Early and late pre-mRNA processing of budgerigar fledgling disease virus 1: identification of viral RNA 5' and 3' ends and internal splice junctions

Dong Luo, Hermann Müller, Xiao-Bo Tang and Gerd Hobom

1 Institut für Mikrobiologie und Molekularbiologie and 8 Institut für Virologie, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany

Budgerigar fledgling disease virus 1 (BFDV-1) is the first avian polyomavirus to be identified, and it possesses uncommon structural and biological properties. Here we present an analysis of the processed viral RNAs in infected chicken embryo fibroblast cells. Two early and 18 late BFDV-1 mRNAs were defined according to their 5' ends and internal splice patterns. In the early region of the genome an incomplete splice reaction covering 195 nt is responsible for creating two mRNAs that could encode small t and large T antigens, which would be initiated from a hypothetical early promoter, P_E. The late mRNA 5' ends define two putative promoter regions (PL1 and PL2), 111 nt apart in the BFDV-1 genome non-coding region. The overall splicing pattern of the late mRNAs is further complicated by an alternative splice reaction of intron 2 (deletion of either 64 nt in intron 2a or of 256 nt in intron 2b) and a splice removing intron 3 (870 nt), resulting in deletion of most of the VP2–VP3 coding region. The positions of the late mRNA 5' ends and the splicing pattern indicate the existence of two open reading frames, putatively encoding two pairs of agnoproteins, in the 5' region of several late mRNAs. These mRNAs appear to be bicistronic and to encode one of the agnoproteins together with one of the viral coat proteins.

Budgerigar fledgling disease virus 1 (BFDV-1) is the first avian isolate of a polyomavirus (Lehn & Müller, 1986; Müller & Nitschke, 1986; Rott et al., 1988). It is closely related to the two well-characterized papovaviruses, simian virus 40 (SV40) and murine polyomavirus (Py) as well as other viruses of this group, with which it shares extensive structural as well as functional similarities (Rott et al., 1988). In contrast to mammalian polyomaviruses, BFDV-1 is associated with an acute disease resulting in death of affected birds (Davis et al., 1981); on the other hand, BFDV-1 is also able to transform primary chick embryo fibroblast (CEF) cells in vitro (Lehn & Müller, 1986).

The genome of BFDV-1 has recently been sequenced and characterized (Rott et al., 1988; Stoll et al., 1993). Similar to other polyomaviruses, the 4981 bp DNA genome can be divided into an early region and a late region. The deduced amino acid sequences from the early region, which encodes the genes for large T and small t antigens, revealed some remarkable differences from other polyomaviruses, including a considerable reduction in size of the large T antigen (Rott et al., 1988). The late region contains a standard pattern of the three genes coding for the structural proteins VP1, VP2 and VP3, as well as open reading frames coding for proteins with largely unknown functions, agnoproteins. SV40, but not Py, also contains an open reading frame coding for an agnoprotein in a similar location within its genome. Finally, the replication origin structures and transcriptional control elements in the non-coding region of BFDV-1 show fundamental deviations from those of all other polyomaviruses (Rott et al., 1988).

Recent advances in PCR have made it possible to analyse RNA processing qualitatively, including determination of 5' and 3' ends as well as the patterns of splicing of introns in viral mRNAs isolated from infected cells. The importance of processing of intron sequences in gene expression and regulation has been demonstrated in many eukaryotic systems (Leff et al., 1986). The correct removal of introns not only produces functionally mature mRNA at a controlled rate, but may also modulate gene expression at a post-transcriptional level in various ways, including splicing-induced enhancement of polyadenylation, mRNA transport, mRNA accumulation and stability of RNA transcripts (Vallarreal &
confluent) were infected with BFDV-1 at a m.o.i. of 10

White, 1983; Ryu & Mertz, 1989; Haung & Gorman, 1990; Niwa et al., 1990; Chiou et al., 1991). Processing of introns, and in particular partial or alternative splicing reactions, are now considered to be an important element in control of gene expression in many viral and eukaryotic systems (Leff et al., 1986).

In order to determine the splice junctions within BFDV-1 early transcripts, CEF cells (10^7 cells, 70% confluent) were infected with BFDV-1 at a m.o.i. of 10

(Lehn & Müller, 1986; Müller & Nitschke, 1986). The total cellular RNA was isolated according to Chomczynski's procedure (Chomczynski & Sacchi, 1987) and cDNA was synthesized for every RNA preparation by reverse transcription initiated with the 3'-d(T)20 adapter. The DNA fragments were amplified by ‘touchdown’ PCR (Don et al., 1991) with two specific primers located upstream and downstream of the expected intron position (Fig. 1a). Electrophoresis on agarose gels revealed 376 bp and 571 bp PCR products in similar amounts, which were expected to correlate with spliced and unspliced mRNA (Fig. 1b), respectively. The amplified DNA fragments were digested with EcoRI (BamHI or SpeI) and XhoI, subcloned into cloning vector pBluescript II SK (Stratagene) and then sequenced. Sequencing of the 571 bp cDNA fragment did not identify any spliced region. However, the sequence of the 376 bp cDNA revealed that a 195 nt intron had been removed from BFDV-1 early pre-mRNA (position 4733-4537) in such a way that both exons of the T antigen gene had become fused, and the resulting spliced early mRNA could serve as a template for translation of BFDV-1 large T antigen. No alternative splice reaction was identified in this region or any other of the early BFDV-1 mRNAs, so, unlike SV40, the BFDV-1 small t antigen mRNA does not contain an intron sequence but must be translated from the unspliced early mRNA. The large T and small t antigens thus share 83 N-terminal amino acids (aa), but differ in their C termini beyond the position of the splice donor signal. Fig. 1(a) illustrates the arrangement of BFDV-1 early gene transcripts and gives a sequence comparison of the BFDV-1 splicing signals with general consensus sequences (Rogers & Wall, 1980). The single intron splicing reaction in BFDV-1 is operating only to a partial extent, which appears to be due to a weak acceptor signal deviating at many positions from the consensus sequence, while the donor sequence appears to be in the normal range. Even though the various species of early mRNAs have not been tested individually for their translational capacities, they are very likely to represent the small t and large T antigen messengers, respectively. This is supported by the results obtained in the baculovirus system using unspliced and spliced cDNA constructs, which code for small t and large T protein (Luo et al., 1994).

Experiments using RT–PCR amplification were carried out to determine the pattern(s) of BFDV-1 late splice junctions. Locations of primers employed in these experiments are shown in Fig. 2(a). The sizes and densities of bands observed after gel electrophoresis reveal the distribution of splicing sites and represent the initial ratios between the amounts of various mRNAs (Hyde-DeRuyscher & Carmichael, 1990). Resulting PCR fragments were subcloned and sequenced. Fig. 2 shows...
the results of PCR amplification experiments covering the main body of the late mRNAs and also shows the sequences of splice junctions as determined by comparison with the published BFDV-1 sequence data (Rott et al., 1988). Altogether four independent splicing reactions were detected in BFDV-1 late mRNAs by sequencing representative molecules of cloned fragments. These include intron 1 with an intervening sequence of 76 nt, intron 2a of 64 nt, intron 2b of 256 nt and intron 3 of 870 nt.

The splicing junction of intron 1 (position 331–408) was first determined by sequencing a cloned PCR fragment amplified using a d(C)12 adaptor and primer #143. Intron 1 was removed in order to avoid the translational stop codons within the intron and to allow open reading frames of putative agnoproteins to be translated continuously (Fig. 2a). Both intron 2a and 2b were correctly removed at an identical splice donor site, GU, at map position 569. However, the positions of the splice acceptors were different. One dinucleotide, AG, was used as splice acceptor for intron 2a at map position 634, and another AG, 192 nt downstream (position 826) served as splice acceptor for intron 2b (Fig. 2a). The splicing junction of intron 3 (position 974–1845) was determined by sequencing of PCR fragments amplified from primers #14 and #11 or #14 and #143, which accounted for more than 80% of the total signal (Fig. 2b), leading us to conclude that splicing of intron 3 might be in a dominant fashion and thus VP1 mRNA would act as the major late message. The most likely possibility is that functional removal of intron 3 significantly boosts the major mRNA accumulation, probably by stabilizing the mRNA. On the other hand, removal of intron 3 is required for VP1 (a major viral structure protein) translation using the AUG codon downstream at position 1899 (Fig. 2a; Rott et al., 1988; Stoll et al., 1993).

In order to analyse the alternative splicing patterns in agnoprotein coding regions in detail we designed two pairs of specific primers based upon 5' end-mapping of late mRNAs (see below). Primers #56 and #71 shown in Fig. 2(a) were used for analysing the BFDV-1 late mRNAs transcribed from cap 1 sites and primers #76 and #71 were used for the late mRNAs transcribed from the cap 2 site as well as the cap 1 mRNAs containing intron 1 (see below). Significantly, a major double-spliced message transcribed from cap 1 sites (introns 1 and 2a removed), which covered up to 80% of total PCR signals (Fig. 3), will generate the biggest deduced agnoprotein, 176 aa in size (agno1c; Fig. 2a). A minor double-spliced message transcribed from cap 1 site (introns 1 and 2b removed) will create an agnoprotein (1d) of 112 aa in size. In contrast, fully spliced mRNAs transcribed from cap 2 site (intron 2a or intron 2b

![Diagram](image-url)

Fig. 2. Determination and characterization of splice junctions in BFDV-1 late transcripts. (a) Genetic map of the BFDV-1 late region and nucleotide sequence comparison of splice boundaries in late transcripts. The organization of putative agnoproteins and viral coat proteins VP1, VP2 and VP3 is shown. The positions and directions of primers used for PCR amplification are indicated below the primary transcript. Oligonucleotide primers for late messengers: #14, 5' GCA-GCCGTCTGGAAATCTTTTCAGAC 3' (EcoRI); #104, 5' ACTAC-ACAGGAGAATITCCTCTTGTG 3' (EcoRI); #9, 5' TTGTTGCTCTCGAGACCCGTTAGGG 3' (Xhol); #11, 5' GTACGGCGACTCCTC-GAGTCAAGGT 3' (Xhol); #143, 5' ATACATGCTTGAACCTCGAG-GAGC 3' (Xhol); #76, 5' ATAAATCTGAGTGACCTACTAGTAG 3' (Xhol); #56, 5' ATAAATCTGAGTGACCTACTAGTAG 3' (Xol); #76, 5' ATGACTGACGCTACTAGTAG 3' (Spel); #76, 5' ATGACTGACGCTACTAGTAG 3' (Spel). (b) Total RNAs were isolated at 36 h p.i. and the resulting RT-PCR products were separated on a 1.5% agarose gel. The PCR fragments (1 to X) were isolated and sequenced. I, IV and VII, unspliced transcripts; II, intron 2a removed; III, intron 2b removed; V, intron 3 or 2a and 3 removed; VI, intron 2b and 3 removed; VII, intron 3 removed; IX, intron 2a and 3 removed; X, intron 2b and 3 removed.
late mRNAs. This result proves that the splice reactions may occur independently from each other. However, no exon skipping ( aberrant) splice reaction product has been detected despite some effort, so no splice reaction occurs between donor 2 and acceptor 3, or donor 1 and acceptors 2 and 3. This result agrees with a scanning model for the nearest splicing partner signal along the pre-mRNA.

The presence of a methylated cap structure at the 5' terminus is an almost general feature of mRNA molecules in eukaryotic cells (Shatkin, 1976). Cap formation has been shown to be associated with the initiation of translation and the regulation of gene expression (Thach, 1992). We used a strategy termed 'rapid amplification of cDNA ends' (RACE) as described (Frohman et al., 1988) to identify the 5' termini of BFDV-1 early and late transcripts. Full-length CDNs obtained by reverse transcription were extended into homopolymer 3' tails using terminal deoxynucleotidyl transferase and substrate dGTP; the G-tailed products were amplified using d(C)12 adaptor and gene-specific amplification primers (#144 for early and #143 for late mRNAs). We found two cap sites at positions 22 and 30 in BFDV-1 early mRNAs and three different 5'-terminal ends at positions 264, 266 and 375 in late mRNAs (Fig. 4). The heterogeneity observed in BFDV-1 early mRNA initiation sites may be at the level of utilization of alternate sites by the basal transcription machinery whereas the late mRNA initiation sites in BFDV-1 are less heterogeneous than those in SV40 and other polyomaviruses (Ghosh et al., 1982; Piatak et al., 1983).

Fig. 4. Mapping the positions of 5' termini of BFDV-1 early and late mRNAs by oligo(dG)/oligo(dC) RACE. (a) Putative TATA boxes, cap sites and ATG codon for translational initiation are indicated. Pz and Pz, putative promoters for early and late transcripts. (b) PCR experiments were performed using d(C)12 adaptor (5' ATAGAATTCCGGATCCCCCCCCCCCC 3') which contains EcoRI and BamHI sites, and a sequence-specific primer (position and orientation indicated in Fig. 1a and Fig. 2a; #144 for early and #143 for late mRNAs). RACE-PCR products were separated on 1.5% agarose gels and sequenced.
Fig. 5. A diagram of BFDV-1 genome structure, containing early, late and non-coding regions. Two early and 18 late RNA transcripts initiated from the putative early and late promoters (P_E and P_L) are shown as thin lines on the left and right of the circle. The gaps with numbers represent deletions of the introns.

Among the clones isolated and sequenced, 12 turned out to be full-length cDNA clones. With 5 and 4 of them being located at positions 264 and 266, respectively, both are correlated with a putative promoter P_L1, while 3 of them were initiated at position 375 which is correlated with a putative promoter P_L2. Identification of mRNA 5' ends was not only based on a number of individual clones (at least four each, often six to nine) extending up to the identical sequence position in the immediate neighbourhood of the C_12 sequence that resulted from the dG/dC tailing and amplification method, but relied also on the crucial observation of an extra, uncoded G in between the C_12 and 5'-terminal sequence position (Hirzmann et al., 1993). This confirms a sequence to be derived from a true mRNA 5' end (in 18 out of 20 full-length clones).

The rather complex patterns of the 18 late mRNAs determined can be divided into one unspliced molecular species, 2738 nt in length, plus 11 varieties (including the major one mentioned above) that have one or more of the three introns deleted (Fig. 5). All of these originated from P_L1 and therefore will have their cap sites either at position 264 or 266. An additional unspliced (2627 nt) and five spliced late mRNAs originated from P_L2, and will have cap sites at position 375 (Fig. 4a). P_L2 is located within intron 1 relative to the P_L1 transcription system and all downstream splicing signals are used for both classes of mRNAs in parallel.

Poly(A) tails at 3' ends of mRNAs have an important function in mRNA stability and translation (Jackson & Standart, 1990). Accurate cleavage and polyadenylation of pre-mRNAs are dictated by two sequence elements in the RNA substrate, one located upstream (which has the sequence AAUAAA and is highly conserved among different pre-mRNAs in vertebrates) and one downstream (which is less well defined and is generally represented by a stretch of residues rich in GU; Proudfoot, 1991). The 3' ends of BFDV-1 early and late messengers were identified by RNA-PCR using the d(T)_1 adaptor and a gene-specific primer (#105 for early and #104 for late mRNAs). (b) Positions of 3' ends of early and late transcripts including cleavage/poly(A) sites and two signal elements (AAUAAA and GU-rich region).
the cleavage reactions occurred consistently at position CA, which is 18 nt downstream of the early AAUAAA signal, and at position CA, which is 15 nt downstream of the late mRNA signal (Fig. 6b). Both of these overlap each other in opposite directions on the genetic map at positions 2948 and 3001, respectively. Both intervening sequences are dissimilar to each other, but extend into downstream GU-containing regions.

We wish to thank U. Ruppert and N. Pohl for excellent assistance. This work was supported by the Deutsche Forschungsgemeinschaft SFB 272.

References


(Received 19 July 1994; Accepted 9 September 1994)