tRNA genes transcribed from the plastid-like DNA of
Plasmodium falciparum

P. Preiser, D. H. Williamson and R. J. M. Wilson

Division of Parasitology, The National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Received August 8, 1995; Revised and Accepted October 4, 1995 EMBL accession nos X90351–X90354 (incl.)

ABSTRACT

Besides their mitochondrial genome, malarial parasites contain a second organellar DNA. This 35 kb circular molecule has a number of features reminiscent of plastid DNAs. Sequence analysis shows that along with other genes the circle codes for 25 different tRNAs all of which are transcribed. Six of the tRNAs have some unusual features, and one has an intron, the only one found so far on the circle. Comparison of codon and anticodon usage indicates that the 25 tRNAs are sufficient to decode all the protein genes present on the circle. The maintenance of such a parsimonious but complete translation system is further evidence for the functionality of the circle.

INTRODUCTION

The malarial parasite Plasmodium falciparum has a circular, 35 kb extrachromosomal DNA which shows some similarities to a vestigial plastid genome. The features suggestive of a plastid ancestry include duplicated tRNA genes on an inverted repeat (IR), a typical chloroplast S10, spc ribosomal protein operon, and three subunits of a prokaryotic type RNA polymerase. There is furthermore an open reading frame with clear homology to an ORF found on the plastids of rhodophytes (1). However, the malarial circle is highly derived and in this respect resembles the reduced plastid genome of the parasitic, non-photosynthetic plant Epifagus virginiana (2). In Epifagus the genes involved in photosynthesis, as well as those coding for the RNA polymerase and some tRNAs, have been deleted. However, the fact that other genes described on this plastid DNA are transcribed and expressed has been taken to suggest that it has retained a function important for the plant's survival. Deletions in the malarial plastid-like genome are more extensive as well as different from those in Epifagus. Nevertheless, most of the genes so far identified are transcribed, implying that the function of the malarial 'plastid' is probably different from that of Epifagus.

In this paper we describe 16 putative tRNA genes located on the 35 kb plastid-like DNA of Pfalciparum which, including the previously described nine duplicated tRNA genes (3) of the inverted repeat (IR), makes a total of 25 species. Analysis of total RNA shows that all the recognized tRNA genes are transcribed. Codon usage is discussed in the light of our knowledge of the tRNAs encoded on the malarial circle, leading to the prediction of an extremely parsimonious but complete translation system. The maintenance of such a system is further evidence that the malarial putative plastid remnant is likely to be functional.

MATERIALS AND METHODS

Cell culture and RNA preparation

The C10 strain of Pfalciparum (4) was cultured in vitro as described (5). RNA was prepared from acid-lysed infected erythrocytes using the RNaid PLUS kit (BIO101) as previously described (1).

DNA analysis

Clones were derived from a shotgun library of purified, sonicated, 35 kb circle DNA that was end repaired and ligated into the Bluescript vector, pBS KS(II)+, as previously described (6). Gaps between sequenced regions were filled from PCR products generated with oligonucleotide primers designed from flanking clones carrying protein-encoding genes. The organization of genes on the circle was determined by constructing contigs ‘walking’ away from both ends of the inverted repeat (IR). The entire circle has been sequenced, apart from a short region in the centre of the inverted repeat, estimated from sized restriction fragments to be only tens of nucleotides (nt) in size. The clones and PCR products were sequenced on both strands and tRNA genes were identified using the Staden algorithm (7) and the Staden-Plus software (Amersham). The standard conserved bases were specified and associated with large score numbers to ‘filter’ the number of candidate tRNA structures.

Primers used

For all experiments primers complementary to the 3' end of the tRNA were used. The size of the expected primer extension product is noted after each primer sequence as follows:

PRE1 Ala GGAAATAAGTAGAATTGAACTAC (70 nt)
PRE2 Asn CTTCTTAAACTAGAATTGAAC (72 nt)
PRE3 Thr GCTAATGGTGAGATTTGAACTC (72 nt)
PRE4 Pro CAGAAAAATAGGATTTGAACC (71 nt)
PRE5 iMet CATTTAAGTAGATTCGAACTAC (85 nt)
PRE6 Asp GAATTTGACGAGAATTGAACTCG (74 nt)
PRE7 Leu TATGAATAAAGAGATTTGAACTC (79 nt)
PRE8 Met GCGAAAAACGGAATTGAACCG (72 nt)
PRE9 Val GTAATATTGGATTCGAACCAATGG (69 nt)
PRE10 Ser GAGAAAAGGGATTCGAACCC (90 nt)

* To whom correspondence should be addressed
PRE11  His  TAAATAAAGAGATTTGAAACTCATATAAAAAG (71 nt)
PRE12  Lys  GAACTACCTGAGATTTGAACTC (72 nt)
PRE13  Arg^  TAAATCTAGTAAAGATTGAACTTAC (72 nt)
PRE14  Arg  TAAACTTGAAAAGAATTGAACTTT (72 nt)
PRE15  Tyr  GTTAAATCAGATTTGAACTGATGTAG (82 nt)
PRE18  Phe  GTTCAAAAATCAGATTTGAACTGATAAC (71 nt)
PRE20B  Gin  TAGAATAATGGGATTTGAACCC (71 nt)
PRE22B  Trp  ATGTCCTAAAGGATTTGAACCC (70 nt)
PRE23B  Leu^  GAGATAAAGGGATTTGAACCC (83 nt)
PRE24  Cys  GATATTCAGATTCAAACTGAAATATAATTTGC (68 nt)
PRE25 Ser^  GAGAAAATGGGATTTGAACCC (81 nt)
PRE26  Gly  CAAATAATGGAAATTGAATCC (70 nt).

All primers were labelled at the 5'-end with T4 polynucleotide kinase and [\gamma-^32P]ATP.

Detection of tRNA transcripts

Two-dimensional gel electrophoresis of total parasite RNA and electroblotting onto Nytran filter (Schleicher and Schuell) was carried out as previously described (8).

Radioactive (\(^{32}P\)-labelled) probes were prepared by random priming (9) using cloned DNA fragments as template, and hybridization was carried out for 16–24 h at 50°C (10). Probing for individual tRNAs was achieved by using specific \(^{32}P\)-labelled primers, hybridization was carried out at 50°C in 5x SSPE, 1% SDS, 100 μg/ml salmon sperm DNA and 10 μg/ml poly(A) (11).

For primer extension analysis, 100 ng of primer and -1 μg of total parasite RNA were hybridized using two different techniques. For primers PRE1–PRE10 and PRE12–PRE15 hybridization and extension reactions were carried out as described by Ragnini and Frontali (12). For primers PRE11, PRE19B–PRE22B and PRE23–PRE26 the recommendations supplied by the manufacturer of the SuperScript II RNaseH\(^{-}\) Reverse Transcriptase (BRL) were used, with the best hybridization temperature being 42°C. Primer extension products were resolved on a 6% (w/v) denaturing polyacrylamide gel, with a sequence ladder being used as a marker.

RESULTS

The 35 kb circle codes for a number of potential tRNAs

Sequence analysis of the 35 kb circle has identified a number of tRNA coding regions (Fig. 1). One of them lies within the inverted repeat between the large and small (LSU and SSU) rRNA genes and codes for seven duplicated tRNAs (3). From the total nucleotide sequence of the 35 kb circle, we have now identified a cluster of 10 potential tRNA genes next to the IR\(^{B}\) region between two ribosomal protein encoding genes (rps4 and rpl4). The clustered tRNA genes are separated from each other by only a few bp and all 10 tRNA genes are located on the same strand, unlike those in the inverted repeat (3). Sequence analysis of this region using the Staden-Plus software shows that nearly all of these genes code for tRNAs with standard conserved sequence elements that fold with minimal free energies into the typical tRNA cloverleaf structure (Figs 2 and 3). The tRNAs coded for by this region are His, Cys, Leu\(^{\text{uua}}\), Met, Tyr, Ser\(^{\text{gcu}}\), Asp, Lys,
Glu and Pro. Two other tRNA genes, Ile and Thr, also have been identified in association with the IR. The first is located between the two inverted repeats while the latter is found downstream of both copies of the LSU tRNA gene (3). Four more tRNAs are located close to the 3' end of the tufl gene; tRNA{sup Phe} is found on the opposite strand from tufl, it is immediately followed by tRNA{sup Gln}, Gly and Trp on the other strand. Two additional tRNAs for Ser{sup (gga)} and Gly{sup (ucc)} are separated by a 268 bp gap from the 3' end of the Clp gene, making a total of 25 different tRNA genes found on the circle. A substitution in the T-stem of the Glu and Pro tRNAs, allowing us to identify all 19 'normal' tRNAs coded for tufA, the opposite strand from rpl4. This probe is expected to hybridize to all the tRNAs between the Met and Glu gene and six spots can be identified accordingly. Assignment of each spot to a tRNA was confirmed by re-probing the filter with {sup P}-labelled oligonucleotides specific for each tRNA. The same filter was then hybridized with restriction fragments or PCR products corresponding to one of the tRNA-containing regions shown in Figure 1. Figure 4A shows such a gel hybridized with a cloned DNA fragment containing part of the tRNA cluster between rps4 and rpl4. This probe is expected to hybridize to all the tRNAs between the Met and Glu gene and six spots can be identified accordingly. Assignment of each spot to a tRNA was confirmed by re-probing the filter with {sup 32P}-labelled oligonucleotides specific for each tRNA. The same filter was then hybridized with probes for other tRNAs, allowing us to identify all 19 'normal' tRNAs coded for by the circle (data not shown).

Analysis of tRNA transcription

Northern blot and primer extension analysis were used to analyze transcription of the tRNA genes on the circle. Total parasite RNA separated on a two dimensional (2D) polyacrylamide gel, followed by electrotransfer onto a Nytran membrane, was hybridized with restriction fragments or PCR products corresponding to one of the tRNA-containing regions shown in Figure 1. Figure 4B shows a gel hybridized with a cloned DNA fragment containing part of the tRNA cluster between rps4 and rpl4. This probe is expected to hybridize to all the tRNAs between the Met and Glu gene and six spots can be identified accordingly. Assignment of each spot to a tRNA was confirmed by re-probing the filter with {sup 32P}-labelled oligonucleotides specific for each tRNA. The same filter was then hybridized with probes for other tRNAs, allowing us to identify all 19 'normal' tRNAs coded for by the circle (data not shown).

In addition, primer extension analysis of total parasite RNA was carried out using kinased primers specific for each tRNA. All the tRNAs of the IR were detected by this method (Fig. 4B) giving products corresponding to the expected size. For tRNAs{sup Ala}, Leu and Arg products 3, 7 and 4 nt smaller than the full length product were also detected. It is not clear if these are primer extension products 3, 7 and 4 nt smaller than the full length product and Al{sup 1}. It is not clear if these are primer extension
Blot analysis was used to show that the correct sized products only one spot corresponding to the full length tRNA was detected band of 71 nt of about equal intensity (data not shown), but again the other tRNAs, representative data for His, Thr, Pro, Gin and artefacts or real stop sites. The same experiment was done for all the other tRNAs, representative data for His, Thr, Pro, Gin and Ser(usage) being given in Figure 5A. Again, the correct products were obtained in all cases, with smaller bands being present in variable amounts. This was especially obvious in the case of the tRNA\textsuperscript{Met}, where the correct sized product was found in very low amounts, with a smaller product being the dominant signal (data not shown). Interestingly on a 2D gel only one spot migrating as the full length (tMet) product is detected (Fig. 4A). (tRNA\textsuperscript{Ser(usage)}) also gave, in addition to the full length product (90 nt), an extra band of 71 nt of about equal intensity (data not shown), but again only one spot corresponding to the full length tRNA was detected (Fig. 4A).

The six unusual tRNAs found on the circle could be non-transcribed pseudogenes, so either primer extension or 2D Northern blot analysis was used to show that the correct sized products were transcribed from these tRNA genes. Primer extension analysis gave the correct sized products for all six tRNAs (Fig. 5B). Probes for tRNAs Gly\textsuperscript{acc}, Gly\textsuperscript{acc} and Glu revealed only the expected 70 nt product, whereas for the Trp tRNA a 70/71 nt doublet was clearly visible, the 70 nt product being of the expected size. The primer extension result for the Cys tRNA was not as clean, with the correct sized 68 nt product being superimposed on a smear of labelled material. Primer extension using the Leu tRNA specific primer was expected to give a 83 nt product if the intron had been removed. The correct sized product was clearly visible (Fig. 5B), in addition to a larger band of equal intensity. This result was confirmed using a different primer for tRNA Leu; in both instances the correct sized band was superimposed on a smear (data not shown). It is not clear whether the strong smearing seen in primer extensions for these two tRNAs was an artefact due to the primers used, as similar smears of lesser extent were also seen with other primers (Figs 4 and 5) or if they truly reflect the state of the RNA in this region. The fact that the 2D Northern blots probed with primers specific for tRNA\textsuperscript{Glu,Met} and tRNA\textsuperscript{Leu} detected only one specific spot (Fig. 4 and data not shown), would support the idea that the smears are an RT artefact.

Long exposures of the primer extension assays revealed a number of potentially larger precursor molecules for tRNAs derived from the tRNA gene cluster (between rps4 and rpl4). Initial analysis indicates that the main stop sites fall between tRNA genes. Hybridizing total RNA run on low percentage polyacrylamide gels with tRNA-specific sequences confirmed these results with a number of larger transcripts detected (data not shown).

### Amino-acid frequencies and codon-usage

Due to their A-T richness, the bias in circle-specific codons is even greater than that reported for the *P.falciparum* nucleus-encoded proteins (19). As shown in Table 1, Ile, Asn, Lys, Leu and Tyr comprise 66.4% of all amino acids (aa) used in identified circle-encoded proteins. A further 11% is accounted for either by Phe or Ser, reducing the remaining 14 aa to only 22.6% of the total examined (~3000). It is striking that the codons for the seven frequent aa all have an A or T at the first position while the five most frequent codons also have an A or T in the second position. Only 2.2% of the codons used to encode these seven aa have a G or C in the third position. Usage of the remaining 13 aa ranges in frequency from 3.5% (Gly) to as low as 0.3% (Trp).

It is of interest to note the relatively high frequency (9%) of Glu (G at first position of codon), and low frequency (4.7%) of Tyr (AT at first and second position of codon) found in nucleus-encoded proteins, compared to those of the circle (2.3 and 10.1%, respectively). Based solely on the A-T content of the circle the frequency of Gly (3.5%) is somewhat high, considering that the two first positions of the codon are GC.

Bias due to the high A-T content of the circle genome is also observed when analysing the frequency of specific codons (Table 2). Only 4.6% of all codons analyzed contain a G or C at the third position, and this is even more striking considering that 2.5% of them are the codons AUG (Met), TGG (Trp) and one of the Lys codons (AAG). Similarly, only 17.3% of all codons contain a G or C at the second position and 15.3% at the first position. The codons most often used all contain only A and T, with AAT (Asn), AAA (Lys), TTA (Leu) and ATA (Ile) being most frequent. There seemed to be little preference in the use of C (342) or G (314) at the first two positions of the codon (Fig. 2). Only 2.2% of the codons used to encode these seven aa have a G or C in the third position. Usage of the remaining 13 aa ranges in frequency from 3.5% (Gly) to as low as 0.3% (Trp).

Due to their A-T richness, the bias in circle-specific codons is even greater than that reported for the *P.falciparum* nucleus-encoded proteins (19). As shown in Table 1, Ile, Asn, Lys, Leu and Tyr comprise 66.4% of all amino acids (aa) used in identified circle-encoded proteins. A further 11% is accounted for either by Phe or Ser, reducing the remaining 14 aa to only 22.6% of the total examined (~3000). It is striking that the codons for the seven frequent aa all have an A or T at the first position while the five most frequent codons also have an A or T in the second position. Only 2.2% of the codons used to encode these seven aa have a G or C in the third position. Usage of the remaining 13 aa ranges in frequency from 3.5% (Gly) to as low as 0.3% (Trp).

It is of interest to note the relatively high frequency (9%) of Glu (G at first position of codon), and low frequency (4.7%) of Tyr (AT at first and second position of codon) found in nucleus-encoded proteins, compared to those of the circle (2.3 and 10.1%, respectively). Based solely on the A-T content of the circle the frequency of Gly (3.5%) is somewhat high, considering that the two first positions of the codon are GC.

Bias due to the high A-T content of the circle genome is also observed when analysing the frequency of specific codons (Table 2). Only 4.6% of all codons analyzed contain a G or C at the third position, and this is even more striking considering that 2.5% of them are the codons AUG (Met), TGG (Trp) and one of the Lys codons (AAG). Similarly, only 17.3% of all codons contain a G or C at the second position and 15.3% at the first position. The codons most often used all contain only A and T, with AAT (Asn), AAA (Lys), TTA (Leu) and ATA (Ile) being most frequent. There seemed to be little preference in the use of C (342) or G (314) at the first two positions of the codon (Fig. 2). Only 2.2% of the codons used to encode these seven aa have a G or C in the third position. Usage of the remaining 13 aa ranges in frequency from 3.5% (Gly) to as low as 0.3% (Trp).

It is of interest to note the relatively high frequency (9%) of Glu (G at first position of codon), and low frequency (4.7%) of Tyr (AT at first and second position of codon) found in nucleus-encoded proteins, compared to those of the circle (2.3 and 10.1%, respectively). Based solely on the A-T content of the circle the frequency of Gly (3.5%) is somewhat high, considering that the two first positions of the codon are GC.
Figure 4. Transcriptional analysis of tRNAs. (A) Northern blot of total parasite RNA separated on a 2-dimensional gel and hybridized with a probe specific for six of the tRNAs found on the circle between rps4 and rpl4. Individual tRNAs are identified by their single letter code. The directions of migration are indicated by arrows. (B) Primer extension analysis of the tRNA cluster found in the inverted repeat. Single letter code is used to identify each tRNA lane. Primer extension product sizes were determined from a sequencing ladder. Products of 70, 72, 79, 72, 69, 72 and 72 nt(*) were those expected, corresponding to tRNAs A, N, L, M, V, R' and R.

Table 1. Amino acid frequency per 100 aa of nuclear and plastid genes in *P.falciparum*

<table>
<thead>
<tr>
<th>Amino-acid frequencies</th>
<th>Pf_nuc</th>
<th>Pf_plas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile</td>
<td>7.1</td>
<td>18.0</td>
</tr>
<tr>
<td>Asn</td>
<td>9.7</td>
<td>13.8</td>
</tr>
<tr>
<td>Lys</td>
<td>10.5</td>
<td>13.0</td>
</tr>
<tr>
<td>Leu</td>
<td>8.1</td>
<td>11.5</td>
</tr>
<tr>
<td>Tyr</td>
<td>4.7</td>
<td>10.1</td>
</tr>
<tr>
<td>Phe</td>
<td>3.6</td>
<td>6.0</td>
</tr>
<tr>
<td>Ser</td>
<td>7.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Gly</td>
<td>4.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Thr</td>
<td>4.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Glu</td>
<td>9.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Asp</td>
<td>6.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Val</td>
<td>4.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Arg</td>
<td>2.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Gln</td>
<td>3.4</td>
<td>1.8</td>
</tr>
<tr>
<td>Pro</td>
<td>3.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Met</td>
<td>1.9</td>
<td>1.3</td>
</tr>
<tr>
<td>Cys</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Ala</td>
<td>4.1</td>
<td>1.2</td>
</tr>
<tr>
<td>His</td>
<td>2.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Trp</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The data for the nuclear genes were taken from Musto el al. (19), while for the plastid DNA, protein genes coding for ~3000 aa were used.

Anti-codon frequency

Genes for a total of 25 tRNAs, with 24 different anti-codons, have so far been identified on the circle, the two identical anticodons being the Met initiator and elongator tRNAs, respectively (Table 2). Unlike the codon frequency, which is clearly driven by the A+T
Table 2. Codon usage of the 35 kb circle

<table>
<thead>
<tr>
<th>Codon</th>
<th>anti number</th>
<th>codon</th>
<th>anti number</th>
<th>codon</th>
<th>anti number</th>
<th>codon</th>
<th>anti number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT-Phe gaa 224</td>
<td>TCT-Ser uga 77</td>
<td>TAT-Tyr gua 379</td>
<td>TGT-Cys gca 45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTC-Phe gaa 3</td>
<td>TCC-Ser uga 2</td>
<td>TAC-Tyr gua 5</td>
<td>TGC-Cys gca 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTA-Leu uaa 425</td>
<td>TCA-Ser uga 48</td>
<td>TAA-*** 113</td>
<td>TGA-*** 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTG-Leu uaa 9</td>
<td>TCG-Ser uga 2</td>
<td>TAG-*** 11</td>
<td>TGG-Trp caa 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTT-Leu uag 0</td>
<td>CCT-Pro ugg 38</td>
<td>CAT-His gua 37</td>
<td>CGT-Arg acg 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTC-Leu uag 0</td>
<td>CCC-Pro ugg 0</td>
<td>CAC-His gua 0</td>
<td>CGC-Arg acg 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTA-Leu uag 2</td>
<td>CCA-Pro ugg 15</td>
<td>CAA-Gln uga 63</td>
<td>CGA-Arg acg 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGT-Leu uag 0</td>
<td>CCG-Pro ugg 0</td>
<td>CAG-Gln ugg 4</td>
<td>CGG-Arg acg 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT-Ile gau 262</td>
<td>ACT-Thr ugu 62</td>
<td>AAT-Asn guu 518</td>
<td>AGT-Ser guu 60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATA-Ile gau 416</td>
<td>ACA-Thr ugu 0</td>
<td>AAC-Asn guu 3</td>
<td>AGC-Ser gcu 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG-Met cau 49</td>
<td>ACG-Thr ugu 1</td>
<td>AAA-Lys uuu 456</td>
<td>AGA-Arg ucu 57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTC-Val uac 28</td>
<td>GCT-Ala ugc 27</td>
<td>GAT-Asp ugu 77</td>
<td>GGT-Gly acc 74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCT-Val uac 4</td>
<td>GCC-Ala ugc 1</td>
<td>GAC-Asp uyc 0</td>
<td>GGC-Gly acc 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGA-Ala ugc 44</td>
<td>GCG-Ala ugc 16</td>
<td>GAA-Glu uuc 81</td>
<td>GGA-Gly uce 49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTG-Val uac 1</td>
<td>GGC-Ala ugc 0</td>
<td>GAG-Glu uuc 6</td>
<td>GGG-Gly ucc 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The codon usage of the plastid circle has been determined by analyzing ~3000 codons from protein coding genes of the circle. Codons are given in capital and lower case letters. The amino acids are given in three letter code. The anticodons found on the circle are aligned with all the codons they can potentially decode, when given in bold it indicates a complete match with the corresponding codon. The ? next to the Ile(ATA) codon indicates the uncertainty about whether the anticodon gau is able to decode it.

content of the circle, the anticodons have somewhat resisted this pressure. This is shown most strikingly at the first nucleotide of the anti-codon, which is the one most at liberty to change due to 'wobble' at this position. Of the 24 different anti-codons eight have G (33.3%) at this positions, two have a C (Met, Trp), while 12 have a U (50.0%) and two have an A (Arg, Gly). Comparing codon to anti-codon frequency a clear trend can be established; while the anticodons containing a U at the first position are matched 81% of the time by their respective codon, this occurs <1% of the time when there is a G. In the cases where there is an A or a C at this position it is matched every time. We conclude that 'wobble' in the first position of the anticodon would provide an effectively complete decoding system as discussed below.

DISCUSSION

The 35 kb circular DNA of *Plasmodium falciparum* codes for nearly 60 genes involved in transcription and translation, including several tRNAs. Very little is known about tRNAs in *Plasmodium*, indeed the only detailed information is limited to those encoded on the inverted repeat of the 35 kb circle (3). Here we present evidence for the identification of 16 more tRNAs located outside the inverted repeat and give a detailed analysis of their transcription. Identification of all tRNA genes was based on their potential to give products that would fold into the correct cloverleaf structure, with conserved motifs characteristic for functional tRNAs. This approach led to the identification of a total of 25 putative tRNAs; of these, nine are duplicated within the IR (3), 10 are found in a cluster just downstream of IRB and an additional six occur in a more distant region of the circle.

Of the 25 tRNAs identified, six are atypical in one way or another. The intron found within the tRNA-Leu gene is the only one so far identified on the malarial plastid DNA. It is an unexpected finding considering the large size reduction observed in this genome. Introns of this type have been identified in Leu tRNA genes from other plastids (13), however, the intron in this malarial gene is only about half the size of the equivalent introns found in other plastids and the overall conservation is very low. It does however have some features which make us believe it is related to the class I intron usually found in the Leu tRNA gene (13,16). As indicated in Figure 3B, the intron sequence can be folded into a putative structure resembling a class I intron containing the conserved stem structures P1, P3, P4, P6, P7, P8 and P9. Some of the stems are significantly reduced in size, like P4 and P8, while two stems are completely missing (P2 and P5).

Nevertheless, the following features are conserved: the sequence can be folded to give a P1 stem-loop along with its UG pair at the splice site. Furthermore, the stem–loop structure P9 followed by the G at the 3' end of the intron is typical of class I introns. Initially, class I introns were thought to be self-splicing requiring no additional protein cofactors, but this is not necessarily the case in plastids (13). The G-C base pair of stem P7 is part of the guanosine binding site thought to play an important role during splicing (20,21) and is conserved in the putative malarial intron structure.

In addition to their typical cloverleaf secondary structure, tRNAs have a number of conserved nucleotides [for review see (17)], including the GG dinucleotide [nucleotides 18 and 19; see (22)] of the D-loop and the GTTC tetranucleotide (nucleotides 53–56) in the T-loop. The GG dinucleotide is thought to be involved in a tertiary interaction with the TC of the T-loop. Five of the tRNAs found on the circle have changes in either one or both of these conserved regions (Fig. 3A). The Glu and Gly tRNAs have a change from GTTC to ATTC which would disrupt both of these conserved regions (Fig. 3A). The Glu and Gly tRNAs have a change from GTTC to ATTC which would disrupt both of these conserved regions (Fig. 3A). The Glu and Gly tRNAs have a change from GTTC to ATTC which would disrupt both of these conserved regions (Fig. 3A).
proper tertiary interactions. Changes like this have also been observed in mitochondrial tRNAs from a number of different organisms (23). While it is not clear why these mutations have occurred, it is feasible that they are a direct result of the pressure, seen in the circle as a whole, of replacing G and C with A and T.

The work reported here shows that all tRNAs on the circle are transcribed. Considering the tight spacing between individual genes it is not surprising that some of the tRNAs appear to be transcribed in a polycistronic fashion, a common feature of plastids. This is suggested by the detection of larger precursor molecules. It is not yet clear whether the smaller and larger transcripts observed in some of the primer extensions are due to mispriming or reverse transcriptase stops, or are due to alternative processing sites in the maturation of a larger precursor molecule.

The 24 different anticodons found so far on the circle have the potential to decode a total of 57 different codons (see Table 2). This is achieved by effective use of the wobble position in combination with the frequency of certain codons. Studies in vitro have shown that one Val anticodon can recognize all four codons (24), and further evidence from a number of organisms has confirmed the ability of one tRNA to decode all four codons in a family group (25). This would mean that the Leu UAG anticodon can decode the four codons (CTN), similarly the Pro UGC, the Thr UGU, the Val UAC, the Arg ACG, and the Ala UGA anticodons should be able to decode the codons CCN, ACN, GTN, CGN and GCN, respectively. In all these cases there is a U in the wobble position of the tRNA identical to that reported in the literature (26). This leads to the conclusion that a ‘two out of three’ mechanism may be acting in these cases (27,28), with a U at the wobble position being least restrictive in a mismatch situation (29).

It is not clear why there are two different anticodons (acc, ucc) used by the circle to decode the four Gly codons GGN, since one anticodon might have sufficed. However, recent results in E. coli and mycoplasma have demonstrated that the nucleotide in position 32 has a significant effect on the decoding capacity of the Gly tRNA; the presence of a U at this position ensures that the UCC anticodon discriminates according to the wobble rules, while C at this position evokes the ‘two out of three’ mechanism (25). In the cases reported here, codon recognition for both Gly tRNAs with U at this position would be according to the wobble rule, with UCC decoding the codons GGA/G while the ACC anticodon can only decode GGT. The fact that only three (GGA, GGG and GGT) out of the four Gly codons are used on the circle implies that the same mechanism is used.

The Ile GAU anticodon is able to recognize the two codons ATT and ATC but it is not clear whether it can also recognize the ATA codon: modification of G to I (I = Inosine) would, according to the ‘wobble’ mechanism (30), allow the recognition of all three Ile codons (31). In such instances, Adenosine is replaced by Inosine in a reaction catalyzed by the enzyme tRNA-hypoxanthine ribosyltransferase (32,33), but it is unclear whether a replacement at a Guanosine nucleotide could also occur. Alternatively some of the GAU anticodons could be modified to LAU (L = Lysidine), allowing it to recognize the AUA Ile codon (34). The Met AUG codon is specifically recognized by the CAU anticodon and both the initiator and elongator tRNA Met have been identified on the circle.

Using the recognition mechanism described above, the 25 different tRNAs found on the malarial plastid DNA are sufficient to decode all codons used. This makes it one of the most parsimonious tRNA systems reported for any translation system and would alleviate the necessity for tRNA import, something thought to happen in mitochondria but not in plastids.

The function of the malarial plastid DNA is not known, but the maintenance of a transcription and translation system suggests it is of some importance to the organism. The data presented here give evidence for a striking co-evolution between the tRNA anticodons and the codons used by predicted proteins encoded by the circle. This system minimizes the number of tRNAs to as few as 25, using a decoding system that may have evolved in response to a drift to an extremely biased nucleotide content.

ACKNOWLEDGEMENTS

We thank Malcom Strath, Andrea Whyte, Daphne Moore and Kaveri Rangachari for technical assistance. This work was funded by the UNDP/World Bank/WHO Special Programme for Research in Tropical Diseases (TDR).

REFERENCES