Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5550–5552, June 1995
Biochemistry

Operational RNA code for amino acids: Species-specific aminoacylation of minihelices switched by a single nucleotide

(aminoacyl-tRNA synthetases/evolution of coding systems/genetic code/human and bacterial glycyl-tRNA synthetase)

DEBORAH HIPPS*, KIYOTAKA SHIBA†, BARRY HENDERSON*, AND PAUL SCHIMMEL*

*Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Department of Cell Biology, Cancer Institute, Japanese Foundation for Cancer Research, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan

Contributed by Paul Schimmel, March 13, 1995

ABSTRACT **The genetic code is based on aminoacylation reactions where specific amino acids are attached to tRNAs bearing anticodon trinucleotides. However, the anticodon-independent specific aminoacylation of RNA minihelix substrates by bacterial and yeast tRNA synthetases suggested an operational RNA code for amino acids whereby specific RNA sequences/structures in tRNA acceptor stems correspond to specific amino acids. Because of the possible significance of the operational RNA code for the development of the genetic code, we investigated aminoacylation of synthetic RNA minihelices with a human enzyme to understand the sequences needed for that aminoacylation compared with those needed for a microbial system. We show here that the species-specific aminoacylation of glycine tRNAs is recapitulated by a species-specific aminoacylation of minihelices. Although the mammalian and *Escherichia coli* minihelices differ at 6 of 12 base pairs, two of the three nucleotides essential for aminoacylation by the *E. coli* enzyme are conserved in the mammalian minihelix. The two conserved nucleotides were shown to be also important for aminoacylation of the mammalian minihelix by the human enzyme. A simple interchange of the differing nucleotide enabled the human enzyme to now charge the bacterial substrate and not the mammalian minihelix. Conversely, this interchange made the bacterial enzyme specific for the mammalian substrate. Thus, the positional locations (if not the actual nucleotides) for the operational RNA code for glycine appear conserved from bacteria to mammals.**

The 20 synthetases are divided into two classes (I and II) of 10 enzymes each (1, 2). Enzymes of each class are approximately comprised of two major domains, where each of the two tRNA domains interacts with a distinct domain of the cognate tRNA synthetase (3–5). The tRNA acceptor–TΨC helix ("minihelix" domain) interacts with the class-defining catalytic domain (and insertions into that domain), and segments outside of the acceptor–TΨC stem–loop, such as the anticodon, interact with a second, highly variable synthetase domain that is joined to the catalytic core. While many synthetases make contact with their tRNA anticodons, specific mutations in the acceptor stems of their tRNAs severely reduce aminoacylation efficiency (6–8) and, even for these "anticodon" examples, RNA oligonucleotides with sequences based on acceptor stems alone are also aminoacylated with their cognate amino acids (9–11). The operational RNA code based on acceptor stems may have predated the genetic code and was possibly incorporated into or combined with the genetic code when the two domains of tRNAs were assembled into a single molecule (4).

Escherichia coli glycyl-tRNA synthetase has an $\alpha_2\beta_2$ quaternary structure, with a 303-amino acid α chain and 689-amino acid β subunit (12, 13). In contrast, the human enzyme is an α_2 dimer of 739-amino acid polypeptides (14, 15), similar to the *Bombyx mori* enzyme (16). Of most significance is the complete lack of sequence relatedness between the human and *E. coli* enzymes, to the extent that the sequence alone provides no evidence that the two proteins have a common origin. In fact, the class-defining motif 2 of the human and *B. mori* enzymes is more related to that of the bacterial threonyl- and prolyl- than to the glycyl-tRNA synthetase (15, 16).

The *E. coli* and human enzymes do not cross-acylate their respective tRNAs (15). This failure to cross-acylate is correlated with many differences in the nucleotide sequences of the acceptor–TΨC stems of the bacterial and mammalian tRNA substrates (17). Of potentially most significance is the replacement of the U73 "discriminator base" in *E. coli* $tRNA^{Gly}$ with an "A" in mammalian glycine tRNAs. The U73 discriminator base together with an acceptor helix C2·G71 base pair is a combination of these three nucleotides that is unique to glycine tRNAs in *E. coli*. Moreover, these three nucleotides are essential for aminoacylation of RNA minihelices with glycine by *E. coli* glycyl-tRNA synthetase (18). In the acceptor stems of mammalian glycine tRNAs, the C2·G71 base pair is conserved (17). Thus, we thought that, in spite of other differences in the sequences of the respective minihelix domains, the failure of the human and *E. coli* enzymes to cross-acylate their respective tRNAs could be due to the difference at position 73. Alternatively, the remarkably different primary structures of the two enzymes raised the possibility that the human enzyme responded to a different set of acceptor stem nucleotides and, additionally or alternatively, that it did not have the capacity to aminoacylate substrates based on acceptor–TΨC sequences alone. To investigate these and related questions, we constructed synthetic RNA minihelices whose sequences were based on the acceptor–TΨC stems of *E. coli* and mammalian glycine tRNAs and on simple sequence variations of these stems.

MATERIALS AND METHODS

Aminoacylation assays were carried out in potassium phosphate buffer (pH 7.2) at 37°C with 20 μM [^{3}H]glycine (specific activity, 5 mCi/μmol; 1 Ci = 37 GBq; New England Nuclear) and 2 mM ATP as described (18). RNA minihelices based on the acceptor stems of *E. coli* and mammalian tRNAs were chemically synthesized from phosphoriboamidites by the methods of Usman *et al.* (19) and Musier-Forsyth *et al.* (20, 21). These substrates were used in assays at concentrations of 100–200 μM. Human glycyl-tRNA synthetase was produced in *E. coli* as a glutathione *S*-transferase fusion protein using the pGEX expression system (Pharmacia) and was purified on glutathione-Sepharose as described by the manufacturer. The enzyme was estimated as at least 90% pure by this procedure and did not contain any *E. coli* glycyl-tRNA synthetase as judged by Western blot analysis with polyclonal anti-*E. coli* glycyl-tRNA synthetase antibodies (22). Enzyme concentration was estimated by ultraviolet absorbance at 280 nm and was used in assays at ≈0.35 μM. The *E. coli* enzyme was prepared

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

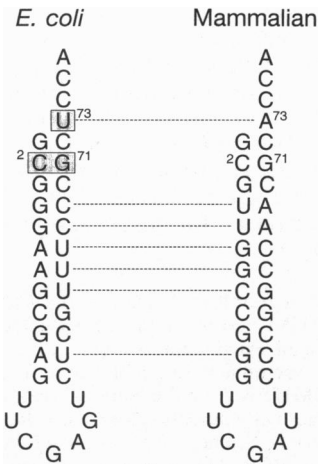

FIG. 1. Synthetic RNA minihelices based on the sequences of acceptor–TΨC stem–loops of the GCC anticodon $tRNA^{Gly}$ from *E. coli* and the UCC cytoplasmic $tRNA^{Gly}$ from rat and mouse (17). Numbers assigned to critical nucleotides are according to the conventional system for a 76-nucleotide tRNA molecule (17). Dashed lines connect positions where the sequence differs between the two minihelices. Shaded boxes enclose nucleotides shown by Francklyn *et al.* (18) to be essential for aminoacylation by *E. coli* glycyl-tRNA synthetase.

as described in an overproducing strain (18) and used at a concentration of ≈0.05 μM.

RESULTS AND DISCUSSION

We chose for our experiments the sequence of the acceptor–TΨC stem for the UCC anticodon rat and mouse cytoplasmic $tRNA^{Gly}$s, which differs from that of the *E. coli* counterpart at N73 and at 6 of 12 positions in the helix (Fig. 1). [Because there are three independent and identical sequences for the acceptor–TΨC stem–loop of the rat and mouse UCC isoacceptor (17), the reliability of this sequence is believed to be high.] *E. coli* glycyl-tRNA synthetase aminoacylates RNA oligonucleotides that encode the U73 discriminator base and a C2·G71 base pair (18). Replacement of either U73 or the C2·G71 base pair is sufficient to eliminate aminoacylation. Of the three critical nucleotides for aminoacylation by the *E. coli* enzyme, C2·G71 is conserved in all eukaryotic cytoplasmic tRNAs (17) but there is a U73 (*E. coli*) → A73 (eukaryote) replacement (Fig. 1).

The cloned human enzyme was expressed in *E. coli* as a fusion protein with glutathione-*S*-transferase and subsequently purified on a glutathione-containing affinity column. The enzyme aminoacylated the mammalian minihelix but did not aminoacylate the *E. coli* minihelix (Fig. 2*A*). (A barely detectable charging of the *E. coli* substrate by the human enzyme was sometimes observed; we cannot eliminate the possibility that this weak aminoacylation arises from *E. coli* enzyme contaminating the recombinant human enzyme isolated from *E. coli*.) Conversely, the *E. coli* enzyme efficiently charged its minihelix but showed no activity toward the mammalian substrates (Fig. 2*B*). Thus, the species specificity observed with the full tRNAs was retained with the minihelix substrates.

To investigate the origins of the species specificity of minihelix aminoacylation, a set of *E. coli* and mammalian minihelix variants was made and tested for cross-aminoacylation. While the mammalian enzyme at best could barely aminoacylate the *E. coli* minihelix, a U73 → A substitution was sufficient to confer strong aminoacylation of the *E. coli* minihelix by the human enzyme (Fig. 2*C*). [As expected from an earlier study (18), we confirmed that this substitution abolished aminoacylation of the *E. coli* minihelix by its homologous enzyme (data not shown).] Similarly, an A73 → U substitution activated the mammalian minihelix for aminoacylation by the *E. coli* enzyme (Fig. 2*D*). We estimate that the *E. coli* enzyme's activity on the

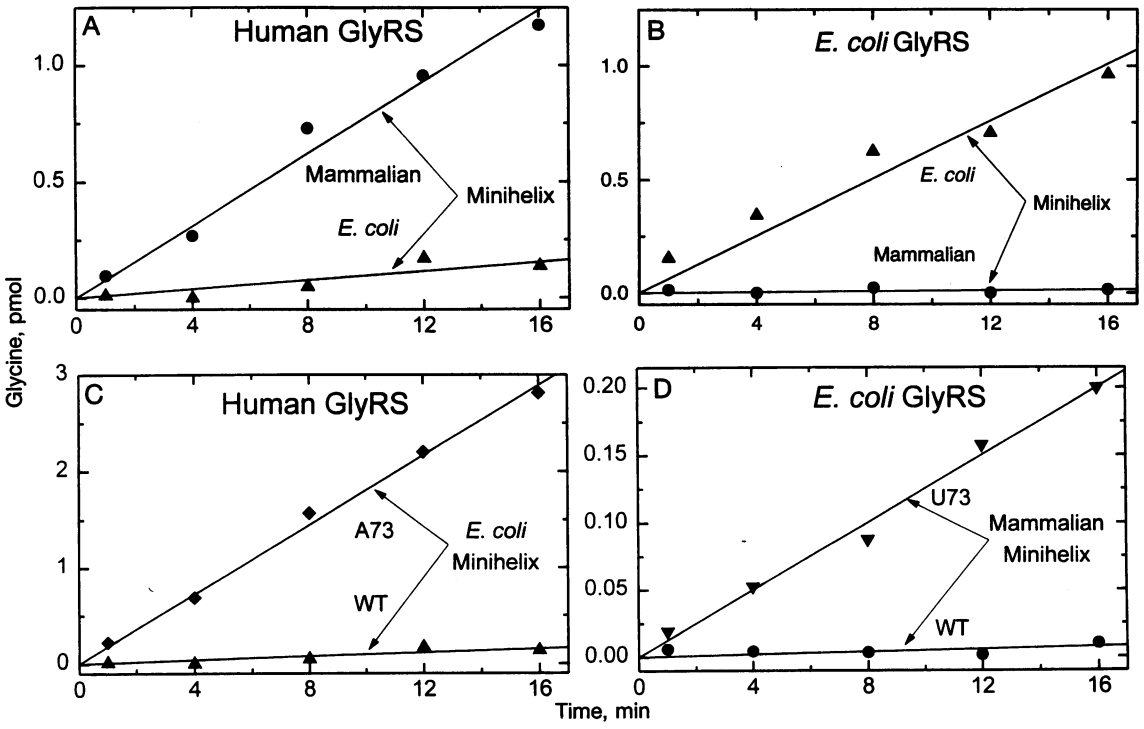

FIG. 2. Aminoacylation of RNA minihelices by *E. coli* and human glycyl-tRNA synthetase. (*A*) Aminoacylation by human glycyl-tRNA synthetase of *E. coli* and mammalian minihelices (Fig. 1). (*B*) Aminoacylation by *E. coli* glycyl-tRNA synthetase of *E. coli* and mammalian minihelices (Fig. 1). (*C*) A U73 → A change confers activity for aminoacylation of an *E. coli* minihelix by human enzyme. (*D*) Aminoacylation of mammalian minihelices by *E. coli* enzyme is activated by an A73 → U change.

U73 mammalian minihelix is within a factor of 5 of its activity on the homologous *E. coli* minihelix substrate.

These results show that position 73 is critical for minihelix aminoacylation by the human enzyme, even though the nucleotide itself differs from that found in the *E. coli* minihelix. To determine whether the 2·71 position is also important for the human enzyme, as it is for the *E. coli* synthetase, we constructed a U2·A71 variant. [This variant of a similar *E. coli* minihelix substrate is not aminoacylated by the *E. coli* enzyme (18).] The U2·A71 mammalian minihelix is weakly charged by the human enzyme (efficiency of aminoacylation reduced by about 10-fold). Thus, while an even greater effect of a U2·A71 substitution on aminoacylation is seen with the *E. coli* enzyme, both enzymes are strongly dependent on the 2·71 position (see below).

A tabulation of results obtained with the two enzymes on five minihelix substrates shows the parallels between the human and *E. coli* systems, when the critical nucleotide positions alone are considered (Fig. 3). For the purposes of this comparison, we designated with + the activity for a wild-type substrate with its homologous enzyme and for other substrates that are within 5-fold of that activity, a (−) for substrates that have aminoacylation rates that are reduced about 10-fold, and a − to ones with no detectable charging. The dependence of aminoacylation by either enzyme on N73 and the C2·G71 base pair is clear in this comparison. Because the two enzymes aminoacylate substrates within the context of either an "*E. coli*" or a "mammalian" minihelix framework, we infer that the sequence differences between the mammalian and *E. coli* minihelix substrates outside of the 73 and 2·71 positions (Fig. 1) are of minor functional significance.

These results suggest that the operational RNA code for glycine has been preserved from bacteria to man, if the positions of the critical nucleotides are considered and not the nucleotides *per se*. The results are striking because of the lack of relatedness of the *E. coli* and human glycyl-tRNA synthetases, in contrast to the clear relatedness of many other *E. coli* and human tRNA synthetases (15). We noted that all eubacterial glycine tRNAs have U73, while archaebacterial and eukaryote cytoplasmic counterparts have A73 (17). If the cytoplasmic mammalian and *E. coli* enzymes had different evolutionary histories, then those histories apparently encountered the same selective pressures from an operational RNA code for glycine that determined the locations of acceptor stem nucleotides needed for aminoacylation. The enzyme structures presumably experienced coadaptations to ensure recognition of the N73 nucleotides regardless of it being U or A and these coadaptations may reflect the historical role of acceptor helix interactions (22).

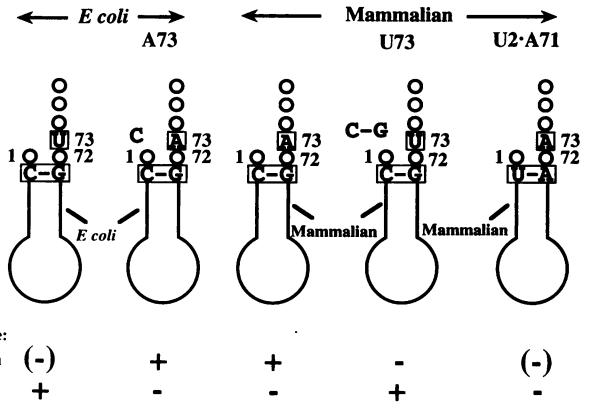

FIG. 3. Summary of minihelix substrates tested with *E. coli* and human glycyl-tRNA synthetases. Activity designations are as follows: +, aminoacylation activity within a factor of five of that of the homologous enzyme on the "wild-type" minihelix substrate; (−), activity reduced about 10-fold from that of wild-type substrate with respect to its homologous enzyme; −, no detectable activity.

We thank H. Motegi and N. Suzuki of the Cancer Institute for construction of the GST fusion of human glycyl-tRNA synthetase. We also thank Dr. Magali Frugier for reading and commenting on the manuscript before submission for publication. This work was supported by Grant GM15539 from the National Institutes of Health, by a grant from the National Foundation for Cancer Research, and by the Ministry of Education, Science and Culture of Japan. D.H. was a Postdoctoral Fellow of the Massachusetts Division of the American Heart Association. B.H. is a Postdoctoral Fellow of the National Institutes of Health.

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203–206.
2. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N. & Leberman, R. (1990) *Nature (London)* **347,** 249–255.
3. Moras, D. (1992) *Trends Biochem. Sci.* **17,** 159–164.
4. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8763–8768.
5. Buechter, D. D. & Schimmel, P. (1993) *Crit. Rev. Biochem. Mol. Biol.* **28,** 309–322.
6. Lee, C.-P., Dyson, M. R., Mandal, N., Varshney, U., Bahramian, B. & RajBhandary, U. L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9262–9266.
7. Putz, J., Puglisi, J. D., Florentz, C. & Giegé, R. (1991) *Science* **252,** 1696–1699.
8. Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., Giegé, R. & Yokoyama, S. (1994) *J. Mol. Biol.* **236,** 710–724.
9. Martinis, S. A. & Schimmel, P. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 65–69.
10. Nureki, O., Niimi, T., Muto, Y., Kanno, H., Kohno, T., Muramatsu, T., Kawai, G., Miyazawa, T., Giegé, R., Florentz, C. & Yokoyama, S. (1993) in *The Translation Apparatus*, eds. Nierhaus, K. H., Franceschi, F., Subermanian, A. R., Erdmann, U. A. & Wittmann-Liebold, B. (Plenum, New York), pp. 59–66.
11. Frugier, M., Florentz, C. & Giegé, R. (1994) *EMBO J.* **13,** 2218–2226.
12. Keng, T., Webster, T. A., Sauer, R. T. & Schimmel, P. (1982) *J. Biol. Chem.* **257,** 12503–12508.
13. Webster, T. A., Gibson, B. W., Keng, T., Biemann, K. & Schimmel, P. (1983) *J. Biol. Chem.* **258,** 10637–10641.
14. Ge, Q., Trieu, E. P. & Targoff, I. N. (1994) *J. Biol. Chem.* **269,** 28790–28797.
15. Shiba, K., Schimmel, P., Motegi, H. & Noda, T. (1994) *J. Biol. Chem.* **269,** 30049–30055.
16. Nada, S., Chang, P. K. & Dignam, J. D. (1993) *J. Biol. Chem.* **268,** 7660–7667.
17. Steinberg, S., Misch, A. & Sprinzl, M. (1993) *Nucleic Acids Res.* **21,** 3011–3015.
18. Francklyn, C., Shi, J.-P. & Schimmel, P. (1992) *Science* **255,** 1121–1125.
19. Usman, N., Ogilvie, K. K., Jiang, M.-Y. & Cedergren, R. J. (1987) *J. Am. Chem. Soc.* **109,** 7845–7854.
20. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991) *Science* **253,** 784–786.
21. Musier-Forsyth, K. & Schimmel, P. (1992) *Nature (London)* **357,** 513–515.
22. Toth, M. J. & Schimmel, P. (1990) *J. Biol. Chem.* **265,** 1000–1004.