Characteristics of the monocistronic genome of extra small virus, a virus-like particle associated with *Macrobrachium rosenbergii* nodavirus: possible candidate for a new species of satellite virus

Joannes Sri Widada and Jean-Robert Bonami

Pathogens and Immunity, CNRS, Université Montpellier II, Case 080, 34095 Montpellier Cedex 5, France

White tail disease (WTD) causes a high mortality rate in the freshwater prawn *Macrobrachium rosenbergii*. The pathogenic agent is a small virus, 25 nm in diameter, *Macrobrachium rosenbergii* nodavirus (MrNV), associated with extra small virus (XSV), a virus-like particle, 15 nm in diameter. Sequencing of the XSV genome showed that it consists of a linear single-stranded RNA of 796 nucleotides, encoding a single structural protein, the capsid CP-17. The genome is in sense orientation, ended by a short poly(A) tail at the 3′-end. Sequence comparison did not allow XSV to be affiliated to known virus families. The hypothesis that XSV is a satellite virus, such as those described in the plant kingdom, is put forward based on its characteristics. It would constitute, therefore, the first satellite virus associated with a nodavirus.

Emerging diseases constitute an increasingly serious health problem in aquaculture. As a result of industrialization and world trafficking of farmed species, diseases are rapidly and easily spread, leading to important economic losses during acute epizootics. The giant freshwater prawn *Macrobrachium rosenbergii* is one of the most economically important farmed palaemonids in the world. A virus-borne disease affecting this crustacean, called white tail disease (WTD), was first observed on the island of Guadeloupe in 1995, then in Martinique (French West Indies) (Arcier et al., 1999) and later in Taiwan (Tung et al., 1999) and the People’s Republic of China (Qian et al., 2003), and more recently in India (A. S. Sahul Hameed, personal communication). The pathogenic agent was *M. rosenbergii* nodavirus (MrNV) (Arcier et al., 1999), a small icosahedral non-enveloped virus, of 25 nm in diameter, which is associated with a virus-like particle, extra small virus (XSV), which is icosahedral in shape and 15 nm in diameter (Qian et al., 2003). Nothing, however, is known about the respective role of each virus in the disease pathogenesis or their interaction.

In this paper we describe the characteristics of the XSV genome and we hypothesize that XSV constitutes a new species of satellite virus.

The genome of XSV consists of a linear single-stranded RNA (ssRNA) of about 0·8–0·9 kb (Qian et al., 2003). In order to determine its nucleotide sequence, the XSV genome was reverse-transcribed and double-stranded (ds)DNA was synthesized. Virus particles were purified from tissue extracts of WTD-infected animals by centrifugation through sucrose gradients, then through CsCl gradients, and finally concentrated by centrifugation as previously described (Romestand & Bonami, 2003). This technique also allowed the separation of XSV from its associated virus, MrNV. Viral RNA was extracted by a combination of phenol/chloroform treatment followed by precipitation with ethanol. Various approaches were used to reverse-transcribe the viral RNA. In the first method, cDNA was synthesized by reverse transcription using a hexanucleotide mix (Roche) and dsDNA was synthesized using the cDNA Synthesis System (Roche). Double-stranded DNA with a single 3′-dA overhang was obtained by Taq DNA polymerase (Promega) treatment at 72 °C for 10 min. DNA was cloned into the pCR2.1-TOPO plasmid using a TOPO TA cloning kit and transfected into *Escherichia coli* TOP10 (Invitrogen). Plasmids were purified using a High Pure Plasmid Isolation kit (Roche) and were analysed by restriction enzymes and sequencing (MilleGen). In order to clone the 3′-end of the genome, reverse transcription was performed with an oligo(dT) anchor primer [GACCACCGTATCGATGTCGACT<sub>16</sub>V], using a 5′/3′ RACE kit (Roche), on the assumption that a poly(A) tail would be present. DsDNA was synthesized by PCR, using a PCR anchor primer (GACCACCGTATCGATGTCGAC) and a primer deduced from sequenced parts of the XSV genome. Cloning and sequencing were carried out as above. In order to walk towards the 5′-end of the virus genome, reverse
transcription was performed using a primer deduced from sequenced parts of the genome. cDNA was then dA-tailed by terminal transferase and dATP. DsDNA was obtained by PCR, using an oligo(dT) anchor primer and a primer deduced from sequenced parts. Cloning and sequencing were carried out as described above.

The virus genome sequencing results showed that it was composed of 796 nucleotides, and a short poly(A) tail of 15–20 nucleotides at the 3′-end (GenBank acc. no. AY247793) (Fig. 1a). There was good agreement between the sequencing result and the genome length established by electrophoresis on an agarose gel (Qian et al., 2003). The presence of a short poly(A) tail was established, because the XSV genome could be reverse-transcribed using an oligo(dT) anchor primer. The occurrence of a poly(A) tail was also suggested by a polyadenylation signal, AAUAAA, found 195 nucleotides downstream of the second stop codon and six nucleotides upstream of the poly(A). The presence of a poly(A) tail indicated that the genome was in a sense orientation. This orientation was further confirmed, because the XSV genome could be reverse-transcribed using a primer complementary to the messenger orientation (see below). Finally, most probably the 3′-end is not terminated by a free 3′-OH, since ligation of this end with a 5′-phosphate oligonucleotide failed to take place.

A search for open reading frames (ORFs) established that the XSV genome contained a unique open reading frame in the sense orientation, 522 nucleotides long (position 63–585), followed by two stop codons. It had the potential coding capacity of 174 amino acids (aa) (Fig. 1a). In the antisense orientation, no significant ORFs were found. The 5′-UTR was 62 nucleotides long, while the 3′-UTR was longer [212 nucleotides in length, excluding the poly(A) tail]. Protein analysis by SDS-PAGE showed that XSV particles contain two polypeptides of almost equimolar ratio that migrate at 17 kDa (CP-17) and 16 kDa (CP-16) (Fig. 1b). In order to establish a correlation with the coding sequence of the genome, the N-terminal ends were sequenced following isolation by SDS-PAGE and transfer to a nitrocellulose membrane. The results indicated that CP-17 was actually encoded by the unique XSV gene. The amino acid sequence obtained, MNKRINNNRTMRSSR, matched perfectly the amino acid sequence deduced from the nucleotide sequence. Sequencing of the N-terminal end of CP-16 resulted in the following sequence: MRSR. Therefore CP-16 is a truncated form of CP-17, lacking the

![Fig. 1. Nucleotide sequence of XSV genome, predicted amino acid sequence and SDS-PAGE protein analysis. (a) Nucleotide sequence obtained from multiple clones and from both strands and deduced amino acid sequence. The Kozak sequence is underlined. Two initiation codons are shadowed. The first amino acid of CP-16 is in bold type. (b) Protein composition was identified by a 15% gel using SDS-PAGE, which was then transferred to membrane. Protein markers in kDa are indicated.](image-url)
first 11 N-terminal amino acids corresponding to a molecular mass of about 1 kDa. It is worthwhile mentioning that the N-terminal end of CP-16 is also a methionine. We asked the question whether the occurrence of CP-16 was fortuitous. Two lines of evidence suggest that CP-16 was not a degraded form of CP-17, obtained during protein preparation. Firstly, protein analyses were performed by directly dissolving intact virus particles in Laemmli buffer prior to electrophoresis, and CP-17 and CP-16 were roughly equimolar. Secondly, an examination of the nucleotide sequence of the XSV genome indicated that the N-terminal amino acid of both polypeptides was methionine, and the ATG codon of CP-16 is localized in as favourable a Kozak context as the ATG codon of CP-17, seven out of ten nucleotides matching with the Kozak consensus sequence GCCRCCCATGG (Kozak, 1987). Therefore CP-16 may be a normal constituent of XSV particles, the normally processed CP-17 found in infectious mature particles or a truncated form found in non-infectious RNA-free particles (ghosts). For the time being, we cannot exclude any of these possibilities.

The present work clearly indicated that the XSV genome is a monocistronic messenger in the sense orientation, although its translation may give rise to two related polypeptides. This posed the question, which enzymes are required for virus development in host cells? Most of them may come from the host cell, such as those involved in the translational machinery, and some from its associated MrNV. In fact, the genome of MrNV encodes an RNA-dependent RNA polymerase (GenBank acc. no. AY222839) that XSV may use for its own development. Several virus associations have been described in plant systems. They involve a small size virus (satellite virus) and a larger virus (helper virus). One of the most studied co-operations concerns the association of panicum mosaic satellite virus (PMSV) and panicum mosaic virus (PMV) (Qiu & Scholthof, 2000, 2001). It has been demonstrated that the smaller partner (satellite virus) needs the larger one for its replication, but that it is the satellite virus that induces severe chlorosis on proso millet plant (Qiu & Scholthof, 2001). XSV, 15 nm in diameter, is associated with MrNV, 25 nm in diameter, in WTD of M. rosenbergii. Further studies are required to understand their interactions and to establish whether the severity of WTD is determined by the smaller partner.

Fig. 2. Sequence alignment of XSV CP-17 with known sequences of structural proteins of satellite viruses. Sequence alignment was done, using an algorithm established by Corpet (1988). SVMWLMV (satellite virus of maize white line mosaic virus, acc. no. NC_003631), TNSV (tobacco necrosis satellite virus, acc. no. NC_001557), PMSV (panicum mosaic satellite virus, acc. no. NC_003847), SSADV (satellite St Augustine decline virus, acc. no. L10083), BMVS (bamboo mosaic virus satellite, acc. no. L72625), STMV (satellite tobacco mosaic virus, acc. no. M25782).
In order to search for an affiliation of XSV to known virus families, the deduced amino acid sequence was submitted to known database comparisons using the Blast2 interface (Altschul et al., 1997). The sequences were aligned using an algorithm established by Corpet (1988). The sequence comparison did not enable us to establish any significant homology with known virus genomes sequenced so far. These viruses include MrNV, with which XSV is associated in WTD. The absence of sequence affiliation to MrNV was also established because probes specific to MrNV failed to hybridize with the XSV genome (Qian et al., 2003). Nevertheless, the characteristics of XSV meet the criteria of satellite viruses, i.e. very small size (subvirus agent); lacking the genes encoding enzymes required for replication and therefore replication dependence on its helper virus; genome distinct from that of the helper virus; and a single gene encoding the structural polypeptide (Zhang et al., 1991). If our hypothesis is confirmed by further work, then MrNV may play the helper virus role and, therefore, XSV would be the first satellite associated with a nodavirus. Except for the chronic bee-paralysis associated satellite virus-like (CBPSV) (Overton et al., 1982), all satellite viruses are plant viruses (Mayo et al., 2000). XSV is more closely affiliated to plant satellites than to CBPSV, as the genome of CBPSV is composed of three species of RNA about 1 kb in size. In all the reported plant satellites, the genome is, as for the XSV genome, one linear sense ssRNA with a size range from around 700 to 1200 bases (Mayo et al., 2000). Sequence comparison of the deduced CP-17 of XSV with known structural protein sequences of satellite viruses indicated that, at the sequence level, CP-17 was not affiliated to any of them (Fig. 2). Sequence comparison also indicated that, apart from their physico-chemical characteristics, satellite viruses do not exhibit significant sequence homology between their structural proteins. The phylogenetic tree, resulting from sequence comparison, shows that satellite viruses are remotely related (data not shown), as also previously reported (Ban et al., 1995; Zhang et al., 1991). Most probably, each satellite–helper virus association is specific to that couple. Nevertheless, satellite viruses are not a diminutive form of their helper. Finally, computer analysis indicated that the N-terminal domain of the capsid protein exhibits a recognized common motif. It contains hydrophilic amino acids and a positively charged arginine. This feature is characteristic of a surface zone in a protein. It may also constitute an interaction domain through a saline bridge allowing, for example, the formation of the viral capsid. In addition, there are doublets, R–R or R–K, that are sites for proteases. Therefore, the structural protein may be synthesized as a precursor that is then cleaved to a mature and functional protein. Shared structural features of the N-terminal domain may suggest a common function.

ACKNOWLEDGEMENTS

We thank Véronique Richard for her technical contribution to this work.

REFERENCES


