NOTE

Complete nucleotide sequence of the S10 genome segment of grass carp reovirus (GCRV)*

Tao Qiu, Ren-Hou Lu**, Jing Zhang, Zuo-Yan Zhu

State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

ABSTRACT: Hemorrhagic disease, caused by the grass carp reovirus (GCRV), is one of the major diseases of grass carp in China. Little is known about the structure and function of the gene segments of this reovirus. The S10 genome segment of GCRV was cloned and the complete nucleotide sequence is reported here. The S10 is 909 nucleotides long and contains a large open reading frame (ORF) encoding a protein of 276 amino acids with a deduced molecular weight of approximately 29.7 kDa. Comparisons of the deduced amino acid sequence of GCRV S10 with those of other reoviruses revealed no significant homologies. However, GCRV S10 shared a putative zinc-finger sequence and a similar distribution of hydrophilic motifs with the outer capsid proteins encoded by Coho salmon aquareovirus (SCSV) S10, striped bass reovirus (SBRV) S10, and mammalian reovirus (MRV) S4. It was predicted that this segment gene encodes an outer capsid protein.

KEY WORDS: Grass carp reovirus (GCRV) · S10 genome segment · Nucleotide sequence

Grass carp reovirus (GCRV), an important fish pathogen involved in hemorrhagic disease (Chen & Jiang 1983), not only infects grass carp Ctenopharyngodon idellus, but also was found capable of infecting black carp Mylopharyngodon piceus, topmouth gudgeon Pseudorasbora parva (Ding et al. 1991) and rare minnow Gobiocypris rarus (Wang et al. 1994), and historically has resulted in large losses in freshwater fish culture in China. The virions are resistant to chloroform and ether, non-sensitive to acid (pH 3) and alkaline (pH 10) treatment, and stable within a certain range of temperature. The virus belongs to the genus Aquareo-

*The nucleotide sequence data for GCRV S10 reported in this paper has been deposited in the GenBank data base under the accession number AF236688.
**Corresponding author. E-mail: rhlu@ihb.ac.cn

© Inter-Research 2001
available data on Aquareovirus, the full-length cDNA of GCHV genome segment S10 was cloned and the characteristics of this sequence are documented here.

Materials and methods. Virus strain, purification and RNA extraction: The CIK cell strain (Zuo et al. 1986) was used to propagate GCRV-873 (Ke et al. 1990). Virus was purified from a continuous sucrose gradient (30 to 60% sucrose in NTE [NaCl-Tris-EDTA] buffer) and centrifuged at 100,000 × g for 2 h. Genomic dsRNA was extracted from the purified virus particles by 1% SDS and 10 µg ml⁻¹ Proteinase K at 37°C for 3 h, as well as by phenol-chloroform extraction. The purified dsRNA was treated with RQ1 RNase-free DNase (Promage) at 37°C for 1 h to remove the contamination of cell DNA, and displayed on 12.5% polyacrylamide gel (Fig. 1A).

Cloning of the S10 gene: The method described by Lambden et al. (1992) for amplification of the rotavirus dsRNA genome segments was used, with modification, to clone the GCRV S10 gene. The oligodeoxyribonucleotide primer 1 (5’-PO4-ATTTACCGCCGAGCCT-GACTT-NH2-3’), which was ligated to both ends of the dsRNA, 50 mM Tris-HCl pH 8.3, 3 mM MgCl2, 75 mM KCl, 1 mM DTT, 0.5 mM each deoxynucleoside of dNTP [Promega], 40 U RNAsin [SABC], and 200 U reverse transcriptase [Gibco BRL]. The mixture was incubated at 37°C for 1 h and the reaction was stopped by addition of EDTA to final 20 mM. The RNA was removed by NaOH and the cDNA was allowed to anneal. After extraction by phenol-chloroform and precipitation by ethanol, the annealed partial duplexes were filled in using DNA Polymerase I Large Fragment (Promage) and purified using a Glassmilk DNA purifying kit (BioStar). The amplification of cDNA was accomplished by PCR with primer 2 only using a PE9600, consisting of a denaturation step at 94°C for 3 min followed by 30 cycles of 30 s at 94°C, 35 s at 60°C, and 2 min at 72°C. Amplified DNA products were separated on agarose gel and the 900 bp fragment, which corresponded in size to that calculated from the GCRV dsRNA molecular weight for S10, was excised and purified using the Glassmilk DNA purifying kit. The purified S10 was ligated directly into pGEM-T vector and transformed into DH5α strain of Escherichia coli (Gibco BRL).

S10 gene sequence determination and analysis: The recombinant plasmids containing the full-length cDNA of S10 were identified according to the size of inserted segment by PCR using 2 M13 primers on pGEM-T and were purified using a plasmid DNA purification mini-kit (Viogene). Two clones that had the full-length cDNA of S10 were used for sequencing. The nucleotide sequence of S10 was determined by Taq dye primer cycle sequencing on an ABI 310 Genetic Analyzer (Perkin Elmer). The sequence of S10 was analysed using Lasergene sequence analysis package (DNASTar).

Results and discussion. Comparing with the GCRV genome dsRNA in Fig. 1A, the clone containing the cDNA of S10 was confirmed by Northern blotting as shown in Fig. 1B, in which the cloned cDNA hybridized to genome segment S10 rather than the other genome segments. Fig. 1C indicates that an expected, approximately 700 bp DNA band was ampli-
fied in the result of RT-PCR using 2 primers within the nucleotide sequence of S10. No product was found in negative control of RT-PCR with genomic DNA of CIK cell as template (data not shown).

The complete nucleotide sequence and deduced amino acid sequence of the S10 genome segment of GCRV are shown in Fig. 2. The nucleotide sequence was obtained from 2 different recombinant plasmids and constitutes a consensus sequence. The S10 gene of GCRV is 909 nucleotides long and contains a large open reading frame (ORF) preceded by a 30 bp untranslated region and followed by a 48 bp untranslated downstream sequence. The ORF starts with an ATG codon at nucleotides 31 to 33 and ends with a TGA at nucleotide 859 to 861. No additional ORF of significant length was detected in either the plus or minus strand RNA. Three AUG triplets, the first (ACGATG) located at 31, the second (CACATG) at 43, and the third (GCTATGG) at 67, were recognized in the same reading frame and near the 5' end of the mRNA strand. Since in most cases the AUG nearest the 5'-terminal cap is exclusively used for initiation for protein synthe-

Fig. 2. Complete nucleotide sequence (presented in the cDNA form) and deduced amino acid sequence of S10 RNA segment of GCRV. The 5'- and 3'-terminal nucleotide sequences are bold and indicated with a grey background and the inverted repeats are bold. Three ATG triplets near the 5'end of nucleotide sequence are underlined. In the amino acid sequence, the putative zinc-finger motif (amino acids 49–71) is also indicated with a grey background. Possible N-glycoslation site (*) and N-myristoylation site (#) are indicated.
sis in a ‘scanning model’ (Kozak 1980), the putative methionine start codon should be the first triplet rather than the others. Interestingly, the third triplet also appeared to be potentially suitable as a functional initiator because it was consistent with the strong initiation sequence of RNATGG (R = purine nucleoside, N = any nucleoside) (Kozak 1981).

The GCRV S10 genome segment displays the terminal sequences 5’ GUUAUU and CAUC 3’, which were recognized to be conservative in the GCRV RNA segments (authors’ unpubl. data). Moreover, the 3’-terminal sequence CAUC 3’ was found to be the same as that of segment S10 of SBRV (Lupiani et al. 1997b), another member of Aquareovirus. In addition, a putative inverted repeat sequence was identified adjacent to the terminal sequence. The 5’ end sequence, GAGCCCCC at 10 to 17, and its 3’ end inverted repeat sequence, GGGGTCTC at 892 to 899, are almost completely complementary.

The conserved terminal sequences are broadly reported in members of the family Reoviridae. They may be important in sorting and packing functions of the virus (Anzola et al. 1987). Analysis of the terminal nucleotide sequences is of interest in relation to understanding mechanisms of transcription and replication. The same 3’ terminal sequence between GCRV S10 and SBRV S10 implied relatively closer relations between them. In addition, like other reoviruses, a domain of inverted repeat adjacent to the 5’ and 3’ terminus always plays an important role in distinguishing this genome segment from other segments (Anzola et al. 1987).

The ORF is considered to encode a protein of 276 amino acids with a deduced molecular weight of approximately 29.7 kDa. The possible modification sites, an N-glycosylation site present at 244 to 247 and an N-myristoylation site at 263 to 268, were indicated and near the C terminus of the polypeptide (Fig. 2). Using the BLAST programs (Altschul et al. 1997), we were unable to find substantial similarity with either the nucleotide or the amino acid sequence of S10 and sequences of characterized animal reovirus and plant-infecting reovirus. However, some similarities in GCRV S10, SCSV S10 and SBRV S10 were observed. They not only have similar segment lengths, but the molecular weights of their deduced peptides are also analogous. Moreover, they are all segment 10 of the viruses in the same genus, i.e. Aquareovirus. We compared the deduced amino acid sequence of GCRV S10 with those of SCSV S10 and SBRV S10 using the Lasergene program (DNASTar). The predicted protein encoded by GCRV S10 showed a similarity index of 17.1 (Lipman & Pearson 1985) with SCSV S10 in a 173 amino acid overlap, and of 30 with SBRV S10 in only a 20 amino acid overlap.

Although it can be proposed that GCRV S10 may correspond to SCSV S10 to a low extent and may not be related to SBRV S10 with respect to amino acid sequences, the hydrophilic plots (Fig. 3) produced according to the method of Kyte & Doolittle (1982) significantly showed that the predicted proteins encoded by GCRV S10, SCSV S10 and SBRV S10 had very similar profiles, particularly in the regions of the N-terminal half. Thus, it can be predicted that their amino acid sequences may have an analogous molecular conformation. Furthermore, it is very likely that this analogous conformation leads to functional similarities among the deduced proteins of the S10 segments of these aquareoviruses.

The deduced amino acid sequence of GCRV S10 possesses a CX2CX15HX1C sequence within the N terminus, from residues 49 to 71, which is analogous to the zinc-finger domain (CCHC) identified within the mammalian reovirus (MRV) σ3 protein encoded by genome S4 gene (Mabrouk & Lemay 1994). Comparable in position, the zinc-finger sequence CX2CX15HX1C is from amino acids 51 to 73 within the MRV σ3 protein. Notably, we also recognized the zinc-finger domains within deduced amino acid sequences of SCSV S10 (CX2CX15HX1C) and SBRV S10 (CX2CX15HX1C) (Fig. 4), although they had not been indicated in the original reports (Lupiani et al. 1997a,b). Moreover, they are located at the same or a position very similar to that of GCRV S10, namely, at residues from 50 to 71 and from 49 to 71, respectively. Thus, the CX2CX15HX1C was found to be the special zinc-finger
sequence shared by the deduced peptide of the S10 segments of these aquareoviruses and MRV σ3. Since the most probable functional motif was identified as being a form CX2CX15HX2C (Mabrouk & Lemay 1994), the function of the zinc-finger motifs indicated here requires further study. In previous reports, the zinc-finger domains may be required by some proteins to maintain proper conformation (McIntyre et al. 1993) and may also be involved in protein-protein interactions (Cunningham et al. 1991).

Since the comparable zinc-finger domain exists in the MRV σ3 protein, the hydrophilic plot of the reovirus serotype-1 (Reo-1) S4 (Seliger et al. 1992), as an ex-ample of MRV, was given together with those of aquareoviruses. Surprisingly, many profiles were also similar to those of aquareoviruses, although MRV belongs to another genus, Orthoreovirus. In particular, the N-terminal 160 amino acids of Reo-1 S4 are very like those of GCRV S10, and the C-terminal is analogous to those of aquareoviruses S10 and SBRV S10, suggesting that the deduced amino acid sequences of the aquareovirus S10 may share some functional similarity with MRV σ3 protein. In addition, it was notable that the terminal sequences of MRV S4, 5' GUCAUU and CAUC 3' are very similar to that of GCRV S10. From an evolutionary standpoint, it is likely that the GCRV S10 and the MRV S4 have evolved from a common ancestral precursor.

Since SCSV S10, SBRV S10, and Reo-1 S4 were previously revealed to encode major outer capsid proteins (Atwater et al. 1986, Lupiani et al. 1997a,b), and since similarities have been shown between the zinc-finger sequences and hydrophilic plots of GCRV S10 and the above viral segments, it can be predicted that the deduced peptide of GCRV S10 is an outer capsid protein. In addition, the smallest segment of the GCRV genome, S11, has already been assigned to encode nonstructural proteins (authors’ unpubl. data), suggesting that the deduced peptide encoded by GCRV S10 may be the smallest outer capsid protein.

We compared the molecular weight of the deduced peptide of GCRV S10 with that of the proteins isolated from GCRV virions as described by Ke et al. (1992). In that report, the smallest outer capsid protein of GCRV is 27 kDa, slightly smaller than the 29.7 kDa protein deduced. Considering previous studies on aquareoviruses (Winton et al. 1987, Hsu YL et al. 1989, Subramanian et al. 1994) which reported that the smallest structural proteins all were approximately 31 to 36 kDa, it is possible that a deviation from reading data on the SDS-PAGE existed in the Ke et al. report. Moreover, the same protein identified in another report (Wang et al. 1990) was 31 kDa, which approximates the size of deduced product of S10. Our study will serve as a preliminary for future investigations. In order to learn more about the structure and function of the protein encoded by GCRV S10 and its similarities with other proteins, additional work including biochemical assays and immunological analysis should be performed.

Acknowledgements. This work was supported by Grant No. ZYJ01-04 from the China-Israel fund for scientific and strategic research and development, and by Grant No. 990205-FB05 from FBBL, the state key laboratory of Freshwater Ecology and Biotechnology.

LITERATURE CITED


Fang Q, Attoui H, Biagini JF, Zhu Z, de Micco P, de Lamballerie X (2000) Sequence of genome segments 1, 2, and 3 of...
the grass carp reovirus (Genus aquareovirus, family reoviridae). Biochem Biophys Res Commun 274(3):762–766