Coding Sequence of Coronavirus MHV-JHM mRNA 4

By M. A. SKINNER*† AND S. G. SIDDELL
Institute of Virology, Versbacher Strasse 7, 8700 Würzburg, Federal Republic of Germany
(Accepted 20 November 1984)

SUMMARY
A coding sequence at the 5' end of mRNA 4 of the coronavirus MHV-JHM was determined by M13/chain-terminator sequencing of cloned cDNA. An open reading frame of 417 bases with the potential to encode a polypeptide of mol. wt. 15200 (139 residues) was identified. The 3' end of the open reading frame overlapped by 16 bases the start of an open reading frame found in mRNA 5. The translation product of mRNA 4 was predicted to be a basic polypeptide rich in threonine. It had a large hydrophobic region near the amino terminus and a basic carboxy terminus. An intracellular, virus-specific polypeptide, which has been previously described as having a mol. wt. of 14000 to 14500 has the size and charge characteristics of such a translation product.

Murine hepatitis virus (MHV) is a member of the Coronaviridae which are cytoplasmic, enveloped, RNA viruses. The molecular biology of this virus group has been reviewed recently (Siddell et al., 1983). Briefly, in cells infected by MHV, in addition to the genomic RNA (mol. wt. 6.0 × 10⁶), six subgenomic mRNAs with mol. wt. ranging from 0.6 × 10⁶ to 3.7 × 10⁶ are produced. The genomic-size RNA which is infectious is termed mRNA 1 and the smallest subgenomic mRNA is mRNA 7. These RNA species form a nested set, each RNA having 3' sequences in common with all smaller RNAs. At the 5' end, each mRNA bears a common leader of about 70 bases derived from the 5' end of the genome (Lai et al., 1983; Spaan et al., 1983; Skinner & Siddell, 1983; Armstrong et al., 1984a). Translation studies in vitro and in oocytes have shown that the primary translation products that give rise to the three virion structural proteins, peplomer (150K), membrane (26K) and nucleocapsid (50K), are produced from mRNAs 3, 6 and 7, respectively (Rottier et al., 1981; Siddell et al., 1981; Leibowitz et al., 1982). The size of these polypeptides corresponds well with the size of the 'unique' RNA of each message, i.e. that portion not found in the next smallest mRNA. Sequencing of mRNA 7 (Skinner & Siddell, 1983) and of mRNA 6 (Armstrong et al., 1984b; M. A. Skinner, unpublished) has indeed shown that most of the available unique sequence is used as coding sequence, except for a long 3' non-coding sequence at the end of mRNA 7. Of the remaining RNAs, mRNA 2 (with a unique coding capacity of 80K) has been shown to produce an intracellular, virus-specific 35/30K polypeptide (Leibowitz et al., 1982; Siddell, 1983) and Leibowitz et al. (1982) have also shown that translation in vitro of genomic RNA (with a unique coding capacity of approximately 200K) produces a series of related polypeptides, with mol. wt. of about 200K, which are assumed to be related to components of the viral polymerase(s), although these polypeptides have not been identified in vivo. Finally, these same experiments allowed one further virus-specific translation product, an intracellular polypeptide of mol. wt. 14000 to 14500 (Siddell et al., 1981; Leibowitz et al., 1982; Siddell, 1983) to be tentatively assigned to either mRNA 4 or 5 (unique coding capacity for each of about 20K) although a definitive assignment was not possible because of the relatively low abundance and similar sizes of these mRNAs. We decided, therefore, to determine the coding sequences of mRNAs 4 and 5, to see which encoded the 14K polypeptide and to obtain the primary sequence of potential

* Present address: Department of Microbiology, University of Leicester, Leicester LE1 7RH, U.K.
product(s) of the other mRNA. In this paper we present the coding sequence of mRNA 4 and show that the predicted translation product has characteristics compatible with those of the previously described, intracellular, virus-specific polypeptide (mol. wt. 14000 to 14500) assigned to mRNA 4 or 5. In the accompanying paper (Skinner et al., 1985) we describe the coding sequence of mRNA 5.

A primer (3'-ATTAGATTTGA-5', Pharmacia P-L Biochemicals), complementary to a sequence just upstream of the initiating codon for the nucleocapsid protein (Skinner & Siddell, 1983) was used to synthesize cDNA from poly(A)-containing RNA isolated from Sac(-) cells infected with MHV-JHM as previously described (Skinner & Siddell, 1983). Methods used for the synthesis and cloning of double-stranded cDNA as well as for the characterization and chain-terminator sequencing of the cloned cDNA have been described previously (Skinner & Siddell, 1983). Sequence data were assembled and analysed by the programs of Staden (1982).

A cDNA clone (in pJMS1010) was isolated and characterized, as described in the accompanying paper, and the sequence of a 2.6 kb region was determined. The sequence illustrated in Fig. 1 is from position 1193 to 1678 of this region and represents the unique sequence of mRNA 4 between 2930 and 3416 bases from the 3' end of the genome. As shown in Fig. 1(b), 80% of the sequence was derived from both strands of the cDNA. The remainder was sequenced from two independent subclones.

At position 5 to 13 of the sequence, the sequence AAUCUAAAC was found. This is identical to the sequence found in the intergenic region of MHV, between genes 6 and 7 (Spaan et al., 1983), and differs by one base from the sequence upstream of gene 6 (AAUCCAAAC; M. A. Skinner, unpublished). A subset of this sequence (UCUAAAC) is also present upstream of the coding sequence of mRNA 5 (Skinner et al., 1985). These sequences are thought to be involved in regulating the initiation of synthesis of the bodies of MHV mRNAs (Armstrong et al., 1984a; Spaan et al., 1983).

Downstream from this sequence, the first AUG codon (position 67) initiates an open reading frame of 417 bases. The sequence around this initiator codon (GNNAUGG) corresponds to the sequence of a functional initiator codon in 10% of the mRNAs surveyed by Kozak (1983). At the 3' end, this open reading frame overlaps, by 16 bases, the start of a long open reading frame (with a +2 frameshift) which is discussed in the accompanying paper. The postulated product of the open reading frame at the 5' end of mRNA 4 would be a basic protein of mol. wt. 15200 (139 residues). Therefore, its size and basic nature are similar to those of a previously reported 14/14.5K polypeptide which was assigned as a translation product of mRNA 4 or 5 (Siddell et al., 1981; Rottier et al., 1981; Leibowitz et al., 1982, Siddell, 1983). The apparent mol. wt. of the polypeptide previously described as 14K to 14.5K was determined more accurately to be 15K to 16K (Skinner et al., 1985). Thus, on the basis of sequence data, it would appear that mRNA 4 encodes the polypeptide described previously. The predicted products of mRNA 5 (Skinner et al., 1985) have molecular weights (12K and 10K) which would allow them to be distinguished from the 15K polypeptide.

The primary sequence shows the protein to be relatively rich in threonine (16-5%, 23 residues) and a hydropathy plot (Fig. 2) shows that it has a very hydrophobic region from residues 8 to 41. Despite the high threonine content of the protein, no threonines are found within this hydrophobic region, even though threonines are not particularly hydrophilic residues. However, between residues 57 and 75, nine of the 19 residues are threonines. Further studies are required to determine the significance of this unusual distribution of threonine residues.

It is speculative, but interesting to note that the hydrophobic N-terminus is compatible with a membrane anchoring sequence and the C-terminal 30 residues form a basic region, which could be involved in RNA binding. This protein is not found as a major component of the envelope of the virus and if the protein is indeed membrane-bound it might, for example, be involved in functions such as siting membrane-bound transcription or replication complexes (Brayton et al., 1984). It will be interesting to see if the protein has a specific localization within the cell and whether the threonine residues are important to its function. Nothing is known of the role played by this protein in MHV infection, but a detailed knowledge of its structure should help in establishing its function, possibly by allowing specific antisera to be raised against synthetic peptides.
Fig. 1. (a) Sequence derived from the cDNA region representing the open reading frame at the 5' end of mRNA 4. The sequence is numbered arbitrarily from 1 (equivalent to 3416 bases from the 3' end of the genome) to 487. The AUG codon (67), initiating the open reading frame of 417 bases, and the termination codon (484) are boxed. The upstream sequence, comparable to that found in other intergenic regions, is underlined. The beginning of the mRNA 5 open reading frame (468) is similarly indicated. (b) Sequencing strategy showing extent and direction of sequencing of MI3 subclones. 82% of the sequence was obtained from both orientations, the remainder was obtained from two independent clones. Restriction enzyme sites are: •, HaeIII; ○, RsaI; ×, AluI.
Fig. 2. Hydropathy plot for the deduced protein sequence of the mRNA 4 product, according to the analysis of Kyte & Doolittle (1982). The vertical scale is the average hydropathy for a frame of seven amino acids. The base line is at -0.49, the average hydropathy of the 20 amino acids. Hydrophobic sequences appear above the base line. Markers along the horizontal scale are at intervals of 25 amino acids.

We thank the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 165, Genexpression) for financial support, Roger Staden for supplying his computer programs, Helga Kriesinger for typing the manuscript and Barbara Schelle-Prinz for skilled technical assistance.

REFERENCES


(Received 24 July 1984)