Synthetic RNA Molecules as Substrates for Enzymes That Act on tRNAs and tRNA-like Molecules

CHRISTOPHER S. FRANCKLYN and PAUL SCHIMMEL*

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received March 19, 1990 (Revised Manuscript Received August 2, 1990)

Contents

Christopher S. Francklyn is an NIH Postdoctoral Fellow in the Department of Biology at Massachusetts Institute of Technology.

Paul Schimmel is Professor of Biochemistry and Biophysics at Massachusetts Institute of Technology.

I. Introduction

Expression of genetic information depends on specific interactions between proteins and nucleic acids. Sequence-specific interactions are critical not only for transcriptional expression and regulation of genes but also for all translation and other posttranscriptional events. The recognition of specific RNA molecules is the basis for all these posttranscriptional events.

The recognition of tRNA and tRNA-like structures in protein synthesis is an example of the kind of problem for which in vitro RNA synthesis has been used to define the basis for specificity. Yeast $tRNA^{Phe}$,[1,2] yeast $tRNA^{Asp}$,[3] *Escherichia coli* $tRNA_i^{Met}$,[4,5] and *E. coli* $tRNA^{Gln}$ [6] are the only RNA molecules for which there are high-resolution three-dimensional X-ray structures. These structures have provided details about the precise geometries of helical domains and unpaired nucleotides, suggesting potential determinants for recognition by proteins such as aminoacyl tRNA synthetases.[7–13]

RNA sequence and length variants can be rapidly generated by in vitro synthesis. From the study of these variants, the contributions of individual nucleotides to the interaction energy and specificity of a specific complex can be tested. In vitro RNA synthesis has been used to investigate tRNA recognition, the mechanism of catalytic RNA,[14–18] RNA splicing,[19–21] and structure–function relationships of ribosomal RNAs.[22–24] With the exception of particular aspects of ribosomal RNA structure, little is known about the tertiary structures of RNAs other than tRNAs, so we focus here on the recognition of tRNAs and tRNA-like structures as model systems for the recognition of more complicated RNAs. Some general features of RNA structure, including the parameters for the basic helical motif and the atomic structures of tertiary interactions, such as triple base pairs, are known from the work on tRNAs.[25] These structural features are a starting point for the consideration of how synthetic RNA substrates have been used to study the recognition of specific tRNAs and tRNA-like structures.

0009-2665/90/0790-1327$09.50/0

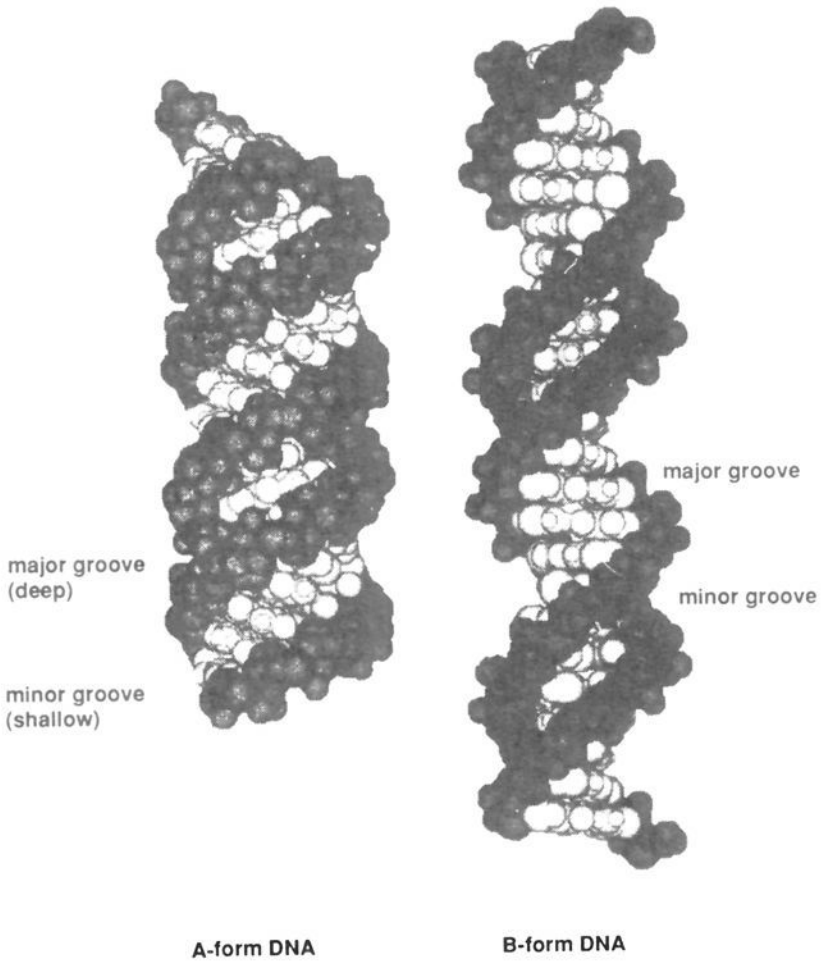

Figure 1. Space-filling models of A-form and B-form DNA. Views are perpendicular to the helical axis. The models are shaded and labeled to emphasize the differences in depth and width between the major and minor grooves, as well as differences in helical pitch.

II. Background and General RNA Structural Considerations

A. RNA Structure

With the exception of certain viral genomes—such as those of reovirus, sheep blue-tongue virus, rice dwarf virus and other unrelated icosohedral animal and plant viruses—RNAs generally are a single covalent strand. Thus, base pairing interactions are intrastranded, in contrast to the interstrand base pairing of DNA. The 2′-hydroxyl group influences the chemical properties of RNA and also imposes stereochemical constraints on RNA structure.[25] This functional group restricts the ribose conformation in oligomeric RNA molecules to the C3′-endo conformation. In DNA, the sugars freely interconvert between the C3′-endo and C2′-endo puckered conformations.

These constraints on ribose conformation determine the helical structures that are accessible. In the absence of modified bases and repeated C·G sequences, DNA molecules in aqueous solutions adopt either the right-handed A- or B-form helix, with complementary bases paired in the Watson–Crick scheme (see Figure 1). The B-form was originally detected in salt-containing DNA fibers at high humidity. The helical pitch is 10.5 base pairs per turn, and the planar bases are perpendicular to the helical axis.[26] The helix has a diameter of 22 Å and has two distinct grooves: a minor groove that lies between the C1′ carbons and a major groove that lies on the opposite side.

The A-form DNA helix is found in fibers of lower humidity and has a diameter (23 Å) greater than that of B-form DNA.[27] The pitch is 11 base pairs per turn, and the planar Watson–Crick base pairs are tilted with respect to the helical axis. An important difference

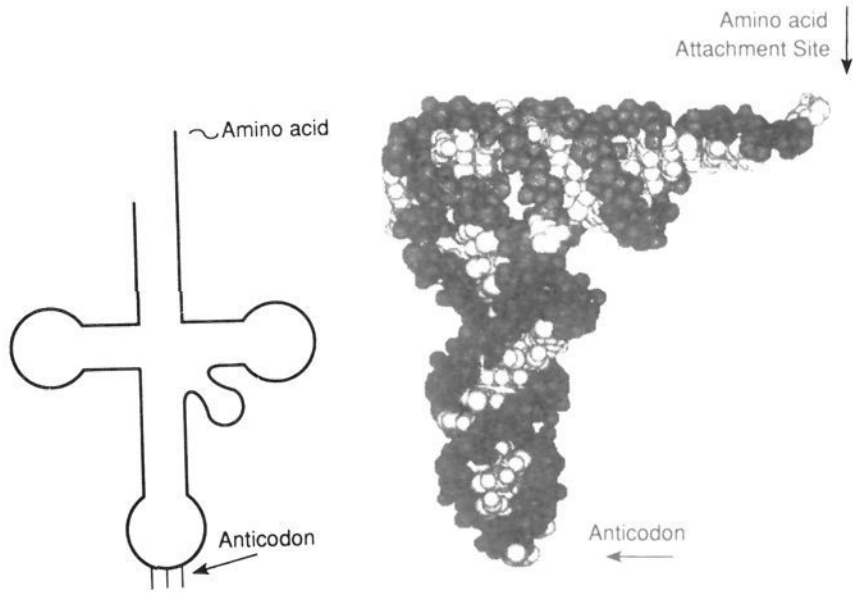

Figure 2. Comparison of a cloverleaf representation with a space-filling model of *E. coli* $tRNA^{Ala}$ whose nucleotide sequence was built into the yeast phenylalanine tRNA coordinates. The amino acid acceptor end and anticodon are indicated.

between A- and B-form helices is that the major groove of the A-form is narrower and deeper, while the minor groove is wider and more shallow. The RNA 11 helix has the same pitch and base pair tilt as A-form DNA, and the shape of the grooves is similar.[28] These two different types of helical structure (B-form and A-form/RNA 11), therefore, present different molecular surfaces to the proteins with which they make sequence-specific contacts.

RNA molecules assume a greater variety of tertiary structures than do DNA molecules, because of the lack of a complementary second strand and because of the potential to form Watson–Crick intrastrand hydrogen bonds between complementary sequences which can be well separated from each other in the linear sequence. In addition, the juxtapositioning of distant bases in the sequence allows for tertiary base pairing schemes that typically are non-Watson–Crick (e.g., Hoogstein pairing). Consequently, in the absence of proteins, double-stranded DNA rarely assumes the globular forms characteristic of transfer RNAs or ribosomal RNAs. The higher order DNA structures that are found in vivo, including those resulting from supercoiling and those associated with the folding of chromosomes, are dependent on topoisomerases and packaging proteins. Even so, the condensation of DNA in chromosomes results in a structure that is more rodlike than globular.

The structures of yeast phenylalanine tRNA[1,2] and the more recently crystallized tRNAs[3-6] have provided important details about the three-dimensional conformations that are possible for RNA molecules. When the sequences of the first tRNA molecules were obtained, the base pairing that gives rise to stems and loops suggested the two-dimensional cloverleaf structure that is now the conventional schematic representation of tRNAs (see Figure 2). It was predicted and confirmed (by extensive physical studies[29]) that base stacking stabilizes the final structure. At the time of the elucidation of the first sequence, the function of conserved unpaired bases in the cloverleaf was unknown. The X-ray structure of yeast $tRNA^{Phe}$ [1,2] revealed the participation of conserved nucleotides, such as U8, A14, G15, G22, G46, and ψ55, in tertiary base pairing schemes that were not anticipated. Included among the nine tertiary interactions found in the yeast $tRNA^{Phe}$ structure are triple base pairs, reverse

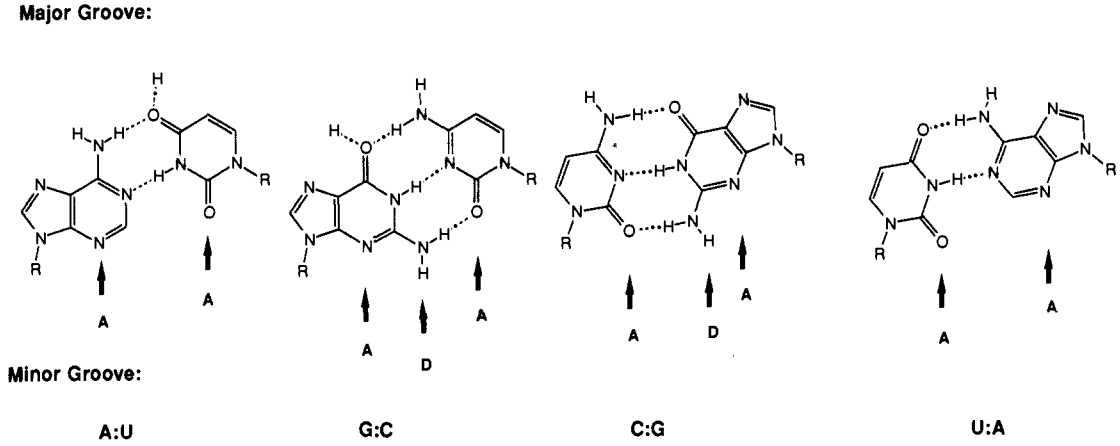

Figure 3. Comparison of the four possible Watson–Crick base pairs in RNA (adapted from ref 40). Functional groups that have the potential to form hydrogen bonds on the minor groove side of the base pair have been marked with arrows. "A" designates a potential hydrogen-bond acceptor, while "D" designates a potential hydrogen-bond donor.

Hoogsteen base pairs, and hydrogen bonds between bases and the sugar–phosphate backbone. Collectively, they establish the compact, L-shaped structure of tRNA, whereby the four helical stems are fused into two helices (the acceptor and TψC stems are stacked together, as is the D-stem with the anticodon stem) and the D- and TψC-loops are annealed together. Thus, the triple base pair between G22, C13, and G46 strengthens the interaction between the TψC and dihydrouridine loops and provides greater resistance to thermal, chemical, and enzymatic degradation. Base pairs can also hydrogen bond with the free 2′-hydroxyl of ribose or, as in the ternary interaction between G18, ψ55, and phosphate 58, with the phosphate oxygen from another portion of the backbone. Thus, the crystallographic structure revealed the influence of complex base pairing and stacking interactions for building a stable globular conformation.

All tRNAs also contain an extra arm with a variable number of nucleotides.[29] The length of this variable arm has been used as a means to classify tRNAs into various groups.[29] The most common type I tRNAs typically have five bases in the variable loop and four bases in the dihydrouridine stem. Type II tRNAs have 13–22 nucleotides in the variable loop and 3 base pairs in the dihydrouridine stem. At present, only the structures of type I tRNAs have been solved by X-ray crystallography.[1–6] Chemical protection studies of tRNAs whose structures have been solved[30,31] suggest that this method can reveal structural aspects of unusual tRNAs[32] or tRNAs with large variable loops.[33]

B. General Issues for Protein–RNA Recognition

1. Protein–DNA Interactions

What is known about DNA recognition[34,35] is instructive as a background for what has been learned about RNA recognition. The binding of proteins to specific sites in double-stranded DNA is an integral part of gene regulation, DNA synthesis, repair, recombination, andd cleavage. X-ray structures have been obtained for several specific complexes, including λ repressor,[36] 434 repressor,[37] trp repressor,[38] and *Eco*RI endonuclease.[39]

All of these proteins make the majority of their sequence-specific contacts with B-form DNA through major groove interactions. An early proposal[40] for how Gln, Asn, Arg, Asp, Gln, and Ser side chains could discriminate between sequences has been confirmed and further elaborated by subsequent structural analysis of protein–DNA complexes. The chemical basis for the discrimination between different base pairs lies in the order of hydrogen-bond acceptor and donor groups across the base pair that is accessible to a protein. As shown in Figure 3, a G·C base pair presents the pattern acceptor (guanine N7)–acceptor (guanine O6)–donor (cytosine N4). In principle, this potential array of hydrogen bonds permits all four base pairs to be distinguished from each other on the basis of major groove interactions. In each protein–DNA complex, the conformation of the protein, sometimes in conjunction with bends or kinks in the DNA conformation, acts to position uniquely the specificity-determining polar side chains with respect to the major groove in an orientation that is idiosyncratic to the complex. The nature of the base pair recognized by any particular amino acid side chain will depend on local geometry; for example, both λ and 434 repressor use glutamines to make sequence-specific contacts. The λ repressor's glutamine hydrogen bonds with adenine,[36] while 434 repressor's glutamines bond to guanines, adenines, and thymines.[37]

As initially suggested by modeling studies[41] based on the uncomplexed proteins and helix swapping experiments,[42] the repressors use a conserved α-helix–β-turn–α-helix to contact DNA, with the second of the two helices lying directly in the major groove. Polar side chains in this structural unit make contact with major groove bases in the operator through a series of hydrogen bonds and, occasionally, through hydrophobic interactions. Variations on this basic theme are found. For example, in λ repressor, the amide NH group of the side chain of Gln44 donates a hydrogen bond to ring N7 in an A·T pair, while the side-chain carboxyl oxygen accepts a hydrogen bond from the exocyclic N6 of adenine.[36] This bidentate interaction is further stabilized by a hydrogen bond from the amide group of Gln44 to the amide carboxyl of Gln33, while the amide amino group of Gln33 donates a hydrogen bond to the phosphate oxygen 5′ to the A·T pair. Thus, amino acid–base pair contacts can be part of a network of specific hydrogen bonds. In the case of trp repressor, tightly bound water molecules are thought to provide specificity by bridging between groups that are not in direct contact.[38] Hydrogen bonds from peptide amide groups

to the phosphate backbone may help to maintain specificity by fixing the orientation of the helix–turn–helix with respect to the major groove.[36,37] Often, subtle features of the DNA sequence influence the specificity of these protein–DNA interactions by modulating the DNA conformation, so as to create a molecular surface that is complementary to the protein.[37]

Like the repressors, *Eco*RI endonuclease also uses α-helices to make hydrogen bonds with the major grooves of its GAATTC recognition sequence, but the recognition helices do not assume a helix–turn–helix structure.[39] The amino-terminal ends of the two recognition helices in each of the two subunits point into the major groove bases of the inner tetranucleotide AATT. This places specificity-determining amino acid side chains in the proper orientation for base recognition: the carboxyl group of Glu144 receives hydrogen bonds from the successive N6 adenine exocyclic amino groups, while the Arg145 guanidinium donates two hydrogen bonds to the imidazole N7 nitrogens of the adenines located across the axis of symmetry. These "bridging" contacts, in which a single amino acid makes hydrogen bonds to functional groups on two successive base pairs, are unique to the *Eco*RI complex. The hydrogen bonds donated by each Arg200 guanidinium group to the O6 and N7 of the outer guanines, by contrast, are typical of the contacts made by the repressors.

2. Protein–RNA Interactions

Given that there are limited restrictions on RNA shape and conformation, there are no simple symmetry considerations that might suggest how proteins recognize RNA sequences. However, the RNA 11 conformation of RNA helices imposes some limits on the potential interactions with protein side chains.[28] In particular, the deep groove of this helical conformation is too narrow for protein structural motifs such as the α-helix to make direct sequence-specific contact. Therefore, the primary basis for sequence discrimination in RNA may be the minor groove.[6] As shown in Figure 3, there are fewer differences in the pattern of potential hydrogen-bond donors and acceptors presented by G·C and A·T (or A·U) base pairs from the face of the minor groove than from the face of the major groove.[40] Because both C and U have the 2-keto group as a potential hydrogen-bond acceptor in the minor groove, discrimination between some of the base pairs may be based on the exocyclic 2-amino group of guanine. This expectation is fulfilled by several examples in the structure of the Gln–tRNA synthetase–$tRNA^{Gln}$ complex.[6]

The three-dimensional structures of transfer RNAs are closely similar. With some specific local features that are idiosyncratic to each tRNA, the molecule features two helical regions, one of which terminates in the amino acid acceptor end, while the other terminates in the anticodon.[29] Thus, the structure of yeast $tRNA^{Phe}$ is a model for interpreting results on the sequence-specific recognition of most tRNAs. After the yeast $tRNA^{Phe}$ structure became available, Rich and Schimmel considered photochemical cross-linking, tritium labeling, and nuclease digestion data on synthetase–tRNA complexes and proposed that recognition is mediated principally through contacts made along the inside surface of the tRNA "L".[43] On this surface, both helical domains are potential sites for sequence-specific recognition through minor groove discrimination. In addition, at the inside of one end of the L the anticodon is a natural site for discrimination because the bases are unpaired and because this sequence codes for the attached amino acid. On the outside of the L, an alternative region is the "variable pocket", which is formed by the interaction of the TψC and D loops.[44,45] The nucleotides that comprise this patch, 16, 17, 59, and 60, are not conserved among tRNAs and are not engaged in Watson–Crick base pairs. Thus, several different regions potentially can contribute recognition determinants, and possible interactions include hydrogen bonds either to the minor groove exocyclic amino or keto groups or to the unpaired bases themselves. Nuclease protection studies carried out on many synthetase–tRNA complexes, including those of *E. coli* $tRNA^{Ile\,46}$, yeast $tRNA^{Phe\,47}$, yeast $tRNA^{Ser\,48}$, *E. coli* $tRNA^{Thr\,49}$, and *E. coli* $tRNA^{Ala\,50}$, have proved useful in identifying potential RNA–protein contact sites.

The three-dimensional structures obtained from crystals of the complexes between yeast $tRNA^{Asp\,51}$, *E. coli* $tRNA^{Gln\,6}$, and their respective synthetases have begun to provide more specific details of synthetase–tRNA interactions. In the cocrystal between *E. coli* $tRNA^{Gln}$ and the glutamine tRNA synthetase, the protein binds along the inside of the L-shaped structure as suggested by Rich and Schimmel.[43] The anticodon and specific acceptor stem nucleotides are in contact with the synthetase. In the acceptor stem, the exocyclic 2-amino group of G2 forms hydrogen bonds to the backbone carboxyl oxygen of Pro181 and to the backbone amide of Ile183. The latter interaction is bridged through a bound water molecule, in a fashion reminiscent of "indirect readout" first suggested in the trp repressor complex.[38] A hydrogen bond to the exocyclic 2-amino group of G3 is made by the carboxyl of Asp235, which also hydrogen bonds to the previously mentioned water molecule (see Figure 4).

A more complex feature is the interaction of the protein with the 3′ end of the acceptor stem and the conformational change by the nucleotides that are located in the 3′ acceptor end. The U1·A72 base pair at the end of the acceptor stem is wedged open by the side chain of Leu136, which protrudes from a β-turn in the acceptor binding domain of the protein. The rate of charging of $tRNA^{Gln}$ variants is influenced by the propensity of this base pair to be melted out.[52] The unpaired nucleotides ($GCCA_{76}$) at the 3′ acceptor end are folded back at a 90° angle with respect to the acceptor stem helix, such that the 3′ end is buried deep within the dinucleotide binding fold, in close proximity to bound ATP and (presumably) the bound amino acid. The 2-amino group of G73 hydrogen bonds to the phosphate oxygen of the previous nucleotide. This interaction stabilizes the unusual conformation of the acceptor arm, and specifically depends on having a G at position 73. There are also contacts between the synthetase and the anticodon which await further refinement of the structure. At present, it is clear that the recognition of $tRNA^{Gln}$ involves, at a minimum, contacts to the exocyclic amino groups in the minor groove and sequence-dependent conformation changes in the tRNA itself. The relative contribution of each of these interactions remains to be determined by

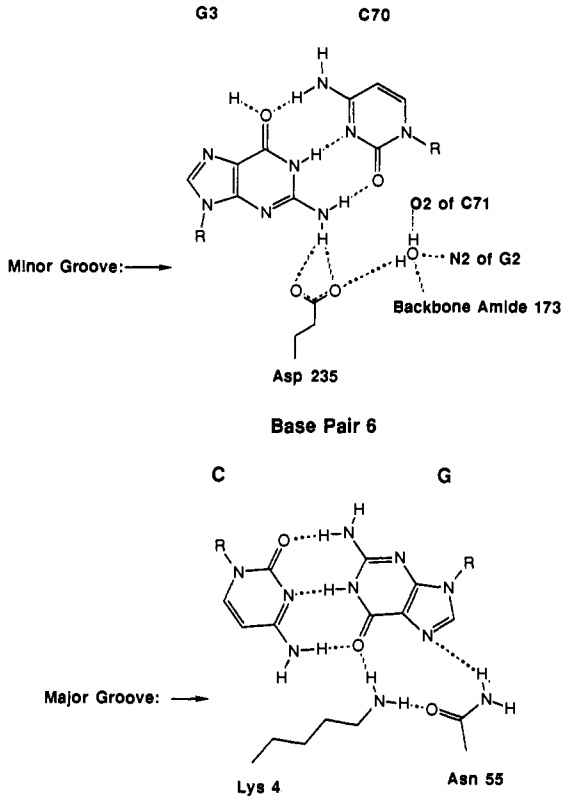

Figure 4. Examples of amino acid–base pair contacts using functional groups present in either the major or minor grooves. (a) The contact between Asp235 of *E. coli* glutamine tRNA synthetase and the minor groove of the G3·C70 base pair of *E. coli* $tRNA^{Gln}$, from ref 6. Hydrogen bonds are indicated by dashed lines. (b) Cooperative contact between Lys4 and Asn55 of λ repressor to major groove functional groups of base pair 6 in the synthetic operator, as described in ref 36.

systematic kinetic and binding studies with synthetic $tRNA^{Gln}$ variants, as well as with mutants of the enzyme.[53]

C. Brief History of Synthetic RNA Methodology

A useful approach for elucidating and testing models for recognition is by investigation of substitution mutations of both the protein and the nucleic acid. For RNA, one obstacle to this approach has been the difficulty in freely generating and isolating mutant and wild-type RNA species from whole cells in amounts that are sufficient for quantitative studies. Originally, only a limited number of methods were available to engineer specific changes in tRNA sequences and structure. Some of the earliest applications of in vitro RNA synthesis technology involved altering the nucleotides proximal to the 3′ acceptor terminus by using snake venom phosphodiesterase and tRNA nucleotidyl transferase.[54-56] Other early work featured the use of partial ribonuclease A digestion to cleave single-stranded regions of the tRNA, such as the anticodon or the unpaired CCA acceptor terminus. The resulting partial tRNA molecules were then tested for aminoacylation.[57-59] A useful technical advance was the introduction of RNA ligase to join small RNA oligonucleotides to larger fragments.[60] This permitted greater flexibility in the synthesis of sequence variants. Methodologies based on RNA ligase have been used to study the roles of specific nucleotides in several tRNA systems, including yeast $tRNA^{Phe}$, *E. coli* $tRNA^{Met}$, and *E. coli* $tRNA^{Ile}$ (discussed in detail below). Synthetic strategies based in part on RNA ligase have also been used in the synthesis of tRNAs from their constituent nucleotides.[61]

An alternative strategy for generating RNA sequence variants is based on the use of RNA polymerases from prokaryotic and eukaryotic sources. Initially, this methodology was a mixed success, owing in part to the use of purified RNA polymerase from *E. coli*, which has three different subunits in the core enzyme (α, β, β′) and a separate one for specific initiation (σ). Each of these subunits must be cloned for optimal use of this system. Frequently, reactions carried out by using this system were characterized by premature termination and the addition of non template encoded polyuridine tracts to the ends of products.[62] In later work, eukaryotic whole cell or nuclear extracts were used that either contained or were supplemented with RNA polymerases and other accessory factors.[63,64] The runoff transcripts obtained from these extracts suffered from some combination of poor yields, incorrect initiation, and premature termination.

Many of these technical barriers were overcome through the use of transcription systems based on the bacteriophages SP6 and T7, which each encode RNA polymerases that are single polypeptide chains. The SP6 system was originally characterized by Chamberlin and co-workers[66] and then used later to great advantage by Melton et al. to produce RNA probes of eukaryotic genes.[67] These in vitro synthesized RNAs were superior (to nick-translated DNA probes) in their ease of synthesis and in their high specific activity. They also were useful for elaborating details about the mechanisms of RNA processing and for providing an efficient means to program in vitro translation.

The T7 RNA polymerase system was first characterized by Studier and co-workers.[68,69] This single-subunit enzyme has a molecular weight of 98 000 and has been cloned and overexpressed in bacteria to aid in its purification. T7 polymerase is highly specific for a 23 base pair promoter sequence that is repeated 17 times in the T7 genome, but which has not been found in *E. coli* or other host DNAs. The viral promoter elements that are required for efficient transcription initiation have been incorporated into high copy vectors with multiple cloning sites for transcription templates.[70] In both the T7 and the SP6 systems, a simple reaction of few components is sufficient to obtain efficient in vitro synthesis.[68,71,72] The T7 system is presently favored because of the greater number of initiations (>100 vs <10) obtainable per template molecule as compared to the SP6 polymerase.[71,72]

The T7 RNA polymerase can initiate transcription from a promoter as small as 18 base pairs. The transcribed sequence can be single stranded, so that transcripts up to tRNA length (about 80 nucleotides) can be obtained from a template which has a double-stranded promoter and single-stranded coding sequence.[71,72] This system has limitations, because T7 RNA polymerase prefers to initiate transcription at a G and, in addition, the sequence of the transcript from +1 to +6 has a marked effect on the yield of product.[71-73] For example, the presence of a cluster of uridines affects the partitioning of the polymerase between

abortive initiation and elongation.[73] The T7 polymerase also has a propensity to add non template encoded nucleotides at the 3′ ends of transcripts, so that products can appear heterogeneous on denaturing polyacrylamide gels.[71] The extent of this non template encoded polymerization appears to be a function of the transcript sequence, but there are, as yet, no simple empirical rules for predicting its occurrence.

A complementary approach to the in vitro synthesis of RNA is the use of chemical synthesis.[74] Early workers in this field were stymied by a number of problems, especially the reactivity of the 2′-hydroxyl and the relative ease of hydrolysis of RNA under mild alkaline conditions. To bring chemical RNA synthesis up to the level of simplicity and repeatability of chemical DNA synthesis, an effective protecting group for the 2′-hydroxyl is required.[75] Usman et al. and others demonstrated the feasibility of the in vitro synthesis of long ribonucleotides by development of 3′-*O*-phosphoramidites that were protected at the 2′-position with a *tert*-butyldimethylsilyl (TBDMS) moiety.[76,77] In conjunction with controlled pore glass supports, the use of these monomers has permitted the complete chemical synthesis of a 77-nucleotide RNA sequence corresponding to $tRNA^{fMet}$.[78] When tested with a purified preparation of methionine tRNA synthetase, the chemically synthesized tRNA had a methionine acceptance of 11% of that of the native tRNA. The chemical approach promises to afford methods for introducing unusual bases into RNA, mixed intrachain RNA–DNA hybrid molecules, and other RNAs not available through enzymatic means.

III. Proteins That Interact with tRNAs and tRNA-like Structures

A. Aminoacyl tRNA Synthetases with tRNAs

1. General Features

The aminoacyl tRNA synthetases are an ancient class of enzymes that catalyze the two-step aminoacylation reaction.[9] There is one enzyme for each amino acid, and that enzyme charges all isoacceptors of its cognate tRNA species. In the first step of the reaction, the amino acid is activated by condensation with ATP to produce a bound adenylate; subsequently, the activated amino acid is transferred to the 3′ end of the cognate tRNA. The esterified tRNA forms a complex with elongation factor Tu, which delivers the charged tRNA to the ribosome. Although all synthetases catalyze the same reaction, they are diverse with respect to sequence, length, and quaternary structure.[9] One structural feature demonstrated to be common to several synthetases is that sequences involved in adenylate synthesis are localized to the amino-terminal part of the protein, while some of the sequences involved in tRNA binding are located in the carboxyl-terminal half.[79,80] The most conserved structure is the dinucleotide binding fold, an alternating arrangement of β-strands and α-helices that contains the sequences responsible for adenylate synthesis.[81]

The recognition problem has been investigated for a number of years by many different approaches.[8-13] An important distinction between this interaction and that between regulatory proteins and DNA is that synthetase discrimination between tRNAs can occur at a binding and at a catalytic step.[13] Unlike the interaction of a repressor with a DNA operator, the tRNA–enzyme complex must dissociate quickly to maintain protein synthesis. Consequently, the interaction is not as tight as repressor–operator interactions, and this limits the extent to which recognition can be achieved at the binding step. Dissociation constants at pH 7.5 are in the range 0.1–1.0 μM, which is at least 4 orders of magnitude weaker than that of a typical repressor–operator complex. The study of numerous cognate and noncognate synthetase interactions (reviewed in ref 13) has shown that, for some complexes, binding discrimination may only contribute a 100-fold preference for the correct tRNAs, while discrimination at the transition state of catalysis may be as high as 10^4 (ref 13).

In one of the earliest systems for studying tRNA recognition, variants of an *E. coli supF* amber suppressor (normally inserts tyrosine at UAG codons) were isolated that were aminoacylated with glutamine.[82] Determination of the minimal sequence changes associated with mischarging identified several positions within the acceptor end of the tRNA. These mutations included A73→G, as well as substitutions for the G1·C72 base pair. The molecular basis of glutamine mischarging with these mutant tRNAs was obscure, as some of these changes did not bring the suppressor sequence into closer agreement with $tRNA^{Gln}$. In the absence of an understanding of the molecular basis of glutamine mischarging, it was not clear how this type of genetic selection could be applied to other tRNAs. With the $tRNA^{Gln}$–GlnRS cocrystal now in hand, the effect of these mutations on mischarging can be at least partially rationalized.

In recent years, the ability to chemically synthesize genes for tRNA amber suppressors and to score their activity in vivo has rekindled interest in systematic studies of tRNA sequence variants. Abelson, Miller, and Normanly synthesized a set of tRNA genes coding for amber-suppressing tRNAs (CUA anticodon), with the object of defining the minimal set of nucleotide substitutions that are required to convert a tRNA from one amino acid identity to another.[83] So far, introduction of the CUA amber anticodon into 11 of 20 tRNAs does not change the amino acid attached in vivo.[10] This set includes Ala, Gly, Cys, Phe, ProH, HisA, Lys, Ser, Gln, Tyr, and Leu. The remaining tRNAs can be divided into two groups; the first, which includes tRNAs for Ile, Gly, Met, Glu, and Trp, are all mischarged with glutamine. The second group comprises those CUA-anticodon tRNAs that are mischarged with lysine tRNA synthetase (Ile, Arg, Met(m), Asp, Thr, and Val). For those tRNAs that are mischarged when their anticodons are substituted, one or more bases in the anticodon may be a recognition determinant for the cognate enzyme. Additionally or alternatively, it may be a determinent for the glutamine or lysine tRNA synthetases.

Those tRNAs unaffected by anticodon changes can be studied through an in vivo "transplantation assay", as devised by Normanly et al.[84] In this method, base substitutions are introduced into an amber-suppressing tRNA gene, which is then transformed into an *E. coli* strain that also carries a plasmid with a reporter gene that bears an amber (UAG) mutation. If the amber

suppressor is functional, the gene product from the reporter gene (typically dihydrofolate reductase[85]) is sequenced to determine the identity of the amino acid that has been inserted at the amber codon. By this method, introduction of 12 nucleotides that are common to a set of serine tRNAs into a leucine tRNA isoacceptor was sufficient to confer some serine acceptance in vivo.[84] Since then, this approach has been extended to the study of $tRNA^{Ala}$, $tRNA^{Phe}$, and $tRNA^{Arg}$ (see below).

Amber suppression is a valuable method for studying how the introduction of nucleotide substitutions into a tRNA sequence affect the amino acid identity of the tRNA. It is restricted, however, to those isoacceptors whose amino acid identities are preserved in the presence of a CUA anticodon. Another problem is that some variants do not accumulate to reasonable intracellular levels, owing to the effect of the nucleotide changes on stability and/or recognition by the processing system.

Another drawback to this approach is that the identity of a tRNA is influenced by competitive reactions between synthetases (cf. Yarus[87]). As shown by Swanson et al.[88] and Hou and Schimmel,[89] some tRNA variants can act as substrates in vivo for more than one aminoacyl tRNA synthetase. Consequently, altering the levels of synthetases by varying their relative gene dosages will change the amino acid acceptor identity of any "dual identity" tRNA. This phenomenon has been analytically treated by calculations that are based on kinetic parameters for aminoacylation in vitro with alanine and tyrosine of a $tRNA^{Tyr}$ variant which encodes the major determinant for alanine identity and is thus charged by tyrosine and alanine.[89] Therefore, the identity of a tRNA may represent the outcome of many potentially competing interactions between a tRNA and the whole set of cognate and noncognate synthetases in the cell. For these reasons, examining the interaction of a tRNA with its cognate synthetase in the absence of competing interactions provides information that is obscured by amber suppression.

Most seriously, amber suppression can, in some cases, occur with substrate variants that charge poorly or not at all in vitro.[90] Thus, suppression can be insensitive to large variations in the efficiency of aminoacylation and cannot be used to make an analytical estimate of the contribution of specific nucleotides to recognition. Studies carried out in vitro circumvent the problems associated with the excess sensitivity of amber suppression, which is influenced by factors in addition to aminoacylation.

2. $tRNA^{Met}$

Schulman and co-workers have used several in vitro techniques to study the recognition determinants of those *E. coli* tRNAs (such as $tRNA_i^{Met}$, $tRNA^{Val}$, and $tRNA^{Arg}$) where the anticodon is thought to play a major role in synthetase recognition. Early work on $tRNA_i^{Met}$ showed that bisulfate-induced conversion of C34→U (first position of anticodon) had a strong negative effect on aminoacylation.[91,92] This observation prompted the use of in vitro RNA synthesis to incorporate all four possible NAU anticodons into $tRNA_i^{Met}$.[93] The substitution was performed by limited digestion with RNase A to remove the native anticodon, followed by the insertion (by ligation with RNA ligase) of the substituted trinucleotides. Later, substitutions were made for A35 and U36.[94] These experiments showed that substitutions at C34 decrease aminoacylation rates with purified methionine tRNA synthetase by 4–5 orders of magnitude, while substitutions of A35 and U36 decreased rates by 1–4 orders of magnitude depending on the specific substitution.[93] Uemura et al. also used the RNA ligase method to show that changing A73 (the fourth base from the 3′ end of the tRNA, i.e., the "discriminator base") had no effect on aminoacylation of $tRNA_i^{Met}$.[95]

The role of the anticodon in determining $tRNA^{Met}$ identity was further investigated through the in vitro T7 RNA polymerase synthesis of anticodon variants of $tRNA_i^{Met}$, $tRNA^{Trp}$, and $tRNA^{Val}$.[96] Introduction of the methionine CAU anticodon into $tRNA^{Val}$ and $tRNA^{Trp}$ conferred aminoacylation with methionine at a rate (expressed as relative V/K_m) that was within 10% of that of $tRNA_i^{Met}$. This suggests that methionine tRNA synthetase recognizes the anticodon of $tRNA_i^{Met}$ and that other regions of the sequence are secondary for specificity. Photochemical cross-linking studies have identified a region of the methionine tRNA synthetase that is within 14 Å of the anticodon,[97] and subsequent site-directed mutagenesis experiments have implicated Trp461 as determining, in part, the specificity for C34.[98] Microinjection studies of *E. coli* $tRNA_i^{Met}$ derivatives into *Xenopus laevis* oocytes suggest that the correct anticodon sequence may also be required for modification of A37 to *N*-[*H*-(9-β-D-ribofuranosylpurine-6-yl)carbamoyl]threonine (t^6A).[99] This modification may increase translational efficiency and stability of the tRNA in vivo.

In work carried out on other synthetases that may recognize the anticodon, replacement of the methionine CAU anticodon with CUA was shown to reduce the rate of methionine acceptance by 3 orders of magnitude and to convert the tRNA into a substrate for glutamine synthetase in vitro,[100] as might be expected from previous in vivo supressor studies.[82] In vitro tRNA synthesis using T7 RNA polymerase was used to show that reciprocal exchange of the valine and methionine anticodons into the respective tRNAs makes them efficient substrates for the reciprocal synthetase.[100] By the same methods, the CCG anticodon and A20 of $tRNA^{Arg}$ were introduced into $tRNA^{Met}$, and the latter was transformed into an excellent substrate for arginine tRNA synthetase.[101] The substitution of A20 by U has been shown to eliminate arginine acceptance in vivo.[102] More recently, Schulman and co-workers have shown that introducing genes for $tRNA_i^{Met}$ with either the phenylalanine GAA or valine GAC anticodons can be used to engineer the initiation of in vivo protein synthesis with phenylalanine or valine.[103]

3. $tRNA^{Phe}$

With the availability of X-ray structural data for yeast $tRNA^{Phe}$, the effect of particular nucleotide substitutions on structure–function can be more accurately modeled than in the case of tRNAs for which no crystal structure yet exists. As in the case of $tRNA^{Met}$, the role of the anticodon in $tRNA^{Phe}$ recognition was initially investigated by use of RNase A digestion and T4 RNA ligase to make anticodon substitutions.[104] Using this

method, Bruce and Uhlenbeck reported that substitution of any one of the three GAA anticodon nucleotides resulted in a 3–12-fold decrease in aminoacylation by purified yeast phenylalanine tRNA synthetase.[105] However, introduction of the GAA anticodon into yeast $tRNA^{Tyr}$ gave a substrate that was only poorly aminoacylated with phenylalanine, suggesting that yeast Phe tRNA synthetase is sensitive to other sites.[106]

A more complete characterization of yeast $tRNA^{Phe}$ was carried out by Samson et al.[107,108] Utilizing the T7 system, they provided the first example of an in vitro tRNA transcript that could be quantitatively aminoacylated in vitro. This result showed that modified bases were not required for aminoacylation of $tRNA^{Phe}$. (A similar observation by Samuelsson et al. showed that a transcript based on the sequence of a glycine tRNA from *Mycoplasma mycoides* was also aminoacylated at a rate in vitro close to that of the native tRNA.[109]) This full-length $tRNA^{Phe}$ transcript was aminoacylated at a rate comparable to that of the native tRNA and had nearly the same temperature stability as the native tRNA.[107]

A series of transplantation experiments utilizing full-length transcripts of $tRNA^{Phe}$, yeast $tRNA_m^{Met}$, and yeast $tRNA^{Arg}$ was used to narrow the yeast $tRNA^{Phe}$ recognition set to G20, G34, A35, A36, and A73.[108] These five nucleotides are outside the conserved set of nucleotides for all tRNAs, but fall within single-stranded regions where the bases are most exposed. In further studies of the properties of unmodified in vitro transcripts, the structures of the yeast $tRNA^{Phe}$ transcript and various mutants have been analyzed by NMR[110] and by lead cleavage.[111] Substitutions at G20 do not produce large structural alterations, suggesting that the poor aminoacylation of $tRNA^{Phe}$ variants at this position may arise from the loss of a specific protein–tRNA contact. Other positions in the variable pocket appear not to be strong determinants, as nucleotide substitutions that preserve the pattern of secondary and tertiary interactions have little effect on aminoacylation kinetic parameters.[112]

The yeast $tRNA^{Phe}$ system is an example whose recognition determinants are distributed in at least three regions of the tRNA structure: the anticodon, the acceptor end (specifically, the discriminator base), and an unpaired base that projects from the surface of the tRNA. This distribution is seemingly in contrast to the recognition of $tRNA^{Met}$ (see Table I). The results of in vivo supression assays of *E. coli* $tRNA^{Phe}$ variants suggest that, as in the case of the yeast $tRNA^{Phe}$, nucleotide 20 is also an important determinant.[113] Other positions, however, are apparently not conserved.

Another novel approach to the $tRNA^{Phe}$ recognition features the chemical synthesis of tDNA substrates. Roe and his co-workers synthesized 76-nucleotide DNA oligomers corresponding to the sequence of *E. coli* $tRNA^{Phe}$ and *E. coli* $tRNA^{Lys}$, made with either ribo- or deoxyadenosines at the 3′ ends.[114] Aminoacylation of $tDNA^{Phe}$, but not of $tDNA^{Lys}$, was dependent on the presence of riboadenosine at the 3′ end, which is consistent with the observation that the 3′ end of *E. coli* $tRNA^{Phe}$ requires a 2′-hydroxyl for aminoacylation. Both of these tDNAs could only be aminoacylated to approximately 15% of the theoretical maximum, which may have been due to incomplete deprotection of some of the bases after the synthesis. Optimal aminoacylation for both substrates was obtained at pH 5.5 and in the presence of 20% dimethyl sulfoxide. These conditions are known to promote misacylation. Kinetic parameters obtained under these conditions for the tDNAs were within a factor of 10 of those of the native tRNA.

TABLE I. Determinants for Recognition of Transfers RNAs Studied with Partially or Completely Synthetic RNAs Molecules

transfer RNAs	nucleotides important for recognition	ref
E. coli alanine (mini- and microhelix)	G3·U70,[a] A73[b]	119, 120
E. coli arginine	anticodon[c] A20	101
E. coli histidine	pG-1·C73[d]	121
E. coli isoleucine	L34[e]	122
E. coli methionine	anticodon[f]	90–94, 96
yeast phenylalanine	anticodon, G20, A73[g]	107, 108, 111, 112
E. coli valine	anticodon[f]	96

[a] G3·U70 is the major determinant for specifying alanine acceptance. G3·U70 confers alanine acceptance in vitro on six different tRNA sequence frameworks which have been tested and on a collection of minihelix and microhelix sequence variants. [b] A73 enhances amnoacylation with alanine of those substrates which encode G3·U70. [c] Transfer of CCG arginine anticodon alone confers arginine acceptance on a methionine tRNA. A20 alone does not confer arginine acceptance, so that anticodon appears to be the major determinant. A20 improves efficiency of aminocylation with arginine of a methionine tRNA that has a CCG anticodon. [d] It has not been shown whether G-1·C73 can confer histidine acceptance on non-histidine tRNAs. [e] The role of L34 in tRNAs other than the LAU-anticodon-containing $tRNA^{Ile}$ has not yet been investigated. [f] Interchange of the methionine CAU and valine UAC anticodon interchanges the amino acid acceptance of the respective tRNAs. [g] Transfer of anticodon, A73, and G20 into *E. coli* $tRNA^{Phe}$, yeast $tRNA^{Arg}$, and $tRNA^{Met}$ confers aminoacylation with phenylalanine by yeast phenylalanine tRNA synthetase.

A possible explanation for these results is that the major determinants for recognition are single-stranded regions (see case of yeast $tRNA^{Phe}$), where the difference between tDNA and tRNA structure is least. In contrast, for substrates where helical regions encode determinants for recognition, the difference between A-form and B-form helices could prevent cross-aminoacylation of tDNA and tRNA substrates.

4. $tRNA^{Ala}$

a. Whole tRNA Substrates. The principal recognition determinants of *E. coli* $tRNA^{Ala}$ were first identified by screening nucleotide sequence variants of an amber-suppressing derivative of $tRNA^{Ala}$. Through systematic mutagenesis of the nonconserved positions in $tRNA^{Ala}$, with an emphasis on variations along the inside of the L-shaped structure, Hou and Schimmel determined that G·C and A·U substitutions at G3·U70 uniquely eliminated alanine acceptance.[115] Introduction of G3·U70 into $tRNA^{Cys}$, $tRNA^{Phe}$, and $tRNA^{Tyr}$ conferred alanine acceptance on each tRNA in vivo.[89,115] The result with $tRNA^{Phe}$ was confirmed by McClain and Foss.[116] Because G3·U70 is unique to $tRNA^{Ala}$ in *E. coli*, the results suggested that $tRNA^{Ala}$ may be discriminated from other tRNAs on the basis of this single base pair. In later studies, Hou and Schimmel showed that eukaryotic alanine tRNAs from *Bombyx mori* and human (which encode G3·U70) were also functional alanine-inserting suppressors in *E. coli*.[117]

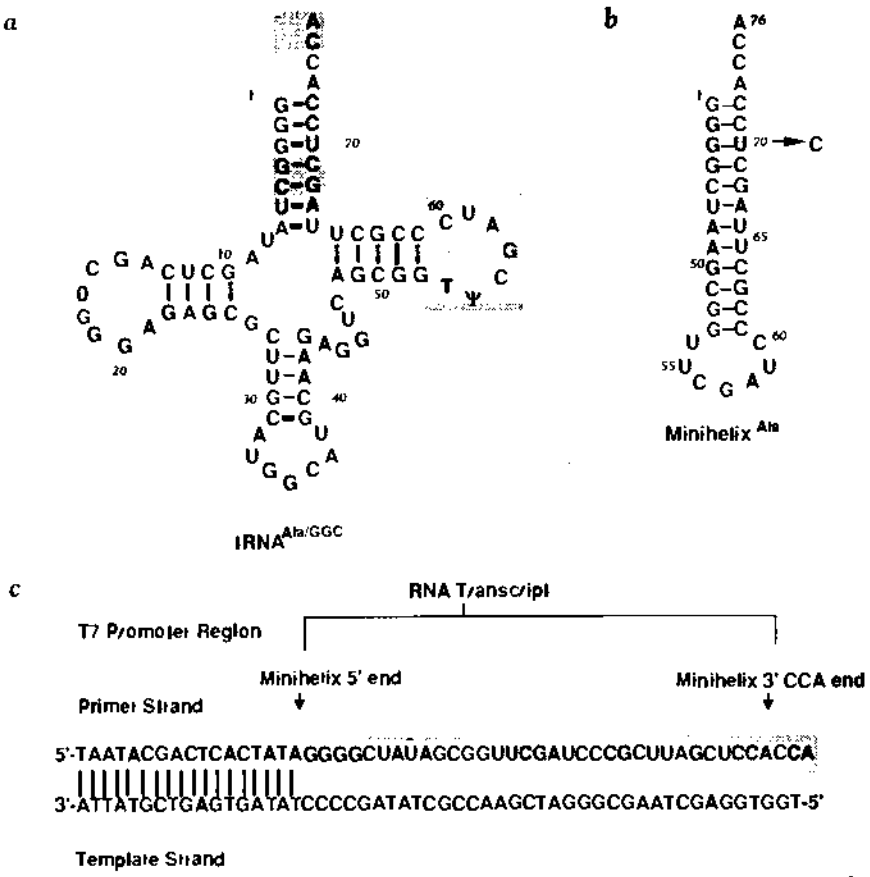

Figure 5. (a) Sequence and cloverleaf structure of tRNA$^{Ala/GGC}$ with nucleotides used in the construction of minihelix^{Ala} indicated by shading. (b) Synthetic hairpin minihelix based on the acceptor TψC helix of tRNA$^{Ala/GGC}$. The G3·C70 variant of minihelix^{Ala} is indicated on this figure; other sequence variants are shown in Figure 6. (c) The template-primer system (see refs 71 and 72) for the synthesis of mini- and microhelices. The synthesis of minihelix^{Ala} is shown as an example. For the synthesis of other mini- and microhelices, a specific single-strand DNA template strand is hybridized to a common 18 base pair promoter strand. Alternatively, minihelix^{Ala} transcripts have been produced from double-stranded templates derived from plasmid DNA preparations that have been linearized with a restriction enzyme, as described in ref 107.

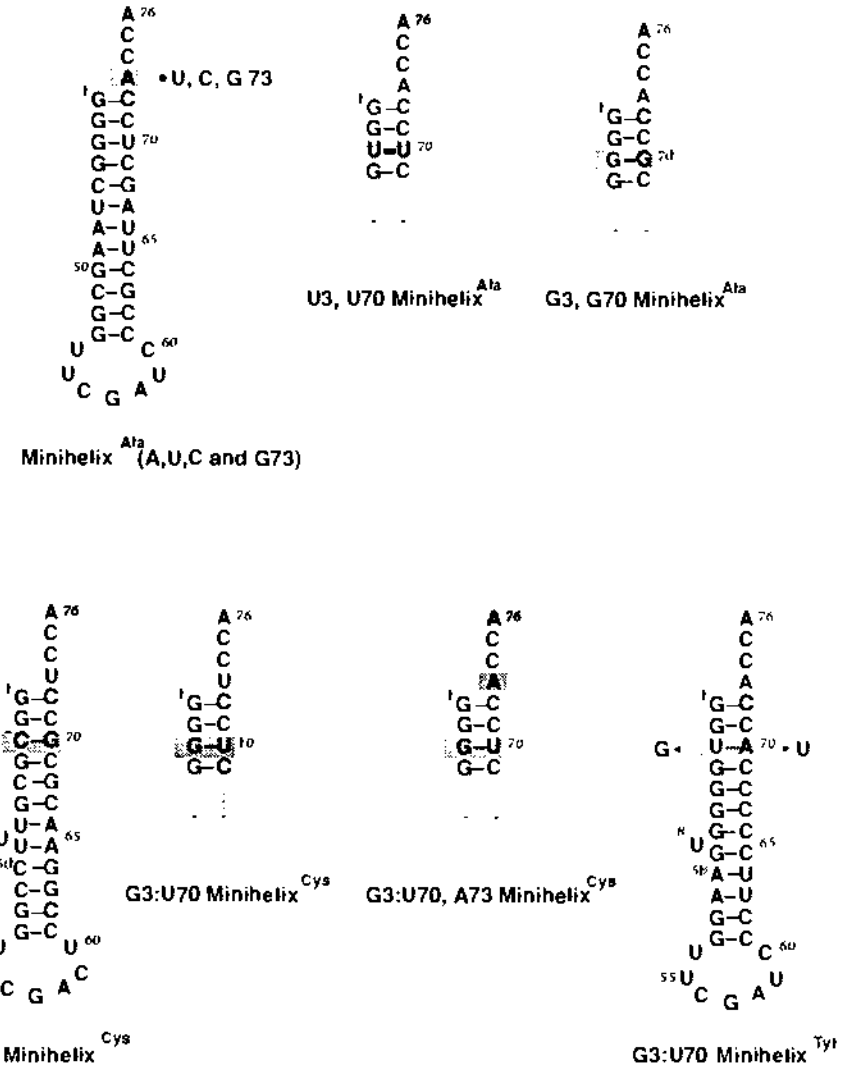

Figure 6. Sequence variants of minihelix^{Ala} minihelix^{Cys}, and minihelix^{Tyr} that have been examined as substrates for amino-acylation by alanine tRNA synthetase. The shaded portions refer to the positions where substitutions have been introduced. (a) Variants of minihelix^{Ala} in which substitutions have been introduced at positions 73 and 3·70. (b) Variants of minihelix^{Cys} in which substitutions have been introduced at positions 73 and 3·70. (c) Minihelix^{Tyr}, showing the U3·A70 to G3·U70 substitution which confers aminoacylation with alanine.

The role of the G3·U70 base pair in alanine tRNA synthetase recognition was also investigated in vitro. It was established that G3·U70 was required for the in vitro alanine acceptance of tRNA^{Ala} and that tRNA^{Cys} and tRNA^{Tyr} became substrates in vitro when G3·U70 was introduced.[115] The G3·U70 tRNA^{Cys} amber suppressor inserts alanine in vivo and no detectable cysteine, but it had a reduced rate and extent of aminoacylation in vitro.[115] In contrast, the G3·U70 tRNA^{Tyr} substrate is efficiently and completely aminoacylated.[89] Alanine tRNA synthetases from the insect *B. mori* and from human cells also demonstrate G3·U70-dependent in vitro aminoacylation of their homologous substrates, suggesting that the role of this base pair has been conserved during evolution.[117]

Park et al. showed that the *E. coli* enzyme recognizes the G3·U70 base pair during both the binding andd catalytic steps of aminoacylation.[118] In particular, when A3·U70 tRNA^{Ala} is bound to the enzyme at a site that competitively inhibits binding of native tRNA^{Ala}, there is no aminoacylation of the A3·U70 species. Thus, the G3·U70 determinant may trigger a conformational change in the transition state of the reaction.

b. Minihelix Substrates. Independent support for the role of the G3·U70 base pair in the catalytic steps of aminoacylation was provided by the analysis of truncated derivatives of tRNA^{Ala} which can be aminoacylated with alanine. Through the use of the in vitro T7 transcription system, short transcripts corresponding to the 12 base pair acceptor–TψC stem and loop of *E. coli* alanine tRNA^{Ala} were analyzed for alanine acceptance.[119] This segment constitutes one domain or "arm" of the L-shaped tRNA structure (see Figure 5). (In a footprint of whole tRNA^{Ala}, alanine tRNA synthetase also protects the acceptor–TψC region from nuclease attack, but does not protect either the D-stem and loop or the anticodon.[50]) This domain is aminoacylated with alanine with a k_{cat} comparable to that of native tRNA; a small elevation in K_m corresponds to a loss of interaction energy of only 1 kcal mol^{-1}. The smallest substrate tested was a seven base pair helix and five nucleotide loop that is based on the sequence of the acceptor stem. Efficient aminoacylation of this substrate showed that sequences outside the acceptor helix are dispensable for charging. In addition, transplantation of G3·U70 into a minihelix based on the acceptor–TψC sequences of tRNA^{Tyr} conferred efficient alanine acceptance in vitro. The kinetic parameters for aminoacylation of this substrate are nearly the same as for the aminoacylation of G3·U70 tRNA^{Tyr}.[89] Thus, the 49 additional nucleotides of tRNA^{Tyr} do not perturb the interaction of the enzyme with the acceptor helix.

The minihelix system has been used to resolve two aspects of tRNA^{Ala} recognition raised by in vivo studies. McClain et al. observed weak suppression of amber codons in β-galactosidase mRNA by tRNA^{Ala} variants encoding alternative bases pairs at the 3·70 position.[90] Among these, variants encoding U·G, G,A, A·C, C,A, and U,U inserted alanine among other amino acids. Using the minihelix system, Shi et al. synthesized

TABLE II. Kinetic Parameters for Aminoacylation with Alanine of Minihelix RNAs and tRNAs at pH 7.5, 37°C[a]

RNA	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m, M^{-1} s^{-1}
$tRNA^{Ala}$ (*E. coli*)	0.93	2.1	4.4×10^5 [b]
G3·C70 $tRNA^{Ala}$			(0)[c]
A3·U70 $tRNA^{Ala}$			(0)[c]
U3·G70 $tRNA^{Ala}$			(0)[c]
G3·U70 $tRNA^{Tyr}$	0.60	14.0	4.3×10^4 [d]
$minihelix^{Ala}$	0.91	9.1	1.0×10^5 [e]
G3·C70 $minihelix^{Ala}$			(0)[b]
U3,U70 $minihelix^{Ala}$			(0)[e]
G3,G70 $minihelix^{Ala}$			(0)[e]
$minihelix^{Cys}$			(0)[e]
G3·U70,A73 $minihelix^{Cys}$	0.28	8.8	3.2×10^4 [e]
G3·U70 $minihelix^{Tyr}$	0.48	8.8	5.3×10^4 [b]

[a] Minihelices are shown in Figure 6. [b] From ref 119. [c] From ref 118. [d] From ref 89. [e] From ref 120.

U3,U70 and G3,G70 variants and found them to be completely inactive for aminoacylation (see Figure 6; Table II).[120] In similar assays utilizing full-length $tRNA^{Ala}$ variants, those encoding U·G, G·C, or A·U base pairs at position 3·70 were also defective for aminoacylation.

Several explanations for the discrepancy between the in vivo and in vitro observations may be proposed. First, other tRNA-binding proteins such as EF Tu may act in vivo to trap the small quantities of aminoacylated product, effectively preventing these poorly aminoacylated tRNAs from acting as substrates in the "posttransfer" editing reaction. Additionally or alternatively, the combination of the overproduction of these variant tRNAs and the high sensitivity of the suppression assay may act to magnify a reaction that is too small to detect in vitro.

A further question addressed with the minihelix substrates concerns the effect of transplanting G3·U70 into $tRNA^{Cys}$. In contrast to G3·U70-encoding substrates that are efficiently aminoacylated by alanine tRNA synthetase, $tRNA^{Cys}$ encodes a U at position 73 instead of an A. This nucleotide was originally called the discriminator, because tRNAs specific for amino acids of a particular chemical type (e.g., hydrophobic) had the same base at position 73 (i.e., an A). Using $minihelix^{Ala}$ and $minihelix^{Cys}$ variants with various nucleotide substitutions at position 73, in vitro charging assays revealed that an A at position 73 is required for efficient aminoacylation by purified alanine tRNA synthetase (see Table II).[120] The substitution of other nucleotides at position 73 sharply decreased the rate and extent of G3·U70 aminoacylation. (These same changes were shown by Hou and Schimmel to have no effect on amber suppressor function in vivo, perhaps for some of the reasons mentioned above.[115]) Thus, G3·U70 alone is sufficient to confer alanine acceptance, but position 73 has a significant modulatory effect.

The idea of "primary" (i.e., G3·U70) and "secondary" (i.e., A73) recognition determinants may turn out to be a feature of tRNA recognition. As described earlier, both the arginine CCG anticodon and A20 must be introduced into $tRNA_i^{Met}$ to achieve efficient in vitro aminoacylation with arginine tRNA synthetase.[101] Of these two determinants, the introduction of the CCG anticodon alone into $tRNA^{Met}$ is 40-fold more effective in raising V_{max}/K_m, as compared to A20 $tRNA^{Met}$. The presence of multiple recognition determinants in a tRNA implies nothing about the degree of interaction between them.

There may be other tRNAs in which the acceptor stem is the primary location for recognition determinants. All sequenced $tRNA^{His}$ molecules contain an additional G at their 5′ ends, making them one nucleotide longer than other tRNAs at this end. This additional nucleotide is paired with C73 in *E. coli* $tRNA^{His}$. Recently, Himeno et al. reported that this G_{-1}·C73 base pair in *E. coli* $tRNA^{His}$ is required for efficient aminoacylation of synthetic transcripts.[121] All substitutions of this base pair that were examined (including a triphosphate variant at the −1-position) had a deleterious effect on aminoacylation, suggesting that the enzyme is sensitive to changes in the tRNA at this position. It is of interest to test a minihelix based on the sequence of the acceptor–TψC stem of $tRNA^{His}$.

5. *$tRNA^{Ile}$ and $tRNA^{Asp}$: Modified Bases Involved in tRNA Recognition*

In the examples discussed above, the high rate and extent of aminoacylation observed with synthetic RNA transcripts suggests that the absence of modified bases has little or no effect on aminoacylation. Muramatsu et al. recently described an example with $tRNA^{Ile}$ in which a modified anticodon base plays a crucial role in synthetase recognition.[122]

There are two *E. coli* isoleucine tRNA isoacceptors that are substrates for isoleucine tRNA synthetase. The major species (GAU anticodon) reads AUU and AUC codons, while the minor species reads AUA codons through an LAU anticodon. L is the modified base lysidine, which has the ε-amino group of a lysine joined to C2 of the pyrimidine ring of cytidine.[123] Through the use of anticodon replacement techniques featuring RNA ligase[60,93,104-106] the substitution of CAU for LAU at the $tRNA^{Ile_2}$ anticodon was demonstrated to abolish aminoacylation by isoelucine tRNA synthetase. Concomitantly, the $tRNA^{Ile}$(CAU) became a good substrate for methionine tRNA synthetase, as expected from the work of Schulman and co-workers.[101] Thus, a G or an L at position 34 specifies isoleucine acceptance in this tRNA. (It is of interest to note that early studies showed that bound Ile–tRNA synthetase blocks hybridization of enzymatically synthesized AUCA and AUCG tetranucleotides to the anticodon of $tRNA^{Ile}$[124] and also blocks cleavage after G35 by T1 ribonuclease.[46]) Examination of the two (G or L) position 35 bases suggests that the ε-nitrogen by lysine may be a surrogate for N3 of guanine to make a portion of L resemble G.

Modified bases apparently play a role in modulating the recognition of yeast $tRNA^{Asp}$ according to a recent paper by Perret et al.[125] Instead of functioning to promote the recognition of this tRNA by the cognate synthetase, the modifications act as blocking determinants to prevent the interaction of yeast $tRNA^{Asp}$ with noncognate enzymes. Kinetic parameters for aminoacylation with aspartate and arginine were determined for both the naturally isolated tRNA and the transcript containing no modified bases. Both of these substrates had the same kinetic parameters for aminoacylation with aspartyl tRNA synthetase, but the unmodified transcript had nearly 300–500 times the specificity (expressed as k_{cat}/K_m) for the arginyl synthetase as the

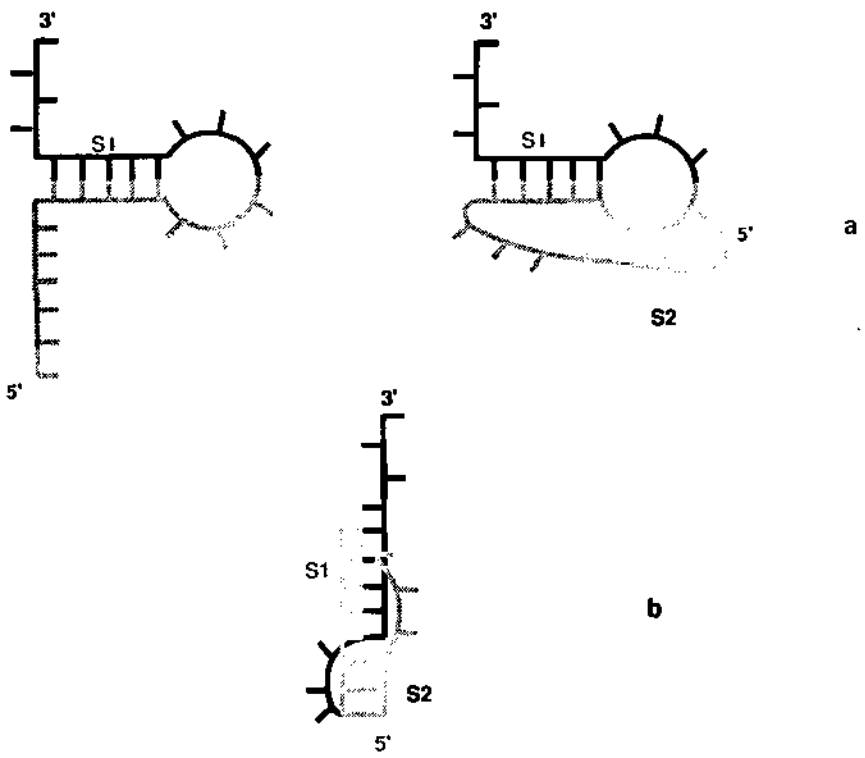

Figure 7. Simplified schematic drawing of a pseudoknot RNA structure. (a) Formation of a pseudoknot where sequences distal to a stem (S1) and loop hydrogen bond to the loop top form a second stem (S2). (b) Alternative representation of the pseudoknot that illustrates the formation of a quasicontinuous helix and the bridging of loop nucleotides across the helix.

modified tRNA. Only three of the modified bases (ψ13, ψ32, and m^{1}G37) are specific for tRNA^{Asp}. Further work will be necessary to explore this phenomenon and determine the contributions of the individual modifications.

B. Interaction of tRNA Synthetases with tRNA-like Structures

1. Plant Viral 3′ Ends

Some aminoacyl tRNA synthetases aminoacylate the 3′ ends of certain genomic and subgenomic plant viral RNAs. This suggests a structural relationship between tRNAs and the 3′ ends of these viral RNAs. Computer predictions of structure were tested experimentally with chemical probes[126] and led to the proposal of an RNA pseudoknot that enables a tRNA-like structure to form at the 3′ end.[127]

In the RNA pseudoknot, bases in a hairpin loop form Watson–Crick pairs with bases that are located outside the hairpin structure.[128] Because less than 11 base pairs form with the loop, there is only partial revolution of one strand about the other, so that a true knot is avoided. In the pseudoknot described by Pleij et al. (see Figure 7) for turnip yellow mosaic virus, there is coaxial stacking of the two different helical stems of the pseudoknot.[129] The stems are joined by two different connecting loops which cross the major and minor grooves, respectively. The pseudoknot structure is supported by the 2-D NMR studies of Puglisi et al. on short synthetic RNA fragments, where the stability of the pseudoknot has been shown to be sensitive to temperature and Mg^{2+} concentration.[130]

Brome mosaic virus (BMV) RNA (aminoacylated by tyrosine) and turnip yellow mosaic virus (TYMV) RNA (aminoacylated with valine) are the most extensively studied plant viral RNAs. For BMV, Dreher et al. demonstrated that a synthetic 135-nt fragment retains aminoacylation function.[131] The sequence requirements for aminoacylation with tyrosine and for viral replication were studied by introducing mutations into the viral 3′ end and at a putative AUA "anticodon" sequence.[131] Those substitutions in which the CCA end was changed had abolished aminoacylation function, but retained at least partial replication function. The sequences at the AUA anticodon, by contrast, were not required for aminoacylation, but severely attenuated replication. The genetic separation of aminoacylation and replication functions in BMV RNA suggests that aminoacylation is not required for virus viability. Aminoacylation may be an evolutionary vestige of a much closer relationship between replication and protein synthesis that predated the appearance of DNA.[132]

Studies on RNAs synthesized from a TYMV cDNA indicate that our understanding of the relationship between the viral 3′ ends and aminoacylation may be incomplete. Dreher et al. synthesized a series of length variants of TYMV RNA in vitro and determined kinetic parameters for their aminoacylation by wheat germ valine tRNA synthetase.[133] Although 82 3′-terminal nucleotides can be folded into a tRNA-like structure that can be aminoacylated in vitro, sequences which lie upstream of this structure (between 82 and 159 from the 3′ end) are required for a maximal rate and extent of aminoacylation. The decreased rate of aminoacylation of fragments shorter than 159 nucleotides is reflected predominantly in a decreased V_{max} rather than K_m, suggesting that the sequence 82–159 affects catalytic rather than binding steps. This is another demonstration of the significance of the transition state for catalysis for recognition by synthetases. Footprinting studies carried out on the fragments with purified synthetase suggest that the enzyme either contacts this region directly or that this region is required for the correct conformation of the tRNA-like domain.[134] Unlike the BMV RNA, the TYMV anticodon is an important determinant for aminocylation, as it is for *E. coli* tRNA^{Val}.

2. Interaction of a Synthetase with a tRNA-like Structure Involved in Translation Repression

Through their regulation of the pool of charged tRNAs, aminoacyl tRNA synthetases influence the rate of protein synthesis. This may create selective pressure for regulation of the intercellular level of synthetases. *E. coli* threonine tRNA synthetase is an example of one that is translationally regulated. The autogenous control of this synthetase was initially demonstrated through the in vivo analysis of *thrS* operon and gene fusions to *lacZ*.[135] Only gene fusions in which the initial portion of the structural gene for *thrS* was joined to *lacZ* were repressed by exogenous copies of *thrS*. Because operon fusions (which lack the initial portions of the structural gene) were not repressed, this provides evidence for regulation at the level of translation. These studies defined the translational operator as a 134 base pair region that is between 10 and 15 nucleotides upstream of the initiator AUG and 100 bases downstream of the transcriptional start site.

The system has been studied in vitro by subjecting the synthetic fragments of the mRNA operator or synthetase–operator complex to various chemical and enzymatic probes. Moine et al. identified four helical loop regions in the operator on the basis of their sen-

sitivity to modification with dimethyl sulfate and diethyl pyrocarbonate.[136] The helix III loop A domain possesses a high degree of sequence similarity to the anticodon region of $tRNA^{Thr1}$. Significantly, the replacement of "G-32" (32 nucleotides before the AUG translational start codon) in the unpaired anticodon loop of this structure leads to a loss of translation regulation; the analogous G in the anticodon loop of $tRNA^{Thr}$ has been shown with chemical probes to be in close contact with threonine tRNA synthetase.

A feedback mechanism has been proposed for regulation by binding of Thr-tRNA synthetase to the anticodon-like structure to block the access of ribosomes to the Shine-Dalgarno sequence. Confirmation and refinement of this model will come through studies of mutations in the operator sequence and characterization of second-site suppressors in *thrS*.[137] Mutations in threonine tRNAs can also be examined, so that basis of "homology" between threonine tRNAs and the operator can be established.

C. RNase P: An Enzyme with an Essential RNA Component That Recognizes tRNA

RNase P is required for maturation of the 5′ ends of tRNA precursors. The enzyme has two different subunits in all organisms investigated so far. In *E. coli*, these consist of a 13.7-kDa protein component (C5) and a 377-nucleotide RNA component known as the M1 subunit.[138] This nuclease distinguishes tRNA precursors from all other RNAs. Mutational analyses of precursor molecules showed that substitutions that disrupt the secondary or tertiary structure of the precursor inhibit the cleavage reaction.[139,140] Thus, the enzyme is sensitive to the structure of the precursor. RNA synthesis of enzyme and substrate component has proved to be an effective way to approach recognition of tRNA precursors.

The essential role of RNA in the catalytic event was first demonstrated when cleavage of the precursor tRNA was shown to be dependent on both M1 RNA and C5 protein.[141,142] Subsequently, Guerrier-Takada et al. showed that the requirement for C5 could be overcome by raising the Mg^{2+} concentration from 10 to 60 mM.[143] Kinetic parameters at 60 mM Mg^{2+} were determined for the holoenzyme reaction and for the reaction with M1 RNA alone. Under these conditions, C5 increased the velocity of the reaction by 2-fold, but had no effect on the K_m. The *rnp A* gene that codes for the C5 protein subunit is essential for viability in *E. coli*, so the operational Mg^{2+} concentration in vivo maybe closer to that (10 mM) used in the original assays. In vitro, it is possible to carry out complementation experiments utilizing the *E. coli* RNA and *Bacillus subtilis* C5 protein.[143] Thus, the protein may recognize features of the RNA structure that have been conserved during evolution.

The C5 protein and M1 RNA components of RNaseP have been cloned and overexpressed.[142,144] Utilizing these reagents, Vioque et al. measured a dissociation constant of 4×10^{-10} M for the binding of M1 to C5.[145] Footprint analysis showed protection of nucleotides between 82–86 and 170–270. A competition assay was used to examine the binding of synthetic truncated derivatives of M1 RNA to C5.[145] A fragment comprising sequences from 94 to 272 effectively competed away binding to nonspecific RNAs, while a fragment spanning either 1–168 or 164–272 did not.

A phylogenetic comparison of nine different sequences from two different eubacterial phyla established a "consensus" RNase P (Min 1) that contained only 263 nucleotides versus the 354–417 nucleotides of the parental structures and incorporated stems, loops, and pseudoknot features that were conserved between all members of the collection.[146] The Min 1 consensus also contained one of the regions implicated by footprinting, i.e., the sequences between 82 and 96 in *E. coli* M1 RNA. In vitro transcripts of the Min 1 structure processed a pre-$tRNA^{Asp}$ substrate at a rate that was only 5-fold slower than that of full-length *E. coli* M1 RNA. The success of this design strategy is consistent with the belief that particular structural determinants of RNase P have been conserved through evolution.

The region from 86 to 92 in M1 has been further implicated by enzyme–substrate cross-linking studies.[147] Mixtures of M1 RNA and a pre-$tRNA^{Tyr}$ were irradiated with UV light at 300 or 254 nm and then resolved on polyacrylamide gels to isolate the specific complexes. Reverse transcriptase was used to establish the points of cross-linking in both the enzyme and substrate. Reverse transcription terminated consistently at C93 in M1 RNA, indicating that C92 is cross-linked to the substrate. The cross-linking experiments also defined points of contact to the substrate. Efficient termination of reverse transcription (using a primer complementary to the 3′ end of the tRNA precursor substrate) occurred at G-2 (two nucleotides before the start of the mature tRNA). This indicates that C92 in M1 is cross-linked to "C-3" in the pre-tRNA. This is within three bases of the cleavage site in the pre-tRNA.

Deletion of C92 in M1 RNA raised K_m by 100-fold and lowered k_{cat} by 6-fold relative to wild type M1, in the absence of C5. However, the specific nucleotide at position 92 is not critical, because a U92 substitution mutant had nearly the same kinetic parameters for processing as wild-type M1. Also, deletion of C92 can be partially overcome by the presence of the C5 subunit. Thus, N92 may influence the local conformation of the RNase P active site, but may be secondary to the influence of the C5 protein subunit.

In parallel with the work on the M1 RNA, in vitro RNA synthesis has also been used to investigate the substrate requirements for the reaction. Truncated versions of *E. coli* $tRNA^{Phe}$ that retain the acceptor–TψC stem and loop are substrates for the enzyme, but the introduction of base substitutions at C74 (the 3′ terminus is A76) eliminated cleavage.[149] As described for alanine tRNA synthetase, RNase P recognizes a limited part of the overall tRNA structure.

There is also evidence to suggest that RNase P recognizes the 3′ CCA sequence of the precursor tRNA molecule.[148] The precursor to *E. coli* $tRNA^{Tyr}$ is three nucleotides longer at the 3′ end than the mature species, such that the sequence is $CCAUCA_{OH}$. Cleavage of this substrate in vitro with M1 RNA or the RNaseP holoenzyme reveals that the turnover number for the reaction with M1 RNA alone is greatly reduced in the absence of the CCA sequence. The wild-type M1 RNA will correctly cleave a pre-$tRNA^{Tyr}$ which lacks the 3′-terminal CCAUCA, although at a slower rate than for the wild-type precursor. A mutant RNase P with

a deletion of C92 also cleaves the mutant precursor, but does so at a site that is four to six bases upstream of the wild-type cleavage site. Reverse transcription of a photo-cross-linked complex between the mutant M1 RNA and the mutant pre-tRNA^{Tyr} gave strong termination at G1 in pre-tRNA. It is noteworthy that high concentrations of exogenous CCA trinucleotide inhibit the reaction of a substrate that contains the CCA group, but stimulates the processing of a substrate that lacks the trinucleotide. Thus, RNase P may have two separate binding sites for the pre-tRNA—one associated with the eventual site of cleavage and one for the CCA end.

IV. Summary and Conclusions

A. Limitations of In Vitro RNA Synthesis

In the case of synthetic tRNAs, one drawback is that the transcripts are unmodified. The lysidine in the *E. coli* tRNA^{Ile_2} isoacceptor is a modified base shown to be essential for aminoacylation,[122] and modifications in yeast tRNA^{Asp} act as negative determinants for other aminoacyl tRNA synthetases.[125] However, unmodified transcripts may become useful in the purification and characterization of tRNA nucleotide modification enzymes, including pseudouridine synthase from *Saccharomyces cerivisiae*[150] and guanine methyltransferase from *Xenopus* oocytes.[151] Microinjection of in vitro transcripts into *Xenopus* oocytes can be used to produce modified tRNAs in vivo.[152] As more of the genes coding for the tRNA modification enzymes are cloned and their gene products characterized, tRNA transcripts produced in vitro can be treated with these enzymes to study the effects of modifications on molecular recognition.

Another obstacle to obtaining the full range of potential sequence variants is the promoter specificity of the T7 RNA polymerase. The enzyme prefers a G at the first position and is highly sensitive to the sequence in the region +1 to +6.[73] Methods to circumvent these nucleotide requirements include the use of purified M1 RNA to process pre-tRNA transcripts in vitro.[153] The minimal substrate requirements for RNase P have been localized to the acceptor–TψC helix and the CCA end.[149] Thus, as long as a transcript possesses these minimal features, a G can be added to the 5′ end to satisfy the promoter requirements for T7 RNA polymerase. The extra G in the resulting transcript can then be removed by the action of M1 RNA in vitro. Alternatively, T7 transcription can be primed by N_pG dinucleotides to obtain RNA transcripts with 5′ ends starting with nucleotides other than G.[72]

The efficient synthesis by T7 RNA polymerase of RNA transcripts shorter than 20 nucleotides is more sensitive than longer transcripts to the sequence of the coding strand[71] and is also hindered by the tendency of the enzyme to undergo an "abortive cycling mode".[73] In this mode, the polymerase does not clear the promoter and synthesizes short transcripts two to six nucleotides in length. Sequence analysis of abortive products shows that termination occurs most frequently after incorporation of UMP.[73] Use of chemical RNA synthesis may be required in those cases where sequence or length constraints decrease the yield and homogeneity of product.

Several laboratories have observed that T7 polymerase will in some circumstances produce a transcript of approximately 70 nucleotides whose synthesis is independent of exogenous DNA template.[154] Instead, the template for this "product X" may copurify with the polymerase itself and be used by the enzyme when the sequence of the exogenous template is unfavorable. The synthesis of the X RNA is initially slow, but the rate accelerates after a lag period of an hour or greater. Thus, X RNA can represent a greater proportion of the product in transcription reactions that are carried out over long (e.g., 3 h) periods. Initial studies on the sequence and structure of X RNA suggest that it is a self-complementary RNA of approximately 35 nucleotides. If the reaction is carried out for an extended period, then multimeric forms of X RNA appear. Clearly, more work is needed to characterize the phenomenon of non template encoded RNA synthesis by T7 RNA polymerase.

B. RNA Recognition

In vitro RNA synthesis is most useful when combined with a complementary method to rapidly identify positions of potential interest in a given RNA structure. A genetic method such as amber suppression, in which a great number of variants can be rapidly screened, is essential when a priori there is no clear rationale for selecting target nucleotides for mutagenesis. Once a mutant is isolated that is defective for a particular function in vivo, the systematic in vitro synthesis and characterization of RNAs that analytically define the mutant phenotype can be undertaken.

However, there are two instances when such a genetic screen may not be necessary. First, there may be prior evidence (such as molecular phylogeny) that points to a particular region as important for a given function. For example, specific nucleotides in predicted helices can be tested explicitly. Second, the regions of functional importance in a large RNA molecule can sometimes be addressed by synthesis of a series of deletion mutants which are tested in an in vitro assay. This approach is particularly effective when domains can be identified, such as the two that make up the L-shaped tRNA structure. In these cases, transcripts can be made that encode a single domain. This approach offers the ability to study those mutants that might not be easily tested in vivo.

The role of nucleic acid conformation in sequence-specific recognition can be studied through the use of chemical synthesis. Ribo- and deoxyribonucleotides can be programmed in predetermined blocks in a sequence, resulting in the formation of mixed RNA–DNA molecules. These hybrid molecules could then be used as substrates for in vitro assays and might allow conclusions to be drawn about the role of minor groove interactions. The aminoacylation of tDNA^{Phe} suggests that not all synthetases require the A-form that is characteristic of RNA helices. As shown in the examples presented above, tRNA recognition nucleotides can be either paired or unpaired. In the case of tRNA^{Met} and tRNA^{Phe}, important bases for recognition are located in the anticodon. In these cases, the helical nature of the tRNA would not be predicted to play an important role in the presentation of the bases to the protein. Such tRNAs are candidates for studies of the

tDNA analogue of a tRNA. In $tRNA^{Ala}$, the G3·U70 base pair is located within a helical region of the tRNA, so that the corresponding DNA analogue might be inactive.

Other protein–tRNA systems not covered in this review can be addressed through the use of in vitro RNA synthesis. These include the CCA nucleotidyl transferase and other tRNA processing and modification enzymes,[155] elongation factor Tu, and initiation factors. Seong and Rajbhandary have used an in vivo system to show that elongator $tRNA^{Met}$ and initiator $tRNA^{Met}$ are distinguished by initiation factors on the basis of a single base pair mismatch at the 5′ end of the initiator tRNA.[52] This is an example of a well-defined system that can be further investigated by in vitro studies with synthetic transcripts.

The insights gained into the recognition of RNAs by protein through the use of in vitro RNA synthesis may be useful for the design of small molecules that bind to specific RNAs that are essential for cell viability. For example, a drug that binds selectively to the G3·U70 base pair of $tRNA^{Ala}$ could arrest protein synthesis. By taking advantage of sequence differences (around G3·U70) between the human $tRNA^{Ala}$ and that of a pathogenic organism, selective drug binding might be achieved.

The interaction between the RNA-dependent reverse transcriptase of retroviruses and the specific tRNA that acts as a primer for reverse transcriptase is another system that might be amenable to the design of drugs that bind to RNA. The annealing of the primer tRNA to the primer binding site is the first step in initiation of cDNA synthesis by reverse transcriptase and thus represents a potential target for the arrest of viral multiplication.[156] HIV reverse transcriptase and primer lysine tRNA form a complex that can be detected by glycerol gradient centrifugation, which may serve as an assay for testing inhibitors of the binding reaction.[157] Further, the bound tRNA also protects reverse transcriptase from thermal denaturation and trypsin digestion. The synthesis of RNA length and sequence variants could be used in this system to define determinants on the tRNA required for binding and serve as initial targets for drug design.

The interpretation of experiments on systems other than tRNAs, such as the M1 RNA of RNase P, is severely hampered by the lack of three-dimensional structural information. The tRNA structure was possible in part because of the availability of relatively large quantities of a specific tRNA, such as $tRNA^{Phe}$, from a convenient natural source (i.e., bakers' yeast). RNA synthesis has developed sufficiently that it can now make available large amounts of otherwise scarce RNA species, such that structural analysis of these molecules is now feasible. Thus, while the first experiments exploited the use of RNA synthesis to generate sequence variants which define determinants for recognition in a defined structure (i.e., transfer RNA), future applications will increasingly use synthesis as the means to generate the materials themselves that will be used for structure determinations. RNAs studied in this manner could include defined elements or domains of M1 RNA, ribosomal RNAs, and cellular and viral RNAs. However, investigators will still be faced with the problem of whether the synthetic materials accurately recapitulate their natural counterparts, where modified bases and folding patterns specific to the cellular environment have an influence on structure and function.

Registry No. RNase P, 71427-00-4; aminoacyl tRNA synthetase, 9028-02-8.

V. References

(1) Kim, S. H.; Suddath, F. L.; Quigley, G. J.; McPherson, A. A.; Sussman, J. L.; Wang, A. H.-J.; Seeman, N. C.; Rich, A. *Science* **1974**, *185*, 435–440.
(2) Robertus, J. D.; Ladner, J. E.; Finch, J. T.; Rhodes, D.; Brown, R. S.; Clark, B. F. C.; Klug, A. *Nature* **1974**, *250*, 546–551.
(3) Westhoff, E.; Dumas, P.; Moras, D. *J. Mol. Biol.* **1985**, *184*, 119.
(4) Schevitz, R. W., et al. *Nature* **1979**, *278*, 188–190.
(5) Woo, N. H.; Roe, B. A.; Rich, A. *Nature* **1980**, *286*, 346–351.
(6) Rould, M. A.; Perona, J. J.; Soll, D.; Steitz, T. A. *Science* **1989**, *246*, 1135–1141.
(7) Rich, A.; RajBhandary, U. L. *Annu. Rev. Biochem.* **1976**, *45*, 805–860.
(8) Schimmel, P. *Biochemistry* **1989**, *28*, 2747–2759.
(9) Schimmel, P. *Annu. Rev. Biochem.* **1987**, *56*, 125–158.
(10) Normanly, J.; Abelson, J. *Annu. Rev. Biochem.* **1989**, *58*, 1029–1049.
(11) Schulman, L. H.; Abelson, J. *Science* **1988**, *240*, 1591–1592.
(12) Yarus, M. *Cell* **1988**, *55*, 739–741.
(13) Schimmel, P. R.; Soll, D. *Annu. Rev. Biochem.* **1979**, *48*, 601–648.
(14) Cech, T. R. *Int. Rev. Cytol.* **1985**, *93*, 3–22.
(15) Zaug, A. J.; Cech, T. R. *Science* **1986**, *231*, 470–475.
(16) Latham, J. A.; Cech, T. R. *Science* **1989**, *245*, 276–281.
(17) Uhlenbeck, O. C. *Nature* **1987**, *328*, 596–600.
(18) Jeffries, A. C.; Symons, R. H. *Nucleic Acids Res.* **1989**, *17*, 1371–1377.
(19) Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. *Cell* **1984**, *37*, 415–427.
(20) Ruskin, B.; Krainer, A. R.; Maniatis, T.; Green, M. R. *Cell* **1984**, *38*, 317–331.
(21) Siebel, C. W.; Rio, D. C. *Science* **1990**, *248*, 1200–1208.
(22) Stern, S.; Powers, T.; Changchien, L.-M.; Noller, H. F. *Science* **1989**, *244*, 783–190.
(23) Burgin, A. B.; Parados, K.; Lane, D. J.; Pace, N. R. *Cell* **1990**, *60*, 405–414.
(24) Preiser, P. R.; Levinger, L. F. *J. Biol. Chem.* **1990**, submitted for publication.
(25) Cantor, C.; Schimmel, P. *Biophysical Chemistry, Part I: the conformation of biological macromolecules*; Freemen: San Francisco, 1980; pp 176–200.
(26) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–2183.
(27) Fuller, W.; Wilkins, M. F. H.; Hamilton, H. R.; Arnott, S. *J. Mol. Biol.* **1965**, *12*, 60–80.
(28) Arnott, S.; Wilkins, M. F. H.; Fuller, W.; Langridge, R. *J. Mol. Biol.* **1967**, *27*, 525–533.
(29) Rich, A.; RajBhandary, U. L. *Annu. Rev. Biochem.* **1976**, *45*, 805–860.
(30) Vlassov, V. V.; Giege, R.; Ebel, J. P. *Eur. J. Biochem.* **1981**, *119*, 51–59.
(31) Romby, P.; Carbon, P.; Westhoff, E.; Ehresmann, C.; Ebel, J.-P.; Ehresmann, B.; Giege, R. *J. Biomol. Struct. Dyn.* **1987**, *5*, 669–687.
(32) De Bruijn, M. H. L.; Klug, A. *EMBO J.* **1983**, *2*, 1309–1321.
(33) Dock-Bregeon, A. C.; Westhof, E.; Giege, R.; Moras, D. *J. Mol. Biol.* **1989**, *206*, 707–722.
(34) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1984**, *53*, 293–321.
(35) Ollis, D. L.; White, S. W. *Chem. Rev.* **1987**, *87*, 981–995.
(36) Jordan, S. R.; Pabo, C. O. *Science* **1988**, *242*, 893–898.
(37) Aggarwal, A. K.; Rodgers, D. W.; Drottar, M.; Ptashne, M.; Harrison, S. C. *Science* **1988**, *242*, 899.
(38) Otwinowski, Z.; Schevitz, R. W.; Zhang, R.-G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. *Nature* **1988**, *335*, 321–329.
(39) McClarin, J. A.; Frederick, C. A.; Wang, B.-C.; Greene, P.; Boyer, H. W.; Grable, J.; Rosenberg, J. M. *Science* **1986**, *234*, 1526–1541.
(40) Seeman, N. C.; Rosenberg, J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.
(41) Lewis, M.; Jeffry, A.; Wang, J.; Ladner, R.; Ptashne, M.; Pabo, C. O. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47*, 435–440.
(42) Wharton, R. P.; Ptashne, M. *Cell* **1984**, *38*, 361–369.
(43) Rich, A.; Schimmel, P. R. *Nucleic Acids Res.* **1977**, *4*, 1649–1665.

(44) Ladner, J. E.; Jack, A.; Robertus, J. P.; Brown, R. S.; Rhodes, D.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4414–4418.
(45) McClain, W. H.; Foss, K. *Science* **1988**, *241*, 1804–1807.
(46) Dickson, L. A.; Schimmel, P. R. *Arch. Biochem. Biophys.* **1975**, *167*, 638–645.
(47) Romby, P.; Moras, D.; Bergdoll, M.; Dumas, P.; Vlassov, V. V.; Westhoff, E.; Ebel, J. P.; Giege, R. *J. Mol. Biol.* **1985**, *184*, 455–471.
(48) Dock-Bregeon, A.-C.; Garcia, A.; Giege, R.; Moras, D. *Eur J. Biochem.* **1989**, 283–290.
(49) Theobold, A.; Springer, M.; Grunberg-Manago, M.; Ebel, J.-P.; Giege, R. *Eur. J. Biochem.* **1988**, *175*, 511–524.
(50) Park, S.-J.; Schimmel, P. *J. Biol. Chem.* **1988**, *263*, 16527–16530.
(51) Ruff, M.; Cavarelli, J.; Mikol, V.; Lorber, B.; Mitschler, A.; Giege, R.; Thierry, J. C.; Moras, D. *J. Mol. Biol.* **1988**, *201*, 235–236.
(52) Seong, B. L.; Lee, C.-P.; Rajbhandary, U. L. *J. Biol. Chem.* **1989**, *264*, 6504–6508.
(53) Perona, J. J.; Swanson, R. N.; Rould, M. A.; Steitz, T. A.; Soll, D. *Science* **1989**, *246*, 1152–1154.
(54) Deutscher, M. P. *J. Biol. Chem.* **1972**, *247*, 469–480.
(55) Deutscher, M. P. *J. Biol. Chem.* **1973**, *248*, 3116–3121.
(56) Tal, J.; Deutscher, M. P.; Littaver, O. Z. *Eur. J. Biochem.* **1972**, *28*, 478–491.
(57) Mirzabekov, A. D.; Lasfity, D.; Levina, E. S.; Bayen, A. A. *Nature, New Biol.* **1971**, *229*, 21–22.
(58) Thiebe, R.; Harbers, K.; Zachau, H. G. *Eur. J. Biochem.* **1972**, *26*, 144.
(59) Chambers, R. W. *Prog. Nucleic Acids Res. Mol. Biol.* **1971**, *11*, 489–523.
(60) England, T. E.; Uhlenbeck, O. C. *Biochemistry* **1978**, *17*, 2069–2076.
(61) Jin, Y.; Qiu, M.; Li, W.; Zeng, K.; Bao, J.; Gong, P.; Wu, R.; Wang, D. *Anal. Biochem.* **1987**, *161*, 453–459.
(62) Rosenberg, M.; Weissman, S.; DeCrombugghe, B. *J. Biol. Chem.* **1975**, *250*, 4755–4769.
(63) Patterson, B.; Rosenberg, M. *Nature* **1979**, *279*, 692–696.
(64) Manly, J. L.; Fire, A.; Cano, A.; Sharp, P. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3855–3859.
(65) Weil, P. A.; Luse, D. S.; Segall, J.; Roeder, R. G. *Cell* **1979**, *18*, 469–484.
(66) Butler, E. T.; Chamberlin, M. J. *J. Biol. Chem.* **1982**, *257*, 5772–5778.
(67) Melton, D. A.; Kreig, P. A.; Rebagliati, M. R.; Maniatis, T.; Zinn, K.; Green, M. R. *Nucleic Acids Res.* **1984**, *12*, 7035–7056.
(68) McAllister, W. T.; Morris, C.; Rosenberg, A. H.; Studier, F. W. *J. Mol. Biol.* **1981**, *153*, 527–544.
(69) Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 2035–2039.
(70) Rosenberg, A. H.; Lade, B. N.; Chui, D.-S.; Lin, S.-W.; Dunn, J. J.; Studier, F. W. *Gene* **1987**, *56*, 125–135.
(71) Milligan, J. F.; Groebe, D. R.; Witherall, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
(72) Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, *180*, 51–62.
(73) Martin, C. T.; Muller, D. K.; Coleman, J. E. *Biochemistry* **1988**, *27*, 3966–3974.
(74) Cedergren, R.; Grosjean, H. *Biochem. Cell. Biol.* **1987**, *65*, 677–692.
(75) Caruthers, M. H.; Dellinger, D.; Prosser, K.; Barone, A. D.; Dubendorf, J. W.; Kierzek, P.; Rusendahl, M. *Chem. Scr.* **1986**, *26*, 25–30.
(76) Usman, N.; Ogilvie, K. K.; Jian, M.-Y.; Cedergren, R. J. *J. Am. Chem. Soc.* **1987**, *109*, 7845–7854.
(77) Chou, S.-H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2422–2435.
(78) Ogilvie, K. K.; Usman, N.; Nicoghosian, K.; Cedergren, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5764–5768.
(79) Blow, D. M.; Brick, P. *Biological Macromolecules and Assemblies*; Jurnak, F., MacPherson, A., II, Eds.; Wiley: New York, 1985; pp 441–469.
(80) Brunie, S.; Mellot, P.; Zelwar, C.; Risler, J. L.; Blanquet, S.; Fayat, G. *J. Mol. Graphics* **1987**, *5*, 18–21.
(81) Burbaum, J. J.; Starzyk, R. M.; Schimmel, P. *Proteins* **1990**, *7*, 99–111.
(82) Ozeki, H.; Inokuchi, H.; Yamao, F.; Kodaira, M.; Sakano, H.; Ikemura, T.; Shimura, Y. In *Transfer RNA: Biological Aspects*; Soll, D., Abelson, J., Schimmel, P., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1980; pp 341–362.
(83) Normanly, J.; Masson, J.-M.; Kleina, L.; Abelson, J.; Miller, J. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6548–6552.
(84) Normanly, J.; Ogden, R. C.; Horvath, S. J.; Abelson, J. *Nature* **1986**, *321*, 213–219.
(85) Baccanari, D. P.; Averett, D.; Briggs, C.; Burchall, J. *Biochemistry* **1977**, *16*, 3566–3572.
(86) McClain, W. H.; Foss, K. *J. Mol. Biol.* **1988**, *202*, 697–709.
(87) Yarus, M. *Nature, New Biol.* **1972**, *239*, 106–108.
(88) Swanson, R.; Hoben, P.; Sumner-Smith, M.; Uemura, H.; Watson, L.; Soll, D. *Science* **1988**, *242*, 1548–1551.
(89) Hou, Y.-M.; Schimmel, P. *Biochemistry* **1989**, *28*, 4942–4947.
(90) McClain, W. H.; Chen, Y.-M.; Foss, K.; Schneider, J. *Science* **1988**, *242*, 1682–1684.
(91) Schulman, L. H.; Pelka, H. *Biochemistry* **1977**, *16*, 4256–4265.
(92) Schulman, L. H. In *Transfer RNA: Structure Properties and Recognition*; Schimmel, P., Soll, D., Abelson, J. N., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1979; pp 311–324.
(93) Schulman, L. H.; Pelka, H. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6755–6759.
(94) Schulman, L. H.; Pelka, H.; Susani, M. *Nucleic Acids Res.* **1983**, *11*, 1439–1455.
(95) Uemura, H.; Imai, M.; Ohtsuka, E.; Ikehara, M.; Soll, D. *Nucleic Acids. Res.* **1982**, *10*, 6531–6539.
(96) Schulman, L. H.; Pelka, H. *Science* **1988**, *242*, 765–768.
(97) Leon, O.; Schulman, L. H. *Biochemistry* **1987**, *26*, 5416–5422.
(98) Ghosh, G.; Pelka, H.; Schulman, L. H. *Biochemistry* **1990**, *29*, 2220–2225.
(99) Grosjean, H.; DeHenau, S.; Doi, T.; Yamane, A.; Ohtsuka, E.; Ikehara, M.; Beauchemin, N.; Nicoghosian, K.; Cedergren, R. *Eur. J. Biochem.* **1987**, *166*, 325–332.
(100) Schulman, L. H.; Pelka, H. *Biochemistry* **1985**, *24*, 7309–7314.
(101) Schulman, L. H.; Pelka, H. *Science* **1989**, *246*, 1595–1597.
(102) McClain, W. H.; Foss, K. *Science* **1988**, *241*, 1804–1807.
(103) Chattapadhyay, R.; Pelka, H.; Schulman, L. H. *Biochemistry* **1990**, *29*, 4263–4268.
(104) Bruce, A. G.; Uhlenbeck, O. C. *Biochemistry* **1982**, *21*, 855–861.
(105) Bruce, A. G.; Uhlenbeck, O. C. *Biochemistry* **1982**, *21*, 3921–3926.
(106) Bare, L.; Uhlenbeck, O. C. *Biochemistry* **1985**, *24*, 2354–2360.
(107) Sampson, J. R.; Uhlenbeck, O. C. *Prog. Natl. Acad. Sci. U. S.A.* **1988**, *85*, 1033–1037.
(108) Sampson, J. K.; DiRenzo, A.; Behlen, L.; Uhlenbeck, O. C. *Science* **1989**, *243*, 1363–1366.
(109) Samuelsson, T.; Boren, T.; Johansen, T.-I.; Lustg, F. *J. Biol. Chem.* **1988**, *263*, 13692–13699.
(110) Hall, K.; Sampson, J. R.; Uhlenbeck, O. C.; Redfield, A. G. *Biochemistry* **1989**, *28*, 5794–5801.
(111) Behlen, L. S.; Sampson, J. R.; DiRenzo, A. B.; Uhlenbeck, O. C. *Biochemistry* **1990**, *29*, 2515–2523.
(112) Sampson, J. R.; DiRenzo, A. B.; Uhlenbeck, O. C. *Biochemistry* **1990**, *29*, 2523–2532.
(113) McClain, W. H.; Foss, K. *Science* **1988**, *241*, 1804–1807.
(114) Khan, A. S.; Roe, B. A. *Science* **1988**, *241*, 74–79.
(115) Hou, Y.-M.; Schimmel, P. *Nature* **1988**, *333*, 140–145.
(116) McClain, W. H.; Foss, K. *Science* **1988**, *240*, 793–796.
(117) Hou, Y.-M.; Schimmel, P. *Biochemistry* **1989**, *28*, 6800–6804.
(118) Park, S.-J.; Hou, Y.-M.; Schimmel, P. *Biochemistry* **1989**, *28*, 2740–2746.
(119) Francklyn, C.; Schimmel, P. *Nature* **1989**, *337*, 478–481.
(120) Shi, J.-P.; Francklyn, C.; Hill, K.; Schimmel, P. *Biochemistry* **1990**, *29*, 3621–3626.
(121) Himeno, H.; Hasegawa, T.; Ueda, T.; Watanbe, K.; Miura, K.; Shimizu, M. *Nucleic Acids Res.* **1989**, *17*, 7855–7863.
(122) Muramatsu, T.; Nishikawa, K.; Nemoto, F.; Kuchino, Y.; Nishimura, S.; Miyazawa, T.; Yokoyama, S. *Nature* **1988**, *336*, 179–181.
(123) Muramatsu, T.; Yokoyama, S.; Horie, N.; Matsuda, A.; Ueda, T.; Yamaizumi, Z.; Kuchino, Y.; Nishimura, S.; Miyazawa, T. *J. Biol. Chem.* **1988**, *263*, 9261–9267.
(124) Schimmel, P. R.; Uhlenbeck, O. C.; Lewis, J. B.; Dickson, L. A.; Eldred, E. W.; Schreier, A. A. *Biochemistry* **1972**, *11*, 642–646.
(125) Perret, V.; Garcia, A.; Ebel, J.-P.; Florentz, C.; Giege, R. *Nature* **1990**, *344*, 787–789.
(126) Dumas, P.; Moras, D.; Florentz, C.; Giege, R.; Verlaan, P.; Belkun, A. V.; Pleij, C. W. A. *J. Biomol. Struct. Dyn.* **1987**, *4*, 707–728.
(127) van Belkum, A.; Verlan, P.; Kun, J.-B.; Pleij, C.; Bosch, L. *Nucleic Acids Res.* **1988**, *16*, 1931–1950.
(128) Pleij, C. W. A.; Reitvald, K.; Bosch, L. *Nucleic Acids Res.* **1985**, *13*, 1717–1731.
(129) Schimmel, P. *Cell* **1989**, *58*, 9–12.
(130) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. *Nature* **1988**, *331*, 283–286.
(131) Dreher, T. W.; Bujarski, J. J.; Hall, T. C. *Nature* **1984**, *311*, 171–175.
(132) Weiner, A. M.; Mazels, N. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7383–7387.
(133) Dreher, T. W.; Florentz, C.; Giege, R. *Biochimie* **1988**, *70*, 1719–1727.
(134) Rietvald, K.; van Poelgeest, R.; Pleij, C. W. A.; van Boom, J. H.; Bosch, L. *Nucleic Acids Res.* **1982**, *10*, 1929.
(135) Springer, M.; Plumbridge, J. A.; Butler, J. S.; Graffe, M.; Dondon, J.; Mayaux, J. F.; Fayat, G.; Lestienne, P.; Blanquet,

S.; Grunberg-Manago, M. *J. Mol. Biol.* **1985**, *185*, 93–104.
(136) Moine, H.; Romby, P.; Springer, M.; Grunberg-Manago, M.; Ebel, J. P.; Ehresmann, C.; Ehresmann, B. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7892–7896.
(137) Springer, M.; Graffe, M.; Dondon, J.; Grunberg-Manago, M. *EMBO J.* **1989**, *8*, 2417–2424.
(138) Altman, S.; Baer, M.; Guerrier-Takada, C.; Vioque, A. *Trends Biochem. Sci.* **1986**, *11*, 515–518.
(139) Smith, J. D. *Brookhaven Symp. Biol.* **1975**, *26*, 1–11.
(140) McClain, W. H.; Seidman, J. G. *Nature (London)* **1975**, *257*, 106.
(141) Kole, R.; Altman, S. *Biochemistry* **1981**, *20*, 1902–1906.
(142) Reed, R. E.; Baer, M. F.; Guerrier-Takada, C.; Donis-Keller, H.; Altman, S. *Cell* **1982**, *30*, 627–630.
(143) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849–857.
(144) Vioque, A.; Altman, S. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5904–5908.
(145) Vioque, A.; Arnez, J.; Altman, S. *J. Mol. Biol.* **1988**, *202*, 835–848.
(146) Waugh, D. S.; Gree, C. J.; Pace, N. R. *Science* **1989**, *244*, 1569–1571.
(147) Guerrier-Takada, C.; Lumelsky, N.; Altman, S. *Science* **1989**, *246*, 1578–1584.
(148) Guerrier-Takada, C.; McClain, W. H.; Altman, S. *Cell* **1984**, *38*, 219–224.
(149) McClain, W. H.; Guerrier-Takada, C.; Altman, S. *Science* **1987**, *238*, 527–530.
(150) Samuelsson, T.; Olsson, M. *Abstracts from the 13th International tRNA Workshop*; 1989.
(151) Grosjean, H.; Trewyn, R. W.; Giege, R. *Abstracts from the 13th International tRNA Workshop*; 1989.
(152) Grosjean, H.; Kubli, E. In *Microinjection and Organelle Transplantation Techniques*; Celis, J. E., Graessman, A., Loyter, A., Eds.; Academic: New York, 1986; pp 304–326.
(153) Baer, M. F.; Reilly, R. M.; McCorkle, G. M.; Hai, T.-Y.; Altman, S.; RajBhandary, U. L. *J. Biol. Chem.* **1988**, *263*, 2394–2351.
(154) Konarska, M. M.; Sharp, P. A. *Cell* **1989**, *57*, 423–431.
(155) Deutscher, M. P. *Crit. Rev. Biochem.* **1984**, *17*, 45–71.
(156) Varmus, H. E. *Science* **1982**, *216*, 812–820.
(157) Sallafranque-Andreola, M. L.; Robert, D.; Barr, P. J.; Fournier, M. J.; Litvak, S.; Sarih-Cottin, L.; Tarrago-Littvak, L. *Eur. J. Biochem.* **1989**, *184*, 367–384.