1μM native $tRNA^{Cys}$ was allowed to charge to

completion using both ^{14}C and ^{35}S labelled cysteine. This could then be used to develop a conversion factor for cpm min^{-1} to nmole·min^{-1}·mg $protein^{-1}$. The data in Table III was converted in this fashion. 2) Using data from two reactions as described in (1) a standard value for k_{cat} of a T7 transcribed $tRNA^{Cys}$ was determined (in nmole·min^{-1}·mg protein $^{-1}$), and a conversion factor then determined from the calculated value of k_{cat} (in cpm·min^{-1}·unit $protein^{-1}$) . The data in Table I, II, and IV was calculated in this way.

Results:

Comparison of the *in vitro* aminoacylation kinetics of native and T7 transcribed $tRNA^{Cys}$

Before progressing to a study of the aminoacylation kinetics of mutant tRNAs, it was necessary to determine if a T7 transcribed tRNA behaved similarly to the native, modified tRNA. Both tRNAs were aminoacylated in 30mM HEPES pH7.4 with 7mM Mg^{++}, and 2mM ATP. A Mg^{++} titration curve indicates that these conditions are optimal for the T7 transcript; the native tRNA is aminoacylated better at lower Mg^{++} concentrations (data not shown). Under these conditions the T7 transcript has a 1.6-fold increase in k_{cat} and a somewhat unusual 9.5-fold increase in K_m compared to the native tRNA (Figure 2 and Table I). Overall, k_{cat}/K_m for the native tRNA is increased 5.5-fold compared to the T7 transcript.

***In vitro* studies of amber suppressor $tRNA^{Cys}$ mutants**

The wild-type amber suppressor tRNA (CysCUA) and the four mutants acm2, acm5, acm7, and acm8 (Figure 3) were transcribed, purified, and aminoacylated. Figure 4 shows the Michaelis-Menten plot for these tRNAs, and Table I shows the values for k_{cat}, K_m, and k_{cat}/K_m (the selectivity constant) associated with the various mutants. Also shown is the specificity change, the degree to which k_{cat}/K_m is reduced or increased compared to a wild-type

$tRNA^{Cys}$. The range of tRNA concentrations used with the amber suppressors was 2-20μM. Since the measured K_m in many cases was greater than 15μM, determining the individual values of K_m or k_{cat} for many of the tRNAs required significant extrapolation of the Michaelis-Menten equation. Although Enzfitter calculated a standard error for these values, another independent determination of the kinetic constants could easily produce calculated values outside of these ranges. The qualitative effect on the individual kinetic constants (i.e., did k_{cat} or K_m decrease or increase) was clear, however. The value of k_{cat}/K_m does not require significant extrapolation of the Michaelis-Menten curve; its value can be calculated from the x-intercept of an Eadie-Hofstee plot.

Compared to a wild-type transcript (CysGCA) CysCUA is reduced 3100-fold in k_{cat}/K_m and acm2 is reduced 2400-fold; while acm5 and acm7 are reduced 2900- and 3500-fold respectively. Mutant acm8 is reduced 1200-fold in k_{cat}/K_m relative to a wild-type transcript. All mutants had a decreased value of k_{cat} and an increased value of K_m relative to a wild-type transcript (Table I).

These values for the selectivity constant, k_{cat}/K_m are not consistent with the *in vivo* identity data. Even if one considers an error of a factor of two in the calculation of k_{cat}/K_m, there seems to be no possible correlation between *in vivo* identity and the *in vitro* selectivity constant under these conditions. An Eadie-Hofstee plot for tRNA acm2 (37% Cys) shows data points arranged in a vertical line, indicating a high K_m but preventing direct measurement of either of the individual values of k_{cat} or K_m. This high value of K_m may be the factor which causes its loss of *in vivo* identity. In any case, the large reductions in k_{cat}/K_m seen in all of the amber suppressor tRNAs suggested that the anticodon plays an important role in the recognition of $tRNA^{Cys}$ by the cysteinyl-tRNA synthetase.

Effects of mutations in the anticodon and discriminator on recognition by cysteinyl-tRNA synthetase:

In order to determine the role of the anticodon nucleotides in recognition by the cysteinyl-tRNA synthetase, all three single base changes were made at each of the anticodon positions. The mutants are labeled CysXXX where XXX indicates the anticodon present in the tRNA. The importance of the discriminator nucleotide in other systems (and the fact that only one other isoaccepting group in *E. coli,* glycine, has the U73 discriminator) induced us to construct the three single base change mutants at this position (these mutants are labeled Cys73X where X is the nucleotide in the discriminator position). After transcription and gel purification, the kinetic parameters for these tRNAs were determined.

Figure 5 shows the Michaelis-Menten plot for CysGCA, the reference for all of the anticodon and discriminator mutations. Table III shows the experimentally determined values of k_{cat}, K_m, and k_{cat}/K_m for the mutants, as well as the degree to which k_{cat}/K_m was reduced relative to a wild-type $tRNA^{Cys}$ transcript (specificity change). As with the amber suppressor tRNAs, the range of tRNA concentrations (1 to 20μM) does not always allow the most accurate determination of individual values of k_{cat} or K_m, nevertheless, the qualitative effect of the mutation on the kinetic parameters is usually quite clear.

Mutations in nucleotide 34, the wobble position, clearly had dramatic effects on the aminoacylation kinetics (Figure 6 and Table III). All three replacements caused reductions in k_{cat}/K_m of three orders of magnitude relative to a wild-type transcript. All three substitutions increased K_m and decreased k_{cat}. Nucleotide substitution at position 35 had highly variable effects (Figure 7 and 8, Table III). Interestingly, the C35A (CysGAA) mutation had essentially no effect on either K_m or k_{cat}. The C35U (CysGUA) and C35G (CysGGA) mutations, on the other hand caused reductions in k_{cat}/K_m of 320-fold and 120-fold respectively. Both seem to cause increases in K_m and substantial reductions in k_{cat}. Less dramatic effects on charging occur when nucleotide 36 is replaced (Figure 9 and

Table III). The three substitutions produce reductions in specificity of 120-, 140-, and 41-fold for A36U (CysGCU), A36G (CysGCG), and A36C (CysGCC) respectively. All of these substitutions cause reductions in k_{cat} and increases in K_m.

The most dramatic effects caused by the replacement of a single nucleotide are seen with the substitutions at the discriminator nucleotide, position 73 (Figure 10 and Table III). The U73C mutation (Cys73C) produces the best substrate, causing a reduction in specificity of 3200-fold compared to a wild-type transcript. The data suggests that this effect is primarily a k_{cat} effect, with only a slight increase in K_m, although this data must be interpreted cautiously due to the poor aminoacylation of this substrate. The U73A (Cys73A) mutant has a 13,000-fold reduction in k_{cat}/K_m, again with a suggestion that the effect is almost entirely caused by a decrease in k_{cat}, with a minor effect (about 2-fold) on K_m. Mutant Cys73G (U73G) could not be aminoacylated to any significant degree, and so it presumably had an even more dramatic reduction in k_{cat}/K_m than Cys73A.

Determination that elements other than the anticodon and discriminator contribute to recognition:

To determine whether the anticodon is sufficient to direct aminoacylation by the cysteinyl-tRNA synthetase, charging reactions were carried out using the tRNAs shown in Figure 11 (AS1 and AS2). These tRNAs were designed and cloned by M. Saks and J. Sampson (pers. comm.) and are based on the $tRNA^{Ala}$ (UGC) sequence with two non-alanine acceptor stems. In addition the following changes were made to the $tRNA^{Ala}$ sequence: the anticodon and base pair 31:39 (in the anticodon stem) of alanine were replaced with those of cysteine, and base pair U49:A65 in the TΨC stem was changed to A49:U65 to promote better stacking interactions with the two new acceptor stems. Both tRNAs have

the C73 discriminator nucleotide, as opposed to the U73 discriminator found in $tRNA^{Cys}$.

The two tRNAs were transcribed, purified and then aminoacylated under the standard reaction conditions. The tRNAs were not charged to any significant degree, despite continuing the reactions for five minutes (twice as long as a standard reaction), at RNA concentrations as high as 20μM. The presence of the C73 discriminator would be expected to cause a 3200-fold reduction in k_{cat}/K_m (see previous section), but this is measurable under the standard conditions. This implies that these tRNAs lack recognition elements for CysRS, or that they lack certain structural features which are required to allow proper interaction between the synthetase and recognition elements .

Localization of other recognition elements:

To help define the recognition elements or structural features missing from the AS1 and AS2 tRNAs, a set of tRNAs were designed in which structural domains of the cysteine tRNA were replaced by the equivalent regions from the AS1 tRNA. The domains (shown in Figure 12) are the acceptor stem, the anticodon stem and loop, and the tertiary domain (comprising the D stem and loop, TΨC stem and loop, extra arm, and propeller twist nucleotides). Dividing the tRNA into these three domains should not affect the interactions which maintain the three-dimensional shape of the tRNA, which occur between nucleotides in the tertiary domain. Hybrid tRNAs should therefore adopt the three-dimensional configuration of the tRNA species which contributed the tertiary domain.

AS1 was used as the source for heterologous sequences for several reasons. First, it obviously lacks recognition elements for CysRS. Second, it differs from $tRNA^{Cys}$ in both a sequence and structure specific fashion. $tRNA^{Cys}$ is a D3V4 tRNA, AS1 is type D4V5; crystalographic evidence from yeast $tRNA^{Asp}$ (see Westhof et al., 1985) suggests that D3V4 tRNAs adopt a

somewhat different three-dimensional conformation than D4V5 tRNAs. In addition, the number and location of the dihydrouridine residues which define the D-loop and the unusual G15:G48 Levitt pair found in $tRNA^{Cys}$ vary between $tRNA^{Cys}$ and AS1. Finally, the acceptor stem of AS1 is different from that of $tRNA^{Cys}$ at more positions (12 of 14) than the AS2 (which is only different at 4 positions).

The five mutant tRNAs which were constructed are shown schematically in Figure 13. Three of the tRNAs are single substitutions of the domains. Two additional tRNAs with substitutions in two of the domains were constructed to investigate whether any recognition elements in the regions acted independently. The tRNAs are designated by the Greek letter delta followed by the name of the domain(s) which have been replaced with AS1 sequences (**acc**eptor stem, **anti**codon stem/loop, or **tertiary**). All of the tRNAs had anticodon GCA and the U73 discriminator, since these are known to be recognition elements for CysRS. Table 4 lists the five tRNAs with their measured kinetic constants, K_m, k_{cat}, and k_{cat}/K_m, and Figure 14 shows the Michaelis-Menten plots for these tRNAs.

The single domain replacement hybrids varied widely in cysteine acceptor activity. Replacement of the cysteine anticodon stem/loop (mutant Δanti) with the equivalent domain from the AS1 tRNA causes a decrease in K_m from 3.8μM to 1.4 μM, with a concomitant reduction in k_{cat} from 1,100 to 680 nmole·min^{-1}·mg $protein^{-1}$. This results in a mild 1.7-fold increase in k_{cat}/K_m for aminoacylation as compared to a wild-type $tRNA^{Cys}$ transcript (CysGCA). Replacement of the tertiary domain with AS1 sequences (mutant Δtertiary) results in an increase in K_m to 26μM, and a decrease in k_{cat} to 230 nmole·min^{-1}·mg $protein^{-1}$. The value of k_{cat}/K_m is thus reduced 33-fold in this mutant from that of a CysGCA transcript. Replacement of the acceptor stem with the AS1-U73 stem in mutant Δacc causes an increase in K_m (from 3.8μM to 10.3μM), and an insignificant

increase in k_{cat}, (from 1,100 to 1,200 nmole·min^{-1}·mg protein^{-1}). Overall, k_{cat}/K_m is reduced a minor 2.4-fold compared to a wild-type cysteine transcript.

Double replacement mutant ΔaccΔanti has decreased values of K_m and k_{cat} when compared to the single mutant Δacc, the same qualitative effects caused by the single mutant Δanti with regard to the wild-type $tRNA^{Cys}$. The value of k_{cat}/K_m is essentially the same as the wild-type transcript, the result which would be expected from two mutations cause respectively an increase of 1.7-fold and a decrease in 2.4-fold in k_{cat}/K_m with regard to CysGCA. (Two mutations acting independently would be expected to cause an increase or decrease in k_{cat}/K_m approximately equal to the product of their individual effects, see Sampson et al., 1992.) The two replacements in ΔtertiaryΔanti cause a reduction in the values of k_{cat} and K_m compared to single mutant Δtertiary, consistent with the effects of the anticodon stem replacement on a wild-type transcript. However the overall effect on k_{cat}/K_m (an 81-fold reduction) caused by the mutations in ΔtertiaryΔanti is not what would be predicted from two mutations which individually cause a 1.7-fold increase and a 33-fold reduction in k_{cat}/K_m.

Discussion:

Discrimination of tRNAs by synthetases *in vivo* is dictated by a combination of two factors; the total catalytic efficiency of the reaction with the cognate synthetase, and the catalytic efficiency of the reaction with the 19 incorrect synthetases. Competition between all of the synthetases found within the cell ultimately defines the amino acid acceptance of a tRNA *in vivo* (see Sherman et al., 1992, Swanson et al., 1988). The elements which define the amino acid acceptance of a tRNA *in vivo* are called the identity set, and the overall process is tRNA identity. The elements which each of the twenty synthetases use to define a cognate tRNA are called the recognition set, and the

process is tRNA recognition. Recognition elements can act by possessing functional groups which interact with synthetase, or by providing a structure which puts another element in the proper orientation to interact with the synthetase. All recognition elements by definition are positive, that is, the presence of the element aids the creation of a productive interaction with that synthetase. Identity elements can be either positive or negative. A positive identity element would be a recognition element, assisting a productive interaction with the cognate synthetase. A negative element would help prevent productive interactions with noncognate synthetases. Negative elements are critical in the cell, since many synthetases use some of the same tRNA nucleotide elements to help recognize cognate tRNAs (Francklyn et al., 1992).

The results of *in vivo* and *in vitro* experiments provide information on different aspects of the tRNA/synthetase interaction. Since *in vivo* experiments occur in a milieu where all twenty synthetases are present, they provide information on the nature of the identity set They cannot however distinguish whether a change in amino acid acceptance is caused by the removal or addition of a positive or negative identity element. Furthermore the necessity of working with amber or opal suppressors prevents direct experimentation on the effects of the anticodon nucleotides (although an alternate method which uses an initiation assay can study anticodon effects, at least within the context of an $tRNA^{fmet}$ body, see Chattapadhyay et al., 1990). *In vitro* experiments can only examine the recognition set, since the element of competition between synthetases has been removed. The weakness of these experiments is that they cannot detect negative identity elements. The two types of experiments complement each other, so that together it is possible to obtain a more detailed view of the process of tRNA/synthetase interactions.

It is clear from the *in vitro* experiments that the cysteinyl-tRNA synthetase

uses a variety of widely spaced elements as part of its recognition set. These elements include the anticodon nucleotides, the discriminator nucleotide, a feature within the tertiary domain of the tRNA, and possibly a modified nucleotide or nucleotides (or a structural feature produced by a modified nucleotide). The results also suggest that the acceptor stem and anticodon stem play little or no role in recognition. Since it was not possible to make all substitutions at all positions in the tRNA, the possibility exists that there are undiscovered recognition elements within one of the domains. tRNA. The probability of unknown recognition elements remaining in the acceptor stem seems slight however, only one base pair (G1:C72) was not changed in tRNA Δacc.

The results of the aminoacylation reactions with native $tRNA^{Cys}$ suggest the possibility that one or more of the seven modified nucleotides (see Figure 1) may be a recognition element. Since the thymidine and pseudouridine in the TΨC stem are found in all tRNAs it seems unlikely that these residues are involved. Any of these modified nucleotides could influence recognition by either possessing functional groups which interact directly with the synthetase or by inducing subtle variations in the folding of the tRNA. Since the native and T7 transcribed tRNAs are aminoacylated best at different $MgCl_2$ levels (data not shown) there is reason to suspect that the latter explanation is the most likely.

The domain replacement experiments suggest that there are no recognition elements in the acceptor stem or anticodon stem. The minor effects on k_{cat}/K_m caused by these mutants are most likely caused by reorienting other important elements (perhaps the discriminator, CCA terminus, or the anticodon) or by making small changes in helix geometry or some other element of the tertiary structure of the molecule. It does appear that there are recognition elements and/or important structural features within the tertiary domain of $tRNA^{Cys}$. This is particularly interesting since the body of *E. coli* $tRNA^{Cys}$ may

have a conformation which is different from that of yeast $tRNA^{Phe}$, the standard model for tRNA structure. *E. coli* $tRNA^{Cys}$ has an unusual G15:G48 Levitt pair, found in no other tRNA from this species (and unique to *E. coli* among other sequenced cysteine tRNAs). Sampson and Uhlenbeck (1990) found that replacing the G15:C48 Levitt pair in yeast $tRNA^{Phe}$ with the other common pair A15:U48 resulted in a small, 1.6-fold decrease in k_{cat}/K_m, while replacement with G15:G48 caused a four-fold reduction in k_{cat}/K_m. There was an associated reduction in the rate of lead catalyzed tRNA cleavage, a process which is highly dependent upon having a yeast phenylalanine tRNA-like structure (Behlen et al., 1990). Furthermore, evidence from crystallographic (Westhof et al., 1985) and functional (Giegé et al., 1990, Perret et al., 1992) studies of yeast $tRNA^{Asp}$ suggest that D3V4 tRNAs (like *E. coli* $tRNA^{Cys}$ and yeast $tRNA^{Asp}$) are structurally distinct from D4V5 molecules like $tRNA^{Phe}$.

The element which contributed the most to cysteine recognition *in vitro* was the discriminator nucleotide, U73. The mildest mutational effect (a 3,200-fold reduction in k_{cat}/K_m) was seen in Cys73C, a pyrimidine/pyrimidine swap. Replacement of the discriminator with purines caused even more dramatic effects; Cys73A was reduced 13,000-fold in k_{cat}/K_m, and Cys73G was not significantly aminocylated under these conditions. Curiously, neither the U73A or U73C mutations seemed to cause large changes in the value of K_m; the bulk of the effect appears to be on k_{cat} (this interpretation must be viewed with some caution due to the poor aminoacylation of these substrates). This is consistent with models by Hou and Schimmel (1989a) which postulate that most discrimination *in vivo* is based on differences in k_{cat}. In the absence of a crystal structure it is impossible to determine the exact role of U73, but there are two obvious possibilities. One is that there is direct recognition of the single stranded pyrimidine. The other possibility is that it plays a role similar to that of the G73 in

E. coli tRNA glutamine, which participates in an intramolecular hydrogen bond which orients the CCA terminus (Rould et al, 1989). Mutations of the discriminator in $tRNA^{Gln}$ (Jahn et al., 1991) cause reductions in k_{cat}/K_m of aminoacylation by the GlnRS of 2- (G73A) to 1600- (G73U) fold. The G73A mutation actually decreases both K_m and k_{cat}, while the other two mutations increase K_m about 10-fold.

The anticodon nucleotides also have relatively large effects on the aminoacylation kinetics. Mutations at nucleotide34 have the most dramatic effects, causing reductions in k_{cat}/K_m of three orders of magnitude, with reductions in k_{cat} and increases in K_m. Nucleotide 36 has lesser effects, with reductions of 120-, 140-, and 41-fold for A to U, G, or C substitutions. As with position 36, the effects appear to be mediated by both k_{cat} and K_m. The effect of substitutions at nucleotide 35 is more complicated. Replacement of C35 with U or G causes a 320- or 120-fold reduction in k_{cat}/K_m respectively, while replacement with A causes essentially no effect in either k_{cat} or K_m. It is ironic that the wobble nucleotide appears to be the most important for recognition by CysRS, since it is the least specific element of the codon. It is also important to note that U34 is generally modified in *E. coli* tRNAs (for review see Björk et al., 1987); extrapolating the effect of the U34 mutation *in vivo* may therefore be difficult.

It is interesting to compare and contrast these results with the *in vivo* results of Pallanck et al. (1992). They studied $tRNA^{Cys}$ recognition using an *in vivo* initiation system (Chattapadhyay et al., 1990) in which the anticodon of interest is incorporated into $tRNA^{fmet}$ DHFR is then prepared from a strain in which the initiation codon for the *fol* gene has been mutated to match the anticodon on the $tRNA^{fmet}$ The protein is then sequenced and the results interpreted just like an amber suppression assay. Their results were consistent with these experiments. They found that there was a 14-fold increase in the

amount of DHFR initiated with cysteine when an A73U mutation was made in their $tRNA^{fmet}$ background; the qualitative effect which would be predicted from these experiments. Using a U73 $tRNA^{fmet}$ they found that changing any of the anticodon nucleotides reduced cysteine acceptance *in vivo*, again consistent with these experiments. They did not test a $tRNA^{fmet}$A73U with the GAA anticodon, but previous experiments (Chattapadhyay et al., 1990) tested it with A73. They found a mixture of Phe, Met, and Val (in the ratio 37:9:5), but no cysteine. Since the protein was not derivatized prior to sequencing it is possible that some amount of cysteine was present and not detected.

The *in vitro* experiments indicate that not all of the mutations act indepedently. tRNAs which contain two independent mutations should affect aminoacylation efficiency (i.e., relative k_{cat}/K_m) in proportion to the product of their effects when present alone. For example, a tRNA with two independent mutations which cause reductions in k_{cat}/K_m of 2- and 5-fold with respect to a wild-type tRNA, would have a value of k_{cat}/K_m which is reduced 10-fold with regard to the wild-type. The specificity constant for the cysteine amber suppressor tRNA suggest that anticodon mutations do not act independently. The reduction in k_{cat}/K_m with regard to a wild-type transcript for the amber suppressor is 3100-fold, much less than the 5-6 orders of magnitude which would be predicted for a double mutant in which the individual mutations cause reductions of 3 ond two orders of magnitude. The Δtertiary and Δanti mutations also did not appear to act independently; in a double replacement mutant they caused a reduction in k_{cat}/K_m of 81-fold, rather than the 19-fold reduction predicted from their individual effects. On the other hand, mutations Δacc and Δanti appear to act independently, the double mutant has a predicted specificity change of -1.4-fold, the observed value 1.0.

The experiments also underscore the fundamental observation that *in vivo*

and *in vitro* experiments study different aspects of the tRNA/synthetase interaction. Surprisingly, amber suppressor tRNA acm2, which retains only 37% of its cysteine identity *in vivo* was found to be a better or equivalent substrate (in terms of k_{cat}/K_m) than other amber suppressor tRNAs which retain their identity *in vivo* (89% to 93% cysteine). One possible explanation for these anomalous results is that the cumulative effects of the three base pair changes in acm2 improve recognition for the glutaminyl-tRNA synthetase (perhaps by repositioning the anticodon, which contains two recognition elements for GlnRS, C34 and U35, Jahn et al., 1991). A marginally improved affinity of acm2 for the glutaminyl-tRNA synthetase allows the GlnRS to successfully compete with the cysteinyl-tRNA synthetase for a poor substrate. Another possibility is that the cumulative effects of the mutations raised the K_m of acm2 for the CysRS so that the glutaminyl-RNA synthetase was able to compete successfully for this substrate. If the latter explanation is correct, it would seem to be an exception to the rule proposed by Hou and Schimmel (1989) that k_{cat} produces the dominant effect in determining identity. Perhaps with tRNAs which are such poor substrates, other factors, such as K_m, can be the dominant factor in determining identity *in vivo.*

Whatever the reason for this anomaly, *in vitro* charging experiments do not include the effects of competition between synthetases, hence the data only reflects the recognition set, not the complete identity set Similarly, *in vivo* experiments study the complete identity set rather than the recognition set Only combined studies give a true picture of the interactions of an aminoacyl-tRNA synthetase and a tRNA.

References:

Abelson, J., Gefter, M. L., Barnett, L., Landy, A., Russell, R. L., & Smith, J. D. (1970) *J. Mol. Biol. 47*, 15-28.

Behlen, L. S., Sampson, J. R., DiRenzo, A. B., & Uhlenbeck, O. C. (1990)

Biochemistry 29, 2515-2523.

Björk, G. R., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jönsson, Y. H., & Wikström, P. M. (1987) *Annu. Rev. Biochem. 56*, 263-287.

Bohmann, K. & Isasson, L. (1979) *MGG, Mol. Gen. Genet 176*, 53-55

Chattapadhyay, R., Pelka, H., & Schulman, L. H. (1990) *Biochemistry 29*, 4263-4268

Eriani, G., Delarue, M., Poch, O., Gangloff, J., & Moras, D. (1990) *Nature 347*, 203-206

Eriani, G., Dirheimer, G., & Gangloff, J. (1991) *Nucleic Acids Res. 19*, 265 -269

Francklyn, C., Shi, J.-P., & Schimmel, P. (1992) *Science 255*, 1121-1125.

Giegé, R., Florentz, C., Garcia, A., Grosjean, H., Perret, V., Puglisi, J., Théobald-Dietrich, A., & Ebel, J.P. (1990) *Biochimie 72*, 453-461.

Hou, Y-. M-., & Schimmel, P. (1989) *Biochemistry 28*, 4942-4947.

Hou, Y.-M., Shiba, K., Mottes, C., & Schimmel, P. (1991) *Proc. Natl. Acad. Sci. USA 88*, 976-980.

Jahn, M., Rogers, M. J., & Söll, D. (1991) *Nature 352*, 258-260.

Miller, J. H. (1972) in *Experiments in Molecular Genetics* , Cold Spring Harbor Laboratory Press, New York.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783-8798.

Moras, D. (1992) *TIBS 17*, 159-164.

Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., & Schimmel, P. (1991) *Science 253*, 784-786.

Normanly, J., Masson, J.-M., Kleina, L., Abelson, J., & Miller, J. H. (1986a) *Proc. Natl. Acad. Sci. USA 83*, 6548-6552.

Normanly, J., Ogden, R.C., Horvath, S. J., & Abelson, J. (1986b) *Nature 321*, 213-219.

Pallanck, L., Li, S., & Schulman, L. (1992) *J. Biol. Chem 267*, 7221-7223.

Perret, V, Florentz, C., Puglisi, J. D., & Giegé, R. (1992) *J. Mol Biol. 226*, 323-333.

Peterson, E. T., and Uhlencbeck, O. C. (1992) *Biochemistry 31*, 10380-10389.

Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135-1142

Sampson, J. R., DiRenzo, A. B., Behlen, L. S., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 2523-2532.

Sampson, J. R., Behlen, L. S., DiRenzo, A. B. & Uhlenbeck, O. C. (1992) *Biochemistry 31,* 4161-4167

Sampson, J. R., & Uhlenbeck, O. C. (1988) *Proc. Natl. Acad. Sci. USA 85*, 1033-1037.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA 74*, 5463-5467.

Schulman, L. H., & Pelka, H. (1983) *Proc. Natl. Acad. Sci. USA 80*, 6755-6759.

Schulman, L. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 23-87.

Sherman, J. M., Rogers, M. J., & Söll, D. (1992) *Nucleic Acids Res. 20*, 2847-2852.

Squires, C. & Carbon, J. (1971) *Nature New Biol. 233*, 274-277.

Swanson, R., Hoben, P., Sumner-Smith, M., Uemura, H., Watson, W., & Söll, D. (1988) *Science 242*, 1541-1551

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Miol. Biol. 184*, 119-145.

tRNA	k_{cat} (nmole min^{-1} mg $protein^{-1}$)	K_m (μM)	k_{cat}/K_m	Specificity change (fold)
Native $tRNA^{Cys}_{GCA}$	680 ± 60	0.4 ± 0.1	1600	+5.5
T7 $tRNA^{Cys}_{GCA}$	1,100 ± 50	3.8 ± 0.4	290	*a*

Table I: Kinetic parameters for native and T7 transcribed *E. coli* cysteine tRNAs. Specificity change indicates the level of decrease or increase in k_{cat}/K_m relative to T7 transcribed $tRNA^{Cys}_{GCA}$ a) not applicable.

	k_{cat} (nmole min^{-1} mg $protein^{-1}$)	K_m (μM)	k_{cat}/K_m	Specificity change (fold)	Suppression efficiency (% wild type)	Suppression of ThyA amber	DHFR (% at pos. 10)	
							Cys	Gln
CysGCA[a]	1,100 ±50	3.8 ± 0.4	290	*b*	*b*	*b*	*b*	*b*
CysCUA	4.8 ± 0.5	51 ± 7	0.093	-3100	11-24	(+)	94	6
acm2	*c*	*c*	0.12	-2400	5-12	(+)	37	63
acm5	3.8 ± 0.5	37 ± 6	0.10	-2900	4	(+)	90	10
acm7	2.3 ± 0.4	28 ± 7	0.082	-3500	9	(+)	93	5
acm8[d]	5.6 ± 0.6	20 ± 3	0.28	-1200	20	(+)	89	11

Table II: Kinetic parameters and results of *in vivo* experiments for mutant cysteine amber suppressor tRNA molecules. Specificity change indicates the level of increase or decrease in k_{cat}/K_m relative to that of a T7 transcribed $tRNA^{Cys}$. See text for details about tRNAs. *a*) This data from Table I is reprinted for convenience. *b*) Not applicable *c*) The Eadie-Hofstee plot for this tRNA shows points in a vertical line indicating a high K_m but preventing the direct measurement of k_{cat} or K_m. *d*) This data is from a separate experiment and its specificity change is derived from a separate calculation of k_{cat}/K_m for a wild-type transcript (CysGCA). Values for glutamine acceptance include glutamic acid (12% for acm2, 2% for all others) which is produced from glutamine during the Edman degradation.

	Codon recognized	k_{cat} (nmole min^{-1} mg $protein^{-1}$)	K_m (μM)	k_{cat}/K_m	Specificity change (fold)
CysGCA	Cys	1,100 ± 40	3.1± 0.4	350	*a*
Cys**A**CA	Cys	7.3 ± 0.6	38 ± 5	0.19	-1800
Cys**U**CA	Opal	4.6 ± 0.2	25 ± 2	0.18	-1900
Cys**C**CA	Trp	4.1 ± 0.4	30 ± 5	0.14	-2500
CysG**A**A	Phe	950 ± 40	3.3± 0.4	290	-1.2
CysG**U**A	Tyr	29 ± 2	28 ± 2	1.1	-320
CysG**G**A	Ser	71 ± 7	25 ± 4	2.8	-120
CysGC**U**	Ser	30 ± 5	10 ± 3	2.9	-120
CysGC**G**	Arg	45 ± 2	18 ± 2	2.5	-140
CysGC**C**	Gly	141 ± 6	16 ± 1	8.6	-41
Cys73**A**	Cys	0.11± 0.02	4.0 ±1.6	0.027	-13,000
Cys73**G**	Cys	*b*	*b*	*b*	*b*
Cys73**C**	Cys	0.76± 0.07	7.1 ±1.5	0.11	-3200

Table III: Kinetic parameters for anticodon and discriminator mutants. Specificity change is the level of decrease or increase in k_{cat}/K_m relative to a T7 transcribed $tRNA^{Cys}_{GCA}$. See text for explanation of mutants. a) not applicable. b) no aminoacylation detected.

	k_{cat} (nmole min^{-1} mg $protein^{-1}$)	K_m (μM)	k_{cat}/K_m	Specificity change (fold)
CysGCA[a]	1,100 ± 50	3.8 ± 0.4	290	*b*
Δanti	680 ± 30	1.4 ± 0.2	480	+1.7
Δtertiary	230 ± 30	26 ± 5	8.7	-33
Δacc	1,200 ± 30	10.3 ± 0.5	120	-2.4
ΔaccΔanti	850 ± 20	3.0 ± 0.2	290	-1.0
ΔtertiaryΔanti	65 ± 5	18 ± 2	3.6	-81

Table IV: Kinetic parameters for body replacement mutants of *E. coli* cysteine tRNA.Specificity change indicates the level of increase or decrease in k_{cat}/K_m relative to a T7 transcribed $tRNA^{Cys}_{GCA}$. See text for details about tRNAs. *a*) Data from Table I repeated here for convenience *b*) not applicable.

Figure legends:

Figure 1: Cloning and expression strategy for mutants of $tRNA^{Cys}$. A) RNA sequence of $tRNA^{Cys}$ from *E. coli*. (sequence from Sprinzl et al., 1991) B) Standard cassette produced from six oligonucleotides for cloning tRNAs. See text for details of the cloning.

Figure 2: Michaelis-Menten plot for native and T7 transcribed $tRNA^{Cys}$.

Figure 3: RNA sequence of T7 transcribed amber suppressor mutant tRNAs. The background sequence is the cysteine amber suppressor tRNA. The marked changes occur in the following mutant tRNAs: all changes, acm2; diamonds, acm5; circles, acm7; squares, acm8.

Figure 4: Michaelis-Menten plot for amber suppressor tRNAs. Note that the data for acm8 is from a separate experiment, and its specificity change is calculated from a separate calculation of k_{cat}/K_m for CysGCA, a wild-type $tRNA^{Cys}$.

Figure 5: Michaelis-Menten plot of CysGCA, the reference for the anticodon and discriminator mutants.

Figure 6: Michaelis-Menten plot of wobble position (nucleotide 34) mutants of $tRNA^{Cys}$.

Figure 7: Michaelis-Menten plot of CysGAA.

Figure 8: Michaelis-Menten plot of CysGUA and CysGGA (nucleotide 35) mutants of $tRNA^{Cys}$.

Figure 9: Michaelis-Menten plot of nucleotide 36 mutants of $tRNA^{Cys}$.

Figure 10: Michaelis-Menten plot of discriminator (nucleotide 73) mutants of $tRNA^{Cys}$.

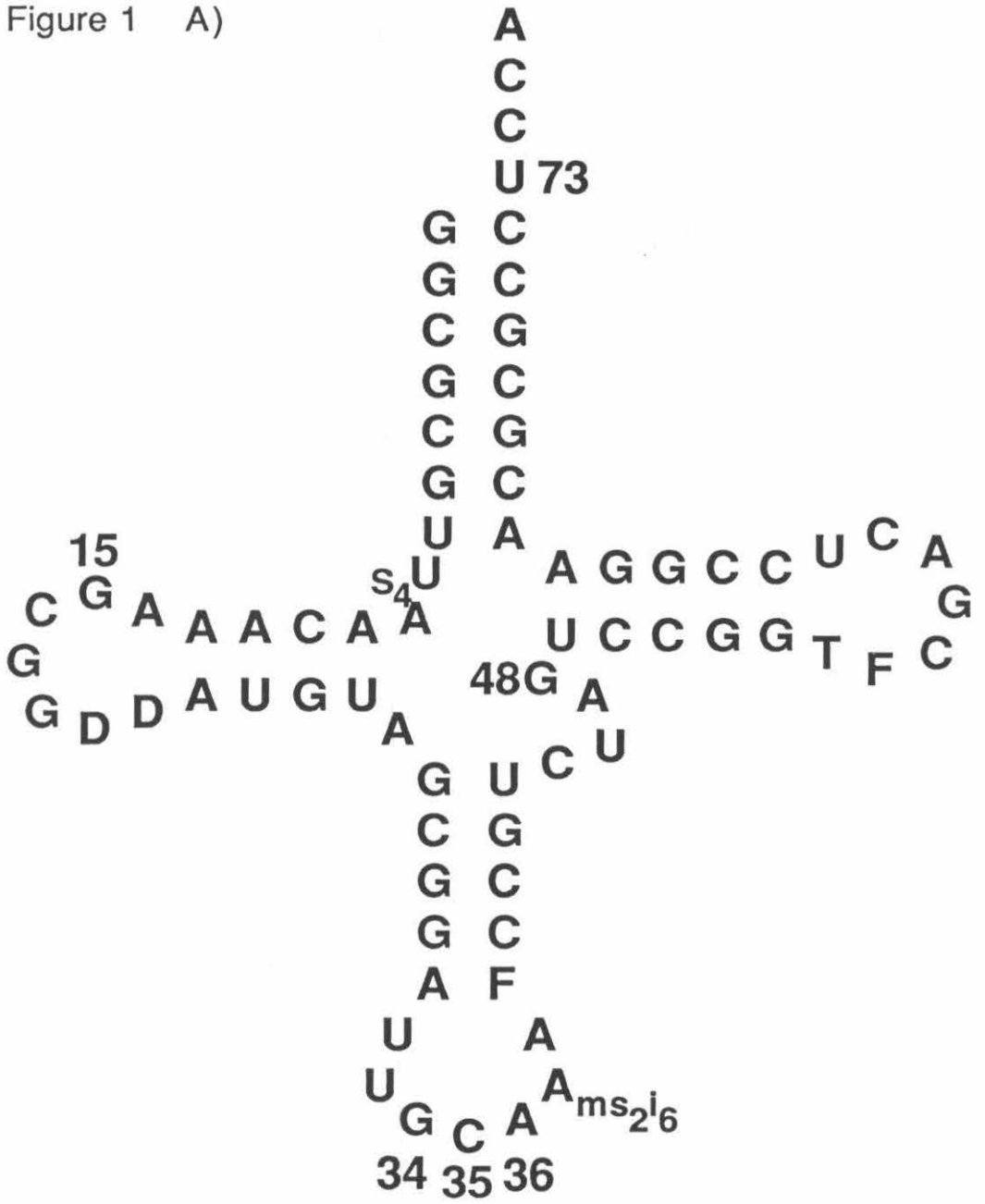
Figure 11: RNA sequence of T7 transcribed AS1 and AS2 tRNAs. See text for details about their design. A) AS1 B) AS2.

Figure 12: RNA sequence of T7 transcribed $tRNA^{Cys}$ showing the breakpoints for the three structural regions used in the domain replacement experiments.

Figure 13: Schematic diagram of the domain replacement mutants. Thick lines represent cysteine sequence, thin lines AS1 sequence.

Figure 14: Michaelis-Menten plot of domain replacement mutants of $tRNA^{Cys}$.

Figure 1 A)


A
C
C
C
U 73
G C
G C
C G
G C
C G
G C
U A
15
G
C
G
A
A A C A
S4U
A
A G G C C U C A
G
U C C G G T F C
G
G
D D A U G U
A
48G A
C U
G U
C G
G C
G C
A F
U A
U
A ms2i6
G C A
34 35 36

B)


EcoRI
T7 promoter
BstNI
PstI
tRNA

Figure 2

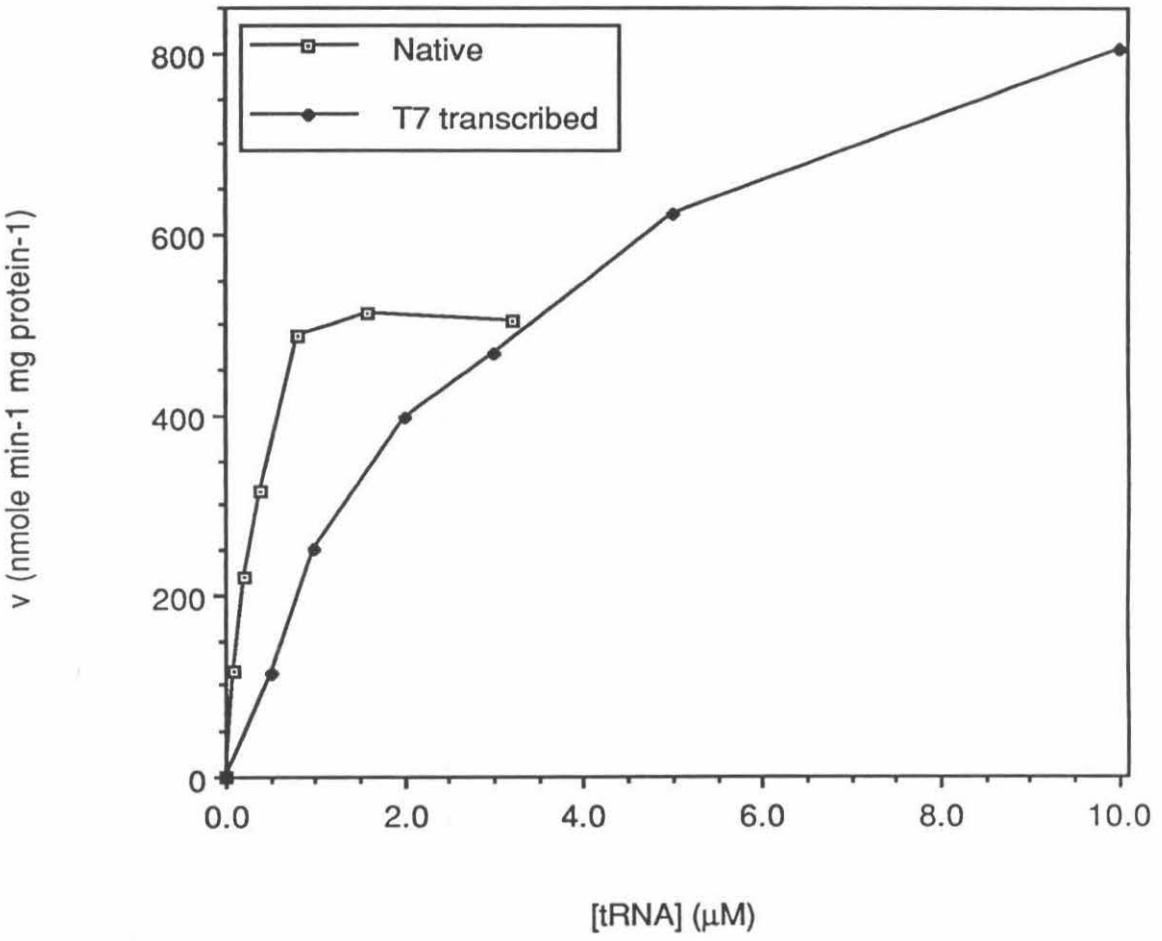

Native
T7 transcribed
800
600
400
200
0
0.0
2.0
4.0
6.0
8.0
10.0
v (nmole min-1 mg protein-1)
[tRNA] (μM)

Figure 3

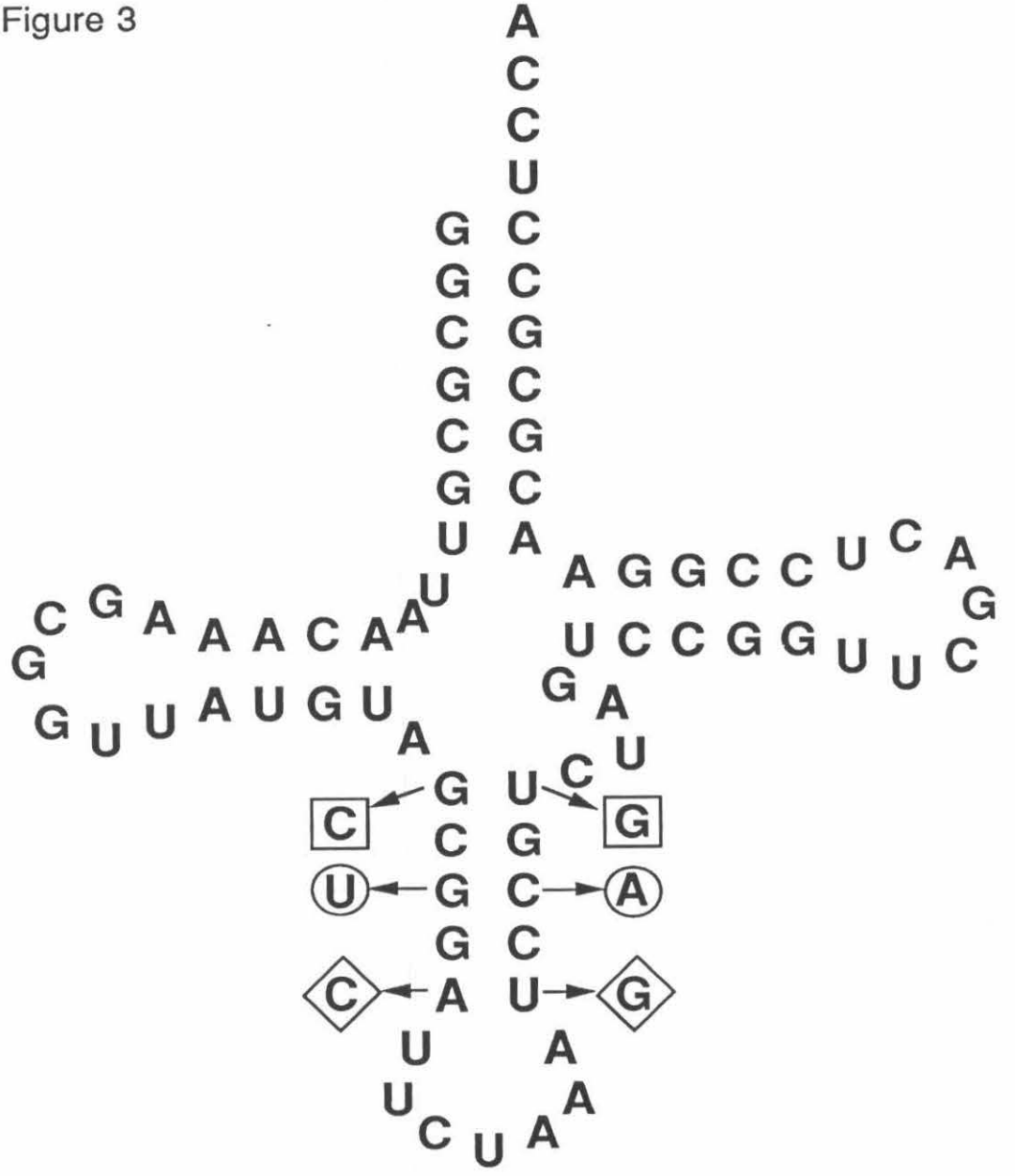

A
C
C
U
G C
G C
C G
G C
C G
G C
U A
A G G C C U C A G
U C C G G U U C
C G A A A C A A U
G
G U U A U G U A
G A
U
C
G U
C G
G C
G C
A U
U
A
U C U A A
C
G
U
A
C
A
G

Figure 4

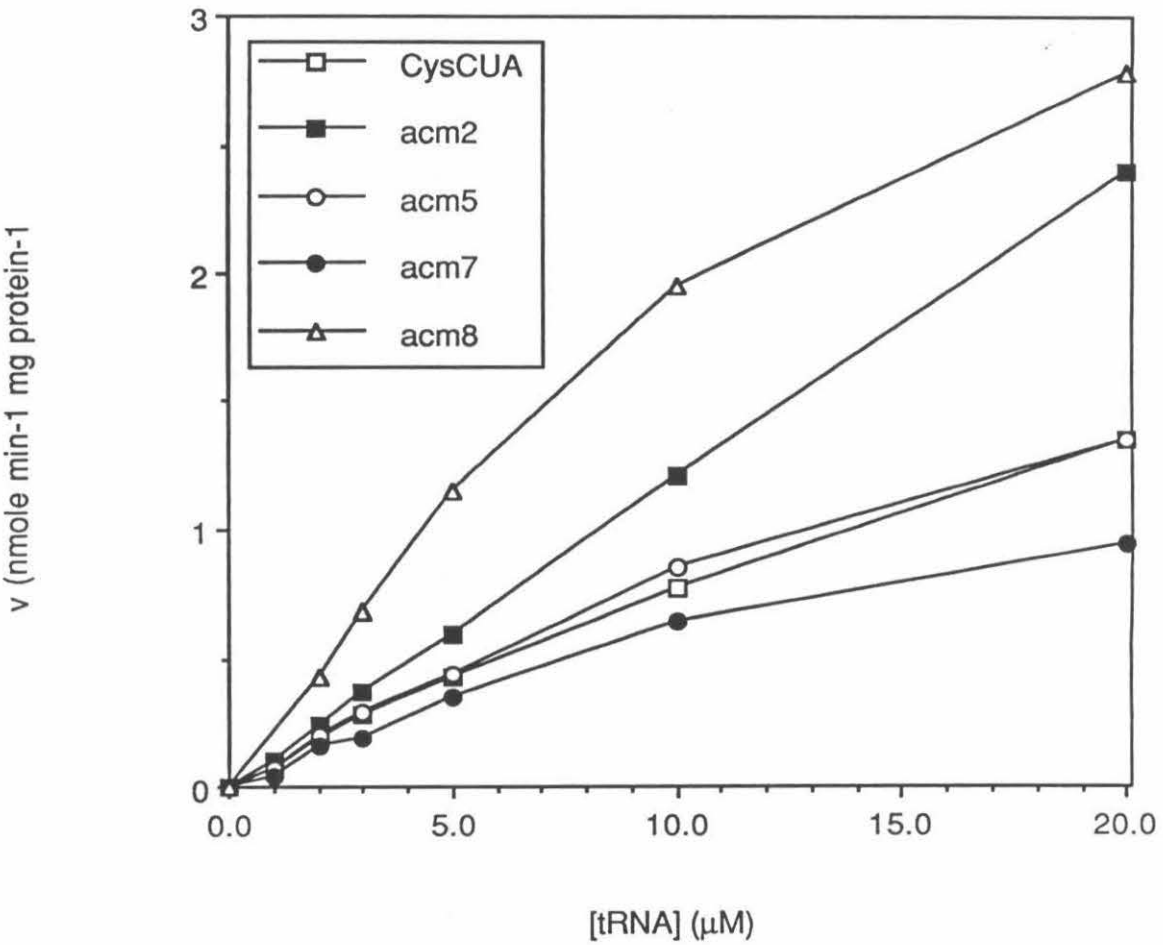

CysCUA
acm2
acm5
acm7
acm8
v (nmole min-1 mg protein-1
[tRNA] (μM)
0
1
2
3
0.0
5.0
10.0
15.0
20.0

Figure 5

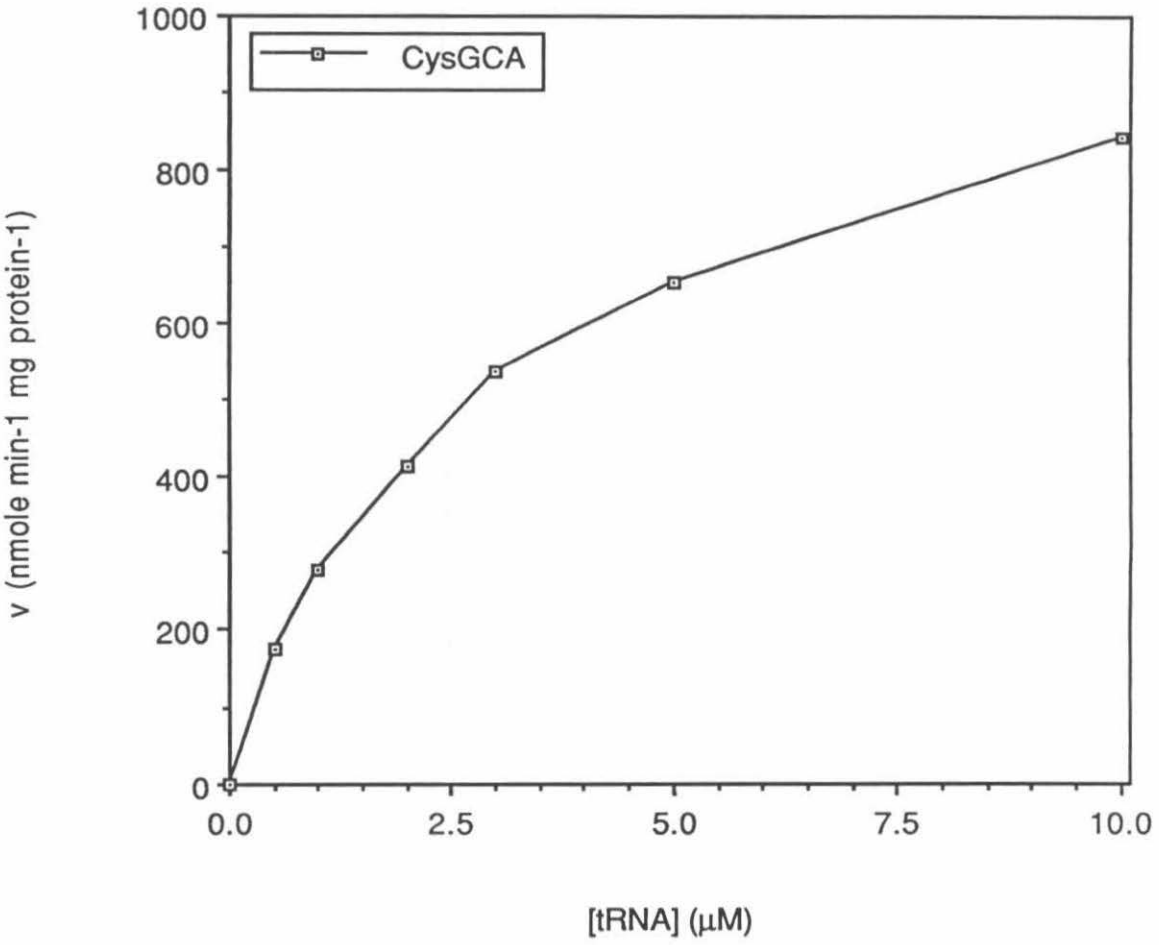

CysGCA
1000
800
600
400
200
0
0.0
2.5
5.0
7.5
10.0
v (nmole min-1 mg protein-1)
[tRNA] (μM)

Figure 6

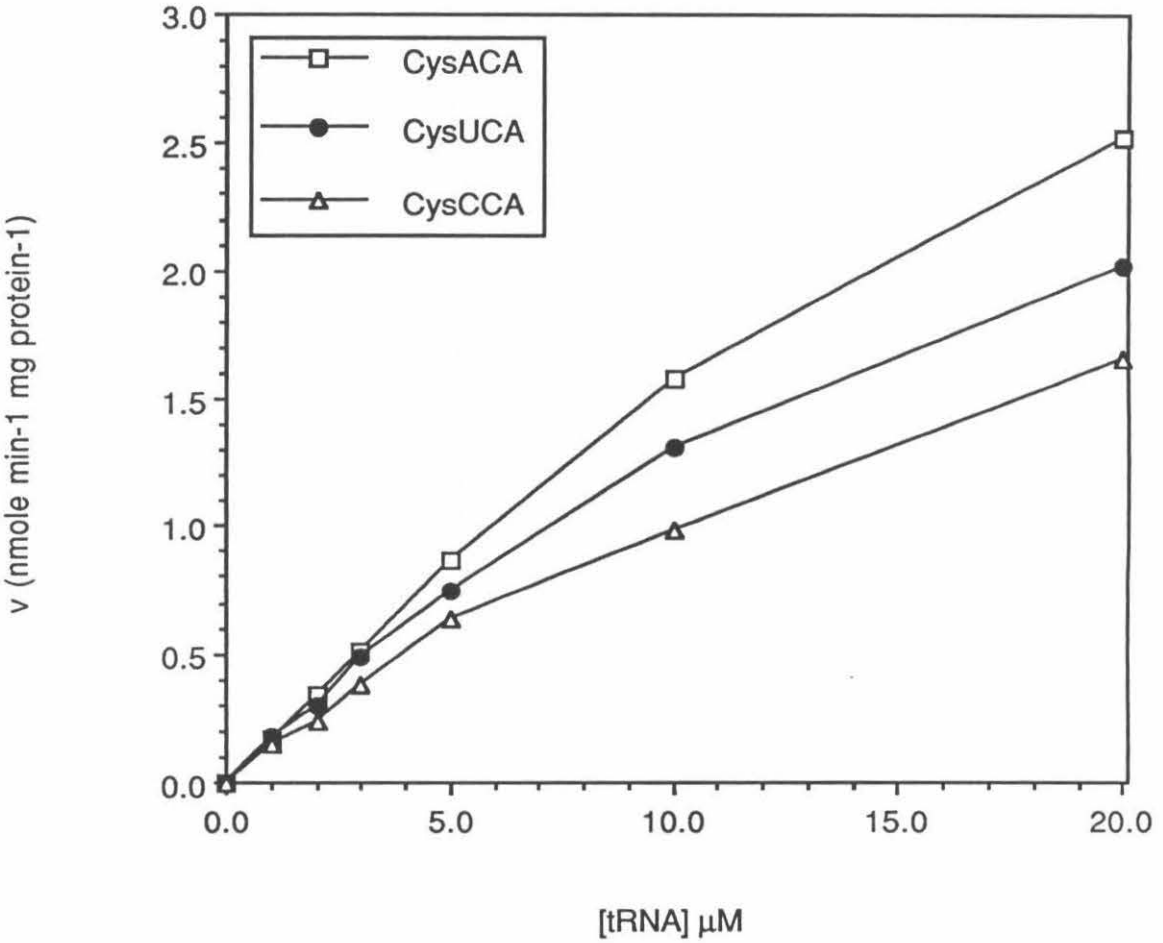

3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
5.0
10.0
15.0
20.0
CysACA
CysUCA
CysCCA
v (nmole min-1 mg protein-1)
[tRNA] µM

Figure 7

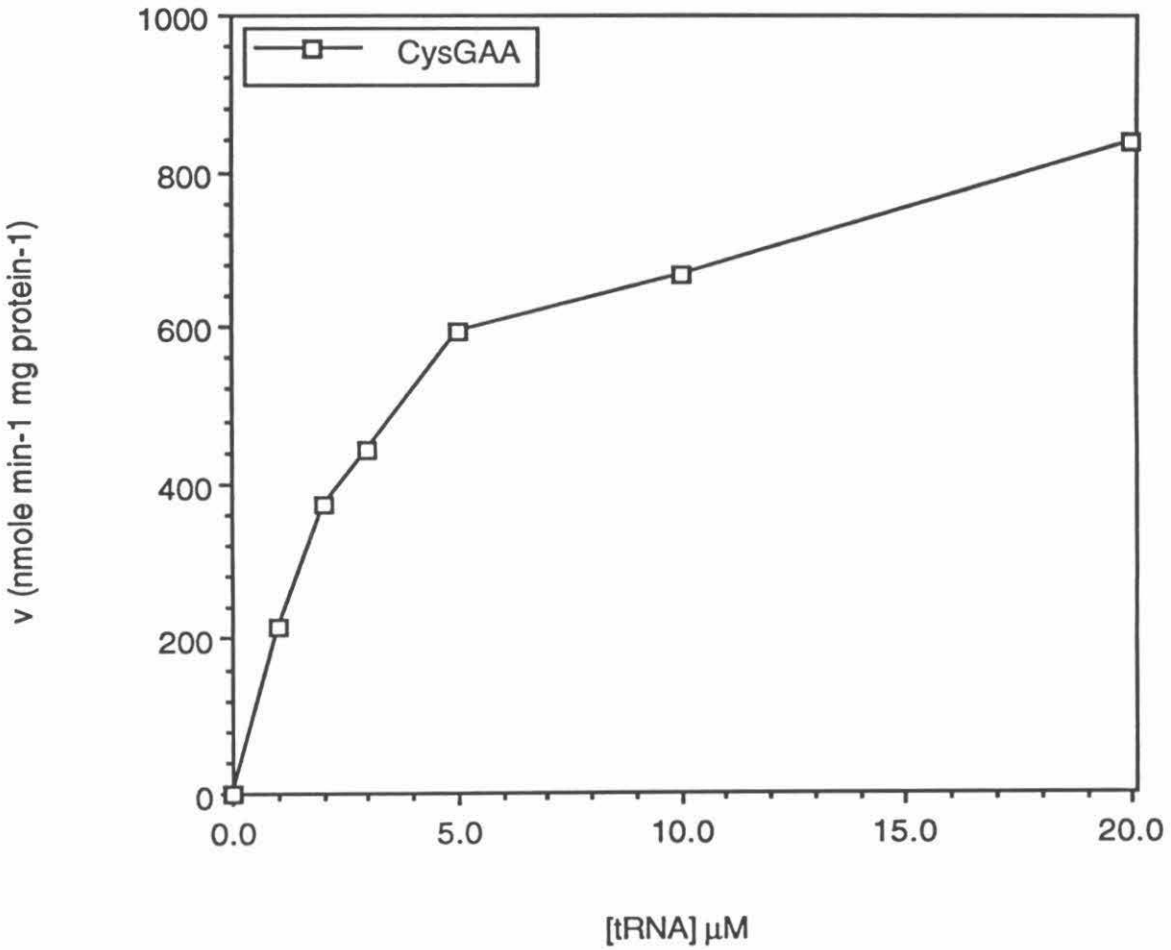

1000
800
600
400
200
0
CysGAA
v (nmole min-1 mg protein-1)
0.0
5.0
10.0
15.0
20.0
[tRNA] µM

Figure 8

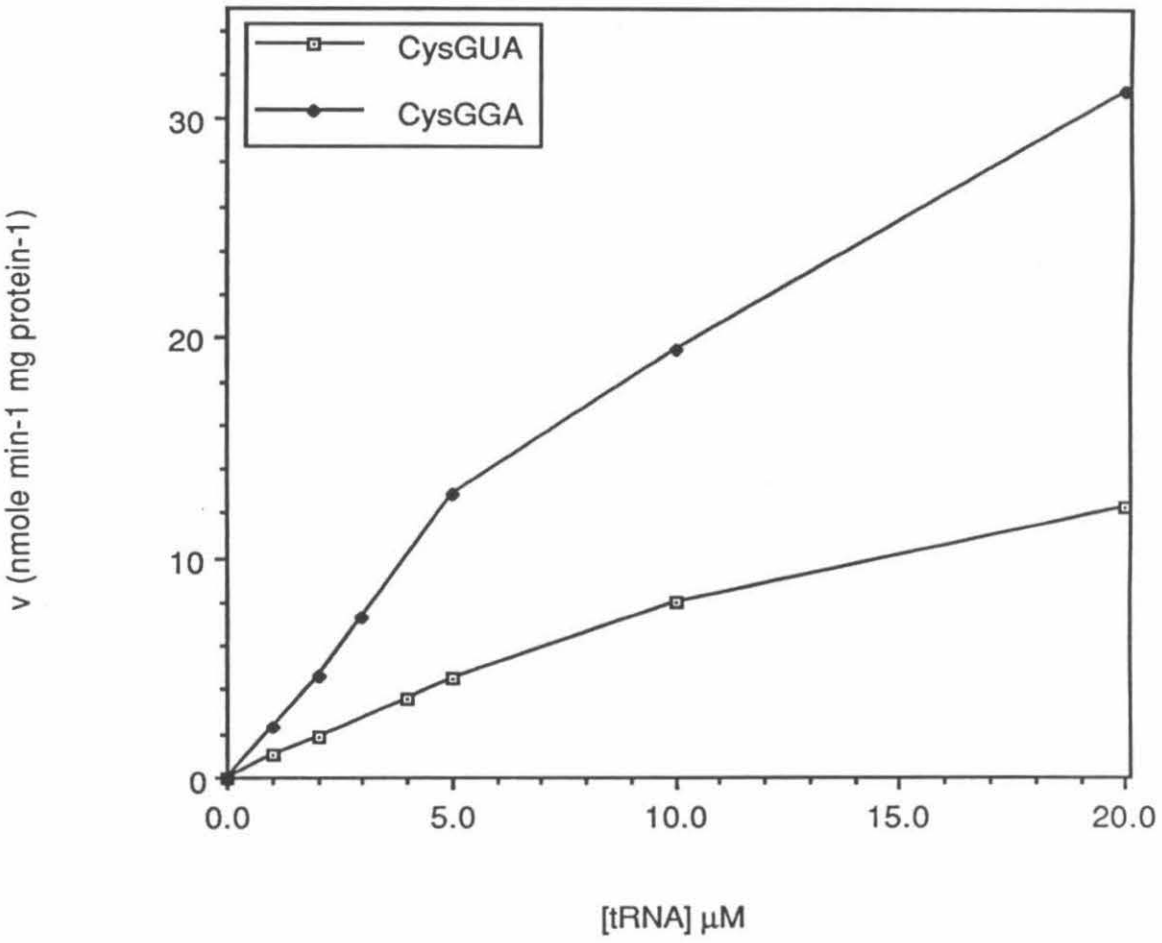

CysGUA
CysGGA
30
20
10
0
0.0
5.0
10.0
15.0
20.0
v (nmole min-1 mg protein-1)
[tRNA] µM

Figure 9

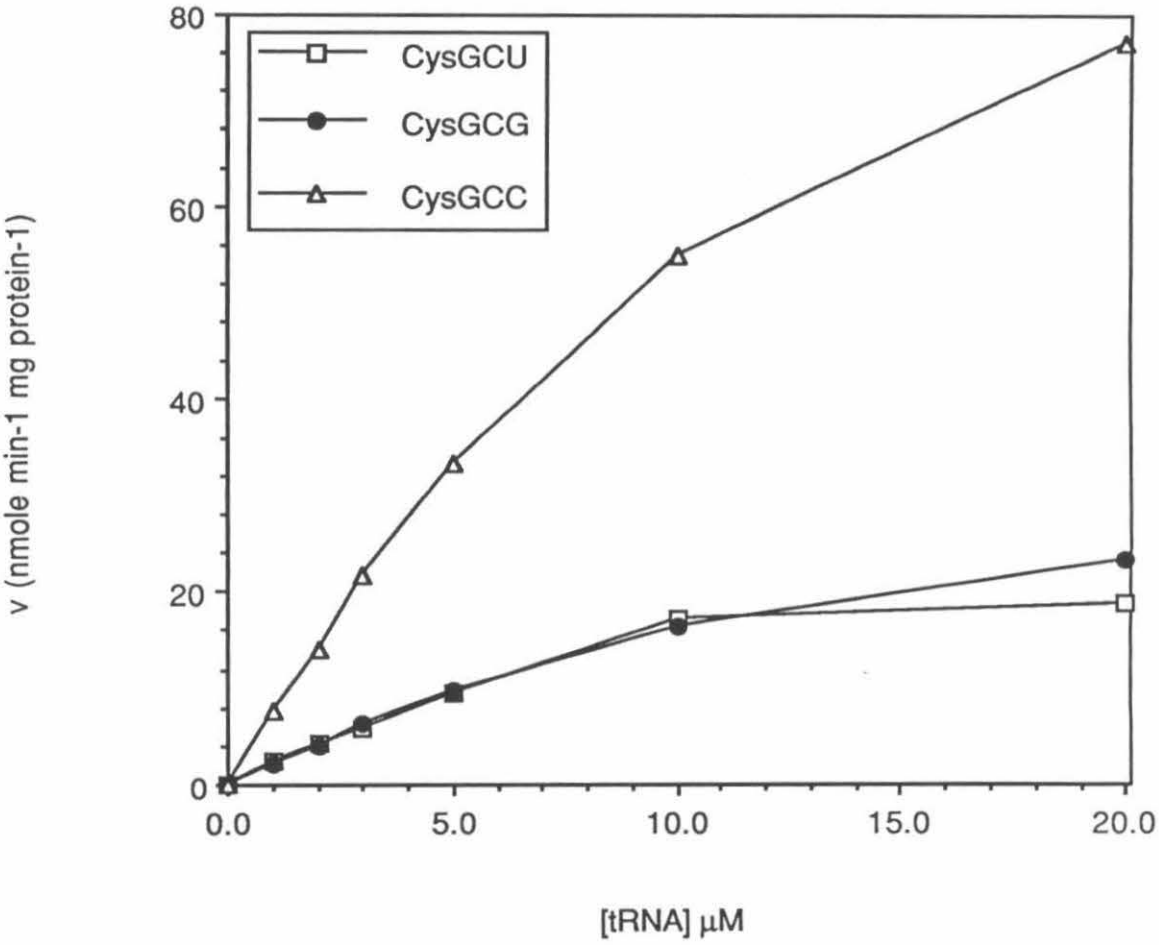

80
60
40
20
0
0.0
5.0
10.0
15.0
20.0
CysGCU
CysGCG
CysGCC
v (nmole min-1 mg protein-1)
[tRNA] µM

Figure 10

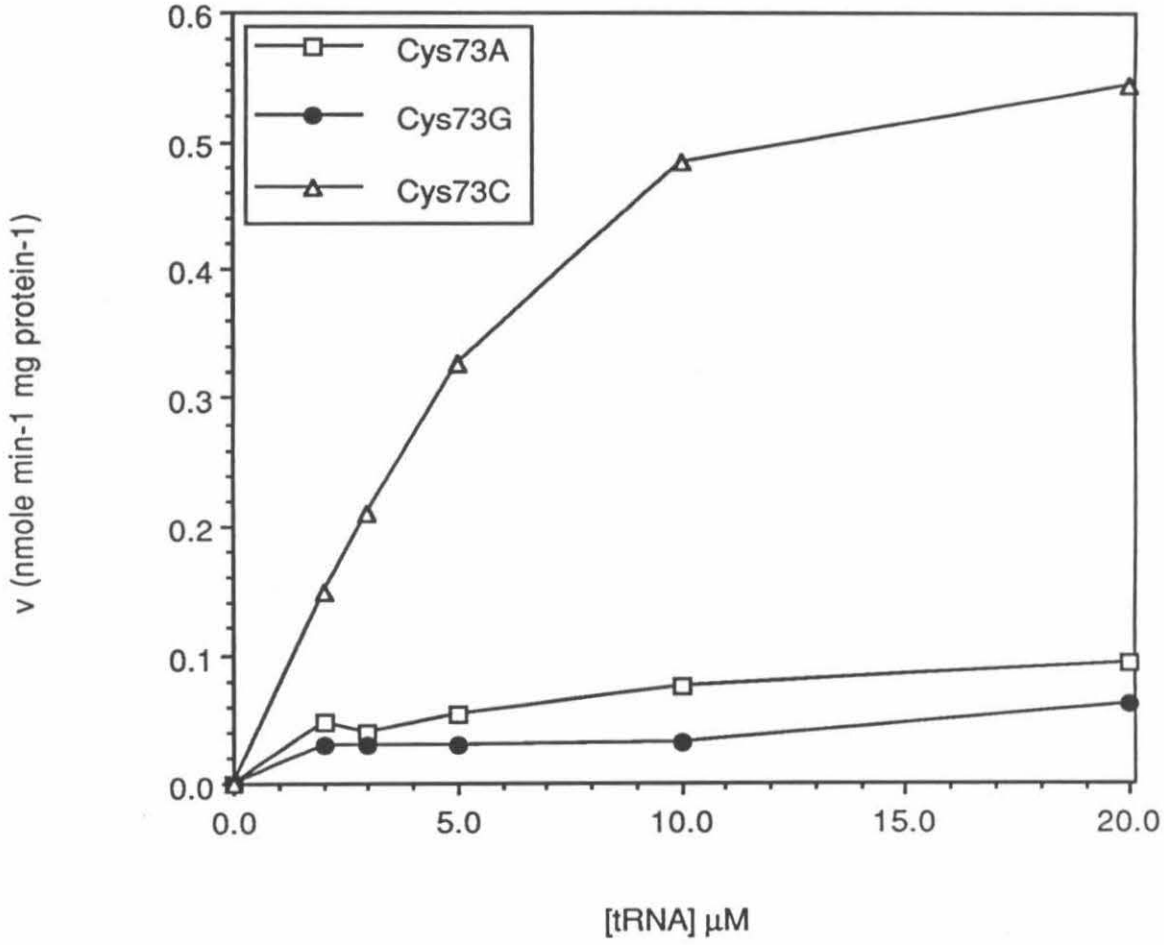
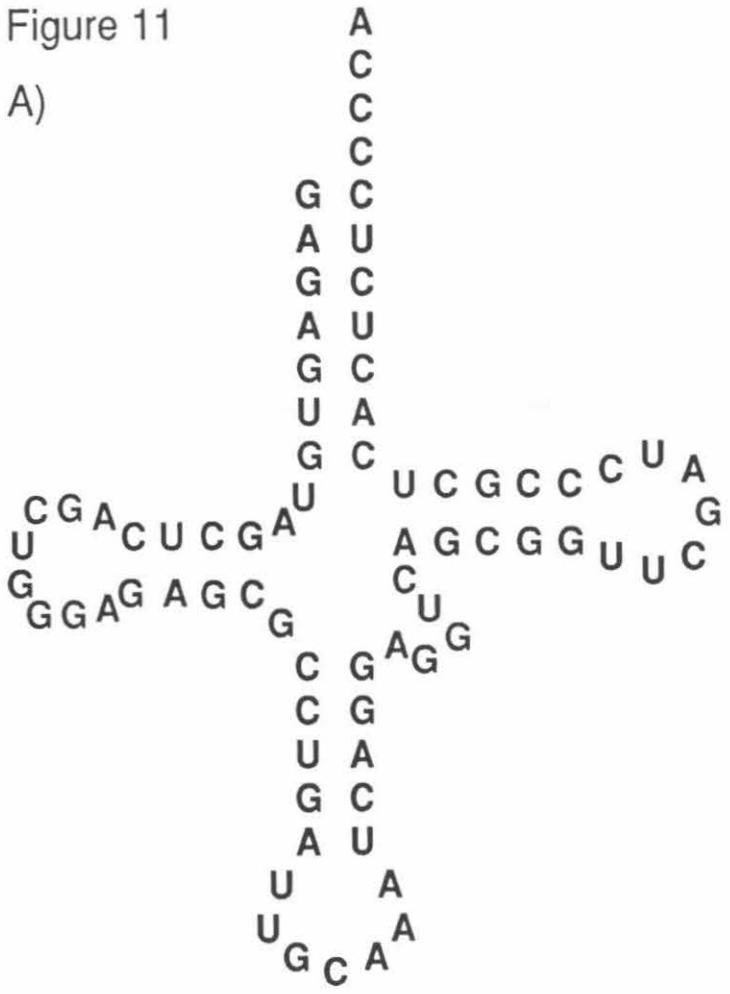

Cys73A
Cys73G
Cys73C
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
5.0
10.0
15.0
20.0
v (nmole min-1 mg protein-1)
[tRNA] µM

Figure 11

A)



B)

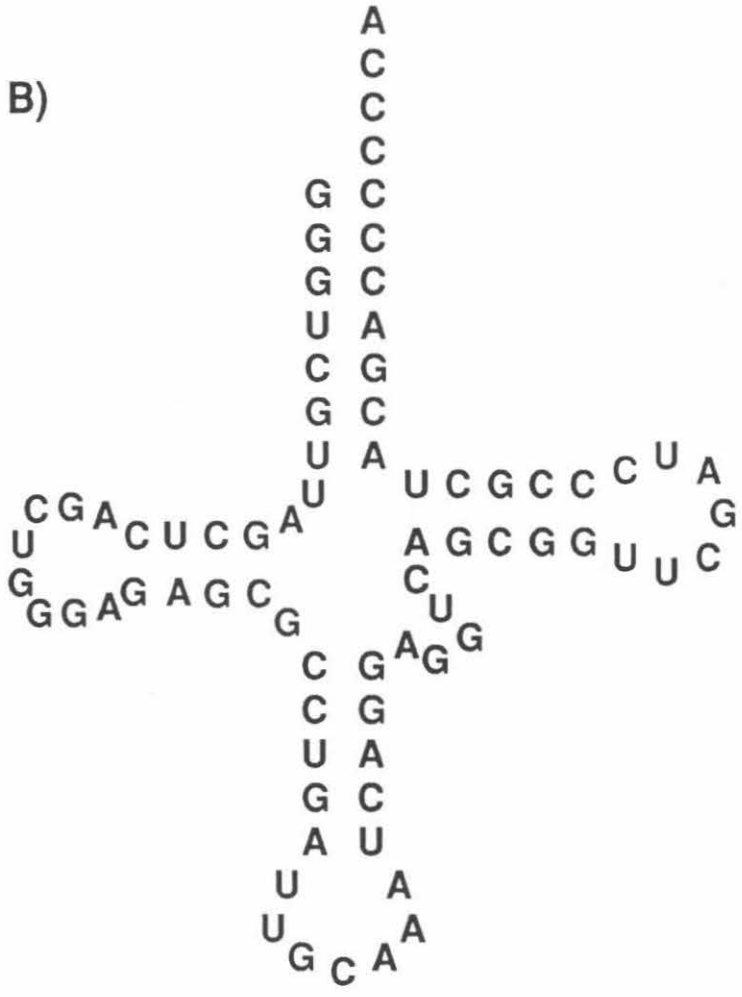


Figure 12

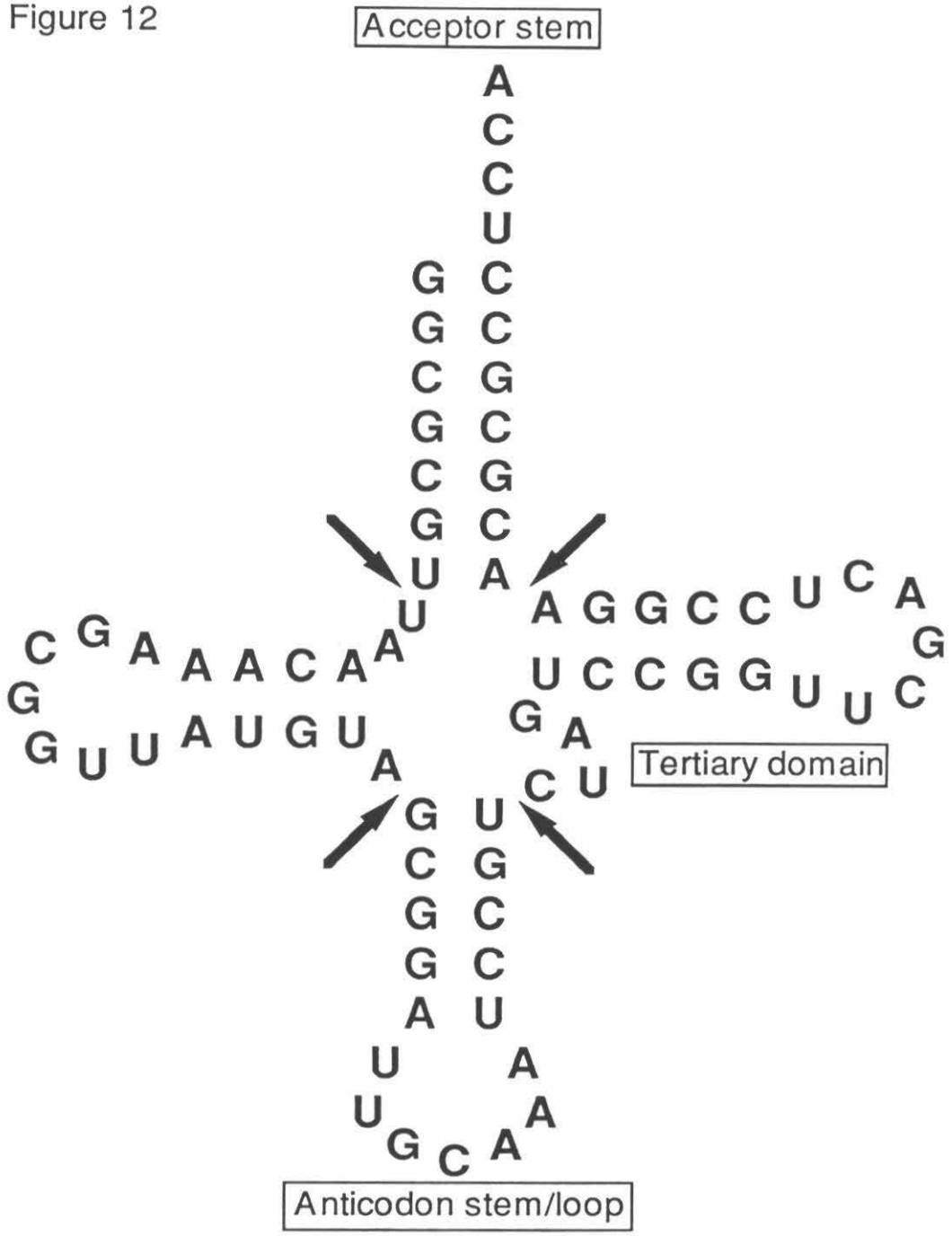


Figure 13

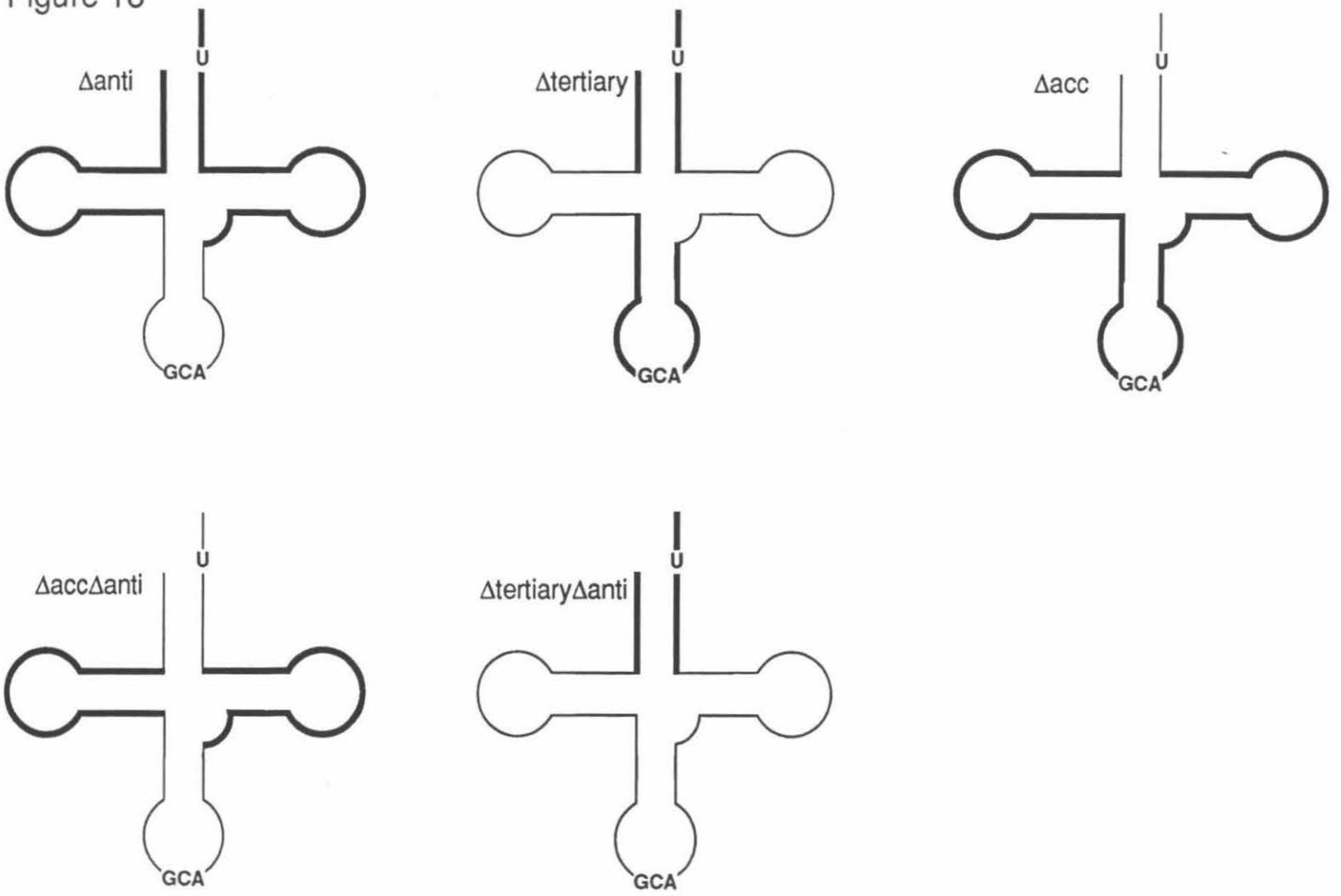
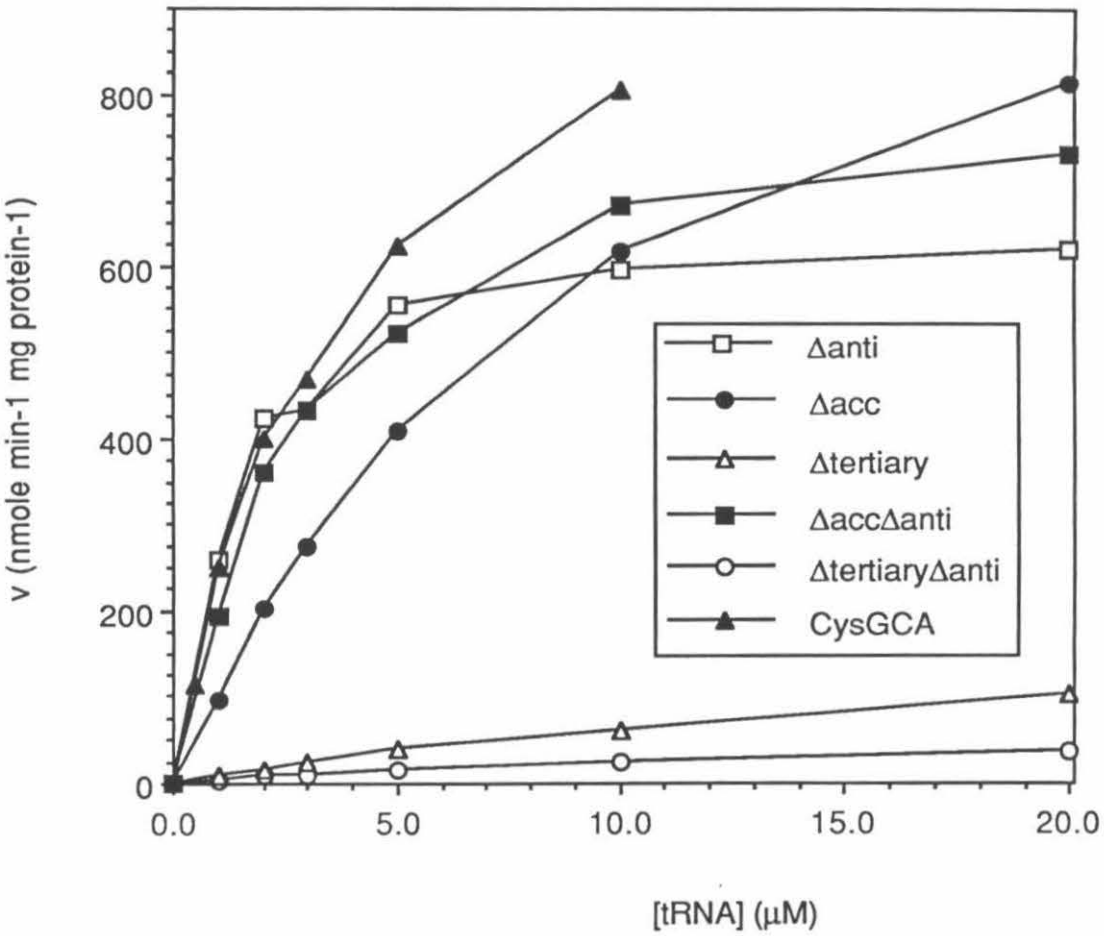

Δanti
U
GCA
Δtertiary
U
GCA
Δacc
U
GCA
ΔaccΔanti
U
GCA
ΔtertiaryΔanti
U
GCA

Figure 14


v (nmole min-1 mg protein-1)
800
600
400
200
0
0.0
5.0
10.0
15.0
20.0
[tRNA] (μM)
Δanti
Δacc
Δtertiary
ΔaccΔanti
ΔtertiaryΔanti
CysGCA

Chapter IV

Conclusions and Future Prospects

Abstract:

The cysteinyl-tRNA synthetase uses the anticodon nucleotides, discriminator (nucleotide 73) and an element(s) within the tertiary domain of the molecule to recognize cognate tRNAs. The identity of the recognition element(s) in the tertiary domain remain to be established, but there are several likely candidates which bear future study, including the two dihydrouridine residues in the D-loop, the A13:A22 mismatch pair in the D-stem, the unusual G15:G48 Levitt pair, and possibly the variable pocket nucleotides (16, 20, 59, and 60). The three-dimensional structure of the tertiary domain might also be a recognition element. Although *in vitro* experiments cannot directly define the identity set, experiments with mutant glycine and phenylalanine tRNAs may be able to help show how these molecules are prevented from being misacylated with cysteine. Finally, the *in vivo* results obtained in this lab and others suggest that tRNA identity is overspecified and hence the notion of a simple set of recognition elements unique to an isoaccepting group (the "second genetic code") is incorrect. Fidelity is guaranteed not by a set of isoacceptor specific elements, but by the sum of many interactions between the synthetase and partially overlapping tRNA elements.

Identification of recognition elements in the tertiary domain:

The tertiary domain of $tRNA^{Cys}$, consisting of the D stem and loop, the TΨC stem and loop, extra arm, and propeller twist nucleotides, contributes to recognition by the cysteinyl-tRNA synthetase. Replacing this region with sequences from a heterologous tRNA which varied in both a sequence and structure specific manner caused a reduction of 31-fold in the selectivity constant (k_{cat}/K_m) for aminoacylation by the cysteinyl-tRNA synthetase (CysRS). The nature of the element(s) which are recognized by the cysteinyl-tRNA synthetase are unknown at this time. Unfortunately, the single isoacceptor of $tRNA^{Cys}$ prevents the use of comparison among cognate tRNAs to help identify recognition elements (see McClain and Nicholas Jr., 1987), but careful study of the sequence of the tRNA suggests several possible elements as targets for future study (see Figure 1).

Unusual sequence elements within $tRNA^{Cys}$ are obvious candidates for recognition elements, and the most unusual element, of course, is the G15:G48 Levitt pair found in no other *E. coli* tRNA. (Komine et al., 1990, Sprinzl et al., 1991). As indicated previously, G15:G48 makes a perfectly reasonable tertiary pair in yeast $tRNA^{Phe}$ but reduces k_{cat}/K_m for the phenylalanyl-tRNA synthetase (FRS) fourfold and appears to subtly alter the three-dimensional structure of the tertiary domain (Sampson et al., 1990). A series of experiments in which the two "normal" Levitt pairs G15:C48 and A15:U48 are inserted into $tRNA^{Cys}$ could provide interesting insight into the structural requirements of the tRNA, even if it turns out not to be the primary recognition element.

The domain replacement experiments modified the sequence and structure of the D stem and loop, suggesting several other possible targets for future experiments. The D stem of $tRNA^{Cys}$ contains only three Watson-Crick base pairs, and an A13:A22 mismatch pair (inferred from the U13:U23

mismatch pair in yeast $tRNA^{Asp}$, Westhof et al., 1985), while the AS1 D stem was composed of four Watson-Crick pairs. The size and arrangement of the D loop also varies between the two tRNAs; in $tRNA^{Cys}$ it is seven nucleotides long and contains two dihydrouridine residues (U in the transcripts) both in the so called ß region as compared to the single D residue in the α region of the 8 nucleotide AS1 loop. In addition, the size of the α region is different between the two tRNAs, AS1 has a four nucleotide α region, $tRNA^{Cys}$ has a three nucleotide region. This implies that nucleotide16 or 17 of the D loop does not exist, and it shifts the location of the conserved G18G19 dinucleotide. A series of tRNAs with various D stem and loop elements could help determine if the recognition elements in the tertiary domain are in this structure. It should be noted that D stem and loop elements are involved in many of the tertiary interactions in tRNAs; mutations in this region may have to be paired with other changes in the variable loop to produce a functional tRNA (see below).

Another interesting set of experiments would be to prepare tRNAs with the variants of the elements of the variable pocket, nucleotides 16, 17 (absent in $tRNA^{Cys}$, present in AS1), 20, 59, and 60 (Ladner et al., 1975). Variable pocket nucleotides are important in several other systems, especially yeast (Sampson et al., 1989) and *E. coli* (Peterson and Uhlenbeck, 1992) $tRNA^{Phe}$ and *E. coli* $tRNA^{Leu}$ (G. Tocchini-Valentini, pers. comm.), making them good candidates for recognition elements in the *E. coli* cysteine system. There is only one common variable pocket nucleotide (C60) between $tRNA^{Cys}$ and AS1, so a series of tRNAs with mutations in the other three residues would be the logical place to begin. Since the two tRNAs have different numbers of variable pocket nucleotides, another interesting experiment could be performed with a $tRNA^{Cys}$ U17 insertion mutant.

It is also possible that the overall structure of the molecule is the important recognition element, or at least important in positioning other recognition elements. The crystal structures of yeast $tRNA^{Phe}$ (D4V5, Kim et al., 1974) and $tRNA^{Asp}$ (D3V4, Westhof et al., 1985) suggest that D3V4 $tRNA^{Cys}$ has a different three-dimensional structure than the D4V5 AS1 tRNA. Furthermore, Sampson et al., (1990) found that the G15:G48 Levitt pair affects the rate of lead cleavage in $tRNA^{Phe}$ suggesting that this too might contribute to a structural difference between AS1 tRNA and $tRNA^{Cys}$. A series of tRNAs which present the presumed AS1 tertiary interactions with or without cysteine sequence elements could also help ascertain the role of the tertiary domain in recognition by the cysteinyl-tRNA synthetase. The example of other systems suggest that structural clues have variable effects. Yeast aspartyl-tRNA synthetase is very tolerant of conformational variability, yeast phenylalanyl-tRNA synthetase is much less so (Giegé et al., 1990, Perret et al., 1992).

***In vitro* methods of studying identity:**

Although *in vitro* methods cannot answer general questions about tRNA identity, it may be possible to use these methods to determine how certain tRNAs are prevented from being misacylated *in vivo*. Specifically, *in vitro* experiments may allow us to determine the mechanisms that prevent *E. coli* tRNAs coding for phenylalanine and glycine from being charged with cysteine *in vivo*.

Cysteine tRNAs with the GAA (Phe) and GCA (Cys) anticodons are equivalent substrates for the cysteinyl-tRNA synthetase, suggesting that some element in $tRNA^{Phe}$ must protect it from being aminoacylated with cysteine. The wild-type $tRNA^{Phe}$ discriminator is A73 which would be expected to reduce the catalytic efficiency of aminoacylation by the cysteinyl-tRNA synthetase by four orders of magnitude, effectively preventing misacylation. By testing the

aminoacylation kinetics of wild-type and U73 $tRNA^{Phe}$ molecules (Figure 2), it will be possible to if the discriminator is the sole defense against cysteine misacylation, or whether other negative elements in the tertiary domain help maintain Phe identity *in vivo*.

There are three glycine specific isoacceptors in *E. coli,* with anticodons UCC ($tRNA^{Gly1}$), GCC ($tRNA^{Gly2}$), and CCC ($tRNA^{Gly3}$). The different anticodon nucleotides would be expected to cause each tRNA to be aminoacylated by the cysteinyl-tRNA synthetase with different catalytic efficiencies. All else being equal $tRNA^{Gly1}$ and $tRNA^{Gly3}$ would be expected to be poor substrates because of the lack of the G34 wobble nucleotide. $tRNA^{Gly2}$ would be expected to be a better substrate (all else being equal) since the GCC anticodon was only reduced 41-fold in total catalytic efficiency in a $tRNA^{Cys}$ background compared to the wild-type GCA anticodon. All glycine tRNAs have a U at position 73, just like $tRNA^{Cys}$; the discriminator therefore cannot be used to prevent misacylation. Curiously, $tRNA^{Gly1}$ is D3V4 like $tRNA^{Cys}$ while $tRNA^{Gly2}$ is D4V5. It is intriguing to speculate that the $tRNA^{Cys}$-like $tRNA^{Gly1}$ is prevented from mischarging with Cys primarily by the presence of U34 in the anticodon, while G34 containing $tRNA^{Gly2}$ is prevented from mischarging by virtue of the structural differences inherent in its D4V5 tertiary domain. This is testable by aminoacylating mutants of $tRNA^{Gly1}$ and $tRNA^{Gly2}$ with both the UCC and GCC anticodons (Figure 3).

Final considerations on tRNA specificity:

The results of studies on *E. coli* $tRNA^{Gln}$ (Jahn et al., 1991, Normanly et al., 1990), $tRNA^{Cys}$ (this work, Normanly et al., 1986), and $tRNA^{Phe}$ (Peterson and Uhlenbeck, 1992, Normanly et al., 1986) suggests that tRNAs are overspecified; mutants in which multiple recognition elements have been replaced reduce the total catalytic efficiency of the aminoacylation reaction but

do not change *in vivo* identity. In the examples above the anticodon is an important recognition element, yet amber suppressors made from these tRNAs retain their *in vivo* identity, despite the dramatic reductions in overall catalytic efficiency caused by the CUA anticodon. This suggests that a paradigm for discrimination must be devised which does not rely on recognition of a few unique elements (the so-called "second genetic code") but rather on competition between synthetases for substrates containing various numbers of somewhat overlapping recognition elements (for instance $tRNA^{Cys}$ and $tRNA^{Phe}$ both recognize A36 as part of their recognition sets: this work, Peterson and Uhlenbeck, 1992). Competition, and the balance between synthetase and tRNA concentrations ensure the fidelity of aminoacylation under this paradigm (general reference: Söll, 1990; specific experiments: Swanson et al., 1988, Sampson and Uhlenbeck, 1992, Sherman et al., 1992). Although not as elegant as the simple model of the "second genetic code," it suggests that tRNA recognition and identity are dynamic processes, robust enough to tolerate the mutations which are the driving force in evolution.

References:

Giegé, R., Florentz, C., Garcia, A., Grosjean, H., Perret, V., Puglisi, J., Théobald-Dietrich, A., & Ebel, J.P. (1990) *Biochimie 72*, 453-461.

Jahn, M., Rogers, M. J., & Söll, D. (1991) *Nature 352*, 258-260.

Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Proc. Natl. Acad. Sci USA 71*, 4970-4974.

Komine, Y., Adachi, T., Inokuchi, H., & H. Ozeki (1990) *J. Mol. Biol. 211*, 579-598.

Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C., & Klug, A. (1975) *Proc. Natl. Acad. Sci. USA 72*, 4414-4418.

McClain, W. H., & Nicholas Jr., H. B. (1987)*J. Mol. Biol. 194*, 635-642.

Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J., & Miller, J. H. (1990) *J. Mol. Biol. 213*, 719-726.

Normanly, J., Masson, J.-M., Kleina, L., Abelson, J., & Miller, J. H. (1986) *Proc. Natl. Acad. Sci. USA 83*, 6548-6552.

Perret, V, Florentz, C., Puglisi, J. D., & Giegé, R. (1992) *J. Mol Biol. 226*, 323-333.

Peterson, E. T., and Uhlenbeck, O. C. (1992) *Biochemistry 31*, 10380-10389.

Sampson, J. R., DiRenzo, A. B., Behlen, K. S., & Uhlenbeck, O. C. (1989) *Science 243*, 1363-1366.

Sampson, J. R., DiRenzo, A. B., Behlen, K. S., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 2523-2532.

Sampson, J. R., Behlen, L. S., DiRenzo, A. B. & Uhlenbeck, O. C. (1992) *Biochemistry 31,* 4161-4167

Sherman, J. M., Rogers, M. J., & Söll, D. (1992) *Nucleic Acids Res. 20*, 2847-2852.

Söll, D. (1990) *Experientia 46*, 1089-1096.

Sprinzl, M., Dank, N., Nock, S., & A. Schön (1991) *Nucleic Acids Res. 19*, 2127-2171.

Swanson, R., Hoben, P., Sumner-Smith, M., Uemura, H., Watson, L., & Söll, D. (1988) *Science 242*, 1548-1551.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119-145.

Figure legends:

Figure 1: RNA sequence of $tRNA^{Cys}$. Presumptive variable pocket nucleotides are circled. The α and ß regions of the D loop are boxed and indicated. The two nucleotides which make the Levitt pair (G15 and G48) are indicated. Modified nucleotides are shown as the parent nucleotide.

Figure 2: RNA sequences of *E. coli* and yeast $tRNA^{Phe}$. Proposed mutations are shown as arrows away from the sequence. The mutation in base 3:70 of the *E. coli* tRNA is to improve transcription by the T7 RNA polymerase. Modified nucleotides are shown as the parent nucleotide. A) *E. coli* $tRNA^{Phe}$ B) Yeast $tRNA^{Phe}$.

Figure 3: RNA sequence of yeast *E. coli* $tRNA^{Gly1}$ and $tRNA^{Gly2}$ showing proposed mutations. Modified nucleotides are shown as the parent nucleotides. A) $tRNA^{Gly1}$ B) $tRNA^{Gly2}$.

Figure 1

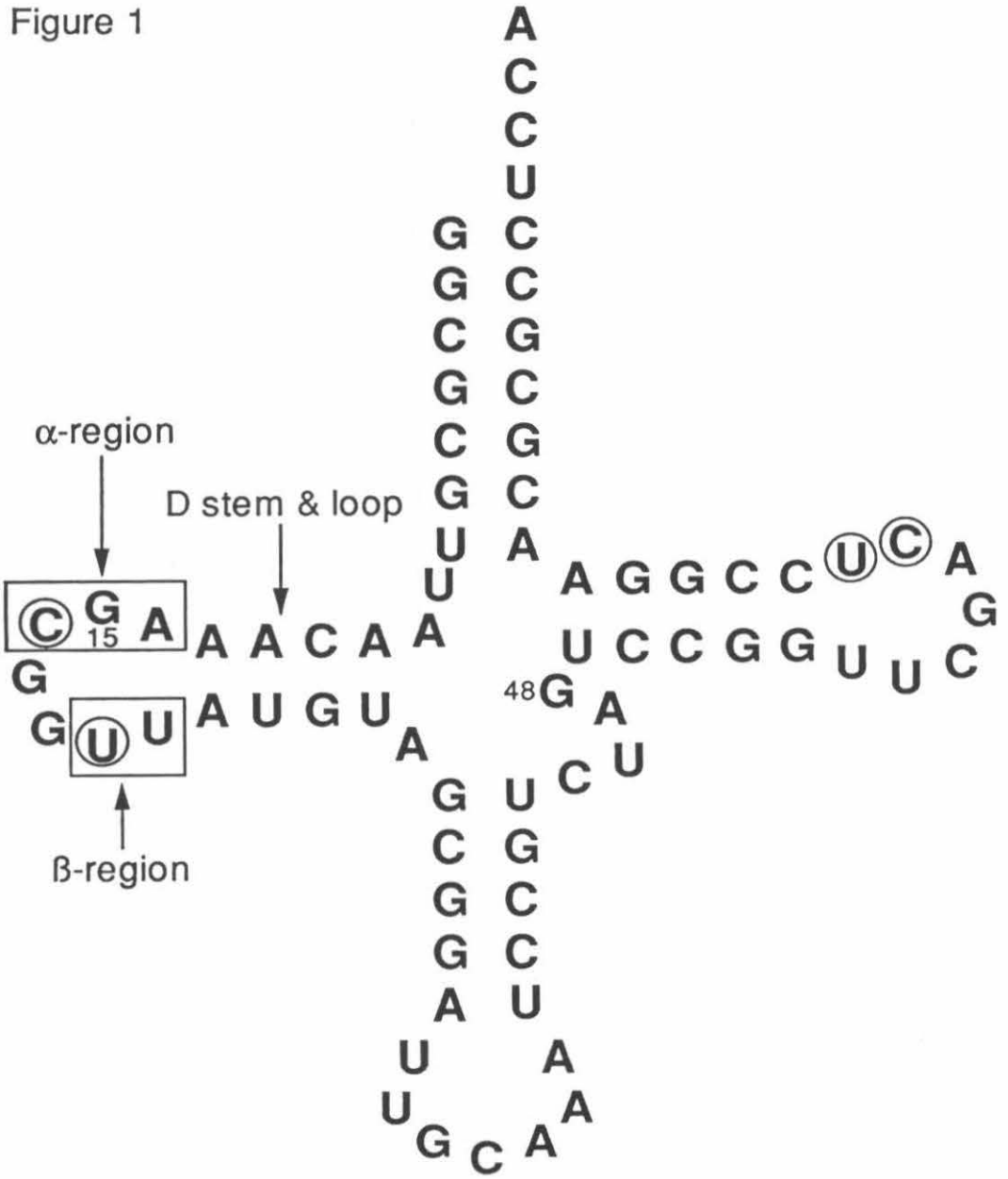
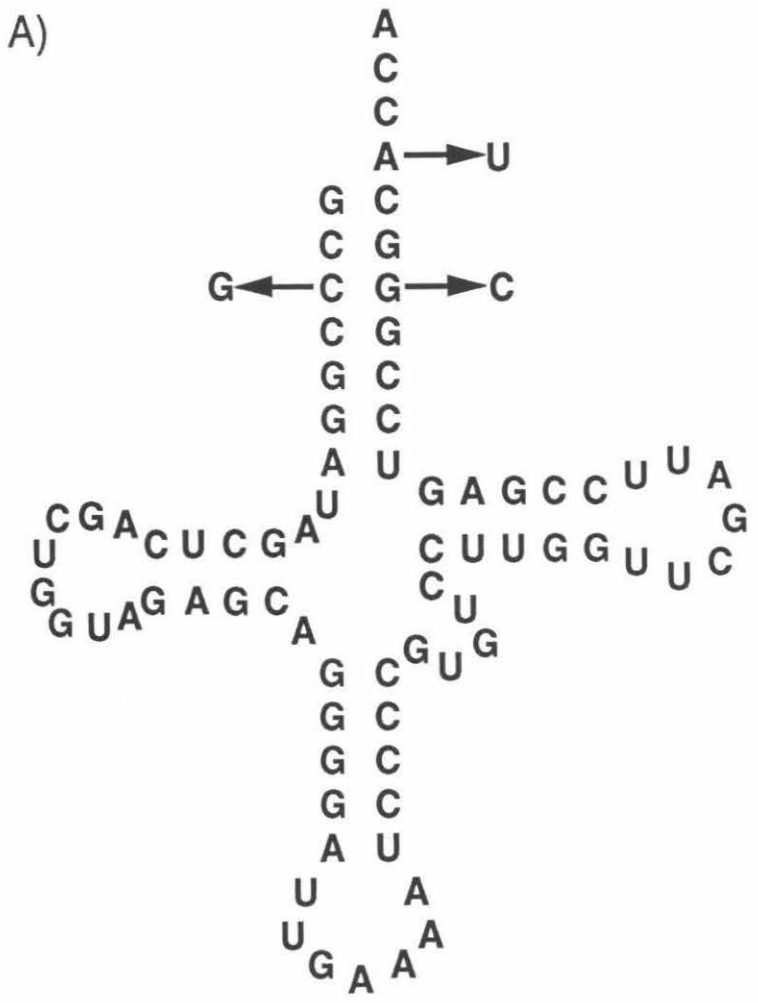


Figure 2


A)
A
C
C
C
A → U
G C
C G
G ← C G → C
C G
G C
G C
A U
U
C G A C U C G A
U
G G U A G A G C A
G A G C C U U A
G
C U U G G U U C
C U
G C G U G
G C
G C
G C
G C
A U
U A
U G A A A

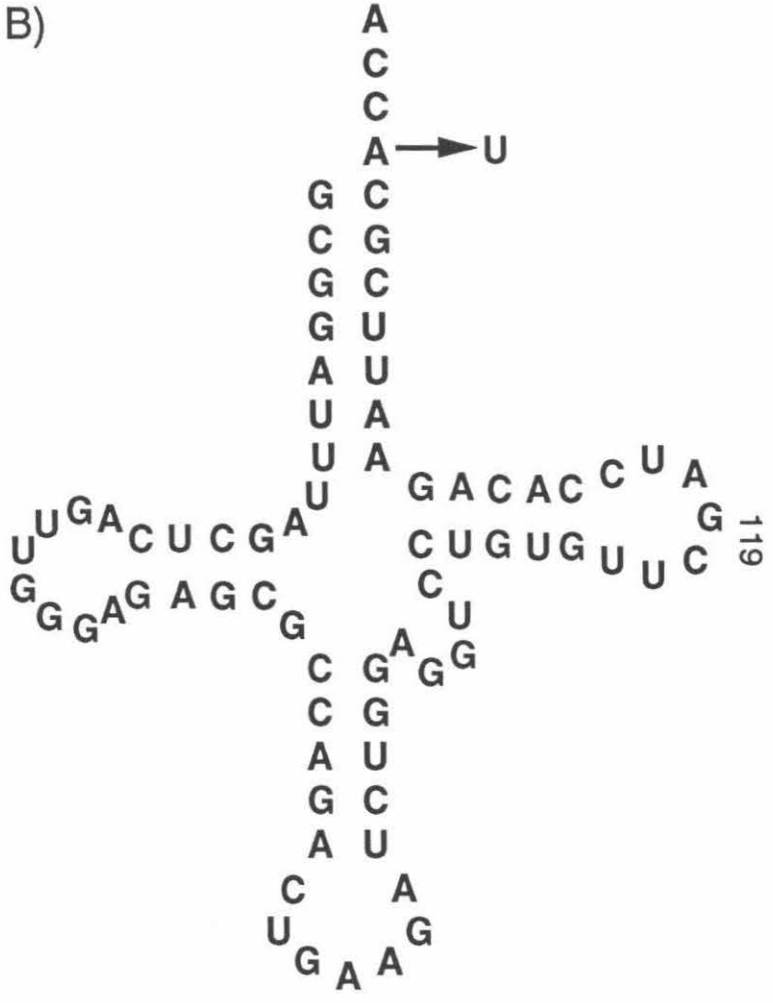
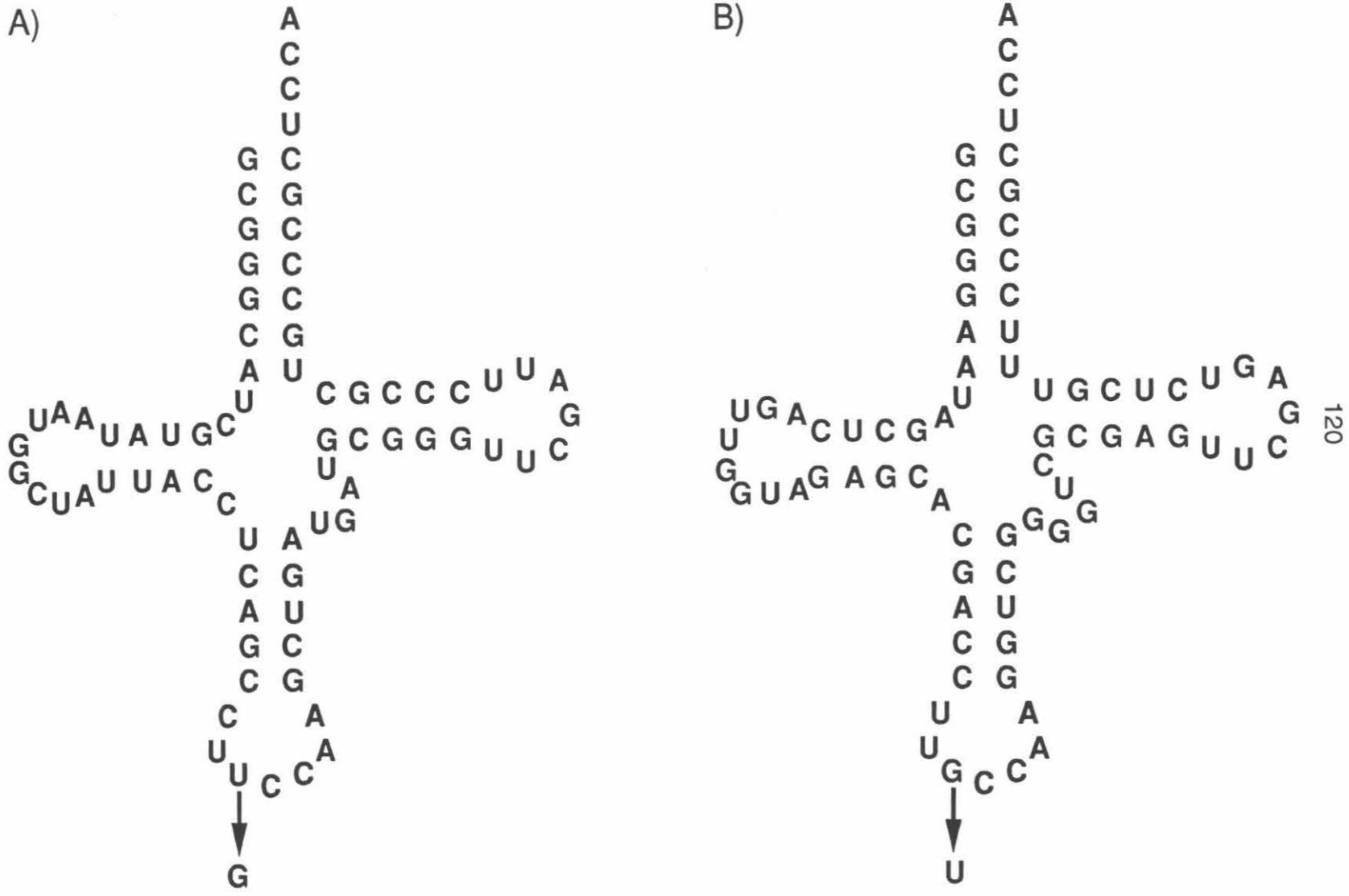

B)
A
C
C
C
A → U
G C
C G
G C
G U
A U
U A
U A
U
U G A C U C G A
U
G G G A G A G C G
G A C A C C U A
G
C U G U G U U C
C U
C G A G G
C G
C G
A U
G C
A U
C A
U G A A G
119

Figure 3



Appendix:

Nucleotide Sequence of ORF2: An open reading frame upstream of the yeast tRNA ligase gene

Portions of the work contained herein were presented in Komatsoulis, G. A., Westaway, S.K., & Abelson, J. N. (1987) *Nucleic Acids Res. 15*, 9079.

Abstract:

ORF2, an upstream open reading frame of the yeast tRNA ligase was sequenced and found to encode a putative protein product of 623 amino acids (molecular weight 71.3kD). Extensive computer analysis was conducted, but its function *in vivo* remains unknown.

Introduction:

In baker's yeast, *Saccharomyces cerevisiae*, nine tRNA species contain small intervening sequences or introns which begin after nucleotide number 37, one base 3' of the anticodon (Ogden et al., 1984, numbering follows Sprinzl et al., 1991). These introns are removed in a multi-step reaction involving two enzymes: tRNA splicing endonuclease and tRNA ligase (Greer et al., 1983) as shown in Figure 1. The endonuclease has been purified and characterized by Rauhut et al. (1990) and the ligase was purified, characterized, and cloned by Phizicky et al. (1986). The tRNA ligase gene was sequenced by Westaway et al. (1988), who found several upstream open reading frames, including one (ORF2) with a transcription start only 125 base pairs from the start of tRNA ligase (but in the opposite direction).

Knockout mutants of the tRNA ligase gene behaved unusually (E. Phizicky and S. K. Westaway, pers. comm.). The tRNA ligase reading frame was insufficient to complement such mutants; rescue of the lethal phenotype required tRNA ligase and several kilobases (kB) of upstream sequence. The ORF2 gene was subsequently found to be a single copy, essential gene; furthermore knockout mutations of ORF2 require the complete tRNA ligase-ORF2 region for complementation (E. Phizicky & S. K. Westaway, pers. comm.) The apparent importance of this gene and its linkage to yeast tRNA ligase prompted the sequencing of the ORF2 gene.

Results and Discussion:

Sequencing of ORF2:

Plasmid pSW5 (Figure 2A, obtained from S. K. Westaway) was the source of ORF2 DNA for sequencing. Figure 2B shows the clones constructed for the sequencing project. Clones initiating with the number 8 were constructed in M13mp18 and those initiating with the number 9 were cloned in M13mp19.

Using these clones, an approximately 2.0 kB region was sequenced by the dideoxy chain termination method(Sanger et al., 1977), covering the remainder of ORF2 (Westaway et al., 1988 had sequenced the first 69 amino acids). The sequence is shown in Figure 3, including part of the tRNA ligase gene (Westaway et al., 1988).

Analysis of the ORF2 gene and the hypothetical gene product:

The gene and hypothetical gene product were analyzed using the University of Wisconsin Genetics Computer Group (GCG) package of sequence analysis programs (Genetics Computer Group, 1991). The ORF2 reading frame encodes a protein of 623 amino acids with a molecular mass of 71.3 kD. Figure 4 shows its amino acid composition. The most striking feature is the large number (55) of lysine residues. Figure 5 shows the isoelectric graph for the putative ORF2 protein; its isoelectric point is 7.52. To help determine if the protein was membrane bound (one of the subunits of the tRNA splicing endonuclease is a membrane protein, Rauhut et al., 1990) a hydropathy plot was generated by the program PLOTSTRUCTURE (Figure 6), however no evidence of a transmembrane domain was seen.

Comparisons of ORF2 with current protein databases were performed using several search algorithms, but no functional similarities were seen. Figure 7 shows the result of a BLAST (Altschul et al., 1990) search. There is some similarity to a tomato protein with a leucine zipper, and a paramyosin gene from a fluke (one region from about amino acid 120-150 matches both proteins). Ultimately, however, the function of this protein is still unclear.

References:

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403-410.

Genetics Computer Group (1991) Program Manual for the GCG Package, Version 7 Madison, WI.

Greer, C. L., Peebles, C. L., Gegenheimer, P., & Abelson, J. (1983) *Cell 32*, 537-546.

Ogden, R. C., Lee, M.-C., & Knapp, G. (1984) *Nucleic Acids Res 12*, 9367-9382

Phizicky, E. M., Schwartz, R. C., & Abelson, J. (1986) *J. Biol. Chem. 261*, 2978-2986.

Rauhut, R., Green, P. R., & Abelson, J. (1990) *J. Biol. Chem. 265*, 18180-18184

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA 74*, 5463-5467.

Sprinzl, M., Dank, N., Nock, S., & A. Schön (1991) *Nucleic Acids Res. 19*, 2127-2171.

Westaway, S. K., Phizicky, E. M., & Abelson, J. (1988) *J. Biol. Chem. 263*, 3171-3176.

Wolf, H., Modrow, S., Motz, M., Jameson, B. A., & Hermann, G. (1988) *CABIOS 4*,187-191.

Figure Legends:

Figure 1: Mechanism of yeast tRNA splicing. After Greer et al. (1983).

Figure 2: Cloning strategy. A) pSW5 (S. K. Westaway, pers. comm.) was the source of ORF2 sequences. B) The HpaI-EcoRI fragment containing ORF2, showing M13mp18 and M13mp19 clones constructed for sequencing. See text for details.

Figure 3: Sequence of ORF2 and presumptive ORF2 protein product. The sequence includes some of the yeast tRNA ligase gene sequenced by Westaway et al. (1988).

Figure 4: Composition and molecular weight of the hypothetical ORF2 protein product.

Figure 5: Isoelectric graph of the hypothetical ORF2 product (from the GCG program ISOELECTRIC)

Figure 6: Output from GCG computer program PLOTSTRUCTURE (Wolf et al., 1988) showing Kyte and Doolittle hydrophilicity, surface probability, flexibility, Jameson-Wolf antigenic index, Chou & Fasman and Garnier-Osguthorpe-Robson secondary structure predictions, and predicted glycosylation sites.

Figure 7: Results of a BLAST search against the current (10-January-1993) protein databank.

Figure 1

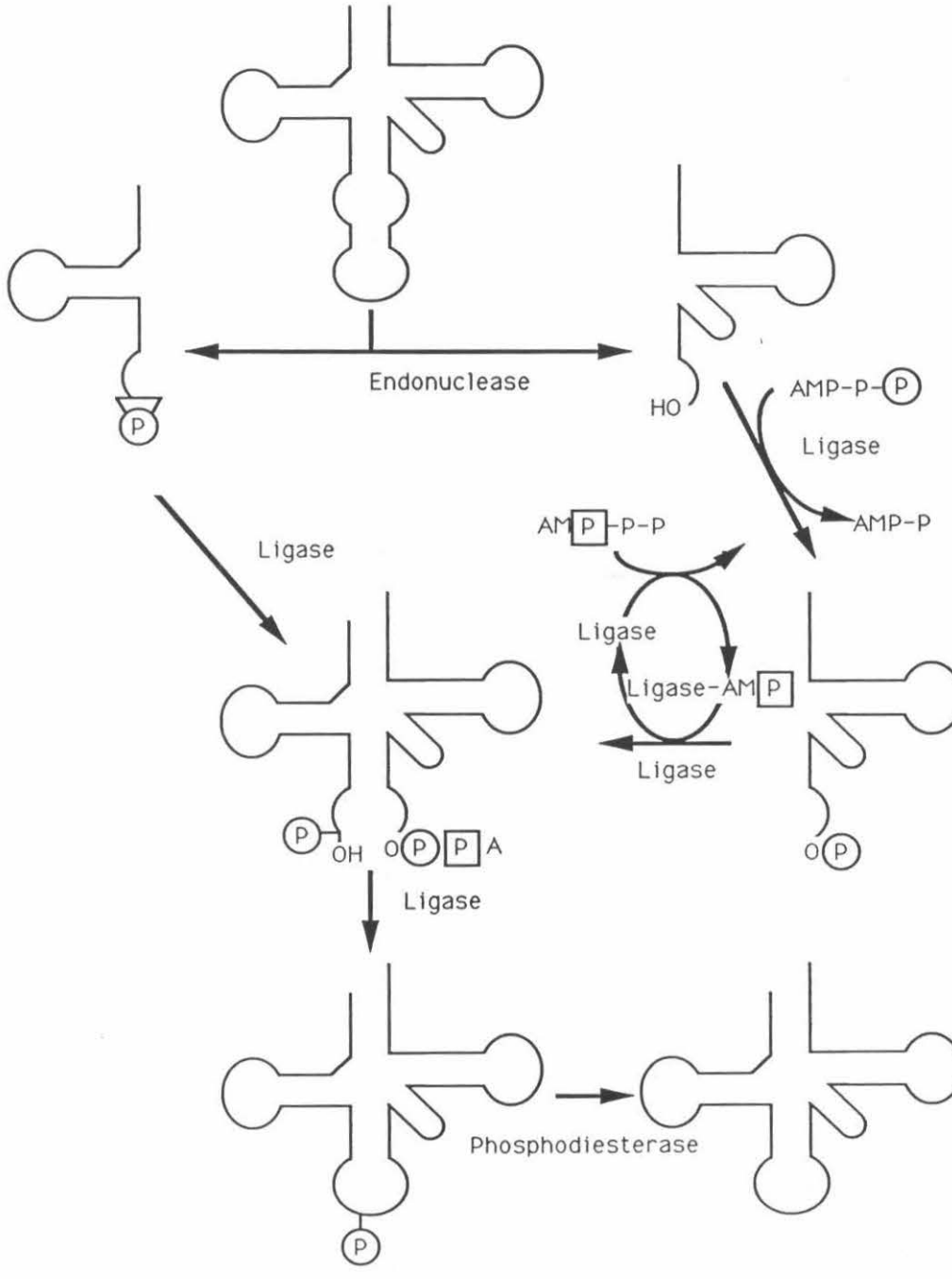

Endonuclease
HO
AMP-P-P
Ligase
AMP-P
Ligase
AMP-P-P
Ligase
Ligase-AMP
Ligase
OH
O
P
P
A
Ligase
O
P
Phosphodiesterase
P

Figure 2

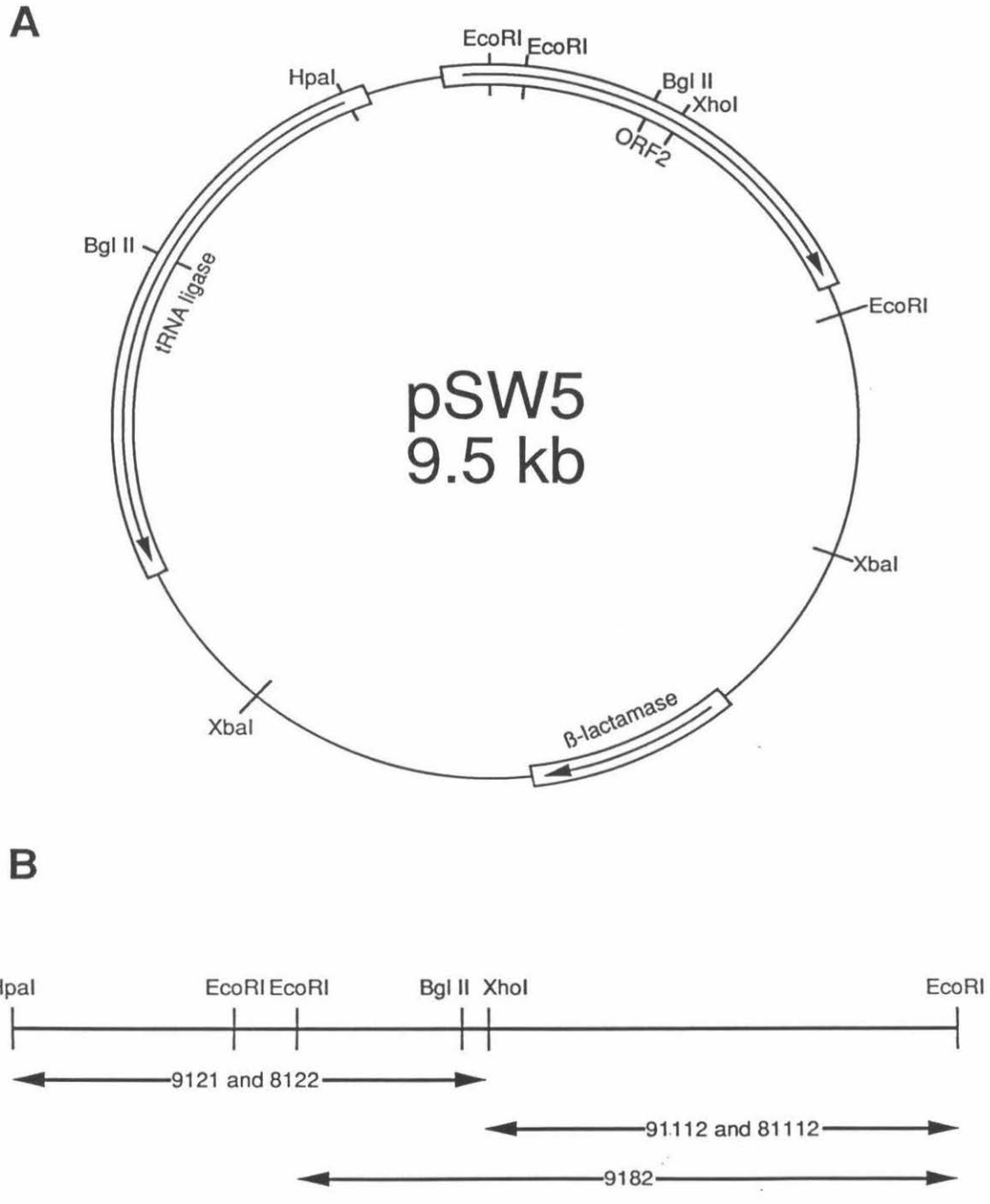

A
EcoRI
EcoRI
HpaI
Bgl II
XhoI
ORF2
Bgl II
tRNA ligase
EcoRI
pSW5
9.5 kb
XbaI
XbaI
ß-lactamase
B
HpaI
EcoRI EcoRI
Bgl II XhoI
EcoRI
9121 and 8122
91112 and 81112
9182

Figure 3

```
   1 TGGGCTAGGC ATCGCTTCTT CGTATGAATA CTTTTATGAT CACTAAAGTC AAAGTGTAGA
  61 TTTTCAAAGA AAGAAAGACT ATGCAAATGC GTAAAATTTG CTCGAAGCAA TTCTCAACAG
 121 AATGGAATGA ATTAGATGCT CTTCTGTATA CTTTCTTTGG GATAAACTCA TAGTGAAATT
 181 TTATCAATGG AAAAGTACGC GGCCGGCCCG CTGGAACATA AAGATAATTT ACTGAAACAG
 241 TGTGCTTAAA CGTATTTGAA AACCAAGGTG AATGTACCCT TAAAGTGCTG GCAGTGAATA
 301 GCAAAGGTCC ATTAAAGTGT TCTGGTTTTA TAATATCAAA AGCATTTATA GAA ATG CCC
   1                                                           Met Pro

 360 GCT GAA ATT GAC ATT GAT GAA GCT GAC GTA TTA GTT TTA TCA CAG GAG
   3 Ala Glu Ile Asp Ile Asp Glu Ala Asp Val Leu Val Leu Ser Gln Glu

 408 TTA CAA AAG ACA AGT AAA CTC ACT TTC GAA ATC AAT AAA TCT TTG AAA
  19 Leu Gln Lys Thr Ser Lys Leu Thr Phe Glu Ile Asn Lys Ser Leu Lys

 456 AAA ATT GCA GCT ACA TCC AAT CAA TCC AGT CAA CTC TTC ACT CCT ATT
  35 Lys Ile Ala Ala Thr Ser Asn Gln Ser Ser Gln Leu Phe Thr Pro Ile

 504 CTT GCT AGA AAT AAT GTT TTA ACC ACA TTA CAA AGA AAT ATT GAA AGT
  51 Leu Ala Arg Asn Asn Val Leu Thr Thr Leu Gln Arg Asn Ile Glu Ser

 552 ACA TTG AAT TCC GTT GCC TCA GTT AAG GAT CTA GCA AAC GAA GCT TCC
  67 Thr Leu Asn Ser Val Ala Ser Val Lys Asp Leu Ala Asn Glu Ala Ser

 600 AAG TAT GAG ATC ATT TTA CAA AAG GGT ATT AAT CAA GTC GGT TTA AAG
  83 Lys Tyr Glu Ile Ile Leu Gln Lys Gly Ile Asn Gln Val Gly Leu Lys

 648 CAA TAC ACC CAA GTA GTA CAT AAG CTA GAT GAT ATG TTG GAA GAC ATT
  99 Gln Tyr Thr Gln Val Val His Lys Leu Asp Asp Met Leu Glu Asp Ile

 696 CAG TCT GGA CAA GCC AAT CGA GAA GAA AAC TCA GAA TTC CAT GGG ATT
 115 Gln Ser Gly Gln Ala Asn Arg Glu Glu Asn Ser Glu Phe His Gly Ile

 744 TTA ACT CAC TTG GAA CAA TTG ATC AAA CGT AGC GAG GCT CAA CTA AGA
 131 Leu Thr His Leu Glu Gln Leu Ile Lys Arg Ser Glu Ala Gln Leu Arg

 792 GTA TAT TTT ATT TCA ATT TTG AAC AGT ATT AAA CCG TTT GAT CCA CAA
 147 Val Tyr Phe Ile Ser Ile Leu Asn Ser Ile Lys Pro Phe Asp Pro Gln

 840 ATC AAT ATC ACC AAA AAG ATG CCA TTT CCA TAC TAC GAG GAC CAG CAG
 163 Ile Asn Ile Thr Lys Lys Met Pro Phe Pro Tyr Tyr Glu Asp Gln Gln

 888 TTA GGC GCT TTA TCG TGG ATT TTA GAT TAT TTT CAT GGA AAT TCA GAA
 179 Leu Gly Ala Leu Ser Trp Ile Leu Asp Tyr Phe His Gly Asn Ser Glu

 936 GGT TCT ATT ATA CAG GAC ATA CTC GTC GGT GAA AGG AGT AAA TTA ATC
 195 Gly Ser Ile Ile Gln Asp Ile Leu Val Gly Glu Arg Ser Lys Leu Ile

 984 CTC AAG TGC ATG GCA TTC CTT GAA CCT TTT GCC AAA GAA ATC AGC ACT1
 211 Leu Lys Cys Met Ala Phe Leu Glu Pro Phe Ala Lys Glu Ile Ser Thr

1032 GCA AAA AAC GCC CCG TAT GAG AAG GGC AGT AGC GGG ATG AAC AGC TAC
 227 Ala Lys Asn Ala Pro Tyr Glu Lys Gly Ser Ser Gly Met Asn Ser Tyr

1080 ACG GAG GCC TTA TTA GGC TTC ATC GCC AAT GAA AAA TCA CTA GTA GAC
 243 Thr Glu Ala Leu Leu Gly Phe Ile Ala Asn Glu Lys Ser Leu Val Asp
```

```
1128 GAC CTC TAC TCT CAA TAT ACA GAA AGT AAA CCC CAC GTC TTG TCA CAG
 259 Asp Leu Tyr Ser Gln Tyr Thr Glu Ser Lys Pro His Val Leu Ser Gln

1176 ATC TTG TCG CCT TTA ATT AGC GCA TAT GCT AAG CTT TTC GGT GCA AAT
 275 Ile Leu Ser Pro Leu Ile Ser Ala Tyr Ala Lys Leu Phe Gly Ala Asn

1224 TTG AAA ATT GTA CGA AGC AAC CTC GAG AAC TTT GGA TTC TTT AGT TTT
 291 Leu Lys Ile Val Arg Ser Asn Leu Glu Asn Phe Gly Phe Phe Ser Phe

1272 GAG CTA GTG GAA AGC ATA AAT GAT GTG AAA AAA TCT CTT CGA GGC AAG
 307 Glu Leu Val Glu Ser Ile Asn Asp Val Lys Lys Ser Leu Arg Gly Lys

1320 GAA CTA CAA AAC TAT AAT TTA TTG CAA GAT TGT ACG CAA GAA GTA CGT
 323 Glu Leu Gln Asn Tyr Asn Leu Leu Gln Asp Cys Thr Gln Glu Val Arg

1368 CAA GTA ACA CAG TCA TTA TTC AGA GAT GCC ATT GAT AGG ATT ATC AAA
 339 Gln Val Thr Gln Ser Leu Phe Arg Asp Ala Ile Asp Arg Ile Ile Lys

1416 AAG GCA AAT TCT ATT TCT ACT ATT CCT TCC AAT AAT GGT GTC ACT GAA
 355 Lys Ala Asn Ser Ile Ser Thr Ile Pro Ser Asn Asn Gly Val Thr Gl

1464 GCA ACT GTA GAT ACC ATG TCA AGA CTA AGA AAG TTC AGT GAG TAC AAA
 371 Ala Thr Val Asp Thr Met Ser Arg Leu Arg Lys Phe Ser Glu Tyr Lys

1512 AAC GGA TGT TTA GGC GCC ATG GAC AAT ATC ACA CGT GAA AAT TGG TTA
 387 Asn Gly Cys Leu Gly Ala Met Asp Asn Ile Thr Arg Glu Asn Trp Leu

1560 CCA TCC AAT TAT AAG GAG AAA GAA TAC ACT TTG CAA AAC GAG GCT TTA
 403 Pro Ser Asn Tyr Lys Glu Lys Glu Tyr Thr Leu Gln Asn Glu Ala Leu

1608 AAT TGG GAA GAT CAT AAT GTA TTA CTA TCA TGT TTT ATA AGT GAT TGC
 419 Asn Trp Glu Asp His Asn Val Leu Leu Ser Cys Phe Ile Ser Asp Cys

1656 ATA GAC ACT TTG GCA GTT AAT CTT GAG AGA AAG GCA CAA ATA GCA CTG
 435 Ile Asp Thr Leu Ala Val Asn Leu Glu Arg Lys Ala Gln Ile Ala Leu

1704 ATG CCT AAC CAA GAG CCA GAT GTG GCT AAT CCT AAT AGC TCT AAA AAT
 451 Met Pro Asn Gln Glu Pro Asp Val Ala Asn Pro Asn Ser Ser Lys Asn

1752 AAG CAC AAG CAA CGT ATT GGA TTC TTC ATT TTA ATG AAC CTG ACG CTT
 467 Lys His Lys Gln Arg Ile Gly Phe Phe Ile Leu Met Asn Leu Thr Leu

1800 GTT GAG CAG ATC GTG GAA AAG TCA GAA TTA AAC TTA ATG TTA GCT GGA
 483 Val Glu Gln Ile Val Glu Lys Ser Glu Leu Asn Leu Met Leu Ala Gly

1848 GAA GGT CAC TCC AGA TTG GAA CGG CTG AAG AAA CGT TAT ATT AGT TAT
 499 Glu Gly His Ser Arg Leu Glu Arg Leu Lys Lys Arg Tyr Ile Ser Tyr

1896 ATG GTA TCA GAT TGG AGA GAT TTA ACT GCA AAT TTG ATG GAT TCT GTG
 515 Met Val Ser Asp Trp Arg Asp Leu Thr Ala Asn Leu Met Asp Ser Val

1944 TTT ATT GAT AGT AGC GGG AAG AAG TCA AAA GAC AAA GAA CAA ATA AAG
 531 Phe Ile Asp Ser Ser Gly Lys Lys Ser Lys Asp Lys Glu Gln Ile Lys

1992 GAG AAA TTT AGA AAA TTC AAT GAA GGA TTC GAA GAT TTA GTA TCA AAA
 547 Glu Lys Phe Arg Lys Phe Asn Glu Gly Phe Glu Asp Leu Val Ser Lys
```

```
2040 ACA AAA CAG TAT AAG CTT TCA GAT CCA TCA TTG AAA GTA ACT TTG AAG
 563 Thr Lys Gln Tyr Lys Leu Ser Asp Pro Ser Leu Lys Val Thr Leu Lys

2088 TCA GAA ATA ATA TCG TTG GTT ATG CCC ATG TAT GAA AGA TTC TAC AGT
 579 Ser Glu Ile Ile Ser Leu Val Met Pro Met Tyr Glu Arg Phe Tyr Ser

2136 AGA TAT AAA GAC TCT TTC AAG AAT CCT AGA AAG CAT ATC AAA TAT ACC
 595 Arg Tyr Lys Asp Ser Phe Lys Asn Pro Arg Lys His Ile Lys Tyr Thr

2184 CCT GAC GAA CTA ACT ACT GTT CTT AAC CAA TTA GTG AGA TAG ATTGCGAACA
 611 Pro Asp Glu Leu Thr Thr Val Leu Asn Gln Leu Val Arg    623

2236 TCACTTTTCA AAAAGCAAAA TTAAAGAAAA GAACGTACCA TAAACAGATA TAAACATATA
2296 TATGTATAAC AATTAGAGTT TCTAATATTT TATGTTTGCT CTGGACGCCA GATGTAAGAA
2356 CAGGAAGGGA TGACGTAGGA ACCTTGTGCT AGACTAAAGG GTACTCTCAT TTATACTTTG
2416 TAATCCAGAT TCATTATCTA ACGTATTAAT CAGTACTTTC TCGGGTGAAG AAAGTCTTGA
2476 AATAAATCTT TCCTCGGAAT CAGAATTC 2503
```

Figure 4

PEPSTATS of: Orf2.Pep check: 3448 from: 1 to: 623

Continuous From: 1 To: 623 Length: 623

Summary for whole sequence:

Molecular weight = 71292.82 Residues = 623
Average Residue Weight = 114.435 Charge = +9

Residue	Number	Mole Percent
A = Ala	31	4.976
C = Cys	5	0.803
D = Asp	32	5.136
E = Glu	48	7.705
F = Phe	26	4.173
G = Gly	22	3.531
H = His	9	1.445
I = Ile	46	7.384
K = Lys	55	8.828
L = Leu	71	11.396
M = Met	14	2.247
N = Asn	43	6.902
P = Pro	19	3.050
Q = Gln	32	5.136
R = Arg	25	4.013
S = Ser	59	9.470
T = Thr	29	4.655
V = Val	31	4.976
W = Trp	4	0.642
Y = Tyr	22	3.531

Small	(A+G)	53	8.507
Hydroxyl	(S+T)	88	14.125
Acidic	(D+E)	80	12.841
Acid/Amide	(D+E+N+Q)	155	24.880
Basic	(H+K+R)	89	14.286
Charged	(D+E+H+K+R)	169	27.127
Small hphob	(I+L+M+V)	162	26.003
Aromatic	(F+W+Y)	52	8.347

Figure 5

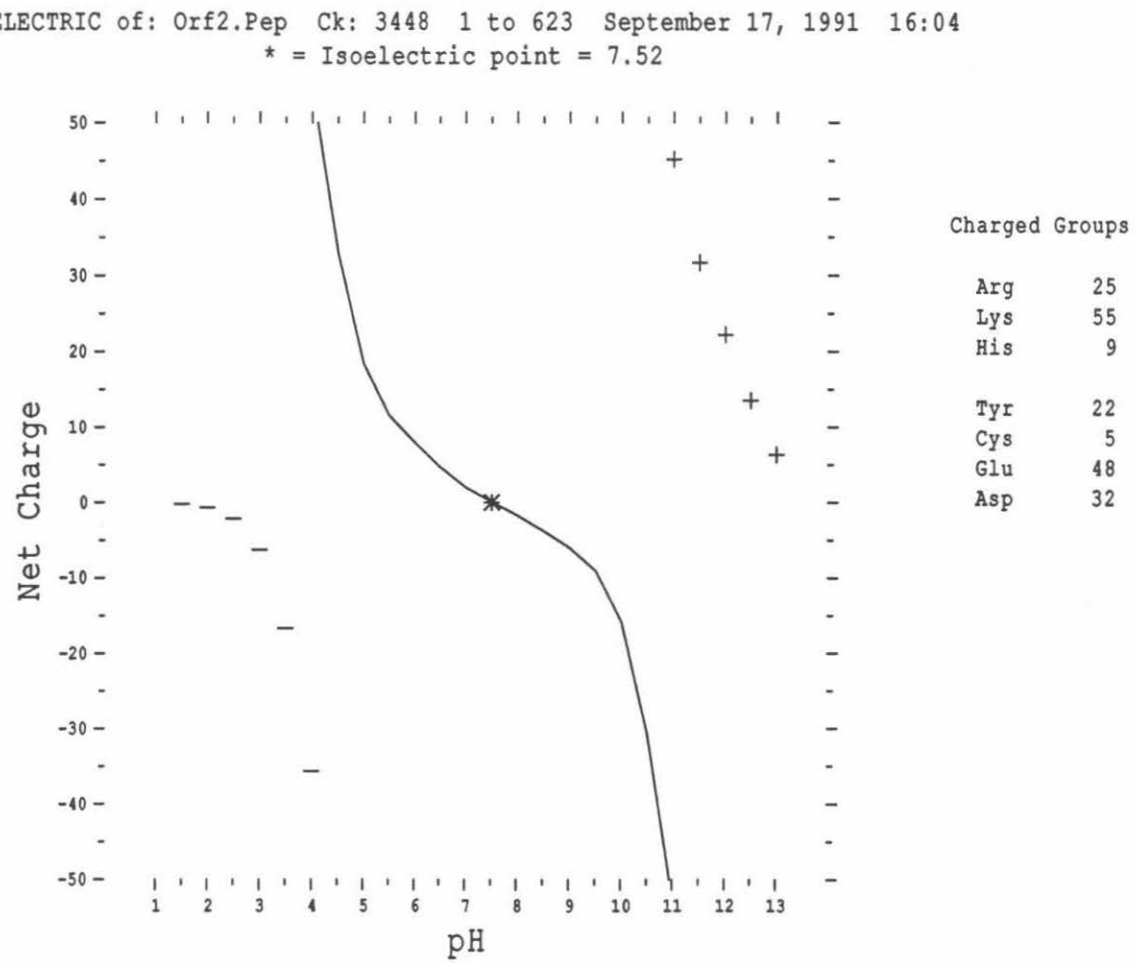

ISOELECTRIC of: Orf2.Pep Ck: 3448 1 to 623 September 17, 1991 16:04
* = Isoelectric point = 7.52
Net Charge
pH
Charged Groups
Arg 25
Lys 55
His 9
Tyr 22
Cys 5
Glu 48
Asp 32

Figure 6

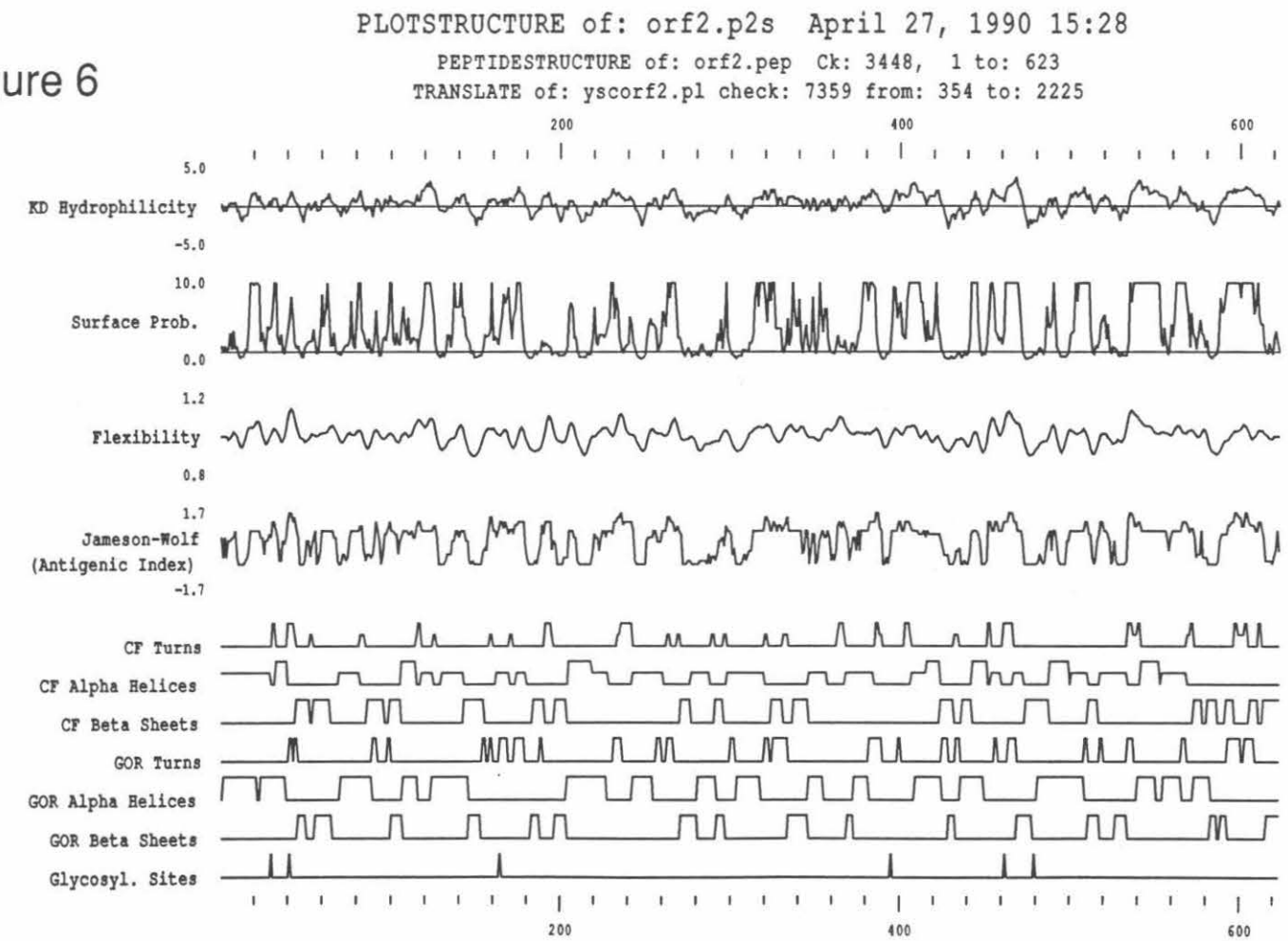

PLOTSTRUCTURE of: orf2.p2s April 27, 1990 15:28
PEPTIDESTRUCTURE of: orf2.pep Ck: 3448, 1 to: 623
TRANSLATE of: yscorf2.pl check: 7359 from: 354 to: 2225
200
400
600
5.0
KD Hydrophilicity
-5.0
10.0
Surface Prob.
0.0
1.2
Flexibility
0.8
1.7
Jameson-Wolf
(Antigenic Index)
-1.7
CF Turns
CF Alpha Helices
CF Beta Sheets
GOR Turns
GOR Alpha Helices
GOR Beta Sheets
Glycosyl. Sites
200
400
600

Figure 7

```
Query=  [.orf2]orf2.pep
        (623 residues)

Database:  Non-redundant SwissProt+PIR+GenPept+GPUpdate, 4:35 AM EST Jan
10,1993

        84,347 sequences; 23,751,181 total residues.
Searching.....................................................done


Smallest

Poisson
                                                                  High
Probability
Sequences producing High-scoring Segment Pairs:          Score P(N)  N

SP:YTRL_YEAST  HYPOTHETICAL 71.3 KD PROTEIN IN TRNA LIGAS.3237  0.0    1
PIR:S21495     *Hypothetical protein - Tomato              51  0.045  4
GP:TOMLEUZIP_1 L.esculentum mRNA for protein with leucine...51  0.046  4
SP:MYSP_SCHMA  PARAMYOSIN. >GP:SCMPMY_1 Schistosoma manso.. 50  0.29   2
PIR:A25993     Paramyosin - Fluke (Schistosoma mansoni)  .. 50  0.29   2
GP:SCMPMYA1_1  S.mansoni paramyosin mRNA, complete cds.    50  0.49   2



>SP:YTRL_YEAST HYPOTHETICAL 71.3 KD PROTEIN IN TRNA LIGASE 3'REGION.
            >PIR:A27300 Hypothetical protein, 71K - Yeast (Saccharomyces
            cerevisiae) >GP:YSCORF2_2 Yeast DNA for ORF2 upstream of
tRNA
            ligase gene. [Saccharomyces cerevisiae]
            Length = 623

 Score = 3237 (1565.3 bits), Expect = 0.0, P = 0.0
 Identities = 623/623 (100%), Positives = 623/623 (100%)

Query:     1
MPAEIDIDEADVLVLSQELQKTSKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTL 60

MPAEIDIDEADVLVLSQELQKTSKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTL
Sbjct:     1
MPAEIDIDEADVLVLSQELQKTSKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTL 60

Query:    61
QRNIESTLNSVASVKDLANEASKYEIILQKGINQVGLKQYTQVVHKLDDMLEDIQSGQAN 120

QRNIESTLNSVASVKDLANEASKYEIILQKGINQVGLKQYTQVVHKLDDMLEDIQSGQAN
Sbjct:    61
QRNIESTLNSVASVKDLANEASKYEIILQKGINQVGLKQYTQVVHKLDDMLEDIQSGQAN 120

Query:   121
REENSEFHGILTHLEQLIKRSEAQLRVYFISILNSIKPFDPQINITKKMPFPYYEDQQLG 180

REENSEFHGILTHLEQLIKRSEAQLRVYFISILNSIKPFDPQINITKKMPFPYYEDQQLG
Sbjct:   121
REENSEFHGILTHLEQLIKRSEAQLRVYFISILNSIKPFDPQINITKKMPFPYYEDQQLG 180
```

```
Query:   181
ALSWILDYFHGNSEGSIIQDILVGERSKLILKCMAFLEPFAKEISTAKNAPYEKGSSGMN 240

ALSWILDYFHGNSEGSIIQDILVGERSKLILKCMAFLEPFAKEISTAKNAPYEKGSSGMN
Sbjct:   181
ALSWILDYFHGNSEGSIIQDILVGERSKLILKCMAFLEPFAKEISTAKNAPYEKGSSGMN 240

Query:   241
SYTEALLGFIANEKSLVDDLYSQYTESKPHVLSQILSPLISAYAKLFGANLKIVRSNLEN 300

SYTEALLGFIANEKSLVDDLYSQYTESKPHVLSQILSPLISAYAKLFGANLKIVRSNLEN
Sbjct:   241
SYTEALLGFIANEKSLVDDLYSQYTESKPHVLSQILSPLISAYAKLFGANLKIVRSNLEN 300

Query:   301
FGFFSFELVESINDVKKSLRGKELQNYNLLQDCTQEVRQVTQSLFRDAIDRIIKKANSIS 360

FGFFSFELVESINDVKKSLRGKELQNYNLLQDCTQEVRQVTQSLFRDAIDRIIKKANSIS
Sbjct:   301
FGFFSFELVESINDVKKSLRGKELQNYNLLQDCTQEVRQVTQSLFRDAIDRIIKKANSIS 360

Query:   361
TIPSNNGVTEATVDTMSRLRKFSEYKNGCLGAMDNITRENWLPSNYKEKEYTLQNEALNW 420

TIPSNNGVTEATVDTMSRLRKFSEYKNGCLGAMDNITRENWLPSNYKEKEYTLQNEALNW
Sbjct:   361
TIPSNNGVTEATVDTMSRLRKFSEYKNGCLGAMDNITRENWLPSNYKEKEYTLQNEALNW 420

Query:   421
EDHNVLLSCFISDCIDTLAVNLERKAQIALMPNQEPDVANPNSSKNKHKQRIGFFILMNL 480

EDHNVLLSCFISDCIDTLAVNLERKAQIALMPNQEPDVANPNSSKNKHKQRIGFFILMNL
Sbjct:   421
EDHNVLLSCFISDCIDTLAVNLERKAQIALMPNQEPDVANPNSSKNKHKQRIGFFILMNL 480

Query:   481
TLVEQIVEKSELNLMLAGEGHSRLERLKKRYISYMVSDWRDLTANLMDSVFIDSSGKKSK 540

TLVEQIVEKSELNLMLAGEGHSRLERLKKRYISYMVSDWRDLTANLMDSVFIDSSGKKSK
Sbjct:   481
TLVEQIVEKSELNLMLAGEGHSRLERLKKRYISYMVSDWRDLTANLMDSVFIDSSGKKSK 540

Query:   541
DKEQIKEKFRKFNEGFEDLVSKTKQYKLSDPSLKVTLKSEIISLVMPMYERFYSRYKDSF 600

DKEQIKEKFRKFNEGFEDLVSKTKQYKLSDPSLKVTLKSEIISLVMPMYERFYSRYKDSF
Sbjct:   541
DKEQIKEKFRKFNEGFEDLVSKTKQYKLSDPSLKVTLKSEIISLVMPMYERFYSRYKDSF 600

Query:   601 KNPRKHIKYTPDELTTVLNQLVR 623
             KNPRKHIKYTPDELTTVLNQLVR
Sbjct:   601 KNPRKHIKYTPDELTTVLNQLVR 623


>PIR:S21495 *Hypothetical protein - Tomato
            Length = 631
```

```
 Score = 51 (24.7 bits), Expect = 98., P = 1.0
 Identities = 16/60 (26%), Positives = 29/60 (48%)

Query:IKEKFRKFNEGFEDLVSKTKQYKLSDPSLKVTLKSEIISLVMPMYERFYSRYKDSFKNPR 604
      +KEK + FN  FE++         + D  LK  L+  + + + P Y  F +R   +  ++R
Sbjct:LKEKLKLFNSYFEEICKTQSTWIIFDEQLKEELRISVAGALSPAYRNFIGRLQSNNDSSR 599

 Score = 46 (22.2 bits), Expect = 5.3e+02, Poisson P(2) = 0.97
 Identities = 7/25 (28%), Positives = 19/25 (76%)

Query:   597 KDSFKNPRKHIKYTPDELTTVLNQL 621
             +DS ++  +HIK++ ++L + +++L
Sbjct:   595 NDSSRHTERHIKFSVEDLEARISEL 619

 Score = 41 (19.8 bits), Expect = 2.8e+03, Poisson P(4) = 0.045
 Identities = 9/20 (45%), Positives = 11/20 (55%)

Query:   128 HGILTHLEQLIKRSEAQLRV 147
             +GI   LE LI+R  A   V
Sbjct:   365 RGIFMELENLIRRDPAKTPV 384

 Score = 41 (19.8 bits), Expect = 2.8e+03, Poisson P(4) = 0.045
 Identities = 9/25 (36%), Positives = 15/25 (60%)

Query:   475 FILMNLTLVEQIVEKSELNLMLAGE 499
             F++ N   + Q V  SEL L+L+ +
Sbjct:   471 FMMNNERYIVQKVKDSELGLLLGDD 495


>GP:TOMLEUZIP_1 L.esculentum mRNA for protein with leucine zipper.
            [Lycopersicon esculentum]
            Length = 632

 Score = 51 (24.7 bits), Expect = 98., P = 1.0
 Identities = 16/60 (26%), Positives = 29/60 (48%)

Query:IKEKFRKFNEGFEDLVSKTKQYKLSDPSLKVTLKSEIISLVMPMYERFYSRYKDSFKNPR 604
      +KEK + FN  FE++         + D  LK  L+  + + + P Y  F +R   +  ++R
Sbjct:LKEKLKLFNSYFEEICKTQSTWIIFDEQLKEELRISVAGALSPAYRNFIGRLQSNNDSSR 599

 Score = 46 (22.2 bits), Expect = 5.3e+02, Poisson P(2) = 0.97
 Identities = 7/25 (28%), Positives = 19/25 (76%)

Query:   597 KDSFKNPRKHIKYTPDELTTVLNQL 621
             +DS ++  +HIK++ ++L + +++L
Sbjct:   595 NDSSRHTERHIKFSVEDLEARISEL 619

 Score = 41 (19.8 bits), Expect = 2.8e+03, Poisson P(4) = 0.046
 Identities = 9/20 (45%), Positives = 11/20 (55%)

Query:   128 HGILTHLEQLIKRSEAQLRV 147
             +GI   LE LI+R  A   V
Sbjct:   365 RGIFMELENLIRRDPAKTPV 384
```

```
 Score = 41 (19.8 bits), Expect = 2.8e+03, Poisson P(4) = 0.046
 Identities = 9/25 (36%), Positives = 15/25 (60%)

Query:   475 FILMNLTLVEQIVEKSELNLMLAGE 499
             F++ N   + Q V  SEL L+L+ +
Sbjct:   471 FMMNNERYIVQKVKDSELGLLLGDD 495


>SP:MYSP_SCHMA PARAMYOSIN. >GP:SCMPMY_1 Schistosoma mansoni (blood
fluke)
            paramyosin mRNA, partial cds. [Schistosoma mansoni]
            Length = 439

 Score = 50 (24.2 bits), Expect = 1.4e+02, P = 1.0
 Identities = 13/43 (30%), Positives = 25/43 (58%)

Query:    23 SKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTLQRNIE 65
             +KLT EI+     I + S ++S+L     A +++ + LQR ++
Sbjct:    33 TKLTLEIKDLQSEIESLSLENSELIRRAKAAESLASDLQRRVD 75

 Score = 49 (23.7 bits), Expect = 1.9e+02, Poisson P(2) = 0.29
 Identities = 9/24 (37%), Positives = 18/24 (75%)

Query:   122 EENSEFHGILTHLEQLIKRSEAQL 145
             +E  E ++ L +LE+L K++E++L
Sbjct:   298 NEIEEIRSTLENLERLRKHAETEL 321


>PIR:A25993 Paramyosin - Fluke (Schistosoma mansoni) (fragment)
            Length = 439

 Score = 50 (24.2 bits), Expect = 1.4e+02, P = 1.0
 Identities = 13/43 (30%), Positives = 25/43 (58%)

Query:    23 SKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTLQRNIE 65
             +KLT EI+     I + S ++S+L     A +++ + LQR ++
Sbjct:    33 TKLTLEIKDLQSEIESLSLENSELIRRAKAAESLASDLQRRVD 75

 Score = 49 (23.7 bits), Expect = 1.9e+02, Poisson P(2) = 0.29
 Identities = 9/24 (37%), Positives = 18/24 (75%)

Query:   122 EENSEFHGILTHLEQLIKRSEAQL 145
             +E  E ++ L +LE+L K++E++L
Sbjct:   298 NEIEEIRSTLENLERLRKHAETEL 321


>GP:SCMPMYA1_1 S.mansoni paramyosin mRNA, complete cds. [Schistosoma
mansoni]
            Length = 866

 Score = 50 (24.2 bits), Expect = 1.4e+02, P = 1.0
 Identities = 13/43 (30%), Positives = 25/43 (58%)

Query:    23 SKLTFEINKSLKKIAATSNQSSQLFTPILARNNVLTTLQRNIE 65
             +KLT EI+     I + S ++S+L     A +++ + LQR ++
Sbjct:   335 TKLTLEIKDLQSEIESLSLENSELIRRAKAAESLASDLQRRVD 377

 Score = 49 (23.7 bits), Expect = 1.9e+02, Poisson P(2) = 0.49
```

```
 Identities = 9/24 (37%), Positives = 18/24 (75%)

Query:   122 EENSEFHGILTHLEQLIKRSEAQL 145
             +E  E ++ L +LE+L K++E++L
Sbjct:   600 NEIEEIRSTLENLERLRKHAETEL 623


Parameters:
  E = 1.7, S = 63 (30.5 bits),  E2 = 0.090, S2 = 36
  W = 3, T = 12 (5.8 bits), X = 21 (10.2 bits)
  M = PAM120
  H = 0, V = 500, B = 250

Statistics:
  Lambda = 0.335 nats/unit score, Lambda/ln2 = 0.484 bits/unit score
  K = 0.182, H = 0.837 bits/position
  Expected/Observed high score = 64 (30.9 bits) / 3237 (1565.3 bits)
  # of residues in query:  623
  # of neighborhood words in query:  10,777
  # of exact words scoring below T:  20
  Database:  Non-redundant SwissProt+PIR+GenPept+GPUpdate, 4:35 AM EST
Jan 10,
             1993
  # of residues in database:  23,751,181
  # of word hits against database:  20,985,520
  # of failed hit extensions:  17,521,674
  # of excluded hits:  3,463,612
  # of successful extensions:  234
  # of overlapping HSPs discarded:  219
  # of HSPs reportable:  15
  # of sequences in database:  84,347
  # of database sequences with at least one HSP:  6
No. of states in DFA:  549 (54 KB)
Total size of DFA:  163 KB (176 KB)
Time to generate neighborhood:  0.12u 0.02s 0.14t
No. of processors used:  8
Time to search database:  30.41u 0.64s 31.05t
Total cpu time:  30.60u 0.71s 31.31t
<END
```