The EMBO Journal vol.12 no.13 pp.5201 – 5208, 1993

Selection of a 'minimal' glutaminyl-tRNA synthetase and the evolution of class I synthetases

Etienne Schwob[1] and Dieter Söll[2]

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA
[1]Present address: Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria
[2]Corresponding author

Communicated by D.Moras

The evolution of the aminoacyl-tRNA synthetases is intriguing in light of their elaborate relationship with tRNAs and their significance in the decoding process. Based on sequence motifs and structure determination, these enzymes have been assigned to two classes. The crystal structure of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS), a class I enzyme, complexed to $tRNA^{Gln}$ and ATP has been described. It is shown here that a 'minimal' GlnRS, i.e. a GlnRS from which domains interacting with the acceptor-end and the anticodon of the tRNA have been deleted, has enzymatic activity and can charge a $tRNA^{Tyr}$-derived amber suppressor (*supF*) with glutamine. The catalytic core of GlnRS, which is structurally conserved in other class I synthetases, is therefore sufficient for the aminoacylation of tRNA substrates. Some of these truncated enzymes have lost their ability to discriminate against non-cognate tRNAs, implying a more specific role of the acceptor-end-binding domain in the recognition of tRNAs. Our results indicate that the catalytic and substrate recognition properties are carried by distinct domains of GlnRS, and support the notion that class I aminoacyl-tRNA synthetases evolved from a common ancestor, jointly with tRNAs and the genetic code, by the addition of non-catalytic domains conferring new recognition specificities.
Key words: E.coli/genetic code/$lacZ_{1000}$/nonsense suppression/protein synthesis

Introduction

A fundamental question in biology is how enzymes recognize their substrates and how specific recognition of similar substrates can be achieved. A related question deals with the evolution of proteins during the billion years separating simple archaic systems from the complex present-day organisms. Aminoacyl-tRNA synthetases are among the oldest proteins, arising early in evolution to establish the rules of the genetic code through their interactions with amino acids and tRNAs (Rich, 1962; Schimmel, 1991). They catalyze the specific attachment of an amino acid to the terminal adenosine of its cognate tRNA. Usually 20 such enzymes, one for each amino acid, constitute the minimal set required for protein biosynthesis. The accuracy of this ATP-dependent reaction is largely responsible for the high fidelity of protein biosynthesis (Schimmel, 1987; Schulman, 1991). Aminoacyl-tRNA synthetases are ubiquitous; their conservation and the universality of the genetic code suggest that they have evolved long before the divergence of pro- and eukaryotes. Their ancientness and the subtlety of their recognition properties create an ideal system for studying the evolution and specificity of proteins.

The genes encoding the 20 aminoacyl-tRNA synthetases from *Escherichia coli*, as well as many others from various sources, have been sequenced. In spite of their common biological role, these enzymes display a large variability in size and oligomeric structure (Schimmel, 1987; Burbaum and Schimmel, 1991). Sequence comparisons indicated that these synthetases can be subdivided into two classes of 10 enzymes (Cusack *et al.*, 1990; Eriani *et al.*, 1990). The determination of the crystal structure of three class I enzymes (Brick *et al.*, 1989; Rould *et al.*, 1989; Brunie *et al.*, 1990) and two class II enzymes (Cusack *et al.*, 1990; Ruff *et al.*, 1991) supports this conclusion by revealing that each class contains a distinct nucleotide binding fold (Moras, 1992). The structure of *E.coli* glutaminyl-tRNA synthetase (GlnRS), a monomer of 553 amino acids (M_r 63 400), has been determined complexed with $tRNA^{Gln}$ and ATP (Rould *et al.*, 1989, 1991). The protein is an elongated molecule of four domains, one domain containing the active site while the three other domains have roles in specific recognition of $tRNA^{Gln}$ and discrimination against non-cognate tRNAs (Rould *et al.*, 1989, 1991; Weygand-Durasevic *et al.*, 1993). The three-dimensional structure of the active site domain of GlnRS, MetRS and TyrRS is conserved (Perona *et al.*, 1991) and suggests that all class I synthetases may have a common ancestor. Here we provide experimental evidence for this hypothesis by showing that the active site domain of GlnRS alone can aminoacylate tRNA substrates. This 'minimal' synthetase could be a prototype of the early enzymes, which later evolved and diverged to recognize specific tRNAs.

Results

Experimental design

The active site domain of GlnRS contains the signature sequence (HIGH, MSK; Starzyk *et al.*, 1987) and displays extensive homologies with other class I synthetases (Breton *et al.*, 1990; Perona *et al.*, 1991). This domain binds the three substrates of GlnRS (ATP, glutamine and tRNA) and therefore may resemble the primordial synthetase from which class I synthetases evolved. To test this hypothesis, we set up domain-deletion experiments based on our knowledge of the GlnRS structure (Rould *et al.*, 1989). As such deletions may inactivate the enzyme by destruction of its catalytically active conformation, we devised a genetic scheme for selecting active truncated forms of GlnRS based on the suppression of the $lacZ_{1000}$ amber mutation (Inokuchi *et al.*, 1984). The main advantage of this system comes from the fact that cells do not rely on the mutated GlnRS for protein

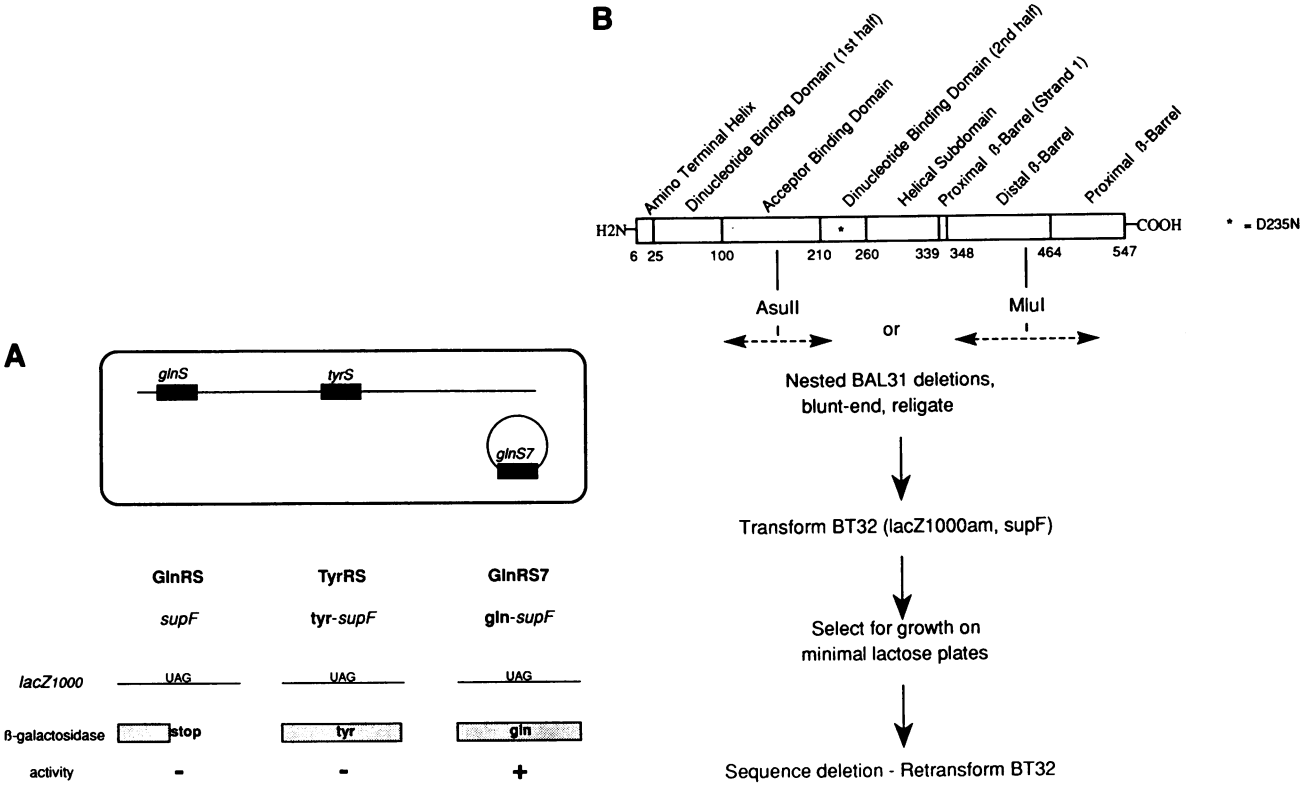

Fig. 1. Selection of truncated GlnRS. (**A**) Suppression of the $lacZ_{1000}$ mutation by mischarging of *supF* by GlnRS7. The wild-type GlnRS encoded by the *glnS* gene, together with the other aminoacyl-tRNA synthetases, ensures normal protein biosynthesis. Wild-type GlnRS cannot misacylate *supF*; an inactive truncated β-galactosidase is produced (left). When tyrosine is inserted at the site of the mutation, a full-length, but inactive, β-galactosidase is produced (middle). The GlnRS7 enzyme can mischarge *supF* with glutamine and generate an active β-galactosidase (right). The activity of truncated GlnRS7 enzymes is therefore directly scorable by suppression of the $lacZ_{1000}$ mutation. (**B**) Experimental design. Tertiary structure domains of GlnRS are indicated along its primary structure. Sites for bidirectional deletions are noted (*Asu*II and *Mlu*I).

Table I. *In vivo* selected clones

Clone	Deletion (residues)	Sequence of junction
Q20A1	109–201 (−1)	CTC CAC GCC-------T GGC AAC AAG L H A108 G202 N K
Q20A2	114–199 (−1)	ATC GAA CTG-------C CAG ACT GGC I E L113 Q200 T G
Q20A6	115–215	GAA CTG ATC A-------GC ATC AGC E L I114 [S]216 I S
Q20M1	381–460	CAG GTG CC-------C GCG GCA CAT Q V P380 A461 A H
Q20M2	369–498 (−1)	CCG AAC CAT C-------AAA CAG GGC P N H368 K499 Q G
Q20M5	359–493 (+1)	TAT CAG GGC G-------G GAA TCG CTG Y Q G358 E494 S L
Q20M6	376–467 (−1)	GAA ATG GGC-------T GAA ATC CGT E M G375 E468 I R
Q20M7	367–490 (+1)	GTT ACC ATG CC-------ATT AAC CCG V T M366 I491 N P
Q20M11	389–461 (+1)	TGG ATT GAT CG-------GCA CAT GCG W I D388 A462 H A
Q20M13	394–465	GAT TTC CG-------G CCG GTT GAA D F R393 P466 V E

The clones giving rise to the smallest GlnRS enzymes still able to mischarge the *supF* $tRNA^{Tyr}$ and suppress the $lacZ_{1000}$ mutation are indicated using the following nomenclature: Q20A for *Asu*II (acceptor-binding domain) deletions, Q20M for *Mlu*I (distal β-barrel domain) deletions. The second column indicates the residues which are deleted (inclusive) and the type of frameshift generated, if any, in parentheses. The nucleotide sequence at the borders of the deletion is indicated

synthesis and survival, therefore allowing the selection of poorly active or otherwise lethal mutants. While previous studies have tested the ability of truncated synthetases to complement null strains or to carry out aminoacylation *in vitro* (Starzyk *et al.*, 1987; Toth and Schimmel, 1990), our procedure detects, in the presence of wild-type GlnRS, the activity of an episomal mutated *glnS7* gene by a specific and sensitive reporter assay (Figure 1A). The *glnS7* gene encodes the mutant GlnRS7 enzyme (D235N) that mischarges *supF* $tRNA^{Tyr}$ with glutamine, whereas the wild-type GlnRS does not (Inokuchi *et al.*, 1984; Uemura *et al.*, 1988). Active β-galactosidase is made only when *supF* is misacylated with glutamine by GlnRS7, and the activity of truncated GlnRS7 enzymes can be directly scored by suppression of the $lacZ_{1000}$ mutation and growth on minimal lactose plates.

Mutagenesis and characterization of the mutants

A set of internal deletions was generated from two sites in the *glnS7* gene that were chosen to gradually remove the tRNA acceptor and anticodon-binding domains of GlnRS. Crystallographic study of the GlnRS/$tRNA^{Gln}$ complex suggested that these domains (acceptor-end-binding domain and proximal + distal β-barrel domains; see Figure 1B) are most likely implicated in tRNA recognition, but not catalysis (Rould *et al.*, 1989). After transformation of the random pool of *glnS7* deletion mutants into a $lacZ_{1000}$ *supF* strain, only mutants giving rise to GlnRS enzymes still able to form Gln-$tRNA^{Tyr}$, and thus active β-galactosidase, will grow on minimal lactose plates. Plasmid DNA pools made from various time points of Bal31 exonuclease treatment were transformed into strain BT32, although only the clones obtained from the longest time points were analyzed (Table I; Figure 2A). These clones may correspond to the smallest active GlnRS enzymes. Only ~0.1–5% of all transformants displayed prototrophic growth on minimal lactose medium after 2 days incubation at 30°C. This indicates a low probability of obtaining catalytically active mutants by random deletion, but a rather efficient suppression of the

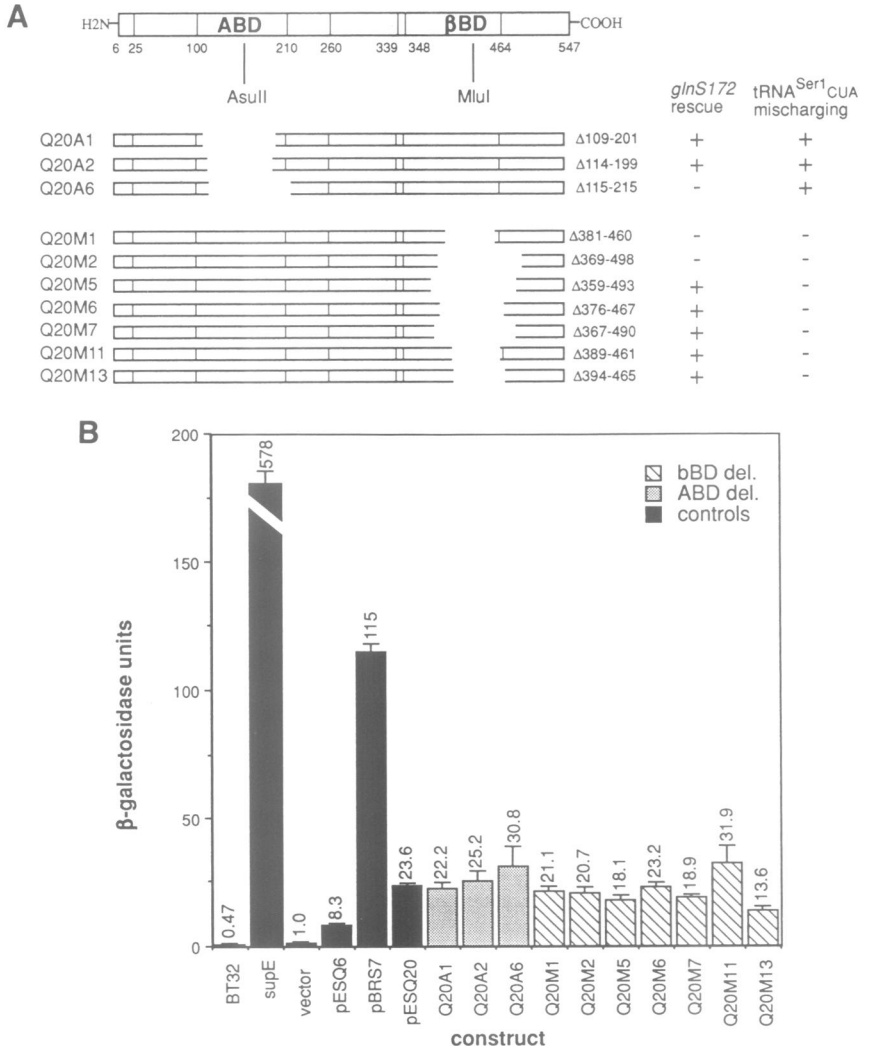

Fig. 2. Characterization of the truncated GlnRS enzymes. (**A**) Deletion range and tRNA recognition specificity of the mutants. Deleted regions are open and missing residues indicated to the right. Charging of $tRNA^{Gln}$ was assessed by the rescue of the thermosensitive *glnS* strain UT172 at non-permissive temperature. Mischarging of other non-cognate tRNAs was tested by the suppression of a $lacZ_{1000}$ strain (BT235) carrying the $tRNA^{Ser}$-derived suppressor on a compatible plasmid (pACYC184). (**B**) Activity of the truncated GlnRS enzymes. The misacylation activity of the mutant enzymes was estimated by the β-galactosidase activity produced upon suppression of the $lacZ_{1000}$ amber mutation by the glutamine-charged *supF* $tRNA^{Tyr}$ in strain BT32. *supE*= BT235 containing *supE* $tRNA^{Gln}$; pESQ6 = wild-type GlnRS in pBluescript; pBRS7 = *glnS7* in pBR322. The mean and standard deviation (error bar) of triplicate experiments are indicated.

lacZ$_{1000}$ amber mutation (Inokuchi *et al.*, 1984). The range of the deletions was determined by sequencing. The phenotype was confirmed to be plasmid dependent by retransformation and sequencing. Although the mutants were stable, reduced plating efficiency (10–50% of prototrophic transformants) seen upon retransformation suggests that these constructs may encode toxic products (Vidal-Cros and Bedouelle, 1992).

More mutants could be obtained for the β-barrel domain (βBD) than for the acceptor-binding domain (ABD), suggesting either a larger dispensable fragment in the C-terminal half of the molecule or a lower structural requirement for refolding into an active conformation. Both are probably true: the β-barrel domain is twice as big as the acceptor-binding domain (220 versus 110 residues) and the ABD is inserted between the two halves of the nucleotide-binding fold. Each of the three ABD deletion mutants (Q20A1, Q20A2 and Q20A6) lacks almost the entire ABD as described by Rould *et al.* (1989), but leaves intact the dinucleotide-binding fold (Figure 2A). Strikingly, these deletions are quite tightly clustered: all three interrupt the polypeptide chain in helix D, the first element of the ABD, and restore it between β-strands 7 and 8 (Q20A1, Q20A2) or at the beginning of helix G (Q20A6), the first element of the second half of the dinucleotide binding fold (see Figure 5). For the βBD mutants, the deletion range varies from 72 amino acids in the distal β-barrel (Q20M13) to 135 residues encompassing most of the proximal and distal β-barrel domains (Q20M5). As for the ABD mutants, the deleted regions coincide with substructures defined by X-ray crystallography (Rould *et al.*, 1989, 1991). For example, the Q20M1 deletion removes amino acids 381–460 (β-strands 16–19); since these two residues are only 9.8 Å apart, it is conceivable that they would be easy to connect without disturbing the entire structure. Similarly, the Q20M13 deletion (from residues 394 to 465) removes the segment from the end of β-strand 16′ to the beginning of strand 20 (Rould *et al.*, 1991). In both cases, the distal β-barrel domain containing the binding pockets for the anticodon bases C34 and G36 of tRNA^{Gln} is entirely lost, most likely leading to the repacking, around U35, of the exposed hydrophobic residues left. Since Arg341, Gln517 and Arg520 are not deleted, it is possible that some recognition of U35 is maintained. It is, however, not clear whether the splayed out conformation of the three anticodon bases seen in the GlnRS/tRNA^{Gln} co-crystal would persist in this artificial situation.

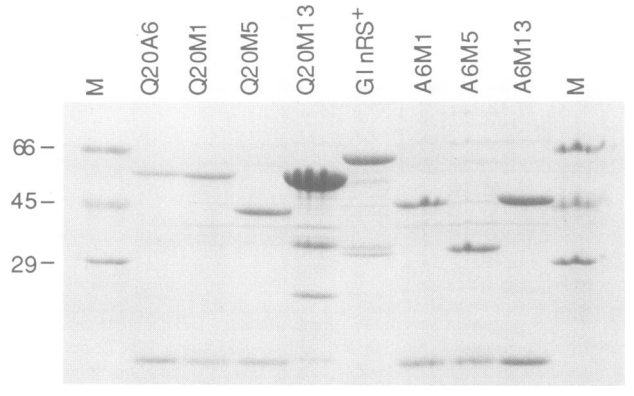

Fig. 3. Expression of the truncated GlnRS. Coomassie-blue stained SDS–PAGE of wild-type and truncated GlnRS molecules overexpressed in the T7 RNA polymerase system. Molecular weight standards and clones are indicated.

Surprisingly, most (7 out of 10) of the *in vivo* selected clones have deletions that produce frameshift mutations (Table I). In such cases, it is expected that translation would terminate prematurely at the next in-frame stop codon. Polypeptides generated by premature termination of the Q20A1 and Q20A2 mutants are, however, unlikely to possess enzymatic activity because they would lack the second half of the nucleotide binding fold. Examination of the sequences between the deletion site and the following termination codon revealed, in the case of the tightly clustered Q20A1 and Q20A2 deletions, sequences favorable for +1 frameshift suppression (Atkins *et al.*, 1991). The occurrence of such a phenomenon was observed previously for MetRS (L.H.Schulman, personal communication) and confirmed by Western analysis which revealed a polypeptide of the expected size at an ~10-fold lower level than the one produced by wild-type GlnRS (data not shown). The tight clustering of both mutants and the fact that the Q20A1 deletion was independently selected twice strongly argues in favor of a positive selection of these mutants. No evidence for frameshift suppression could be found for the +1 or −1 βBD mutants. Since the estimated (on SDS–PAGE) molecular weights of these mutants are in good agreement with those predicted for carboxy-terminal truncations, we propose that these mutant proteins, generated from internal gene deletions, are actually C-terminal truncations of GlnRS (e.g. Q20M5, see Figure 3).

Activity of the truncated GlnRS mutants

To have an idea on how active the truncated GlnRS might be, we concentrated on their mischarging *in vivo* by assaying the β-galactosidase activity of the *lacZ*$_{1000}$ gene suppressed by mischarging (Inokuchi *et al.*, 1984; Rogers and Söll, 1988). Wild-type GlnRS encoded by the chromosomal *glnS* gene does not mischarge *supF* to a detectable level (Figure 2B; Swanson *et al.*, 1988). Thus, in strain BT32 transformed with the mutated *glnS7* alleles, the sole source of β-galactosidase is the activity of the plasmid-encoded truncated GlnRS7 enzymes. The level of β-galactosidase activity measured in these strains (20–30 Miller units; Figure 2B) correlates with their growth rate on minimal lactose plates (2 days at 30°C). This level is similar to that produced by the parental pESQ20 plasmid (encoding the full-length GlnRS7 enzyme), although we cannot conclude that the truncated enzymes are as active as their full-length parent. Hence, β-galactosidase activity obtained by mischarging suppression is saturable and also depends on GlnRS's mischarging potential, as seen when the GlnRS7 enzyme is expressed to moderate or high levels (compare pBRS7 and pESQ20 in Figure 2B). The relationship between β-galactosidase activity and the synthetase level (or mischarging potential) is not linear (E.Schwob, I.Weygand-Durasevic and D.Söll, manuscript in preparation), so the activity of the mutants cannot be deduced from their relative β-galactosidase levels. Presumably, too much mischarging leads to translational errors and general reduction of enzymatic activity. Nevertheless, the β-galactosidase levels measured in the strains transformed with the various deleted *glnS* clones are reproducible and demonstrate biochemically the activity of the truncated GlnRS enzymes.

Recognition of cognate tRNA^{Gln}

In order to determine whether the truncated GlnRS enzymes still recognize and charge efficiently their cognate tRNA^{Gln}, we tested their ability to rescue a strain that is thermosensitive for GlnRS (strain UT172; Englisch-Peters *et al.*, 1991). Figure 2A shows that most of the mutants were able to substitute for the defective GlnRS and sustain growth of strain UT172 at non-permissive temperature, indicating that they have retained some specificity towards tRNA^{Gln}. It is not clear to us why clones Q20A6, Q20M1 and Q20M2 do not complement strain UT172; either the corresponding enzymes do not recognize tRNA^{Gln} any longer, or their enzymatic activity is too weak to sustain growth as sole GlnRS source. Alternatively they may encode toxic misacylating enzymes that would become lethal at high temperature and in the absence of wild-type GlnRS, as is the case for the parental pESQ20 plasmid (data not shown).

Mischarging of tRNA^{Ser1}

In vivo misacylation of tRNA by a synthetase depends on two factors: the affinity of the synthetase for the non-cognate tRNA and its ability to compete with the cognate synthetase, the latter being directly proportional to their relative concentrations (Rogers and Söll, 1988; Swanson *et al.*, 1988). Compared with *supF*, the amber suppressor derived from tRNA^{Ser1} is a poor substrate for *in vivo* misacylation by GlnRS; it is not mischarged by high levels of wild-type GlnRS (Weygand-Durasevic *et al.*, 1993) or by normal level of the mischarging GlnRS7 mutant (Rogers and Söll, 1988). This negative recognition, carried out by the acceptor-binding domain of GlnRS, can however be overcome by over-expressing GlnRS7 *in vivo* (Swanson *et al.*, 1988), by modifying the acceptor stem of tRNA^{Ser} (Rogers and Söll, 1988) or the acceptor-binding domain of GlnRS (Weygand-Durasevic *et al.*, 1993). When we tested if the truncated GlnRS7 mutants could also mischarge the tRNA^{Ser}-derived suppressor, a clear pattern was observed. The GlnRS mutants in which the acceptor-binding domain was deleted, but had an intact anticodon-binding domain, could all mischarge the tRNA^{Ser}-derived amber suppressor, whereas all the mutants deleted in their anticodon-binding domain had lost this ability. This indicates either that the anticodon-binding domain of GlnRS is required for mischarging of tRNA^{Ser}, or that the removal of the acceptor-binding domain of GlnRS relaxes its tRNA specificity. For the former, the accurate recognition of the tRNA^{Ser} CUA anticodon by the mutant GlnRS would provide an anchor point for positioning the non-cognate tRNA in the active site of GlnRS. The second hypothesis proposes that the acceptor-binding domain of GlnRS is involved in discriminating against non-cognate tRNAs, and its removal would permit productive interactions with non-cognate tRNAs. There is good evidence for this second hypothesis since both the structural (Rould *et al.*, 1989; Perona *et al.*, 1991) and mutational analyses (Weygand-Durasevic *et al.*, 1993) of the acceptor-binding domain of GlnRS are consistent with a role of this domain in the negative recognition of non-cognate tRNAs. In this model, tRNA^{Gln} adopts a sequence-specific conformation compatible with the structure of the acceptor-binding domain, and its removal (this study) or structural alteration (Weygand-Durasevic *et al.*, 1993) would allow non-cognate tRNAs to enter into productive interactions with GlnRS. This is supported by acceptor stem mutants derived from the tRNA^{Ser} suppressor charged with glutamine (Rogers and Söll, 1988).

To see whether removal of the acceptor-binding domain relaxes the tRNA specificity of wild-type GlnRS, we performed deletions from the *Asu*II site of the pESQ6 plasmid and tested the ability of the truncated GlnRS to now recognize the tRNA^{Ser1}-derived suppressor in the *lacZ*$_{1000}$ assay. When these clones were transformed in strain BT235 containing the tRNA^{Ser1} suppressor, they allowed the formation of microcolonies on lactose minimal plates, whereas the full-length wild-type GlnRS did not (data not shown). Although the level of suppression did not allow sustained growth, this result suggests that removing GlnRS's acceptor-binding domain may suffice for relaxing its tRNA recognition. As proposed by previous studies (Rould *et al.*, 1989; Weygand-Durasevic *et al.*, 1993), we conclude that the acceptor-binding domain of GlnRS is not involved in catalysis, but rather in the accurate selection of the cognate tRNAs.

Expression of the truncated GlnRS mutants

In order to determine the kinetic parameters of the mutant GlnRS proteins, we wanted to overexpress and purify the mutant proteins away from the wild-type enzyme. The λ*glnS* genes from clones Q20A6, Q20M1, Q20M5 and Q20M13 were cloned and expressed to high levels in the T7 RNA polymerase system (Studier *et al.*, 1990); polypeptides of the expected sizes (52, 54, 42 and 55 kDa, respectively) were prominent in crude extracts analyzed by SDS−PAGE (Figure 3). Western analysis confirmed that these polypeptides are indeed derived from GlnRS (data not shown). However, this very high level of expression resulted in the production of insoluble aggregates which precluded the purification and kinetic analysis of the mutant proteins. Unfortunately, at this point, we cannot provide *in vitro* kinetic data on the catalytic parameters of these GlnRS mutants.

'Minimal' GlnRS mutants

It is assumed that in evolution simple RNA molecules were aminoacylated by a few primordial synthetases (Rich, 1962; Schimmel, 1991; Moras, 1992). Strong evidence for such a system comes from the fact that AlaRS, HisRS and GlyRS can aminoacylate RNA microhelices that correspond to the tRNA acceptor stem (Francklyn and Schimmel, 1989; Musier-Forsyth *et al.*, 1991; Francklyn *et al.*, 1992). Even minihelices based on tRNA^{Met} whose major identity elements are in the anticodon (Schulman, 1991) are aminoacylated with methionine in a sequence-specific manner (Martinis and Schimmel, 1992). Likewise, RNA minihelices are substrates for the tRNA-processing enzyme, RNase P (McClain *et al.*, 1987). This suggests that primordial synthetases may consist primarily of a catalytic core domain able to bind and charge RNA helices, and that other domains of the present-day synthetases are dispensable for catalysis. To test this hypothesis, we constructed a 'minimal' GlnRS by fusing the truncated N-terminal domain of the Q20A6 mutant to the C-terminal domains of the Q20M1, Q20M5 and Q20M13 mutants described above. Corresponding polypeptides of 42, 30 and 43 kDa were produced (Figure 3). These combined deletion mutants conferred weak prototrophy to strain BT32 on minimal lactose plates (colonies appeared after 5 days at 30°C,

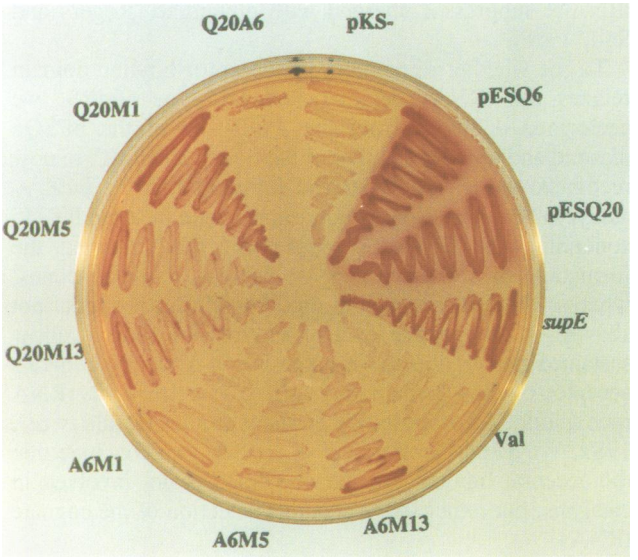

Fig. 4. Suppression efficiency of the 'minimal' GlnRS enzymes on McConkey indicator plates. Aminoacylation activity was assessed by the ability of the minimal GlnRS mutants to produce a red pigment on McConkey indicator plates. Strain BT32 ($lacZ_{1000}$, *supF*) was transformed with the plasmids carrying the indicated *glnS* constructs and transformants streaked on McConkey plates; strain BT235 ($lacZ_{1000}$) transformed with either *supE* $tRNA^{Gln}$ or synthetic $tRNA^{Val}$am were included as positive and negative controls, respectively. Intensity of the red color developing after 24 h incubation at 30°C is proportional to the level of β-galactosidase activity produced by suppression.

whereas controls did not) indicating their ability to mischarge *supF* tRNA with, however, a much reduced efficiency compared with their parent plasmids. Since the β-galactosidase activity of these strains was too low for direct measurements, we tried to detect the mischarging activity of the minimal GlnRS mutants on the very sensitive McConkey indicator plates where Lac^+ colonies develop a red pigment (Miller, 1972). Figure 4 shows that the pigmentation of cells containing the 'minimal' GlnRS enzymes (A6M1, A6M5 and A6M13) is intermediate between cells containing the single deletions and cells that contain a $tRNA^{Val}$ amber suppressor which is not mischarged by GlnRS. The 'minimal' GlnRS enzymes could not rescue the thermosensitive UT172 strain probably because their enzymatic activity is too low. We conclude that 'minimal' GlnRS enzymes which lack all the idiosyncratic domains interacting with the tRNA acceptor-end and anticodon stem–loop still possess enzymatic activity (Figure 5).

Discussion

The smallest native aminoacyl-tRNA synthetase from *E.coli* is CysRS, a monomer of 461 amino acids (Eriani *et al.*, 1991; Hou *et al.*, 1991). Its comparison with other class I synthetases revealed that size differences are mainly accounted for by the shortening of 'connective polypeptides' (CP1 and CP2). These connective polypeptides are insertions of various sizes that were found between conserved stretches of homologies in the core domain of related synthetases

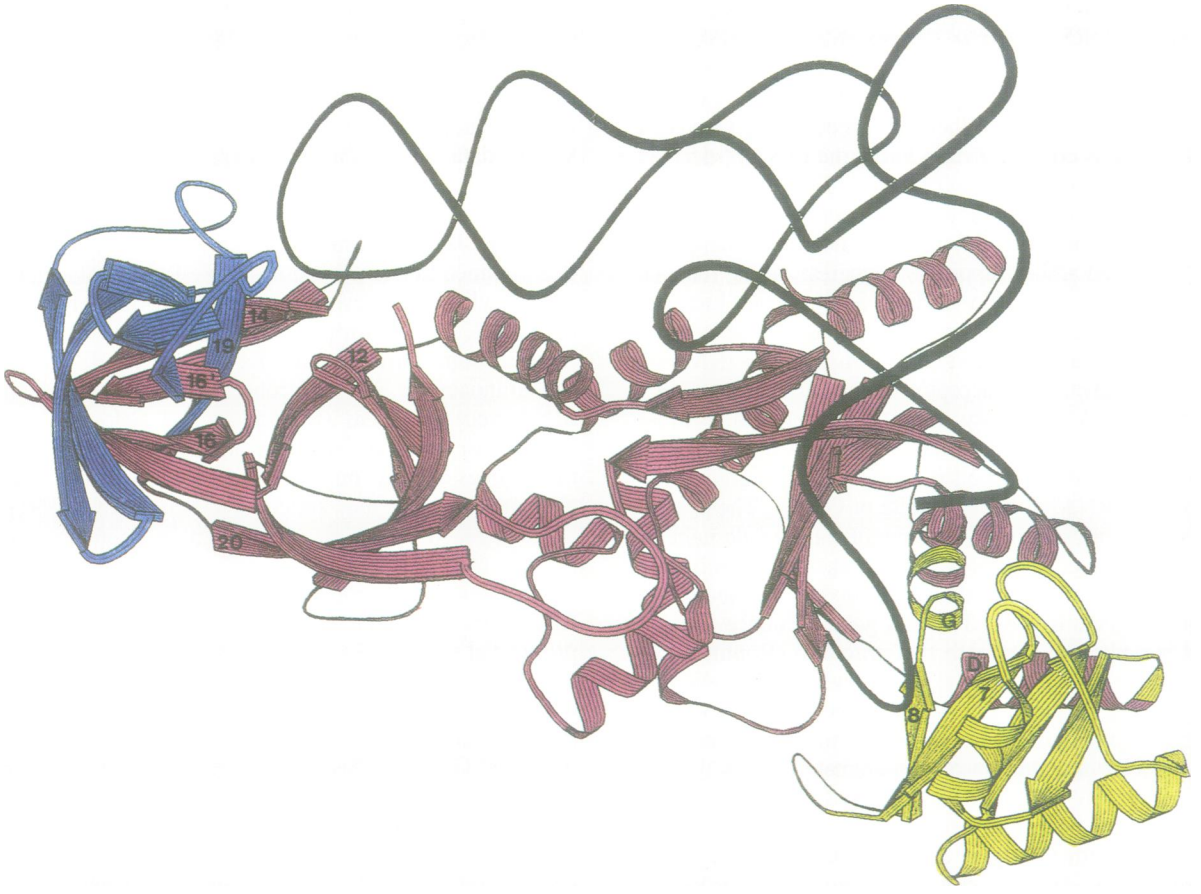

Fig. 5. Ribbon view of the GlnRS enzyme complexed to $tRNA^{Gln}$. The phosphate–sugar backbone of $tRNA^{Gln}$ is represented in black with its acceptor-end at the lower right and its anticodon loop at top left. The secondary structure domains that are removed in the Q20A6 and Q20M13 enzymes are shown in yellow and blue, respectively.

(Starzyk *et al.*, 1987; Burbaum and Schimmel, 1991). CP2 is absent in TyrRS and CP1 from IleRS could be reduced by genetic manipulation from 300 to 150 amino acids without loss of activity (Starzyk *et al.*, 1987). Other proteolysis or deletion experiments showed that discrete domains are dispensable for catalysis (Cassio and Waller, 1971; Koch and Bruton, 1974; Jasin *et al.*, 1983; Toth and Schimmel, 1990), but none could generate a monomeric enzyme of <461 amino acids displaying catalytic activity *in vivo*. The GlnRS mutants from this study range from 361 to 481 amino acids for the single domain deletions, and from 260 to 380 amino acids for the 'minimal' GlnRS enzymes. The size of CP1 in the Q20A6 mutant is reduced from 110 to 5 residues. The other GlnRS mutants have deletions in domains unique to GlnRS interacting with the anticodon stem/loop of tRNA. We believe that the ability to isolate such large deletion mutants lies in the sensitivity of our assay and the random deletion approach followed by genetic selection employed in our study.

The anticodon of tRNAs identifies, through the rules of the genetic code, the amino acid being attached to the acceptor-end, but it rarely serves as the sole determinant of the recognition by the corresponding synthetase (Schulman, 1991). In the case of $tRNA^{Ala}$, a single base pair in the acceptor stem is responsible for the specific recognition by AlaRS (Hou and Schimmel, 1988; McClain and Foss, 1988), whereas the identity of $tRNA^{Gln}$ resides in both the anticodon loop and acceptor stem (Jahn *et al.*, 1991). Our study indicates that the usual recognition of the anticodon of $tRNA^{Gln}$ is not required *in vivo* as the mutant enzymes Q20M1 and Q20M5 which lack either the distal (from strand 16 to 19) or both β-barrel domains (from strand 14 onwards) are still able to aminoacylate tRNA substrates. All mutants, however, retain Arg341 of strand 12 which is involved in the recognition of U35 (Rould *et al.*, 1991). Similarly, the acceptor-binding domain of GlnRS, corresponding to CP1 of other synthetases, is not required for aminoacylation, but rather for docking the 3'-end of $tRNA^{Gln}$ towards the active site or to prevent the access of non-cognate tRNAs to the active site. We therefore conclude, as suggested by the crystal structure, that the acceptor- and anticodon-binding domains of GlnRS are not required for catalysis, but rather for the specific recognition of tRNAs.

We propose that aminoacyl-tRNA synthetases have evolved from a catalytic core (the active site domain) by the subsequent addition of idiosyncratic domains responsible for the correct recognition of the cognate tRNA among the growing number of tRNAs created during the expansion of the genetic code. This hypothesis is consistent with several observations: (i) primordial tRNAs were probably short RNA helices or duplexes which did not require large enzymes for their charging; (ii) the existence of a common structural motif in the active site domain of all class I synthetases; (iii) the presence, at the same location between the two halves of this domain, of 'connective polypeptides' of various sizes for each member of this class; (iv) the low degree of homology between the anticodon-binding domains of related synthetases [e.g. this domain is predominantly α-helical in MetRS, but β-sheet in GlnRS (Perona *et al.*, 1991)]; (v) the common branching, based on sequence comparisons, and subsequent divergence of synthetases specific for similar amino acids (Nagel and Doolittle, 1991).

The existence of two classes of synthetases which differ in the structure of their catalytic site, their mode of recognition of tRNA, and site of attachment of the amino acid to the terminal ribose is intriguing (Moras, 1992). We do not provide evidence for the mode of evolution of class II synthetases, but the similar structure of their phylogenetic tree suggests a concerted evolution to accommodate newly available amino acids (Nagel and Doolittle, 1991). One specific feature of class II synthetases is their oligomeric structure. Based on yeast mitochondrial PheRS, which is a monomer instead of the usual $\alpha_2\beta_2$ tetramer, it was proposed (Sanni *et al.*, 1991) that a second subunit could serve to improve tRNA specificity by providing negative interactions with non-cognate tRNAs. This advantage would not be necessary in the mitochondrial environment where competition among tRNAs may be less. This role is analogous to the one we propose for the acceptor-binding domain of GlnRS and may underlie a general theme for protein evolution, i.e. the intra- or inter-molecular recruitment of additional sequences for a new function. Consistent with this idea, it has been shown that synthetases acquire new functions (Akins and Lambowitz, 1987) and that their oligomeric state is not a prerequisite for function (Shiba and Schimmel, 1992).

Materials and methods

Strains and plasmids

Strains BT32 (*lacZ*$_{1000}$, *supF* $tRNA^{Tyr}$), BT235 (*lacZ*$_{1000}$, su$^-$) and BT235-*supE* (*lacZ*$_{1000}$, *supE* $tRNA^{Gln}$) have been described previously (Inokuchi *et al.*, 1984; Swanson *et al.*, 1988). Strain UT172 contains the *glnS172* allele producing a thermosensitive GlnRS (Englisch-Peters *et al.*, 1991). The genes for the synthetic $tRNA^{Val}$am (Normanly and Abelson, 1989) and for $tRNA^{Ser1}$(CUA) were cloned in pACYC184 as described previously (Weygand-Durasevic *et al.*, 1993).

The pESQ20 plasmid was constructed as follows: the 2.0 kb *Dra*I fragment containing the *glnS* gene (Yamao *et al.*, 1982), originally cloned in the *Hin*cII site of M13mp11, was transferred as an *Eco*RI–*Hin*dIII fragment to the corresponding sites of pBluescript KS+ (Stratagene) to yield pESQ6. A novel *Bsp*HI site was introduced at the initiator codon (CGATGATCATGA) by site-directed mutagenesis, giving plasmid pESQ14. This plasmid gives rise to an increased expression of GlnRS (E.Schwob, unpublished data). The *Asu*II–*Hin*dIII subfragment of pESQ14 was exchanged with the corresponding fragment of the *glnS7* gene (Uemura *et al.*, 1988), to yield plasmid pESQ20. Standard molecular biology techniques were as described by Sambrook *et al.* (1989).

Mutagenesis and selection

The pESQ20 plasmid (20 μg) coding for the mischarging GlnRS7 (D235N) mutant enzyme was linearized either at a unique *Asu*II site (at residue 157, in the acceptor-binding domain) or a unique *Mlu*I site (at residue 420, in the distal β-barrel domain) and submitted to progressive Bal31 exonuclease digestion (3.75 U; 30 s time points for 7 min; deletion range of $\sim$80 bp/min in each direction). The various time points were combined in three pools (30''–1'30'', 2'–4' and 5'–7') and the DNA blunt-ended with T4 and Klenow polymerases. The religated pools were transformed into strain BT32 (relevant genotype: *lacZ*$_{1000}$, *supF*) and clones giving rise to an enzyme able to mischarge *supF* $tRNA^{Tyr}$ with glutamine were selected on minimal lactose plates after 2 days at 30°C. From several thousand transformants (scored on glucose plates), only seven, three and four Lac$^+$ colonies from the respective *Asu*II pools (acceptor-binding domain) grew on lactose minimum plates, whereas 20, 30 and eight colonies were recovered from the *Mlu*I deletion pools (β-barrel domain). Only the clones obtained from the longest exonuclease time points (5–7 min) were analyzed further. Plasmid DNA was prepared from cultures grown for 2 days in selective lactose minimum medium (Miller, 1972) containing ampicillin and sequenced using specific primers and dideoxynucleotides (Sambrook *et al.*, 1989). This DNA was used to retransform strain BT32 to yield Lac$^+$ colonies from which plasmid DNA was sequenced to confirm that the phenotype is plasmid borne and that the mutants are stable. Q20A5 is an independent clone found to be identical to Q20A1.

Activity of the mutants

The activity of the mutants was estimated by measuring the β-galactosidase activity of *lacZ*$_{1000}$ strains suppressed by *supF* glutaminylated by the

truncated GlnRS enzymes. Precultures of transformants were grown in lactose minimum medium containing ampicillin, then diluted and grown for 4 h at 30°C in 1 × A medium (Miller, 1972) containing 100 μg/ml ampicillin and 1 mM IPTG. β-Galactosidase activity was measured in triplicate, according to Miller (1972).

The ability of the mutants to complement a thermosensitive allele of the *glnS* gene was tested by transforming strain UT172 (Englisch-Peters *et al.*, 1991). Aliquots were plated on LB plates containing ampicillin and incubated at either 30 or 42°C. A (+) in Figure 2A indicates that no less than 20% of the transformants grew at the non-permissive temperature.

Relaxed tRNA specificity of the mutants was tested by their ability to mischarge *in vivo* an amber derivative of $tRNA^{Ser1}$ and suppress the $lacZ_{1000}$ mutation. Strain BT235 containing $tRNA^{Ser1}$(CUA) in pACYC184 (Weygand-Durasevic *et al.*, 1993) was grown in LB + Tet, transformed with the various λ*glnS* plasmids and tested for growth on lactose minimal plates containing ampicillin. A (+) in Figure 2A indicates that >20% of the transformants (scored on LB + Tet + Amp plates) could form large colonies on lactose plates after 2.5 days at 30°C.

Expression of the truncated GlnRS in the T7 polymerase system

The regions comprising the coding sequences of the deleted pESQ20 clones were cut as *Bsp*HI–*Stu*I fragments and cloned into the *Nco*I–*Bam*HI sites of pET11d (Studier *et al.*, 1990). The recombinant proteins were expressed in strain BL21(DE3) as described (Studier *et al.*, 1990) and total cellular extracts analyzed on a 7–22% SDS–polyacrylamide gradient gel.

Minimal GlnRS mutants

To combine the acceptor-binding and β-barrel domain deletions, the *Bss*HII–*Stu*I fragments of Q20M1, Q20M5 and Q20M13 were used to replace the corresponding Q20A6 fragment. All constructs were checked by restriction site analysis and sequencing. Activity of the minimal GlnRS mutants in strain BT32, as judged by their ability to mischarge *supF*, was detected on McConkey plates where Lac^+ colonies develop a red pigment (Miller, 1972).

Acknowledgements

This paper is dedicated to the memory of Prof. J.-P.Ebel. We thank H.Inokuchi, L.Isaksson, J.Miller and W.Studier for strains and plasmids. We are grateful to J.Perona, M.Rould and T.A.Steitz for help and expertise in the interpretation of the crystallographic data, and to J.Rogers and I.Weygand-Durasevic for many discussions. E.S. had support from the Fondation pour la Recherche Médicale. This work was supported by a grant from the NIH.

References

Akins,A. and Lambowitz,A. (1987) *Cell*, **50**, 331–345.
Atkins,J.F., Weiss,R.B., Thompson,S. and Gesteland,R.F. (1991) *Annu. Rev. Genet.*, **25**, 201–228.
Burbaum,J.J. and Schimmel,P. (1991) *J. Biol. Chem.*, **266**, 16965–16968.
Breton,R., Watson,D., Yaguchi,M. and Lapointe,J. (1990) *J. Biol. Chem.*, **265**, 18248–18255.
Brick,P., Bhat,T.N. and Blow,D.M. (1989) *J. Mol. Biol.*, **208**, 83–98.
Brunie,S., Zelwer,C. and Risler,J.-L. (1990) *J. Mol. Biol.*, **216**, 411–424.
Cassio,D. and Waller,J.P. (1971) *Eur. J. Biochem.*, **20**, 283–300.
Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) *Nature*, **347**, 249–265.
Englisch-Peters,S., Conley,J., Plumbridge,J., Leptak,C., Söll,D. and Rogers,M.J. (1991) *Biochimie*, **73**, 1501–1508.
Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) *Nature*, **247**, 203–206
Eriani,G., Dirheimer,G. and Gangloff,J. (1991) *Nucleic Acids Res.*, **19**, 265–269.
Francklyn,C. and Schimmel,P. (1989) *Nature*, **337**, 478–481.
Francklyn,C., Shi,J.P. and Schimmel,P. (1992) *Science*, **255**, 1121–1124.
Hou,Y.M. and Schimmel,P. (1988) *Nature*, **333**, 140–145.
Hou,Y.M., Shiba,K., Mottes,C. and Schimmel,P. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 976–980.
Inokuchi,H., Hoben,P., Yamao,F., Ozeki,H. and Söll,D. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 5076–5080.
Jahn,M., Rogers,M.J. and Söll,D. (1991) *Nature*, **352**, 258–260.
Jasin,M., Regan,L. and Schimmel,P. (1983) *Nature*, **306**, 441–447.
Koch,G.L. and Bruton,C.J. (1974) *FEBS Lett.*, **40**, 180–182.
Martinis,S.A. and Schimmel,P. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 65–69.
McClain,W.H. and Foss,K. (1988) *Science*, **240**, 793–796.
McClain,W.H., Guerrier-Takada,C. and Altman,S. (1987) *Science*, **238**, 527–530.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Moras,D. (1992) *Trends Biochem. Sci.*, **17**, 159–164.
Musier-Forsyth,K., Usman,N., Scaringe,S., Doudna,J., Green,R. and Schimmel,P. (1991) *Science*, **253**, 784–786.
Nagel,G.M. and Doolittle,R.F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8121–8125.
Normanly,J. and Abelson,J. (1989) *Annu. Rev. Biochem.*, **58**, 1029–1049.
Perona,J.J., Rould,M.A., Steitz,T.A., Risler,J.-L., Zelwer,C. and Brunie,S. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2903–2907.
Rich,A. (1962) In Kasha,M. and Pullman,B. (eds), *Horizons in Biochemistry*. Academic Press, New York, pp. 103–126.
Rogers,M.J. and Söll,D. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 6627–6631.
Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) *Science*, **246**, 1135–1142.
Rould,M.A., Perona,J.J. and Steitz,T.A. (1991) *Nature*, **352**, 213–218.
Ruff,M., Krisnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.-C. and Moras,D. (1991) *Science*, **252**, 1682–1689.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanni,A., Walter,P., Boulanger,Y., Ebel,J.P. and Fasiolo,F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8387–8391.
Schimmel,P. (1987) *Annu. Rev. Biochem.*, **56**, 125–158.
Schimmel,P. (1991) *Trends Biochem. Sci.*, **16**, 1–3.
Schulman,L.H. (1991) *Prog. Nucleic Acid Res. Mol. Biol.*, **41**, 23–87.
Shiba,K. and Schimmel,P. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1880–1884.
Starzyk,R.M., Webster,T.A. and Schimmel,P. (1987) *Science*, **237**, 1614–1618.
Studier,W.F., Rosenberg,A.H., Dunn,J.J. and Dudendorf,J.W. (1990) *Methods Enzymol.*, **85**, 60–89.
Swanson,R., Hoben,P., Sumner-Smith,M., Uemura,H., Watson,L. and Söll,D. (1988) *Science*, **242**, 1548–1551.
Toth,M.J. and Schimmel,P. (1990) *J. Biol. Chem.*, **265**, 1000–1004.
Uemura,H., Rogers,M.J., Swanson,R., Watson,L. and Söll,D. (1988) *Protein Eng.*, **2**, 293–296.
Vidal-Cros,A. and Bedouelle,H. (1992) *J. Mol. Biol.*, **223**, 801–810.
Weygand-Durasevic,I., Schwob,E. and Söll,D. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2010–2014.
Yamao,F., Inokuchi,H., Cheung,A., Ozeki,H. and Söll,D. (1982) *J. Biol. Chem.*, **257**, 11639–11643.

Received on June 17, 1993; revised on August 30, 1993