

Heterogeneity at the 5' end of the circumsporozoite protein gene of *Plasmodium falciparum* is due to a previously undescribed repeat sequence

Vidal F.de la Cruz and Thomas F.McCutchan

National Institute of Allergy and Infectious Diseases, Laboratory of Parasitic Diseases, Malaria Division, National Institutes of Health, Bethesda, MD 20892, USA

Submitted 23 April 1986

We have cloned and sequenced the 5' end of the Circumsporozoite Protein (CSP) gene from two African strains of *P. falciparum*, LE5 (Liberia) and B11 (Uganda). The polymorphism noted (1) in this area of the gene results from the presence of a previously undescribed repeating sequence that is separated from the immunodominant repeat (IDR) domain. The sequence was determined in both orientations (arrows). There are no base substitutions in the area indicated by the arrows as compared to the previously published sequence of the CSP gene from a Brazilian isolate (7G8 from the IMTM22 isolate) (2). The B11 gene however, has an insertion of 30 nucleotides (in brackets below) that is not found in the other two sequences (position marked by a "*" in the figure below; nucleotide position numbers are as indicated in reference 2). The insertion appears to be two imperfect copies of an existing 15mer. Two complete copies of the 15mer appear in the Liberian and Brazilian strains while four copies appear in the Ugandan strain. Definition of the extent and type of polymorphism in the CSP gene of *P. falciparum* should help us assess the potential for long term effectiveness of a CSP based antimalarial vaccine.

```

AAT TAT TAT GGG AAA CAG GAA AAT TGG TAT AGT CTT AAA AAA AAT AGT AGA TCA CTT 293
  N  Y  Y  G  K  Q  E  N  W  Y  S  L  K  K  N  S  R  S  L
                                GGA GAA AAT GAT GAT 308
                                G  E  N  D  D
                                [GGA GAT AAT GAT AAT
                                G  D  N  D  N
                                GGA GAT AAT AAT AAT]
                                G  D  N  N  N
                                GGA AAT AAT AAT AAT 323
                                G  N  N  N  N
                                GGA GAT AAT GGT CGT 338
                                G  D  N  G  R

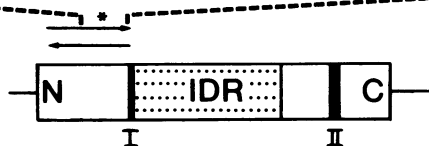
```

395

```

GAA GGT AAA GAT GAA GAT AAA AGA GAT GGA AAT AAC GAA GAC AAC GAG AAA TTA AGG
  E  G  K  D  E  D  K  R  D  G  N  N  E  D  N  E  K  L  R

```



1. Weber, J.L. and Hockmeyer, W.T. (1985) *Mol. Biochem. Parasitol.* **15**, 305-316
 2. Dame, J.B. et al. (1984) *Science* **225**, 593-599.