

Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 unique regions

Thomas Raabe and Stuart Siddell

Institute of Virology, Würzburg University, Versbacherstrasse 7, 8700, FRG  
 Submitted July 3, 1989

EMBL accession no. X15654

The coronaviruses are a group of positive-strand RNA viruses that cause a wide spectrum of disease in mammals and birds. Human coronaviruses are thought to cause about 15% of all common colds (1). Using cDNA cloning and M13-based sequencing we have determined the nucleotide sequence of the HCV 229E genomic region corresponding to the unique regions of the subgenomic mRNAs 4 and 5.

```

ATCAACTAAACTTCCTTATTACGACGTTGAAAAGATCCACATACAGTAATGGCTCTAGGTTTGTTCACATTGCAACTTGTCTGCTGTT 90
                ORF 4 ▶ M A L G L F T L Q L V S A V
AATCAATCGTTAGCAATGCGAAAGTTAGTGTCTGAAGTTTCACGACAGTTATCCAAGACGTGAAAGATGGCACTGTACCTTCAACTTG 180
N Q S L S N A K V S A E V S R Q V I Q D V K D G T V T F N L
CTAGCGTATACACTAATGAGCCTCTTTGTGTGTATTTTGCCTTATTAAAGCAAGATCACACCGTGGCAGAGCTGCTCTTATAGTGTTT
L A Y T L M S L F V V Y F A L F K A R S H R G R A A L I V F
AAAATTCATCCTTTTCGTTTATGTGCCATTGCTGATTGGTCTCAAGCATATATTCGCAACTTTGATTGCTGTAATTTGCTTGGA 360
K I L I L F V Y V P L L Y W S Q A Y I Y A T L I A V I L L G
AGATTTTCCATACAGCTTGGCACTGTGGCTCTACAAGACATGGGATTTTCATTGCTTCAATGTAACCACACTTTGCTATGCAAGGTAA 450
R F F H T A W H C W L Y K T W D F I V F N V T T L C Y A R *
                ORF 5A ▶ M Q G K
GTGTGGTTCCTGAAAATAAGGCTCTGAAACCATTCGTTTGTTCACGAGGGGATCAATTCCTTTACATAGGCGCAGAAATGTTTC 540
C W F L E N K A L K P F V C F Y G G D Q F L Y I G D R I V S
TTATTTCAACTAACGACTTGTACGTTGCTCTTAGAGGACGTATTGATAAAGACCTCAGCCTTTCTAGAAAGTTGAGTTATATAACGG 630
Y F S T N D L Y V A L R G R I D K D L S L S R K V E L Y N G
TGAATGTATACTTGTTCGTAACACCCAGCTGTGGAAATAGTCAACACAGATTTCAAATAGAAATCCAATAAGATGTTCCCTTAAGC 720
E C V Y L F C E H P A V G I V N T D F K L E I H *
                ORF 5B ▶ M F L K L
TAGTGGATGATCATGCTTTGGTTGTTAATGTACTACTCTGGTGTGGTGCCTTATAGTGATACTAGTGTGATTACAATAATTAAC 810
V D D H A L V V N V L L W C V V L I V I L L V C I T I I K L
TAATTAAGCTTTGTTTCACTTGCCATGTTTTGTAATAGAACAGTTTATGGCCCCATAAAAATGTGTACCACATTTACCAATCATATA 900
I K L C F T C H M F C N R T V Y G P I K N V Y H I Y Q S Y M
TGACATAGACCCTTCCCTAAACGAGTTATTGATTTCTAAACTAAACGACAATGTCAAATGACAATTGTACGGGTGACATTGTCCACCCA 990
H I D P F P K R V I D F *
    
```

**Figure legend.** The regions of homology involved in regulating the synthesis of the subgenomic mRNAs 4 and 6 are underlined.

An ORF of 399 nucleotides, that encodes a polypeptide of 15,300 mol.wt. was identified in the unique region of mRNA 4. The unique region of mRNA 5 contains two ORFs (5A and 5B) of 264 and 231 nucleotides respectively. The 5'-proximal ORF (5A) encodes a protein of 10,200 mol. wt. which lacks internal methionine residues. The downstream ORF (5B) encodes a protein of 9,100 mol. wt. which is characterised by a hydrophobic aminotermis. The arrangement of the ORFs and the translation products in this region of the HCV 229E genome closely parallel those of the murine coronavirus, MHV-JHM (2).

**References**

1. Hierholzer and Tannock, (1988), in "Laboratory Diagnosis of Infectious Diseases: Principles and Practice", Springer Verlag, Vol II, pp 451-483.
2. Spaan et. al., (1988), J. gen Virol. 69, 2939-2952