THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 257, No. 21, Issue of November 10, pp. 12795-12799, 1982
Printed in U.S.A.

Euglena gracilis Chloroplast Transfer RNA Transcription Units

NUCLEOTIDE SEQUENCE ANALYSIS OF A $tRNA^{Tyr}$-$tRNA^{His}$-$tRNA^{Met}$-$tRNA^{Trp}$-$tRNA^{Glu}$-$tRNA^{Gly}$ GENE CLUSTER*

(Received for publication, June 7, 1982)

Margaret J. Hollingsworth and Richard B. Hallick‡

From the Department of Chemistry, University of Colorado, Boulder, Colorado 80309

A tRNA coding locus in the *Bam-Sal* 9 region of *Euglena gracilis* Pringsheim strain *Z* chloroplast DNA was chosen for detailed study. This DNA contains the previously mapped tRNA coding sequences of the adjacent *Euglena* chloroplast *Eco*RI products of *Eco*V and *Eco*H (Orozco, E. M., Jr., and Hallick, R. B. (1982) *J. Biol. Chem.* 257, 3258-3264). The 3.2-kilobase pair *Bam-Sal* 9 fragment was cloned into the *Bam*HI and *Sal*I cut plasmid vector pBR322, resulting in the recombined plasmid pPG76. The tRNA coding locus was mapped to a region of *Bam-Sal* 9 that contains portions of both *Eco*H and *Eco*V. The DNA sequence of 1-kilobase pair *Bam-Sal* 9, containing the entire tRNA coding locus, was determined. A cluster of six tRNA genes was found. The gene organization is as follows, where bp is base pair: $tRNA^{Tyr}_{GUA}$-64 bp spacer-$tRNA^{His}_{GUG}$-14 bp spacer-$tRNA^{Met}_{CAU}$-4 bp spacer-$tRNA^{Trp}_{CCA}$-27 bp spacer-$tRNA^{Glu}_{UUC}$-6 bp spacer-$tRNA^{Gly}_{UCC}$. The $tRNA^{Met}_{CAU}$ is believed to be an elongator tRNA. The first four genes are within *Eco*V. The *Eco*RI cleavage site that separates *Eco*V and *Eco*H is in the $tRNA^{Glu}$ gene. The $tRNA^{Gly}$ gene is in *Eco*H. This is the largest known chloroplast tRNA gene cluster.

The chloroplast DNA of *Euglena gracilis* has been shown to have at least nine loci which hybridize with chloroplast tRNAs (1-3). At least some of these loci contain multiple clustered tRNA genes that may be part of polycistronic transcription units (1). In *Euglena*, two tRNA loci have been characterized. One, containing the $tRNA^{Ile}$ and $tRNA^{Ala}$, is found between the 16 and 23 S ribosomal genes (4, 5). A larger cluster in *Eco*G contains the four genes $tRNA^{Leu}$, $tRNA^{Arg}$, $tRNA^{Asn}$, and $tRNA^{Val}$ (6). Among higher plants, maize chloroplast DNA has a locus between the 16 and 23 S rRNA genes containing split $tRNA^{Ile}$ and $tRNA^{Ala}$ genes (7). However, both the chloroplast $tRNA^{Val}$ and the $tRNA^{His}$ genes of maize are found as isolated genes (8, 9), as are the $tRNA^{Val}$ and $tRNA^{Asn}$ of tobacco chloroplast DNA (10, 11).

The *Bam-Sal* 9 fragment of chloroplast DNA was chosen for further study because it was a good candidate for having a large polycistronic tRNA cluster (1). *Bam-Sal* 9 contains *Eco*V and a part of both *Eco*G and *Eco*H. It has previously been shown that both *Eco*V and the portion of *Eco*H internal to *Bam-Sal* 9 hybridized ^{32}P-labeled *Euglena* chloroplast tRNAs (1-3). In this study, a restriction endonuclease cleavage map of *Bam-Sal* 9 was constructed, the tRNA coding locus was identified, and the DNA sequence of approximately 1 kb[1] containing the entire coding locus was determined via the Maxam-Gilbert method (12). Six tRNA genes were identified. The organization of the cluster as determined from the DNA sequence data is $tRNA^{Tyr}_{GUA}$-64 bp spacer-$tRNA^{His}_{GUG}$-14 bp spacer-$tRNA^{Met}_{CAU}$-4 bp spacer-$tRNA^{Trp}_{CCA}$-27 bp spacer-$tRNA^{Glu}_{UUC}$-6 bp spacer-$tRNA^{Gly}_{UCC}$. This is the largest known chloroplast tRNA gene cluster. All of the genes have the same polarity with respect to each other, to the rRNA transcription units, and to the previously characterized $tRNA^{Val}$-$tRNA^{Asn}$-$tRNA^{Arg}$ gene cluster of *Eco*G.

EXPERIMENTAL PROCEDURES[2]

RESULTS AND DISCUSSION

Mapping—The location of the 3.2-kb *Bam*HI-*Sal*I digestion product of *Euglena* chloroplast DNA designated *Bam-Sal* 9 (or "BS9") has been previously reported (1, 2, 13). With reference to the *Eco*RI map of this genome (1, 2, 14), BS9 contains all the *Eco*V, as well as portions of both *Eco*G and *Eco*H. A detailed restriction endonuclease map for this DNA was determined (Fig. 1). The enzymes used were *Eco*RI, *Ava*I, *Ava*II, *Alu*I, *Msp*I, *Hha*I, *Hae*III, *Taq*I, *Hinf*I, and *Dde*I. Maps were deduced mainly by standard methods of single and double digestion. When ambiguities were possible, specific fragments were isolated for redigestion (4). In addition, 31 of the 42 cleavage sites shown in Fig. 1 were confirmed by DNA sequence analysis. There is one possible anomaly in the restriction endonuclease map. According to the nucleotide sequence data, there is an internal *Taq*I site in the 257-bp *Taq*I fragment (Fig. 1*c*) that is not cut by *Taq*I in enzyme digestion reactions. The reason for this uncut *Taq*I site is unknown. All other sites predicted from the nucleotide sequence data have been confirmed during the mapping study.

The nucleotide sequence of the right-hand 367 bp of BS9

* This work was supported by Grant GM28463 from the National Institutes of Health. This is paper III in a series *Euglena gracilis Chloroplast Transfer RNA Transcription Units*. Papers I and II are Orozco, E. M., Jr., and Hallick, R. B. (1982) *J. Biol. Chem.* **257**, 3258-3264; 3265-3275. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of National Institutes of Health Research Career Development Award KO4 GM00372. To whom correspondence should be addressed at, Department of Chemistry, University of Colorado, Campus Box 215, Boulder, Colorado 80309.

[1] The abbreviations used are: kb, kilobase pair; bp, base pair; R, a purine nucleoside; Y, a pyrimidine nucleoside; DTT, dithiothreitol.

[2] Portions of this paper (including "Experimental Procedures" and a portion of "Results and Discussion") are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document 82 M-1489, cite the authors and include a check or money order for $1.60 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

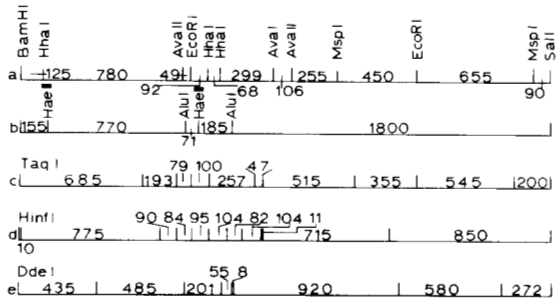

FIG. 1. **Restriction endonuclease map.** Map of the *Bam-Sal* 9 fragment of *E. gracilis* chloroplast DNA. This fragment contains all of *Eco*V and sections of *Eco*H and *Eco*G. Distances between cleavage sites are given in base pairs.

(Fig. 1) has previously been reported (6). This is a portion of the data from 1.6 kb of *Euglena* chloroplast DNA containing a cluster of four tRNA genes (6).

Location of Transfer RNA Genes within BS9—In order to locate the tRNA coding locus (or loci) within BS9, Southern (15) hybridizations were used. Total *Euglena* chloroplast tRNAs were purified. They were radioactively labeled at the 3′-CCA termini with [α-^{32}P]ATP as previously described (16). The tRNA probe was hybridized to membrane filter blots of BS9 restriction endonuclease fragments (1, 6, 16). The region of tRNA hybridization was localized to a 600-bp region between the two *Ava*II sites of BS9 (data not shown). This is a larger coding locus than was previously described for this region for the *bacillaris* strain of *E. gracilis* (3). In another experiment, the synthetic oligonucleotide 5′-CTACCAACTGAGCT was found to hybridize to both *Eco*H and the fragment *Hind*IIIB (6) which contains all of BS9.[3] This synthetic probe is complementary to the D-stem and loop of *Euglena* chloroplast tRNA^{Phe} (17), tRNA^{Ala} (4, 5), and tRNA^{Val} (6). Following nucleotide sequence analysis (described below), the identical 14-base sequence was identified within the D-stem and loop of a tRNA^{Gly} gene. The *Dde*I sequence of the synthetic 14-mer (5′-CTGAG) is the *Dde*I cleavage site between the 485- and 201-bp *Dde*I fragments (Fig. 1*e*).

DNA Sequence Analysis of the tRNA Coding Locus in BS9—The region within BS9 that hybridized with chloroplast [^{32}P]tRNAs was sequenced by the method of Maxam and Gilbert (12). The portion of BS9 analyzed and the sequence strategy used is shown in Fig. 2. DNA sequence was determined for fragments 5′-end-labeled at the following sites: *Eco*RI, *Ava*II, *Ava*I, *Hha*I, *Taq*I, *Alu*I, *Hinf*I, and *Dde*I. For 100% of the data obtained, the sequence of both strands was obtained and/or the same fragment was sequenced more than once. The continuous sequence of 889 bp from BS9, containing 637 bp of the 1.2-kb *Eco*V and 252 bp of the 1.1-kb *Eco*H, is given in Fig. 3.

Transfer RNA Gene Organization—There are six tRNA genes within the tRNA coding locus of BS9 identified by tRNA hybridization. These six genes, which all have the same polarity, are encoded within 557 bp of *Euglena* chloroplast DNA. The DNA sequence of this tRNA gene cluster, as well as of the proximal and distal flanking DNA, is given in Fig. 3. The tRNA genes were identified by comparison of the positions of invariant and semi-invariant bases and secondary structure to the known general structure of other tRNAs (described below) (18, 19). The identity of the genes was predicted from the anticodon sequence and the sequence homologies to known tRNAs (described below). This gene organization is as follows: tRNA$^{Tyr}_{GUA}$-64 bp spacer-tRNA$^{His}_{GUG}$-14 bp spacer-tRNA$^{Met}_{CAU}$-4 bp spacer-tRNA$^{Trp}_{CCA}$-27 bp spacer-tRNA$^{Glu}_{UUC}$-6 bp spacer-tRNA$^{Gly}_{UCC}$.

None of the genes code for the 3′-CCA termini, and none of the genes are interrupted by intervening sequences. The close proximity and common polarity of these genes is consistent with the possibility that some or all of the genes might be initially transcribed as a polycistronic transcription unit. This seems most likely for the tRNA^{His}-tRNA^{Met}-tRNA^{Trp} and the tRNA^{Glu}-tRNA^{Gly} genes, which are separated by intergenic spacers of only 4–14 bp. We are currently attempting to identify the primary transcripts of these genes.

This is the second tRNA gene cluster identified in *Euglena* chloroplast DNA. The coding location is 1.6 kb from the previously characterized four gene tRNA cluster of chloroplast fragment *Eco*G. With these six genes in BS9, the total number of different known *Euglena* chloroplast tRNA genes is now 12, the largest number for any chloroplast DNA. Since the tRNA^{Ile} and tRNA^{Ala} genes are each present three times/genome (1), each gene located in each of three tandemly repeated rRNA transcription units (1), the total number of known *Euglena* chloroplast tRNA genes is now 16. All of these genes, except the tRNA^{Leu} of *Eco*G (6), are of a common polarity.

tRNA^{Met} Gene—The tRNA^{Met} gene most likely encodes an elongator tRNA rather than an initiator tRNA for several reasons. First, the predicted tRNA^{Met} secondary structure (Fig. 4) contains 7 bp in the amino acid acceptor expected for an elongator, rather than the 6 bp expected for an initiator tRNA (23). Second, the base sequence homology of the *Euglena* tRNA^{Met}, although similar for both the tRNA^{Met} and tRNA^{Met} initiator of spinach chloroplasts (59 and 62%, respectively), is significantly higher for *Escherichia coli* tRNA^{Met} (70%) than for tRNA^{Met} initiator (56%). Finally, a putative *Euglena* chloroplast tRNA^{Met} initiator gene has recently been sequenced.[4] This tRNA would have a 6-bp amino acid acceptor stem, 84% sequence homology with the spinach chloroplast tRNA^{Met} initiator, and 68% homology with the *E. coli* tRNA^{Met} initiator.

tRNA Secondary Structures—The secondary structures predicted for the primary transcripts of the six tRNA genes are shown in Figs. 4 and 5. All six genes can be represented in the tRNA cloverleaf structure proposed by Holley *et al.* (20). However, two of the tRNAs would have single mismatched base pairs. The tRNA^{Gly} would be predicted to have a U_{30}–U_{40} mismatch internally in the anticodon stem. The tRNA^{Met} would be predicted to have an A_{31}–G_{39} mismatch at the closure of the anticodon loop (Fig. 4). One of the base assignments for this A_{31}–G_{39} mismatch, A_{31}[5] in the tRNA^{Met} gene, is confirmed as part of the *Hinf*I cleavage site $G_{30}A_{31}T_{32}T_{33}C_{34}$, which was located by restriction endonuclease analysis (Figs. 1 and 4). An autoradiogram of a 20% DNA sequencing gel for a DNA fragment from the tRNA^{Met} gene radioactively labeled at this *Hinf*I cleavage site is shown in Fig. 6. The DNA sequence is 5′-$T_{36}C_{34}A_{35}T_{36}A_{37}A_{38}G_{39}C_{40}C_{41}C_{42}$ ... and contains both the anticodon 5′-$C_{34}A_{35}T_{36}$ and base G_{39}. The G_{39} assignment is apparent.

There are other examples of either tRNAs or tRNA genes with mismatches at these sites in the anticodon stem. The phage T_4 tRNA^{Arg} contains a U_{30}–ψ_{40} mismatch (21); the *Saccharomyces cerevisiae* tRNA^{Glu} gene, a C_{30}–T_{40} mismatch (22), and the phage T5 tRNA^{Asn}, a C_{30}–T_{40} mismatch (23). There is no precedent for an A_{31}–G_{39} pair, but the yeast (24, 25) and mammalian (26, 27) elongator tRNA^{Met} and the *Neu-*

[3] J. Nickoloff and R. B. Hallick, unpublished observations.

[4] G. Karabin and R. B. Hallick, unpublished observations.

[5] The nucleotide positions are numbered according to the positions assigned yeast tRNA^{Phe} (23).

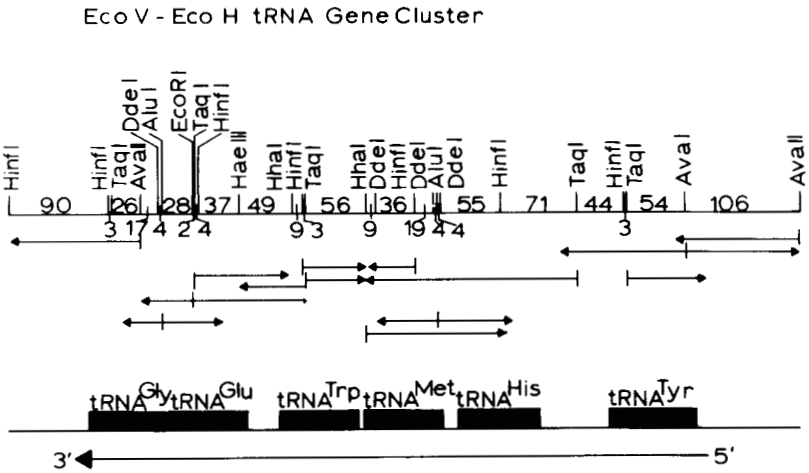

FIG. 2. **Sequencing strategy for the tRNA gene cluster region of BS9.** Fragments were sequenced from the sites indicated in the direction of the *arrows*. tRNA genes are shown *below* in alignment with the map. They would be transcribed from *right* to *left*.

```
                    20                  40                  60                  80                 100
GGAATTTTCTTTTCTTTGATGTTTTATTTGTAAGACTTGGAAGTTTTCCTATTCTACGAACTATTCGCAATCGTGGTCCTCTATATCTTGACATAGTTTT
CCTTAAAAGAAAAGAAACTACAAAATAAACATTCTGAACCTTCAAAAGGATAAGATGCTTGATAAGCGTTAGCACCAGGAGATATAGAACTGTATCAAAA
                   120                 140                 160          tRNA^Tyr→ 180                 200
ATCTATATTTTTTTTGATCTTAAATAACTTTTTCTATGTTAATTGTTTTTGTTGTTTAACATCTAATTGCGAGTTGTTGCCCGAGTGGTTAATGGGGGCG
TAGATATAAAAAAAACTAGAATTTATTGAAAAAGATACAATTAACAAAAACAACAAATTGTAGATTAACGCTCAACAACGGGCTCACCAATTACCCCCGC
                   220                 240                 260                 280                 300
GATTGTAAATCCGCAGTTCATCTTTCGCTGGTTCGAATCCAGCACGACTCAAAATATTTTTATATTTAATCAGACAGTTTTCGAGTTATAATAAATTTTT
CTAACATTTAGGCGTCAAGTAGAAAGCGACCAAGCTTAGGTCGTGCTGAGTTTTATAAAAATATAAATTAGTCTGTCAAAAGCTCAATATTATTTAAAAA
          tRNA^His→ 320                 340                 360                 380                 400
ATTATATTTTTTATAGGTGGGTGTAGCCAAGTGGTAAGGCAAAGGACTGTGACTCCTTCATTCGCGGGTTCGATCCCCGTCATTCACCTTCTATTAATTT
TAATATAAAAAATATCCACCCACATCGGTTCACCATTCCGTTTCCTGACACTGAGGAAGTAAGCGCCCAAGCTAGGGGCAGTAAGTGGAAGATAATTAAA
 tRNA^Met→                  420                 440                 460          480 tRNA^Trp→          500
TTAGGCTCAGTAGCTCAGAGGATAGAGCAGGGGATTCATAAGCCCTTGGTCACAGGTTCAAATCTTGTCTGAGCCAAACTGCGCTTTTAGTTCAATTGGT
AATCCGAGTCATCGAGTCTCCTATCTCGTCCCCTAAGTATTCGGGAACCAGTGTCCAAGTTTAGAACAGACTCGGTTTGACGCGAAAATCAAGTTAACCA
                   520                 540                 560          580 tRNA^Glu→          600
AGAACGTAGGTCTCCAAAACCTGATGTAGTAGGTTCGAATCCTACAGAGCGCGTTTGTTTTTTTTCTTTATCTTAAATTTGCCCCCATCGTCTAGAGGCC
TCTTGCATCCAGAGGTTTTGGACTACATCATCCAAGCTTAGGATGTCTCGCGCAAACAAAAAAAAGAAATAGAATTTAAACGGGGGTAGCAGATCTCCGG
                   620                 640          660 tRNA^Gly→          680                 700
TAGGACATCTCCCTTTCACGGAGGCAACGGGGATTCGAATTCCCCTGGGGGTAATTTATGCGGGTATAGCTCAGTTGGTAGAGCGTGGTCCTTCCAAGTC
ATCCTGTAGAGGGAAAGTGCCTCCGTTGCCCCTAAGCTTAAGGGACCCCCATTAAATACGCCCATATCGAGTCAACCATCTCGCACCAGGAAGGTTCAG
                   720                 740                 760                 780                 800
CAATGTTGCGTGTTCGAATCACGTTACCCGCTTTTAACTTTTTGATCTTTATAGAAAGTTAATTTTTAATTTTTATGATATTTTTAATATACTATGGGAT
GTTACAACGCACAAGCTTAGTGCAATGGGCGAAAATTGAAAAACTAGAAATATCTTTCAATTAAAAATTAAAAATACTATAAAAATTATATGATACCCTA
                   820                 840                 860                 880
GTGCGATTCGTAAAATTTGGATCTTATTTTTAAAAAAATTTTTTGCCATGTAAATTAAATTTTTACTCTTAATTTTCCGATATTTTAGT
CACGCTAAGCATTTTAAACCTAGAATAAAAATTTTTTTAAAAAACGGTACATTTAATTTAAAAATGAGAATTAAAAGGCTATAAAATCA
```

FIG. 3. **Sequence of the tRNA gene cluster and flanking DNA of restriction fragment BS9 from *Euglena* chloroplast DNA.** The top strand of the DNA is written *left* to *right* in the 5′ to 3′ direction. The tRNA-like sequences are in *italics*. *Italics* on both strands represent restriction sites used for 5′-end labeling in Maxam-Gilbert (12) sequence analysis.

rospora crassa mitochondrial $tRNA^{Thr}$ (28) contain the pair ψ_{31}–ψ_{39}. The phage T_5 $tRNA^{Leu}$ has a C_{31}–ψ_{39} pair (23). In addition, the *S. cerevisiae* (29) and *Aspergillus nidulans* (30) mitochondrial $tRNA^{Thr}$ genes contain the pair T_{31}–T_{39}, and a rat mitochondrial $tRNA^{Leu}$ gene (31) has the bases A_{31}–C_{39}.

Invariant and Semiconserved Bases—Based on an analysis of procaryotic, eucaryotic, and phage tRNAs involved in protein synthesis, Singhal and Fallis (19) have described both invariant and semiconserved bases in tRNAs that occur with greater than 90% probability. The 16 invariant bases are U_8, G_{10}, A_{14}, G_{18}, G_{19}, A_{21}, U_{33}, G_{53}, T_{54}, ψ_{55}, C_{56}, A_{58}, C_{61}, C_{74}, C_{75}, and A_{76}. The 20 semiconserved bases are (G or C)$_2$, R_9, Y_{11}, R_{15}, Y_{16}, R_{22}, R_{24}, Y_{25}, R_{26}, (G or C)$_{30}$, Y_{32}, R_{37}, (A or C)$_{38}$, (G or C)$_{40}$, Y_{48}, R_{52}, R_{57}, Y_{60}, Y_{62}, and R_{73}. Some of the six *Euglena* chloroplast tRNA genes have different bases at invariant and semiconserved positions. The $tRNA^{Tyr}$ gene codes for C_{10}–G_{25} in the D-stem rather than G_{10}–Y_{25}. This reciprocal base pair change also occurs in *E. coli* $tRNA^{Tyr}$ (32). The $tRNA^{Glu}$ gene codes for A_{53}–U_{61} base pair rather than the invariant G_{53}–C_{61}. These base assignments for $tRNA^{Glu}$ are confirmed by restriction endonuclease analysis. The A_{53} and U_{61} are within *Hinf*I and *Eco*RI cleavage sites, respectively. This is a very rare base change among known tRNAs and tRNA genes. Only the $tRNA_{1,2}^{Gly}$ and $tRNA_2^{Met}$ genes of *A. nidulans* mitochondria are

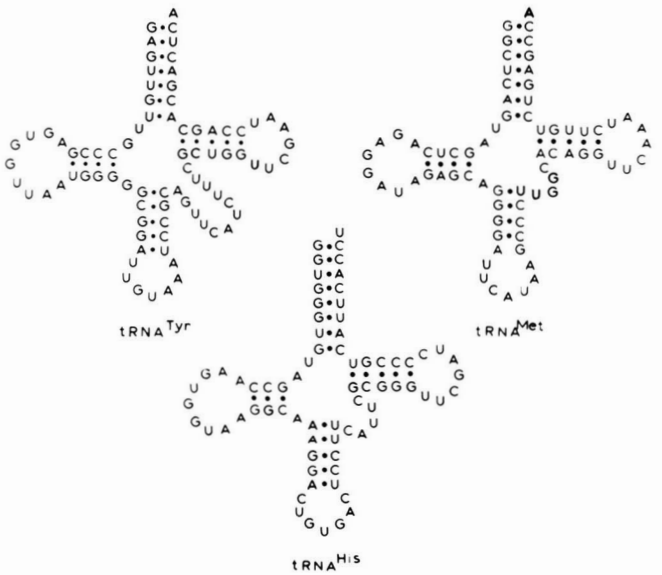

FIG. 4. **Nucleotide sequence of tRNA^{Tyr}, tRNA^{His}, and tRNA^{Met}.** The primary structures are deduced from the DNA sequence and shown in the normal tRNA cloverleaf secondary structures.

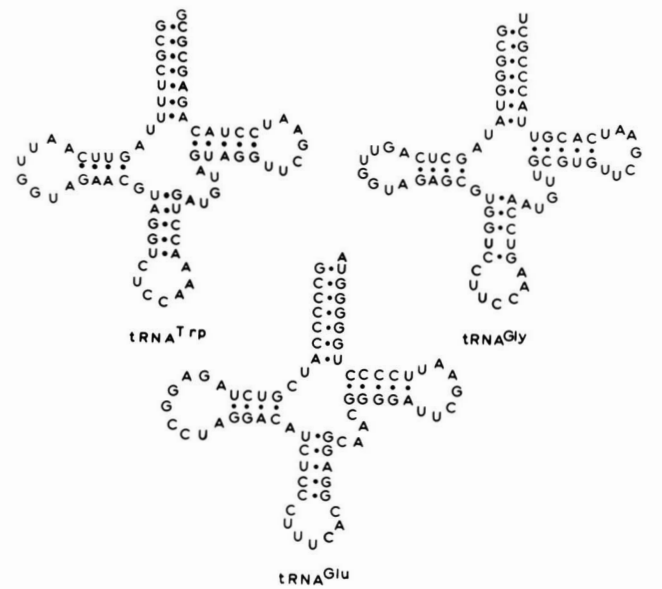

FIG. 5. **Nucleotide sequence of tRNA^{Trp}, tRNA^{Glu}, and tRNA^{Gly}.** The primary structures are deduced from the DNA sequence and shown in the normal tRNA cloverleaf secondary structures.

similar (30). In addition, the tRNA^{His} and tRNA^{Met} genes code for G_{16} and A_{16}, respectively, rather than Y_{16}; and the tRNA^{Trp} gene codes for A_{48}, rather than the Y_{48} which is normally part of the correlated semiconserved pair R_{15}–Y_{48}. Finally, the tRNA^{His}, tRNA^{Met}, and tRNA^{Gly} genes code for U or C_{73}, rather than R_{73}.

Sequence Homology with tRNAs from Other Organisms—The sequence homology of the six tRNAs from the BS9 fragment of *E. gracilis* chloroplast DNA to the corresponding tRNA species from higher plant chloroplasts, *E. coli,* mitochondria, and mammalian cytoplasm is shown in Table I. The same comparison was previously made for seven additional *Euglena* chloroplast tRNAs (6). Some generalizations may be made about the relatedness of *Euglena* chloroplast tRNAs to those encoded in other genomes. First, in most cases, *Euglena* chloroplast tRNAs are most closely related to the same tRNA species from plant chloroplasts. For the eight possible comparisons between *Euglena* and plant chloroplast tRNAs (Table I and Ref. 6), the average primary sequence homology is 76%, with a range of 49–93%. Second, the *Euglena* chloroplast tRNAs are next most similar to procaryotic tRNAs. The six tRNAs of BS9 have an average sequence homology of 68% with the same species from *E. coli.* The average for 12 *Euglena* chloroplast tRNAs (Table I and Ref. 6) is 70%, with a range of 62–78%. Third, the homology of chloroplast tRNAs to those of the mitochondria and cytoplasm of eucaryotes is much lower than the homology to procaryotic and other chloroplast tRNAs. Additional analysis of the tRNA cluster DNA sequence data is presented in the Miniprint Supplement to this paper.

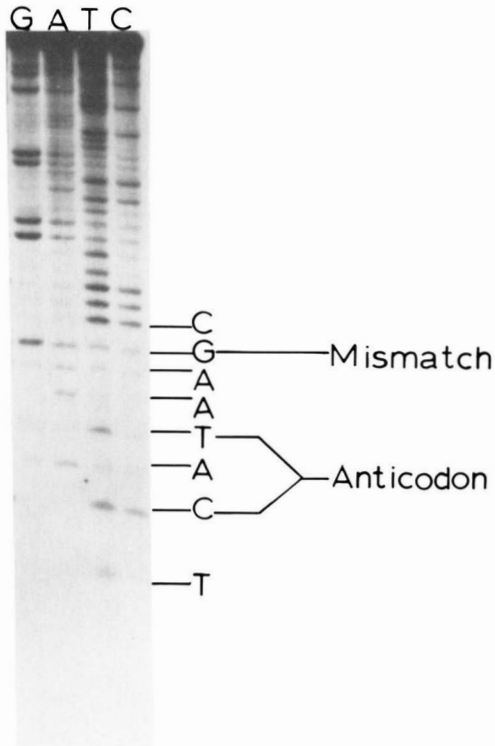

FIG. 6. **Autoradiogram of a 20% sequencing gel of a portion of the tRNA^{Met}.** The DNA was 5′-end labeled at the *Hinf*I cleavage site in the anticodon stem. The data are for the tRNA-like strand of the gene. The G-base-labeled *Mismatch* is G_{39}, which is opposite A_{31} of the *Hinf*I cleavage site in the normal tRNA cloverleaf secondary structure.

TABLE I

Sequence homologies (per cent) between tRNAs specific for the same amino acid

Most sequence data are from Ref. 23. Only tRNAs with the same anticodon are compared.

E. gracilis chloroplast	Plant chloroplast	*E. coli*	*S. cerevisiae* mitochondria	*A. nidulans* mitochondria	Mammal cytoplasm
Tyr		64	42		
His	74[a]	70	58		47
Met	59[b]	70		42	55
Trp	88[b]	66	47	49	56
Glu		78		58	62
Gly UCC		62	53		51

[a] Chloroplast from maize.
[b] Chloroplast from spinach.

Acknowledgments—We would like to thank Drs. Patrick W. Gray and Emil M. Orozco, Jr. for the initial cloning and characterization, respectively, of pPG76. We also thank Tim Heuser for expert advice in computer programming and Alice Sirimarco for preparation of the manuscript.

REFERENCES

1. Orozco, E. M., Jr., and Hallick, R. B. (1982) *J. Biol. Chem.* **257,** 3258–3264
2. Hallick, R. B., Rushlow, K. E., Orozco, E. M., Jr., Stiegler, G. L., and Gray, P. W. (1979) *ICN-UCLA Symp. Mol. Cell. Biol.* **15,** 127–141
3. El-Gewely, M. R., Lomax, M. L., Lau, E. T., Helling, R. B., Farmerie, W., and Barnett, W. E. (1981) *Mol. Gen. Genet.* **181,** 296–305
4. Orozco, E. M., Jr., Rushlow, K. E., Dodd, J. R., and Hallick, R. B. (1980) *J. Biol. Chem.* **255,** 10997–11003
5. Graf, L., Kössel, H., and Stutz, E. (1980) *Nature (Lond.)* **286,** 908–910
6. Orozco, E. M., Jr., and Hallick, R. B. (1982) *J. Biol. Chem.* **257,** 3265–3275
7. Koch, W., Edwards, K., and Kössel, H. (1981) *Cell* **25,** 203–213
8. Schwarz, Z., Kössel, H., Schwarz, E., and Bogorad, L. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4748–4752
9. Schwarz, Z., Jolly, S. O., Steinmetz, A. A., and Bogarad, L. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 3423–3427
10. Todoh, N., Shinozaki, K., and Sugiura, M. (1981) *Nucleic Acids Res.* **9,** 5399–5406
11. Kato, A., Shimada, H., Kusada, M., and Sugiura, M. (1981) *Nucleic Acids Res.* **9,** 5601–5607
12. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
13. Gray, P. W., and Hallick, R. B. (1977) *Biochemistry* **16,** 1665–1671
14. Hallick, R. B. (1982) in *The Biology of Euglena* (D. Buetow, ed.) Vol. IV, Academic Press, New York
15. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
16. Orozco, E. M., Jr., Gray, P. W., and Hallick, R. B. (1980) *J. Biol. Chem.* **255,** 10991–10996
17. Chang, S. H., Brun, C. K., Silberklang, M., RajBhandary, U. L., Hecker, L. I., and Barnett, W. E. (1976) *Cell* **9,** 717–723
18. Rich, A., and RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45,** 805–860
19. Singhal, R. P., and Fallis, P. A. M. (1979) *Prog. Nucleic Acids Res. Mol. Biol.* **23,** 227–290
20. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. F., and Zamir, A. (1965) *Science (Wash. D. C.)* **147,** 1462–1465
21. Mazzara, G. P., Seidman, J. G., McClain, W. H., Yesian, H., Abelson, J., and Guthrie, C. (1977) *J. Biol. Chem.* **252,** 8245–8253
22. Nobrega, F. G., and Tzagoloff, A. (1980) *FEBS Lett.* **113,** 52–54
23. Sprinzl, M., and Gauss, D. H. (1982) *Nucleic Acids Res.* **10,** r1–r55.
24. Gruhl, H., and Feldmann, H. (1976) *Eur. J. Biochem.* **68,** 209–217
25. Koiwai, O., and Miyazaki, M. (1976) *J. Biochem. (Tokyo)* **80,** 951–959
26. Piper, P. W. (1975) *Eur. J. Biochem.* **51,** 283–293
27. Petrissant, G., and Boisnard, M. (1974) *Biochimie* **56,** 787–789
28. Heckman, J. E., Sarnoff, J., Alzner-de Weerd, B., Yin, S., and RajBhandary, U. L. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 3159–3163
29. Berlani, R. E., Pentella, C., Macino, G., and Tzagoloff, A. (1980) *J. Bacteriol.* **141,** 1086–1097
30. Köchel, H., Lazarus, C. M., Basak, N., and Küntzel, H. (1981) *Cell* **23,** 625–633
31. Saccone, G., Cantatore, P., Gadateta, G., Gallerani, R., Lanave, C., Pepe, G. and Kroon, A. M. (1981) *Nucleic Acids Res.* **9,** 4139–4148
32. Goodman, H. M., Abelson, J., Landy, A., Brenner, S., and Smith, J. D. (1968) *Nature (Lond.)* **217,** 1019–1024
33. Boliver, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., and Boyer, H. W. (1977) *Gene (Amst.)* **2,** 95–113
34. Rodriguez, R. L., Bolivar, R., Goodman, H. M., Boyer, H. W., and Betlach, M. C. (1976) *ICN-UCLA Symp. Mol. Cell. Biol.* 471–477
35. Guerry, P., LeBlanc, D. J., and Falkow, S. (1973) *J. Bacteriol.* **116,** 1064–1066
36. Cohen, S. N., and Miller, C. A. (1970) *J. Mol. Biol.* **50,** 671–687
37. Wienand, U., Schwarz, Z., and Feix, G. (1979) *FEBS Lett.* **98,** 319–323

SUPPLEMENTARY MATERIAL TO

Euglena gracilis Chloroplast Transfer RNA Transcription Units. Nucleotide Sequence Analysis of a $tRNA^{Tyr}$-$tRNA^{His}$-$tRNA^{Met}$-$tRNA^{Trp}$-$tRNA^{Glu}$-$tRNA^{Gly}$ Gene Cluster

M. J. Hollingsworth and Richard B. Hallick

EXPERIMENTAL PROCEDURES

Materials: [γ-^{32}P]ATP (3000Ci/mmole) was obtained from New England Nuclear and Amersham. T-4 polynucleotide kinase was purchased from NEN and New England Biolabs. Restriction endonucleases were also purchased from New England Biolabs. Low melting point agarose, agarose, and E. coli alkaline phosphatase were purchased from Sigma. Acrylamide and bis-acrylamide were obtained from Fisher, Kodak and Polysciences. Dimethyl sulfate was also from Kodak. Hydrazine and piperdine were from Baker. Polypropylene columns with fiberglass disks were purchased from Isolab, Inc.

Plasmid DNA Construction, Isolation and Mapping: The plasmid pPG76 was constructed by digesting Euglena chloroplast DNA and pBR322 (33) with Bam H-I and Sal I. The products were ligated with T-4 DNA ligase and transformed into E. coli HB101 (34). Ampicillin resistant, tetracycline sensitive clones were selected by replica plating. Sizes of the inserts were determined by restriction endonuclease digestion and gel electrophoresis. Plasmid DNA was isolated by the cleared lysate method (35) and purified by ethidium bromide-cesium chloride density centrifugation (36). Restriction endonuclease digestions and electrophoretic techniques were performed as previously described (23,24). Salt concentrations used for the endonucleases were those recommended by New England Biolabs.

Purification of DNA Fragments: Three methods were used to purify individual DNA fragments. Initially, fragments were purified from low melting point agarose gels. The gel slice containing the fragment was melted in buffer. The resulting solution was extracted once with phenol and twice with ether. The DNA was collected from the aqueous phase by ethanol precipitation. DNA fragments were also purified by an adaptation of the Maxam-Gilbert crush and soak procedure (12). Instead of filtering the acrylamide solution through glass wool, the acrylamide was pelleted by centrifugation, and the liquid was gravity filtered through polypropylene columns (Isolab) plugged with glass fiber disks. DNA fragments were also eluted from gels electrophoretically (37).

Kination: The reaction mixture consisted of 20 µM [γ-^{32}P]ATP (3000 ci/mmole), 2-8 pmole DNA that had been previously treated with bacterial alkaline phosphatase, 2.5 units of T-4 polynucleotide kinase, 50 mM Tris (pH 7.6), 10 mM MgCl , 5 mM DTT, 0.1 mM spermidine, and 1 mM EDTA. The reaction was allowed to proceed at 37°C for 15 min. Another 2.5 units of enzyme were added and the reaction was continued at 37°C for 15 min. The mixture was then heated to 68°C to inactivate the kinase. Alternately, 75 picomoles of [γ-^{32}P]ATP and 20 units of kinase in the same reaction mixture were incubated at 37°C for 90 min.

Phosphatase Reaction: DNA (20-200 pm 5'-ends) was incubated with 0.01 units phosphatase in 10 mM Tris (pH 8) for 45 min at 68°C. The phosphatase was inactivated by phenol extraction. The DNA was collected by ethanol precipitation.

Sequence Analysis: Sequencing was carried out via the method of Maxam and Gilbert (12).

SUPPLEMENTARY RESULTS AND DISCUSSION

Comparison of Euglena gracilis strain Z and strain bacillaris tRNA coding loci: El-Gewely et al. (3) recently reported a Taq I/Ava I map of the Eco X fragment of chloroplast DNA of the bacillaris strain of Euglena gracilis. Eco X of the bacillaris strain is the same size as Eco V of the Z strain. It also has the same chloroplast DNA map location and internal Ava II cleavage site as Eco V. Eco X maps within a Bam HI-Sal I fragment equivalent to BS9 (3) and hybridizes with chloroplast tRNAs. It is likely that Eco V of the Z strain and Eco X of the bacillaris strain are the same DNA. It has been proposed that all of the significant differences between the chloroplast DNAs of Euglena gracilis Z and bacillaris occur in the rRNA coding region (14). Surprisingly, the restriction nuclease map and the tRNA coding region described by El-Gewely et al. (3) for Eco X do not agree with our data on Eco V. We obtained a different arrangement of Taq I fragments, although the number of cleavage sites and fragment sizes were the same. Another difference is that the Eco X tRNA coding locus was mapped to a single 120 nucleotide Taq I fragment, which is only large enough for one complete tRNA gene, as opposed to the 557 base pair region of Eco V. Also, no tRNA hybridization was described for Eco H of bacillaris chloroplast DNA (3), which is probably identical to Eco H of Z strain chloroplast DNA. In our study, we have mapped the Taq I sites without ambiguity, and confirmed the location of tRNA genes and most cleavage sites by DNA sequence analysis. Either the Z strain and bacillaris chloroplast DNAs are more different than previously believed, or the analysis of tRNA coding loci for the region of bacillaris DNA corresponding to BS9 needs to be re-evaluated.