

CHARACTERIZATION OF THE HUMAN CORONAVIRUS 229E (HCV 229E) GENE 1

Jens Herold, Thomas Raabe, and Stuart G. Siddell

Institut für Virologie
Universität Würzburg
Versbacherstr.7
8700 Würzburg

SUMMARY

The sequence of the HCV 229E gene 1 has been determined and compared with the homologous sequences of the murine hepatitis virus and the avian infectious bronchitis virus. The coding sequence of gene 1 is 20 273 nucleotides in length. Within this coding region are two large open reading frames, ORF 1a (4 086 codons) and ORF 1b (2 687 codons) which overlap by 40 nucleotides. In the overlapping region, the genomic RNA can be folded into a pseudoknot structure, an element which is known to mediate -1 ribosomal frame-shifting in other coronaviruses. Assuming that -1 frame-shifting occurs at the HCV sequence UUUAAAC (nucleotides 12 514 - 12 520), the ORF 1a - ORF 1b product is predicted to be 6 758 amino acids in length. Our sequence analysis of the HCV 229E gene 1 has revealed a high degree of similarity within the ORF 1b of HCV, MHV and IBV, whereas ORF 1a is much less conserved. Elements which are believed to be necessary for specific (e.g. frame-shifting) and general (e.g. NTP-binding/helicase) transcriptional functions have been identified. This study completes the genomic sequence of HCV 229E which is 27.27 kb long and one of the largest known RNA genomes.

INTRODUCTION

The coronavirus genome is a single stranded, polyadenylated RNA of positive polarity (1). It is 27.6 kb in length in the case of avian infectious bronchitis virus (IBV) and about 31 kb in the case of mouse hepatitis virus (MHV)(2, 3). Expression of the coronavirus genome is mediated by the genomic RNA (referred to as mRNA 1) and a set of 3' coterminal subgenomic mRNAs. The mRNAs code for the coronavirus structural and nonstructural proteins. The known structural proteins are the S (spike), the M (membrane), the N (nucleocapsid), the SM (small membrane) and, for some coronaviruses, the HE (haemagglutinin esterase) proteins (4, 5, 6). The remaining mRNAs encode nonstructural proteins, most of which have not yet been detected in virus-infected cells.

To initiate infection, the genomic RNA is translated into a RNA-dependent RNA polymerase which subsequently uses its own mRNA as a template to create a negative stranded full-length copy of the viral genome and a set of positive and negative strand subgenomic RNAs (7). The RNA-dependent RNA polymerase has to be a multifunctional protein. It should contain the activities (i) to transcribe the genome into a negative stranded copy, (ii) to transcribe a set of subgenomic RNAs of both polarities, (iii) to transcribe the leader sequence and (iv) to produce progeny virion RNA. Furthermore, it is likely that it also possesses capping and polyadenylation activity. Here we present the sequencing strategy and a sequence analysis of the HCV 229E gene 1 and compare it with the homologous sequences of IBV and MHV.

MATERIALS AND METHODS

Virus and cells

HCV 229E virus was obtained from Dr. D.A.J. Tyrrell of the MRC Common Cold Unit, Salisbury, U.K. and cloned and propagated in C16 cells at 33°C as previously described (8). The cytoplasmic RNA from 10⁸ cells which had been infected 48 h previously at an m.o.i. of 3 was extracted by standard procedures and the poly A-containing fraction selected by hybridization to poly U-Sepharose.

cDNA cloning

cDNA synthesis was performed by the method of Gubler and Hoffman (9) using either sequence specific oligonucleotides or random hexamers as reverse transcription primers. Double-stranded cDNA molecules were blunt-ended with T4 DNA polymerase and ligated into SmaI-linearized, phosphatase-treated pBluescript II KS⁺ (Stratagene) and transformed into competent *E.coli* TG-1 cells (10). The cDNA clones encompassing the gene 1 of HCV 229E were obtained by screening 3 independent cDNA libraries with specific, ³²P-labelled oligonucleotides.

DNA sequencing and PCR Amplification

PCR was performed using a GeneAmp / RNA PCR Kit (Perkin Elmer Cetus) according to the manufacturers procedures. A 5'-biotinylated, synthetic oligonucleotide (Oli B4: 5'-Biotin-GCCTATGAAAGTGCTGTTGTTAATGG-3') was used as an upstream primer and a non-biotinylated oligonucleotide (Oli 36: 5'-TTAGATTTAAGAACAGCCTGTGACGC-3') served as the downstream and reverse transcription primer. Oli B4 extends from nucleotide 5336 to 5371 of the genomic sequence, Oli 36 is complementary to nucleotides 6230 to 6255. The resulting cDNA strands were separated with streptavidin-coupled magnetic beads according to the manufacturers protocol (Dynal) and the nucleotide sequence of both strands was determined using a T7 DNA Polymerase Sequencing Kit (Pharmacia).

Double and single-stranded DNA sequencing was carried out by the dideoxynucleotide chain termination method (11). The sequence of all cDNA clones was determined on both strands.

Computer analysis of the sequence data

Sequence data were assembled with the programs of Staden (12) and analyzed using the UWGCG (University of Wisconsin, Genetic Computer Group) sequence analysis software package.

RESULTS

Selection of cDNA clones representing the gene 1 of HCV 229E

Most of the clones which have been used to obtain the sequence of gene 1 of HCV 229E were produced by a random priming method. Clone T5B5 was produced by priming with an 18-mer oligonucleotide which is complementary to a sequence at the 5' end of the HCV 229E S gene. Clone 35D5 was identified as coming from the 5' end of the genome by screening a random primed cDNA library with a leader-specific oligonucleotide. The other random-primed cDNA clones were found by probing the different cDNA libraries with ³²P-labelled oligonucleotides, walking either from the 5' end (clone 35D5) in a 3' direction or from the 3' end (clone T5B5) in a 5' direction. The sequence information of the region between clone 30F7 and J13D8, where no cDNA clone was found in the libraries, was obtained by PCR amplification and direct solid-phase sequencing of the PCR product. In this way, a set of overlapping cDNA clones representing the complete gene 1 of HCV 229E was obtained (Fig.1).

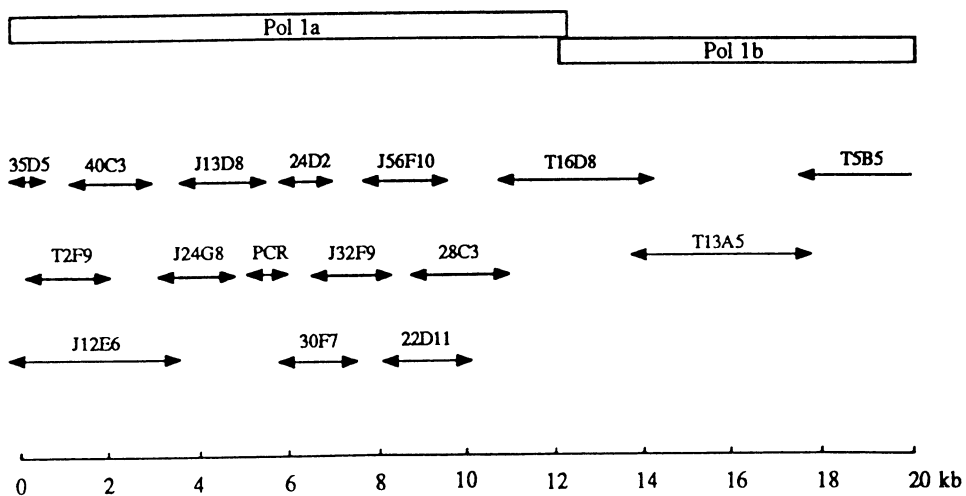


Figure 1. Diagram showing the positions of all cDNA clones used in obtaining the nucleotide sequence of the gene 1 of HCV 229E.

Analysis of the gene 1 nucleotide sequence

15 cDNA clones and a PCR derived DNA fragment were sequenced to obtain the complete sequence of HCV 229E gene 1. The coding sequence of gene 1 is 20 273 nucleotides in length. Analysis of the sequence revealed two large ORFs, 1a and 1b (Fig. 2). ORF 1a is 4 086 codons in length, ORF 1b, which overlaps ORF 1a by 40 nucleotides and is located in a -1 reading frame, is 2 687 codons in length.

In the ORF 1a - ORF 1b overlapping region the genomic RNA can be folded into a pseudoknot structure. This structure is analogous to elements known for IBV and MHV (3, 13, 14). It has been shown by mutational analysis that the pseudoknot structure and the 5'-located "slippery" sequence, UUUAAC, are necessary for the coronavirus frame-shifting event *in vitro* and *in vivo* (15, 16).

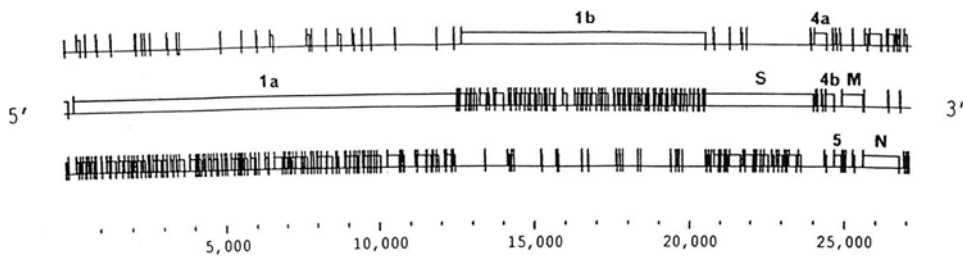


Figure 2. ORFs of the HCV 229E genomic RNA . GCG program (FRAMES).

Sequence similarities between ORF 1a and 1b of HCV 229E, MHV and IBV

After optimal alignment with the GCG program GAP, the overall similarity between the HCV 229E, MHV and IBV 1a ORFs is about 50% and for the 1b ORFs about 70%. However, when analogous domains in ORF 1a i.e. the 3C-like protease or the growth-factor-like domain (3) are compared, much higher similarities are determined.

Table 1. Similarity of particular domains in the 1a and 1b ORFs of HCV 229E, MHV and IBV

		HCV	MHV	IBV
Papain I ¹	Pos. ⁵ % ⁷	1012-1272	1038-1319 43	
X-domain ²	Pos. ⁵ %	1299-1396	1358-1459 54	1039-1139 58
Papain II ³	Pos. ⁵ %	1611-1866	1635-1990 45	1378-1612 42
3Cl pro ³	Pos. ⁵ %	2964-3267	3349-3653 62	2779-3086 60
MP2 ³	Pos. ⁵ %	3268-3546	3654-3941 55	3087-3379 54
GFL-domain ³	Pos. ⁵ %	3934-4069	4338-4475 64	3784-3929 64
Pol-module ⁴	Pos. ⁶ %	530-832	534-836 78	540-842 77
Hel-motif	Pos. ⁶ %	921-996	925-1000 84	931-1006 79
MBD ¹	Pos. ⁶ %	1199-1325	1201-1327 74	1208-1334 79

¹ (17); ² (18); ³ (3); ⁴ (19); ⁵ amino acid position in ORF 1a; ⁶ amino acid position in ORF 1b; ⁷ % similarity to the HCV 229E sequence

DISCUSSION

The sequence analysis of gene 1 completes the genomic sequence of HCV 229E which, excluding the poly(A) tail at the 3' end, is 27 270 nucleotides in length. Gene 1 is divided into two ORFs which can apparently be translated into a Pol 1a product of about 400 000 mol.wt. or a Pol 1a - Pol 1b product of about 700 000 mol. wt. The structures necessary for -1 ribosomal frame shifting, i.e. the "slippery" sequence (UUUAAAC) and a potential pseudoknot structure have been identified in the polymerase gene of HCV 229E.

Gene 1 is believed to code for an RNA dependent RNA polymerase(s) which carries out replication and transcription functions. ORF 1b, which is highly conserved among the coronaviruses sequenced to date, contains polymerase, helicase and metal-binding motifs (3, 17), suggesting that this region may be directly involved into RNA synthesis.

It seems likely that the coronavirus gene 1 is translated as a polyprotein which is subsequently proteolytically cleaved into functionally distinct polypeptides, as is the case for alphaviruses (20) and picornaviruses (21). In this respect, Gorbalenya *et al.* (17) and Lee *et al.* (3) have identified several protease motifs in the ORF 1a sequences of IBV and MHV, respectively. Firstly a papain-like protease motif which is duplicated in MHV and which is thought to be responsible for the cleavage of the N-terminal p28, at least *in vitro* (22, 23). And secondly, a highly conserved 3C-like protease which is predicted to have a preferred specificity for QS or QA dipeptides. Both domains could also be identified in the HCV 229E genome.

In the long term, the availability of the cDNA clones and nucleotide sequence for the polymerase gene of HCV 229E will allow us to produce specific immunological reagents which can be used to elucidate the expression, processing and function of this complex locus.

REFERENCES

1. Wege, H., Muller, A., and ter Meulen, V., 1978, *J. Gen. Virol.*, 41:217
2. Bournsnel, M.E.G., Brown, T.D.K., Foulds, I.J., Tomley, F.M., and Binns, M.M., 1987, *J. Gen. Virol.*, 68:57
3. Lee, H.-J., Shieh, C.-K., Gorbalenya, A.E., Koonin, E.V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M.M.C., 1991, *Virology*, 180: 567
4. Cavanagh, D., Brian, D.A., Enjuanes, L., Holmes, K.V., Lai, M.M.C., Laude, H., Siddell, S.G., Spaan, W., Taguchi, F., and Talbot, P.J., 1990, *Virology*, 176:306
5. Liu, D.X., and Inglis, S. C., 1991, *Virology*, 185:911
6. Godet, M., L'Haridon, R., Vautherot, J-F., and Laude H., 1992, *Virology*, 188:666
7. Sawicki, S.G., and Sawicki, D.L., 1990, *J. Virol.*, 64:1050
8. Phillpotts, R.J., 1983, *J. Virol. Methods*, 6:267
9. Gubler, U., and Hoffman, B.J., 1983, *Gene*, 25:263
10. Hanahan, D., 1985, in: "DNA cloning Vol.I", D.M. Glover, ed., IRL Press, Oxford
11. Sanger, F., Nicklen, S., and Coulson, A.R., 1977, *Proc. Natl. Acad. Sci.*, 74:5463
12. Staden, R., 1982, *Nucleic Acids Res.*, 10:4731
13. Brierley, I., Bournsnel, M.E.G., Binns, M.M., Bilimoria, B., Blok, V.C., Brown, T.D.K., and Inglis, S.C., 1987, *EMBO J.*, 6:3779
14. Bredenbeck, P.J., Pachuk, C.J., Noten, A.F.H., Charité, J., Luytjes, W., Weiss, S.R., and Spaan W.J.M., 1990, *Nucleic Acids Res.*, 18:1825
15. Brierley, I., Digard, P., and Inglis, S.C., 1989, *Cell*, 57:537
16. Brierley, I., Rolley, N.J., Jenner, A.J., and Inglis, S.C., 1991, *J. Mol. Biol.*, 220:889
17. Gorbalenya, A.E., Koonin, E.V., Donchencko, A.P., and Blinov, V.M., 1989, *Nucleic Acids Res.*, 17:4847
18. Gorbalenya, A.E., Koonin, E.V., and Lai, M.M.C., 1991, *FEBS Letters*, 288:201
19. Koonin, E.V., 1991, *J. Gen. Virol.*, 72:2197
20. Strauss, E.G., Rice, C.M., and Strauss, J.H., 1984, *Virology*, 133:92
21. Summer, D.F., and Maizel, J.V., 1968, *Proc. Natl. Acad. Sci.*, 59:966
22. Baker, S.C., Shieh, C.-K., Soe, L.H., Chang, M.-F., Vannier, D.M., and Lai, M.M.C., 1989, *J. Virol.*, 63:3693
23. Denison, M.R., Zoltick, P.W., Hughes, S.A., Giangreco, B., Olson, A.L., Perlman, S., Leibowitz, J.L., and Weiss, S.R., 1992, *Virology*, 189:274