NUCLEOTIDE SEQUENCE OF A GLYCINE TRANSFER RNA CODED BY BACTERIOPHAGE T4

B.G. BARRELL, A.R. COULSON and William H. McCLAIN
Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge, England

and

Department of Bacteriology, The University of Wisconsin, Madison, Wisconsin 53706, USA

Received 20 August 1973

1. Introduction

Upon infection of *Escherichia coli*, bacteriophage T4 directs the synthesis of eight tRNA species [1] that include acceptor activities for the amino acids arginine, glycine, isoleucine, leucine and proline [2, 3]. The nucleotide sequence of a T4-coded leucine tRNA has recently been determined [4]. The amenability of the T4 tRNA system to a variety of analyses has allowed extensive study of T4 tRNA genetic organisation [1, 5–7], physiological role [8, 9] and biosynthesis [10, 11].

In the present work we report the nucleotide sequence of a T4 glycine tRNA. It consists of 74 nucleotide residues that can be arranged in the usual cloverleaf structure. A minor nucleotide located in the first position of the anticodon may allow selective recognition of GGA, or GGA and GGG glycine codons.

2. Results

2.1. *Isolation of tRNA^{Gly}*

When *E. coli* is infected with T4 and pulsed with [^{32}P]orthophosphate, and the low molecular weight RNA is extracted with phenol and analysed on a 10% polyacrylamide gel, eight components are seen (fig. 1). Because the synthesis of host RNA is completely terminated upon infection, all eight of these RNA components originate by transcription from T4 DNA [1]. Material in the position of tRNA^{Gly} (fig. 1) was eluted from the gel as previously described [1], and subjected to sequence analysis which employs electrophoretic fractionations on paper followed by autoradiography, according to the methods of Sanger [12]. The tRNA^{Gly} obtained in this manner is 80–90% pure, judging from the yields of products obtained after digestion with ribonuclease A or T_1.

2.2. *Sequence of the small ribonuclease T_1 products*

The sequences of the ribonuclease T_1 products and their molar yields are shown in fig. 2. The sequences of the smaller products (Gp, C–Gp, A–Gp, pGp, U–Gp and A–U–Gp) were determined by their position on the fingerprint and from the products of digestion with ribonuclease T_2 and ribonuclease A.

2.3. *C–U–C–C–A_{OH}*

Ribonuclease T_2 digestion of this product gave Cp(3), Up(1) whereas complete digestion with snake venom phosphodiesterase gave pC(2), pU(1), pA(1), giving the partial sequence C(U, C, C,)A_{OH}. Limited digestion with snake venom phosphodiesterase gave the products C–U–C–C_{OH} and C–U–C_{OH}, whose sequences were determined by complete digestion with snake venom phosphodiesterase and alkaline hydrolysis.

 North-Holland Publishing Company – Amsterdam

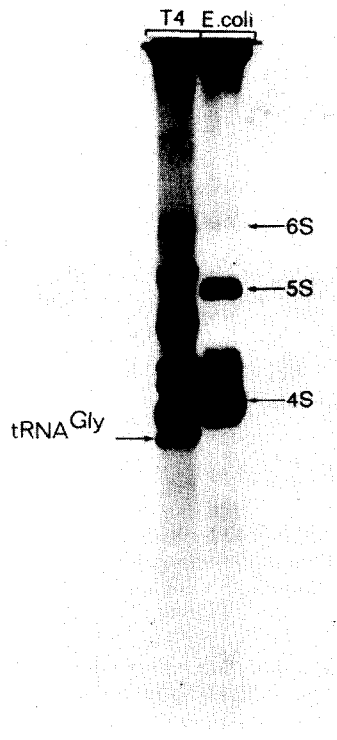


Fig. 1. Autoradiograph of ^{32}P-labelled RNA from T4-infected and uninfected *E. coli* cells fractionated by electrophoresis on a 10% polyacrylamide gel [1]. As extracted from this gel, $tRNA^{Gly}$ was estimated to be 80–90% pure, as judged by molar yields of products obtained after digestion with ribonucleases A or T_1.

2.4. *T–ψ–C–Gp*

Ribonuclease T_2 digestion followed by chromatography on thin-layer plates using the two-dimensional system of Nishimura [13] gave one mole each of Tp, ψp, Cp and Gp. Dephosphorylation with bacterial alkaline phosphatase followed by limited digestion with snake venom phosphodiesterase gave the products $T–ψ–C–G_{OH}$ and $T–ψ–C_{OH}$, which were characterised by complete digestion with snake venom phosphodiesterase and ribonuclease T_2.

2.5. *A–U–A–U–C–Gp*

Ribonuclease T_2 and ribonuclease A digestions gave the composition A–Up (2), Cp (1), Gp (1). The sequence was established by ribonuclease U_2 digestion, which gave products of Ap, U–Ap and (U, C) Gp.

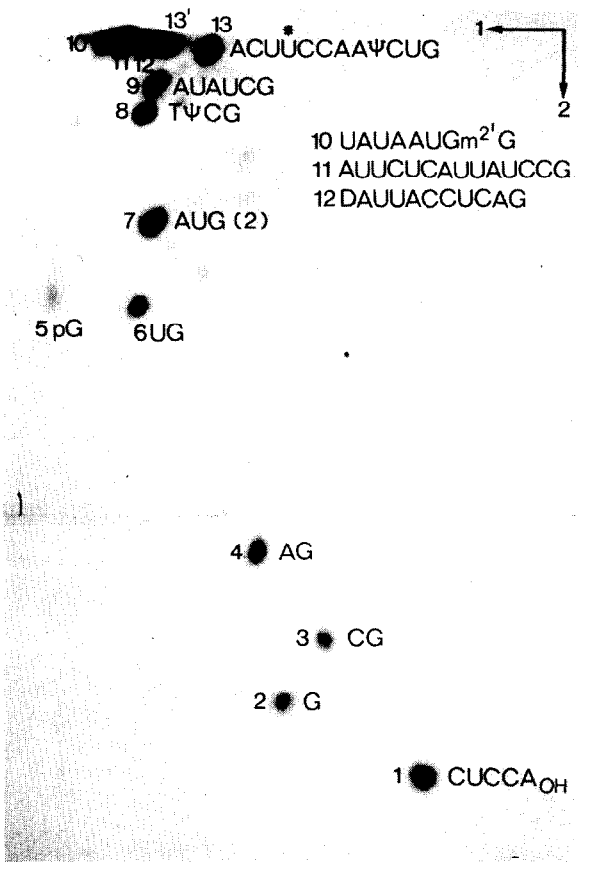


Fig. 2. Autoradiograph of a two-dimensional fractionation of ribonuclease T_1 products of $tRNA^{Gly}$. Molar yields of the products are one, based on the final sequence, unless otherwise indicated. The position of Ůp in product 13 has not been unequivocally established (see text). Electrophoresis [12] on cellulose acetate in pyridine acetate, 7 M urea, pH 3.5, from right to left (first dimension); and on DEAE paper in 7% formic acid (v/v) from top to bottom (second dimension).

2.6. *U–A–U–A–A–U–Gm$^{2'}$–Gp*

This product was also present in some preparations as U–A–U–A–A–U–Gp, presumably arising by undermethylation of the $Gm^{2'}$–Gp, thereby allowing cleavage by ribonuclease T_1. The identity of $Gm^{2'}$–Gp was established by chromatography in the system of Nishimura [13]. Ribonuclease T_2 digestion gave a composition of Up (3), Ap (3), $Gm^{2'}$–Gp (1). Ribonuclease A digestion gave A–Up (1), A–A–Up (1) and $Gm^{2'}$–Gp (1). Limited digestion with ribonuclease A (enzyme to substrate ratio of 1:400 for

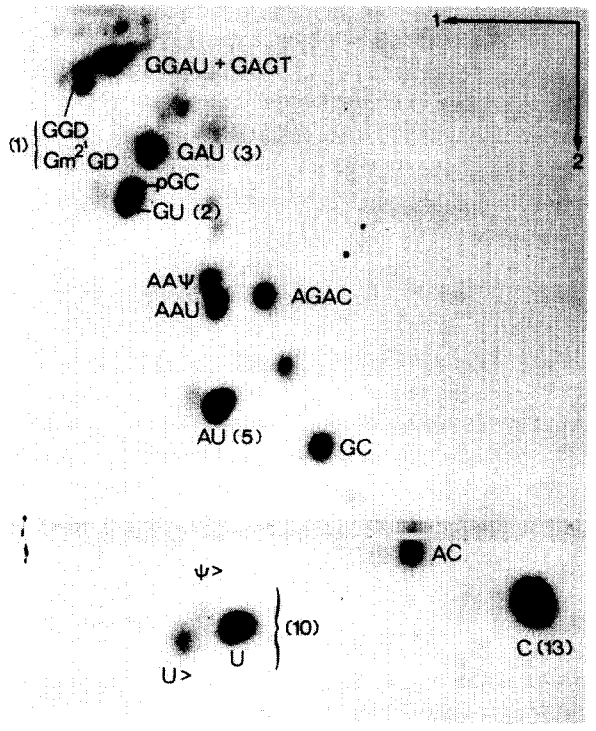


Fig. 3. Autoradiograph of a two-dimensional fractionation of ribonuclease A products $tRNA^{Gly}$. Other conditions were as for fig. 2.

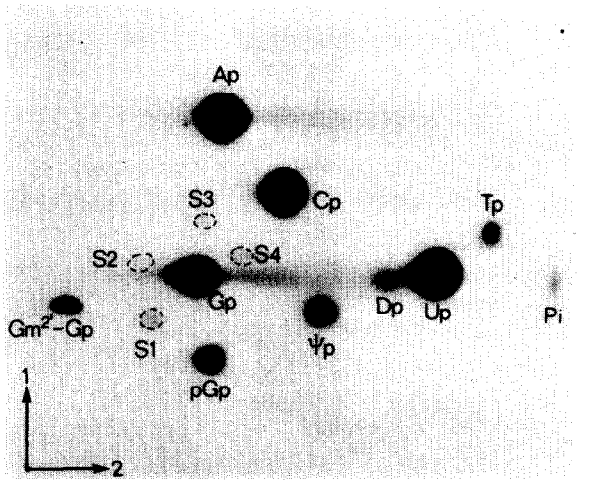


Fig. 4. Autoradiograph of a two-dimensional thin-layer chromatographic fractionation of ribonuclease T_2 digestion products of $tRNA^{Gly}$. 5 μg of RNA containing 10^5 cpm of ^{32}P were digested (4 hr at 37°C in 10 μl of a solution containing 3 μg/ml ribonuclease T_2 and 0.25 mg/ml each of ribonucleases A and T_1 buffered in 0.05 M ammonium acetate, pH 4.5) and then subjected to two-dimensional ascending chromatography on a thin-layer cellulose plate (20 × 20 cm); first run in isobutyric acid–0.5 M NH_3 (5:3, v/v); second run at right angles to the first in isopropanol–HCl–water (70:15:15, v/v/v) [13]

15 min at 0°C) gave (Up, A–Up) and A–A–U–$Gm^{2'}$–Gp. Ribonuclease U_2 digestion gave the products U–Ap, U–A–Ap and U–$Gm^{2'}$–Gp. These products uniquely define the sequence.

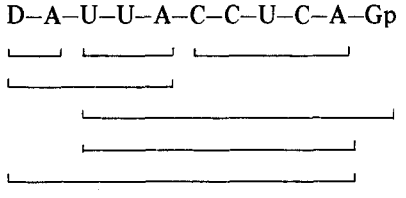
2.7. *A–U–U–C–U–C–A–U–U–A–U–C–C–Gp*

Alkaline hydrolysis and ribonuclease A digestion gave the composition A–Up (3), Up (3), Cp (4), Gp (1). Modification with carbodiimide, followed by ribonuclease A digestion, deblocking with ammonia, and then digesting with ribonuclease U_2 gave the following products: A–U–U–Cp, U–Cp, (Ap, U–U–Ap) U–Cp, Cp and Gp. Ribonuclease U_2 digestion followed by carbodiimide modification and ribonuclease A cleavage gave the products Ap, (U–U–Cp, U–Cp) Ap, U–U–Ap and (U–Cp, Cp) Gp. These three sets of data enable the sequence to be deduced.

2.8. *D–A–U–U–A–C–C–U–C–A–Gp*

This oligonucleotide was obtained in some preparations with Up instead of Dp (identified by two-dimensional chromatography [13]) at the 5′ end. Ribonuclease T_2 and ribonuclease A digestions gave the following compositions: Dp (1), Up (2), Cp (2), A–Up (1), A–Cp (1), and A–Gp (1). Limited digestion with ribonuclease U_2 (0.1 units/ml for 30 min at 21°C), followed by complete digestion of the products with ribonuclease U_2, gave the following set of products:

D–A–U–U–A–C–C–U–C–A–Gp



The structure of D–Ap and U–U–Ap was established by digestions with ribonuclease T_2. The sequence of the ribonuclease U_2 product C–C–U–C–Ap

Fig. 5. Nucleotide sequence of $tRNA^{Gly}$ showing the products obtained by limited digestion with ribonuclease A. The material used for this analysis contained normal amounts of all minor nucleotides except $\overset{*}{U}p$, which was completely unmodified.

was established as follows. Carbodiimide modification followed by ribonuclease A digestion gave a U–Cp. The location of U–Cp was then established by treating C–C–U–C–Ap with polynucleotide phosphorylase in the presence of bacterial alkaline phosphatase and inorganic phosphate; the product of this digestion co-electrophoresed at pH 3.5 on DEAE paper with known $C–C–U_{OH}$ and not with $C–U–C_{OH}$, the only other possible sequence isomer. Alkaline hydrolysis of $C–C–U_{OH}$ gave Cp as the only product.

2.9. *A–C–U–$\overset{*}{U}$–C–C–A–A–ψ–C–U–Gp*

This product was located in one of three positions on the fingerprint, depending on the state of modification of the $\overset{*}{U}p$ residue, which varied with different RNA preparations. It migrated with T_1 product 12 in the completely unmodified form, in which case it could be further purified by subsequent homochromatography. The product could also be found in positions 13 and 13′ on the fingerprint. Most of the sequence analysis was performed on the completely unmodified form. Ribonuclease T_2 and ribonuclease A digestions gave the composition A–Cp (1), Up (3), Cp (3), A–A–ψp (1) and Gp (1). Ribonuclease U_2 digestion followed by carbodiimide modification and then ribonuclease A digestion established the following sequence: A–C–(U–U–Cp, Cp) A–A–ψ–C–U–Gp. Limited digestion of the intact tRNA (fig. 5) with ribonuclease A gave a product terminating in –G–A–C–Up, and thereby established the deduced sequence. We placed $\overset{*}{U}p$ in the first position of the anticodon by analogy with other known tRNA sequences [13].

2.10. *Sequences of ribonuclease A products*

A fingerprint of a ribonuclease A digest of T4 $tRNA^{Gly}$ is given in fig. 3, together with the sequences of the products and their molar yields. Alkaline hydrolysis and ribonuclease T_1 digestion established the sequence of all the products except that of (Gp, A–Gp) Tp. Limited digestion of this product with spleen phosphodiesterase gave products of A–G–Tp and G–Tp, thus establishing the sequence G–A–G–Tp.

2.11. *Analysis of minor nucleotides*

These were characterised by completely digesting $tRNA^{Gly}$ with ribonuclease T_2 and analysing the products by two-dimensional thin-layer chromatography [13]. The identity of the standard and known minor nucleotides are given in fig. 4. Products S1, S2, S3 and S4 probably represent derivatives of $\overset{*}{U}p$, although the data do not eliminate the possibility that they are modified forms of other nucleotide residues in ribonuclease T_1 product 13. S3 is near the position of 5-methylaminomethyl-2-thiouridylic acid [13]. Products S1, S2 and S4 are in positions that have not previously been characterised. We speculate that they represent different modification

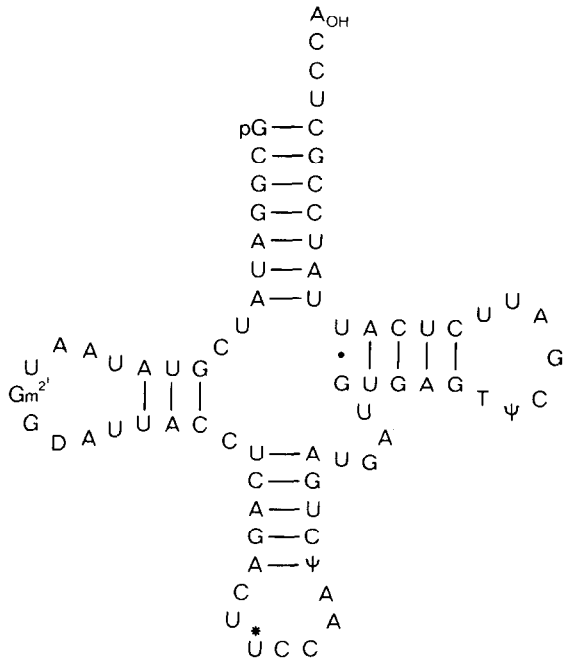


Fig. 6. Cloverleaf structure of $tRNA^{Gly}$.

states of S3 or degradation products. Additional characteristics include the observations that alkali hydrolysis or digestion with ribonuclease T_2 of T_1 product 13′ (fig. 2), which contains S1, followed by electrophoresis at pH 3.5 on Whatman 540 paper, gives a minor nucleotide migrating slightly ahead of Gp (R_U = 0.89). Similar analysis of T_1 product 13, which contains S2, S3 and S4, gives a minor nucleotide that migrates just off the origin (R_U = 0.08).

2.12. *Limited digestion to obtain overlapping fragments*

A comparison of the ribonuclease T_1 and ribonuclease A digestion products allows much of the sequence of the intact tRNA to be deduced as a result of the fortuitous positions of the minor nucleotides and the presence of unique ribonuclease A products. Thus, only a small number of overlaps were needed to complete the structure. This was achieved by limited digestion of $tRNA^{Gly}$ with ribonuclease A, separating the products by homochromatography and analysing them by complete digestion with ribonuclease T_1 and A; these digestion products were subsequently analysed by digestion with ribonuclease A

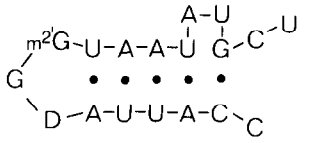


Fig. 7. Alternate base pairing scheme for the dihydrouridine loop of $tRNA^{Gly}$.

and T_1, respectively. The sequences of the products used in deducing the final structure are shown in fig. 5.

3. Discussion

The nucleotide sequence of $tRNA^{Gly}$ is arranged in the cloverleaf pattern in fig. 6. While this molecule conforms to the structural features associated with other known tRNA sequences, we have not as yet tested its aminoacyl acceptor activity or its triplet binding capacity. The chromatographic properties of the 'wobble' nucleotide Ûp suggest that it may be related to 5-methylaminomethyl-2-thiouridylic acid, in which case the tRNA would recognise the glycine code word GGA, or possibly GGA and GGG [13]. This speculation is consistent with other work [3] that has demonstrated the stimulation by T4 infection of a glycine tRNA that is capable of recognising GGG, but not GGU or GGC, in the oligonucleotide binding assay.

E. coli contains at least three distinct tRNA species for glycine [14]. The nucleotide sequences of the GGU/C and the GGG specific glycine tRNA's have been reported [14, 15]. A comparison of these with the T4 $tRNA^{Gly}$ sequence reveals that approximately 60% of the residues in corresponding positions are identical. The regions of greatest homology are the 5′–3′-amino acid acceptor stem and the T–ψ–C loop with its supporting stem. Structural studies of other *E. coli* glycine tRNAs with different coding properties will possibly give us information about the phylogenic relationship between host and T4 glycine tRNA genes. Similar comparisons of T4 and *E. coli* serine tRNAs has revealed a close evolutionary relationship between them (Barrell and McClain, manuscript in preparation).

A distinctive feature of the T4 $tRNA^{Gly}$ is its rapid mobility during gel electrophoresis (fig. 1). A chain length of 66 residues is estimated using the empirical

formula that gel mobility is inversely proportional to the logarithm of the molecular weight; the structure presented here (fig. 6) is 74 residues in length. If there is substance to the alternate base pairing scheme for the dihydrouridine loop presented in fig. 7, it may be responsible for condensing the molecule sufficiently to cause more rapid gel migration.

Acknowledgements

We thank Dr. F. Sanger for continuous support. These studies were supported by grants from the Jane Coffin Childs Memorial Fund for Medical Research (No. 61–207), The University of Wisconsin Graduate School Research Committee and USPHS Research Grant AI 10257.

References

[1] McClain, W.H., Guthrie, C. and Barrell, B.G. (1972) Proc. Natl. Acad. Sci. U.S. 69, 3703.
[2] Daniel, V., Sarid, S. and Littauer, U.Z. (1970) Science 167, 1682.
[3] Scherberg, N.H. and Weiss, S.B. (1972) Proc. Natl. Acad. Sci. U.S. 69, 1114.
[4] Pinkerton, T.C., Paddock, G. and Abelson, J. (1972) Nature New Biol. 240, 88.
[5] McClain, W.H. (1970) FEBS Letters 6, 99.
[6] Wilson, J.H. and Kells, S. (1972) J. Mol. Biol. 69, 39.
[7] Wilson, J.H., Kim, J.S. and Abelson, J.N. (1972) J. Mol. Biol. 71, 547.
[8] Guthrie, C. and McClain, W.H., J. Mol. Biol., in press.
[9] Wilson, J.H. (1973) J. Mol. Biol. 74, 753.
[10] Guthrie, C., Seidman, J.G., Altman, S., Barrell, B.G., Smith, J.D. and McClain, W.H., Nature New Biol., in press.
[11] Barrell, B.G., Seidman, J.G., Guthrie, C. and McClain, W.H. (1973) Proc. Natl. Acad. Sci. U.S., in press.
[12] Barrell, B.G. (1971) in: Procedures in Nucleic Acid Research (Cantoni, G.L. and Davies, D.R., eds.), Vol. 2, p. 751, Harper and Row, New York.
[13] Nishimura, S. (1972) in: Progress in Nucleic Acid Research and Molecular Biology (Davidson, J.N. and Cohn, W.E., eds.), Vol. 12, p. 49, Academic Press, New York.
[14] Squires, C. and Carbon, J. (1971) Nature New Biol 233, 274.
[15] Hill, C.W., Combriato, G., Steinhart, W., Riddle, D.L. and Carbon, J. (1973) J. Biol. Chem. 248, 4252.