Volume 17 Number 17 1989 Nucleic Acids Research

Tandem trinucleotide repeats throughout the nucleotide sequence of a cDNA encoding an *Eimeria tenella* sporozoite antigen

P.A.Liberator, J.Hsu and M.J.Turner

Department of Biochemical Parasitology, Merck Sharp & Dohme Laboratories, Rahway, NJ 07065, USA

Submitted July 31, 1989 EMBL accession no. X15898

Polyclonal anti-*Eimeria tenella* sporozoite antibody was used to isolate cDNA clone S07 from a λgt11 expression library constructed using mRNA from *E. tenella* sporulating oocysts(1). The nucleotide sequence of clone S07 was determined by dideoxy sequencing of ExoIII generated deletion mutants(2) from both ends of the cDNA in pUC119 and is shown here. The sequence is 957 nucleotides long and the deduced amino acid sequence predicts the occurrence of only one in-frame methionine residue at nucleotide position 65. This ATG codon is in an environment that is favorable for translation initiation by eukaryotic ribosomes(3). Initiation at this position would result in an open reading frame of 216 amino acids with a predicted molecular weight of 22.4 kD. A striking feature of the sequence is the tremendously high frequency of the trinucleotide AGC dispersed in a tandemly repeated fashion throughout the length of the clone, i.e., in both protein coding and untranslated regions. Interestingly, those trinucleotide repeats located within the protein coding region are not in the same reading frame. The majority code for either serine (AGC) or glutamine (CAG) residues. In both of these, the third position is a G or C nucleotide which is in good agreement with the overall 69.5% G/C content for this cDNA and the 83% third position G/C bias within the open reading frame.

```
  1  CTC GCC CCA ACT TTT TCC CCC GCG CTC CGC AGC AGC AGC AGC AGC AGC AGC AGC AGC AGC AAA ATG GCA GAC CTC TTC AGC GGA CTC GTG GGC GGC GTC GTC GGC GCT  108
                                                                                          M   A   D   L   F   S   G   L   V   G   G   V   V   G   A

109  GTT GCT GCA GCA GAT TTG CCT GCG GAG GGC GAG AGG GCC CCC CGC CCC GCC CCC GGC ACT GCC TGG ACT TGC TGC TGC AGC AAA CTG CAA GAA GGG GCC CGC GAG CTG  216
      V   A   A   A   D   L   P   A   E   G   E   R   A   P   R   P   A   P   G   T   A   W   T   C   C   C   S   K   L   Q   E   G   A   R   E   L

217  GAG GGT TTT GTG CAG CAG CTG AGT TTT GTT GCA GGG AAG CTG GCC TGC TGC CTG CGG GTG GGG GCG GAG CAG CTG GCG CGC TGC GCT GCG GAG GGG CGG CTG CCC AGC  324
      E   G   F   V   Q   Q   L   S   F   V   A   G   K   L   A   C   C   L   R   V   G   A   E   Q   L   A   R   C   A   A   E   G   R   L   P   S

325  AGC AGC AGC AGC AGC AGC TGC TGC GCG CTG CTG CAG CTC GAG AAG CAG GAC CTC GAG CAG AGC CTC GAG GCC GGC AAG CAG GGC GCG GAG TGC CTC TTG AGG AGC AGC  432
      S   S   S   S   S   S   C   C   A   L   L   Q   L   E   K   Q   D   L   E   Q   S   L   E   A   G   K   Q   G   A   E   C   L   L   R   S   S

433  AAA CTG GCC CTC GAG GCC CTC CTC GAG GGG GCC CGC GTT GCA GCA ACG CGG GGT TTG CTG CTG GTC GAG AGC AGC AAA GAC ACG GTG CTG CGC AGC ATT CCC CAC ACC  540
      K   L   A   L   E   A   L   L   E   G   A   R   V   A   A   T   R   G   L   L   L   V   E   S   S   K   D   T   V   L   R   S   I   P   H   T

541  CAG GAG AAG CTG GCC CAG GCC TAC AGT TCT TTC CTG CGG GGC TAC CAG GGG GCA GCA GCG GGG AGG TCT CTG GGC TAC GGG GCC CCT GCT GCT GCT TAC GGC CAG CAG  648
      Q   E   K   L   A   Q   A   Y   S   S   F   L   R   G   Y   Q   G   A   A   A   G   R   S   L   G   Y   G   A   P   A   A   A   Y   G   Q   Q

649  CAG CAG CCC AGC AGC TAC GGG GCG CCC CCC GCC TCC AGC CAG CAG CCC TCC GGC TTC TTC TGG TAG CCC TGC AGC AGC AGC AGC AGC AGC AGC AGC AGC AGC AGC GCG  756
      Q   Q   P   S   S   Y   G   A   P   P   A   S   S   Q   Q   P   S   G   F   F   W   *

757  GGC GGC AGC CGC GGC GGG GCC GGG GCG CCG CTG CAG CAA CAG CAG CAG CCG GCC CGG CTA GCG CCG CGG AGC ACT CGC AGG GAA CTC CAC AGG CAG CGG GAG AGC AGC  864

865  AGG GAC GAG AAG CAG GTC ATG TAG CGC AGG CAG CAG CGC CAG CTG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CTC CTG CAC CG
```

REFERENCES:

1. Profous-Juchelka, H., Liberator, P., Turner, M.J. (1988). Molec. Biochem. Parasit. 30, 233-242.
2. Henikoff, S. (1984) Gene 28, 351-359.
3. Kozak, M. (1986) Cell 44, 283-292.