Over the past few years, the transcription profiles of many of the parvoviruses have been determined at a detailed level. In general, these profiles are similar, but also display many variations on a theme. In each case, parvoviruses have adapted a complex pattern of alternative splicing and polyadenylation to maximize the information from their small genomes. These complex patterns of expression generally result in a set of mRNAs that express two to three capsid proteins from overlapping open reading frames (ORF), and one large and to three small non-structural proteins. In addition, adeno-associated virus (AAV) gene expression has evolved to be dependent on helper-virus functions.

Most alternative processing of parvovirus pre-mRNAs is constitutive, and even minor variations in the levels of the various final products can have dramatic effects on virus replication. However, there have been reported examples, primarily with the human parvovirus B19, in which these events are regulated in a cell-type specific manner and have been implicated in viral tropism.

Thus far the only parvovirus protein that has been shown to be required for efficient RNA processing is the AAV2 Rep78/68 protein. Whether this role is direct, or facilitates the function of Ad gene products in this regard is not known.

Although the characterization of parvovirus alternative splicing and polyadenylation has been extensive, very little is known about the export of parvovirus RNAs. Two interesting cases highlight the importance of this process to parvovirus biology. The cytoplasmic export of single-spliced NS1-encoding MVM RNAs prevents their further splicing to NS2-encoding R2, and thus export helps govern the relative ratio of the two non-structural proteins. In the case of AAV2, two critical Rep proteins (Rep78 and Rep52), are generated from completely unspliced RNAs, which are exported to the cytoplasm at high efficiency.

Parvovirus RNA processing is an excellent model to study alternative RNA processing of complex overlapping transcription units in a small but complete biological system. These investigations have also yielded insight into the basic relationship between splicing and polyadenylation, and into the co-transcriptional nature of these processes.

PARVOVIRUSES OF NON-PRIMATE MAMMALS (PARVOVIRUS AND AMDOVIRUS)

Rodent parvoviruses (genus Parvovirus)

The rodent parvoviruses minute virus of mice (MVM) and H1 have very similar transcription profiles. The most extensive work in this area has been done for MVM (previously reviewed in Pintel et al., 1995), which we will describe in detail (see Figure 18.1). The transcription maps of feline/canine parvovirus (FPV/CPV), porcine parvovirus (PPV) and Aleutian mink disease parvovirus (AMDV) show differences from the MVM model and will be discussed below.

GENERAL TRANSCRIPTION ORGANIZATION OF MINUTE VIRUS OF MICE

MVM RNAs are generated from two promoters, one at map unit 4 (the RNA initiation site is at approximately nucleotide 201) and one at map unit 38 (the RNA initiation site is at approximately nucleotides 2005–2010; reviewed in Pintel et al., 1995). (Nucleotide numbers refer to GenBank accession number NC_001510.) All MVM RNAs are polyadenylated at the far right-hand end of the genome, approximately 16–20 nucleotides downstream of the final AAUAAA motif that lies at nucleotide 4885 for
MVMp and nucleotide 4821 for MVMi (Clemens and Pintel, 1987). All MVM pre-mRNAs contain a small intron in the center of the genome that is alternatively spliced using two donors and two acceptors (Jongeneel et al., 1986; Morgan and Ward, 1986; Cotmore and Tattersall, 1990; Clemens et al., 1990) and a portion of the P4-generated R1 RNAs are additionally spliced between nucleotides 514 and 1989 to generate R2 (Cotmore et al., 1983; Pintel et al., 1983; Jongeneel et al., 1986). Polyadenylation of MVM RNAs precedes splicing in the nucleus (unspliced polyadenylated molecules can be detected; Clemens and Pintel, 1988) and there is no detectable accumulation of unspliced MVM RNAs in the cytoplasm of infected cells (Clemens and Pintel, 1988).

There is temporal phasing to the production of MVM RNA. Products of the P4 promoter, R1 and R2, which encode the non-structural proteins NS1 and NS2, are generated prior to the R3 RNAs produced from P38, which encode the capsid proteins VP1 and VP2 (Clemens and Pintel, 1988). MVM mRNAs are very stable (>6 hours) in infected cells (Schoborg and Pintel, 1991) and can comprise close to 25 percent of the total mRNA at late times following infection (Pintel et al., 1983).

P38-GENERATED RNAs (R3)
The P38-generated pre-mRNAs (R3) are alternatively spliced to generate two mRNAs, which separately encode the two capsid proteins VP1 and VP2 (Jongeneel et al., 1986; Labieniec-Pintel and Pintel, 1986; Clemens et al., 1990). Therefore, alternative splicing of the P38-generated pre-mRNA governs the relative ratio of accumulation of the capsid proteins during infection (see Figure 18.2). RNAs that splice using the first donor and first acceptor (between nucleotides 2280 and 2377; D1-A1) are predominant and have been given the M (major) designation (in this case R3M; Clemens and Pintel, 1988; Schoborg and Pintel, 1991). The VP2 protein is translated from this message, beginning at an AUG codon at nucleotide 2794 (Labieniec-Pintel and Pintel, 1986). P38-generated pre-mRNAs are also spliced between the second donor and second acceptor donors and acceptors as discussed in the text. The locations of the promoters (P4 and P38) are indicated. The large intron is excised using a non-consensus donor (ncD), and a poor polypyrimidine tract ([Py]n) at its 3′ splice site. The different ORFs that are used are shown in different shading patterns.

Figure 18.1 Transcription Map of MVM. The three major transcript classes (R1, R2, R3) are shown relative to the 5 kb genome diagrammed below. The proteins that they encode are listed on the right of each transcript, and the open reading frames (ORFs) used for each are indicated. The small intron in the center of the genome is spliced using two donors and two acceptors as discussed in the text. The locations of the promoters (P4 and P38) are indicated. The large intron is excised using a non-consensus donor (ncD), and a poor polypyrimidine tract ([Py]n) at its 3′ splice site. The different ORFs that are used are shown in different shading patterns.

Figure 18.2 Nucleotide sequence of the small intron donors, acceptors, and intron splicing enhancer motif (IES), of various parvoviruses. The nucleotide sequences of the donor and acceptor sites for the small intron of MVM H1, CPV, and PPV are shown, relative to published consensus sequences. Also shown is the homology to the experimentally defined IES of the MVM small intron. CPV and PPV use a single acceptor for small intron excision. (Py)n indicates the polypyrimidine region of the 3′ splice sites.

MVMp and nucleotide 4821 for MVMi (Clemens and Pintel, 1987). All MVM pre-mRNAs contain a small intron in the center of the genome that is alternatively spliced using two donors and two acceptors (Jongeneel et al., 1986; Morgan and Ward, 1986; Cotmore and Tattersall, 1990; Clemens et al., 1990) and a portion of the P4-generated R1 RNAs are additionally spliced between nucleotides 514 and 1989 to generate R2 (Cotmore et al., 1983; Pintel et al., 1983; Jongeneel et al., 1986). Polyadenylation of MVM RNAs precedes splicing in the nucleus (unspliced polyadenylated molecules can be detected; Clemens and Pintel, 1988) and there is no detectable accumulation of unspliced MVM RNAs in the cytoplasm of infected cells (Clemens and Pintel, 1988). There is temporal phasing to the production of MVM RNA. Products of the P4 promoter, R1 and R2, which encode the non-structural proteins NS1 and NS2, are generated prior to the R3 RNAs produced from P38, which encode the capsid proteins VP1 and VP2 (Clemens and Pintel, 1988). MVM mRNAs are very stable (>6 hours) in infected cells (Schoborg and Pintel, 1991) and can comprise close to 25 percent of the total mRNA at late times following infection (Pintel et al., 1983).

P38-GENERATED RNAs (R3) The P38-generated pre-mRNAs (R3) are alternatively spliced to generate two mRNAs, which separately encode the two capsid proteins VP1 and VP2 (Jongeneel et al., 1986; Labieniec-Pintel and Pintel, 1986; Clemens et al., 1990). Therefore, alternative splicing of the P38-generated pre-mRNA governs the relative ratio of accumulation of the capsid proteins during infection (see Figure 18.2). RNAs that splice using the first donor and first acceptor (between nucleotides 2280 and 2377; D1-A1) are predominant and have been given the M (major) designation (in this case R3M; Clemens and Pintel, 1988; Schoborg and Pintel, 1991). The VP2 protein is translated from this message, beginning at an AUG codon at nucleotide 2794 (Labieniec-Pintel and Pintel, 1986). P38-generated pre-mRNAs are also spliced between the second donor and second acceptor donors and acceptors as discussed in the text. The locations of the promoters (P4 and P38) are indicated. The large intron is excised using a non-consensus donor (ncD), and a poor polypyrimidine tract ([Py]n) at its 3′ splice site. The different ORFs that are used are shown in different shading patterns.

Figure 18.1 Transcription Map of MVM. The three major transcript classes (R1, R2, R3) are shown relative to the 5 kb genome diagrammed below. The proteins that they encode are listed on the right of each transcript, and the open reading frames (ORFs) used for each are indicated. The small intron in the center of the genome is spliced using two donors and two acceptors as discussed in the text. The locations of the promoters (P4 and P38) are indicated. The large intron is excised using a non-consensus donor (ncD), and a poor polypyrimidine tract ([Py]n) at its 3′ splice site. The different ORFs that are used are shown in different shading patterns.

Figure 18.2 Nucleotide sequence of the small intron donors, acceptors, and intron splicing enhancer motif (IES), of various parvoviruses. The nucleotide sequences of the donor and acceptor sites for the small intron of MVM H1, CPV, and PPV are shown, relative to published consensus sequences. Also shown is the homology to the experimentally defined IES of the MVM small intron. CPV and PPV use a single acceptor for small intron excision. (Py)n indicates the polypyrimidine region of the 3′ splice sites.
the total protein in viral capsids (Tattersall et al., 1977, 1990). This RNA encodes the VP1 protein from an AUG codon at nucleotide 2286 (Labieniec-Pintel and Pintel, 1986). R3 pre-mRNAs that splice using D1 and A2 have also been identified (Morgan and Ward, 1986). This species, designated R3r (rare), would be predicted to encode VP2. The major D1-A1, minor D2-A2, and rare D1-A2 splicing patterns are found, respectively, in approximately 70 percent, 25 percent, and approximately 5 percent of P4-generated mRNA, respectively (Schoborg and Pintel, 1991). Interestingly, while R3m, which uses D2 and encodes VP1, comprises approximately 25 percent of the total capsid protein in infected cells, reflecting the relative levels of the R3m mRNA, it only makes up approximately 10 percent of the total protein in viral capsids (Tattersall et al., 1977), suggesting there may exist an additional constraint on the morphogenesis of the MVM virion.

The two donor sites of the MVM small intron compete for the splicing machinery: experiments in which the donor sites were exchanged have shown that the position of D1 favors its usage, while the primary sequence of D2 must be more like the consensus sequence than the upstream donor to be used efficiently (Haut and Pintel, 1998). However, recognition of the elements comprising the small intron acceptors is likely to be the dominant determinant in alternative small intron excision (Haut and Pintel, 1998; Haut and Pintel, 1999). Although the downstream acceptor position seems to be preferred, the splicing machinery preferentially uses A1, which has a better polypyrimidine tract (PPT), two functional potential branch points, and a more consensus 3′ acceptor (Haut and Pintel, 1998). Experiments in which the acceptor signals have been switched have shown that the sequence of A1 must be stronger than that of A2 for it to be used more efficiently than A2 (Haut and Pintel, 1998). Recognition of A1 by splicing factors is a critical determinant that leads to preferential use of D1 and generation of the major D1/A1 spliced form, since D2 is too close to be efficiently spliced to A1 (Haut and Pintel, 1998, 1999).

SPlicing of the minor spliced form D2/A2 requires the presence of an ISE downstream of D2 (Haut and Pintel, 1998, 1999). The relatively greater strength of A1 seems to impose a requirement for the IES on A2, and consequently, on production of the minor spliced form. This was concluded because, when the sequence of A2 was improved, usage of A1 was decreased and usage of the downstream acceptor no longer required the IES (Haut and Pintel, 1998, 1999). Consequently, debilitation of A1 also rendered usage of the downstream acceptor independent of the IES (Haut and Pintel, 1998).

The small intron appears to be defined primarily by an ‘intron definition’ model, in which all the elements that control its excision are contained wholly within the intron (Haut and Pintel, 1998). Consistent with this hypothesis, increasing the size of the intron to a length no longer compatible with intron-defined excision was shown to change the relative use of the donors and acceptors (Haut and Pintel, 1998, 1999). Because the small intron defines the 3′ border of the NS2-specific exon, it also plays an important role in definition of the NS2-specific exon, and hence excision of the large intron (Haut and Pintel, 1999).

Increasing the size of the small intron diminishes its help in this regard (Haut and Pintel, 1999).

**REGULATION OF SPlicing OF THE SMALL INtron FROM P38-GENERATED PRE-mRNAs**

The determinants that govern the alternative excision of the small intron from P38-generated pre-mRNAs are complex (Haut and Pintel, 1998, 1999). Factors that govern splicing of the small intron include similarity to consensus of the donors and acceptors, the position of the two donors relative to each other, the size of the intron, and the presence of an intronic splicing enhancer sequence (Haut and Pintel, 1998). R3M pre-mRNAs that splice using D1 and A2 comprise approximately 75 percent and 25 percent of P38-generated mRNA, respectively (Schoborg and Pintel, 1991). No splicing is detected between D2 and A1 for any MVM-generated pre-mRNA, probably, at least in part, because the intron is very small (60 nucleotides; Morgan and Ward, 1986; Schoborg and Pintel, 1991).

**NS2-EXON DEFINITION AND EXCISION OF THE LARGE UPSTREAM INtron**

Splicing of the large intron from P4-generated pre-mRNAs is a complicated process that involves small intron excision, NS2-specific exon definition, and selective RNA export (Zhao et al., 1994; Zhao et al., 1995a, 1995b; Gersappe and Pintel, 1999; Haut and Pintel, 1999). The relative rates at which these RNAs are processed help determine the relative ratio of the non-structural proteins NS1 and NS2. At steady-state levels, there is approximately twice the amount of R2 as R1 in MVM-infected murine cells (Schoborg and Pintel, 1991). Alternative splicing of the small intron does not affect expression of NS1, which terminates prior to D1. However, alternative splicing of R2 RNAs at the small intron generates three NS2 isoforms that vary at their COOH termini (Clemens et al., 1990; Cotmore and Tattersall, 1990).

All P4-generated RNAs that are found in the cytoplasm are spliced at the small intron; no unspliced P4-generated RNAs are exported from the nucleus of infected cells (Clemens and Pintel, 1988). Following small intron splicing, a portion of the P4-generated RNAs are exported from the nucleus (comprising the R1 class of MVM RNAs); however, a portion is spliced again, removing the large intron before export. These are the R2 mRNAs. Export of R1 from the nucleus thus prevents further splicing to R2. The determinants that govern export of R1 versus its nuclear retention and further splicing to R2 – a critical feature in determining
the ultimate accumulated levels of NS1 and NS2 – is not known. However, it seems at least in part to be related to the efficiency of excision of the large intron. When splicing of the large intron is artificially made more efficient, less mature R1 accumulates in the cytoplasm (Choi and Pintel, unpublished). The large intron has a non-consensus 5’ donor site and a poor 3’ acceptor. If either of these signals is improved, large intron excision is considerably more efficient (Zhao et al., 1995a). Excision of the wild-type large intron, however, can best be understood based on exon definition models (see Figure 18.3).

As mentioned above, the first step in splicing of the P4-generated pre-mRNAs is likely to be splicosomal engagement and excision of the small intron to generate R1. This has been inferred because no molecules in which only the large intron has been excised are detected in infected cells (Clemens and Pintel, 1988). Subsequent splicing of the large intron, from a subset of spliced R1 molecules, requires the definition of the NS2-specific exon, which is flanked by a poor 3’ splice site at its upstream border (Zhao et al., 1995b; Gersappe and Pintel, 1999). Efficient recognition of this 3’ splice site by the splicing machinery requires an exon splicing enhancer (ESE) that lies within the NS2 specific exon (Zhao et al., 1995b; Gersappe and Pintel, 1999). The NS2-specific ESE is bipartite; mutation of either element alone has only a minor effect on excision of the upstream large intron (Gersappe and Pintel, 1999). Mutation of both elements, however, leads to loss of definition of the NS2-specific exon with its subsequent skipping; almost all of the P4-generated RNA in such ESE mutants is spliced between the large intron donor and the small intron acceptor A1 (Gersappe and Pintel, 1999). That these mutations can be suppressed by improvement of the polypyrimidine tract of the large intron 3’ splice site suggests that the ESE acts to strengthen this site (Zhao et al., 1995b; Gersappe and Pintel, 1999). The sequence within the NS2-specific exon is tightly constrained. In addition to a sequence-specific ESE, this region is translated in two ORFs. The splicing factor(s) that bind(s) to the NS2-specific ESE has not yet been identified.

As well as defining differences in their capsid sequence, the two allotropic variants of MVM, MVMp and MVMi, have different 3’ splice sites, which affect the splicing of the large intron (Gardiner and Tattersall, 1988). The 3’ splice site of MVMp is much more efficient in mouse fibroblasts than the 3’ splice site of MVMi. This difference can be attributed to an A residue at nucleotide 1970 in MVMp, which may aid branch point selection, rather than the G residue found in MVMi (Choi, Burger and Pintel, unpublished). The 3’ splice site of MVMi is as efficient in murine lymphocyte cell lines, as is the 3’ splice site of MVMp in murine fibroblasts (Choi, Burger and Pintel, unpublished).

Definition of exons also requires a strong downstream 5’ donor site. In the case of the NS2-specific exon, the strength, or efficiency, of usage of the small intron donors (which constitute the 3’ end of the NS2-specific exon), is dependent on the complex regulation of the small intron (Haut and Pintel, 1999). Mutation of the small intron ISE, or expansion of the small intron in such a way that small intron excision is altered, results in loss of NS2-specific exon definition and so P4-generated mRNAs are spliced from the large intron donor to the small intron acceptor A1, skipping the NS2-specific exon (Haut and Pintel, 1999).

Although MVM (and in fact all the rodent parvoviruses) use a non-consensus large intron 5’ donor site, each has a strongly consensus 5’ donor site four to five nucleotides downstream of this site that is not used. Use of this downstream donor would join the NS2-specific exon out of frame, and an understanding of the molecular constraints that prevent the use of this seemingly more favorable site will be of general interest. Recent evidence has suggested that both an ESE in the upstream NS1/NS2-shared exon as well as intronic sequences downstream of the donor governs usage of the non-consensus large intron donor (Choi and Pintel, unpublished).

![Figure 18.3](image-url) Detailed map of the MVM NS2-specific exon. The region of the NS2-specific exon in both R1 and R2 is shown, relative to the genome below. The large intron 3' splice site, and the small intron donors, which comprise the 5' and 3' termini of the NS2-specific exon, respectively, are shown. Also shown are the SX1 C/A-rich and HS2 purine-rich elements of the NS2-specific exon bipartite exonic splicing enhancer. In addition, the small intron intronic splicing enhancer (IES), as well as the location of the poor polypyrimidine tract (Py)n of the large intron 3' splice site are shown.
Porcine (PPV) and canine/feline paroviruses (CPV/FPV; genus Parovirus)

A detailed picture of the primary structure of the mRNAs generated by PPV has been obtained (Bergeron et al., 1993); however, the relative steady-state abundances of these mRNAs has not yet been fully determined. In contrast to MVM, the PPV small intron has two donor sites and only a single acceptor site (see Figure 18.2). Both the NS1 and NS2 coding regions of PPV terminate prior to the small intron splice sites and so alternative splicing of the small intron affects only production of the capsid proteins. Perhaps the most striking difference between PPV and MVM is the presence of an internal exon. During infection, but as described above, is generated after mutagenesis of the small intron donors or the NS2-specific splice site is very strong, relative to the 3′/H11032 splice site. PPV R3 message is not produced by other paroviruses in non-primate mammals (Alexandersen et al., 1988; Storgaard et al., 1997). P3- and P36-generated transcripts, which terminate near the right hand end of the molecule (AMDV also uses a splicing pattern of MVM and PPV, however, is that PPV has been shown to encode an mRNA (R3), which joins the large intron donor to the small single intron acceptor, skipping the NS2-specific exon. The small non-structural protein NS3, which is encoded by the PPV R3 mRNA, may be required to compensate for the absence of multiple isoforms of PPV NS2 encoded by the PPV R3 message is not produced by other paroviruses in this group and because its only identification has been by the cloning of PCR-generated fragments, it bears additional validation by other laboratories. For MVM, an mRNA in which the NS2-specific exon is skipped, is not seen during infection, but as described above, is generated after mutagenesis of the small intron donors or the NS2-specific exon itself: Both the PPV large intron donor (AC/GGCAAG) and 3′ polypyrimidine tract (TAAATACACCAACAG/AC) are further from consensus than the comparable MVM sequences and, as is the case for MVM, the PPV small intron acceptor appears significantly stronger than the PPV large intron acceptor. Finally, the PPV NS2-specific exon is only 75 percent the size of the MVM NS2-specific exon and computer-aided structural predictions (Jacobson and Zuker, 1993) of the sequences within this exon are quite different from the other RV-like paroviruses. Some or all of these features probably favor the skipping of the NS2-specific exon of PPV.

CPV is similar to PPV in that its small intron uses two donors and a single acceptor (Wang et al., 1998). The CPV donors are identical to MVM; however, in contrast to MVM, the polypyrimidine tract of the CPV 3′ small intron splice site is very strong, relative to the 3′ splice site of the large intron. Therefore, while MVM, H1, CPV, and PPV have similarly consensus small intron donor sites, their small intron acceptor sites – and thus the relative strength of these sites compared with their large intron 3′ splice sites – vary significantly. Interestingly, although R2-like molecules can be detected, a role for the NS2 protein during CPV infection has not yet been demonstrated (Wang et al., 1998). Accumulation of RNAs generated by the related mink enteritis virus (MEV) during a highly synchronous infection of CRFK cells is qualitatively and quantitatively similar to the profile seen for MVM (Storgaard et al., 1997).

A detailed picture of the primary structure of the mRNAs generated by PPV has been obtained (Bergeron et al., 1993); however, the relative steady-state abundances of these mRNAs has not yet been fully determined. In contrast to MVM, the PPV small intron has two donor sites and only a single acceptor site (see Figure 18.2). Both the NS1 and NS2 coding regions of PPV terminate prior to the small intron splice sites and so alternative splicing of the small intron affects only production of the capsid proteins. Perhaps the most striking difference between PPV and MVM is the presence of an internal exon. During infection, but as described above, is generated after mutagenesis of the small intron donors or the NS2-specific splice site is very strong, relative to the 3′/H11032 splice site. PPV R3 message is not produced by other paroviruses in non-primate mammals (Alexandersen et al., 1988; Storgaard et al., 1997). P3- and P36-generated transcripts, which terminate near the right hand end of the molecule (AMDV also uses a splicing pattern of MVM and PPV, however, is that PPV has been shown to encode an mRNA (R3), which joins the large intron donor to the small single intron acceptor, skipping the NS2-specific exon. The small non-structural protein NS3, which is encoded by the PPV R3 mRNA, may be required to compensate for the absence of multiple isoforms of PPV NS2 encoded by the PPV R3 message is not produced by other paroviruses in this group and because its only identification has been by the cloning of PCR-generated fragments, it bears additional validation by other laboratories. For MVM, an mRNA in which the NS2-specific exon is skipped, is not seen during infection, but as described above, is generated after mutagenesis of the small intron donors or the NS2-specific exon itself: Both the PPV large intron donor (AC/GGCAAG) and 3′ polypyrimidine tract (TAAATACACCAACAG/AC) are further from consensus than the comparable MVM sequences and, as is the case for MVM, the PPV small intron acceptor appears significantly stronger than the PPV large intron acceptor. Finally, the PPV NS2-specific exon is only 75 percent the size of the MVM NS2-specific exon and computer-aided structural predictions (Jacobson and Zuker, 1993) of the sequences within this exon are quite different from the other RV-like paroviruses. Some or all of these features probably favor the skipping of the NS2-specific exon of PPV.

CPV is similar to PPV in that its small intron uses two donors and a single acceptor (Wang et al., 1998). The CPV donors are identical to MVM; however, in contrast to MVM, the polypyrimidine tract of the CPV 3′ small intron splice site is very strong, relative to the 3′ splice site of the large intron. Therefore, while MVM, H1, CPV, and PPV have similarly consensus small intron donor sites, their small intron acceptor sites – and thus the relative strength of these sites compared with their large intron 3′ splice sites – vary significantly. Interestingly, although R2-like molecules can be detected, a role for the NS2 protein during CPV infection has not yet been demonstrated (Wang et al., 1998). Accumulation of RNAs generated by the related mink enteritis virus (MEV) during a highly synchronous infection of CRFK cells is qualitatively and quantitatively similar to the profile seen for MVM (Storgaard et al., 1997).

Aleutian Mink Disease parovirus (AMDV; genus Amdovirus)

Aleutian Mink Disease parovirus has a more complex transcription profile than other paroviruses of non-human mammals (Alexandersen et al., 1988; Storgaard et al., 1997; see Figure 18.4). Two promoters have been identified, at map unit 3 and map unit 36; however, the splicing and polyadenylation of AMDV RNA is quite different from other autonomous paroviruses (Alexandersen et al., 1988; Storgaard et al., 1997). P3- and P36-generated transcripts, which terminate near the right hand end of the molecule (AMDV also uses a splicing pattern of MVM and PPV, however, is that PPV has been shown to encode an mRNA (R3), which joins the large intron donor to the small single intron acceptor, skipping the NS2-specific exon. The small non-structural protein NS3, which is encoded by the PPV R3 mRNA, may be required to compensate for the absence of multiple isoforms of PPV NS2 encoded by the PPV R3 message is not produced by other paroviruses in this group and because its only identification has been by the cloning of PCR-generated fragments, it bears additional validation by other laboratories. For MVM, an mRNA in which the NS2-specific exon is skipped, is not seen during infection, but as described above, is generated after mutagenesis of the small intron donors or the NS2-specific exon itself: Both the PPV large intron donor (AC/GGCAAG) and 3′ polypyrimidine tract (TAAATACACCAACAG/AC) are further from consensus than the comparable MVM sequences and, as is the case for MVM, the PPV small intron acceptor appears significantly stronger than the PPV large intron acceptor. Finally, the PPV NS2-specific exon is only 75 percent the size of the MVM NS2-specific exon and computer-aided structural predictions (Jacobson and Zuker, 1993) of the sequences within this exon are quite different from the other RV-like paroviruses. Some or all of these features probably favor the skipping of the NS2-specific exon of PPV.

CPV is similar to PPV in that its small intron uses two donors and a single acceptor (Wang et al., 1998). The CPV donors are identical to MVM; however, in contrast to MVM, the polypyrimidine tract of the CPV 3′ small intron splice site is very strong, relative to the 3′ splice site of the large intron. Therefore, while MVM, H1, CPV, and PPV have similarly consensus small intron donor sites, their small intron acceptor sites – and thus the relative strength of these sites compared with their large intron 3′ splice sites – vary significantly. Interestingly, although R2-like molecules can be detected, a role for the NS2 protein during CPV infection has not yet been demonstrated (Wang et al., 1998). Accumulation of RNAs generated by the related mink enteritis virus (MEV) during a highly synchronous infection of CRFK cells is qualitatively and quantitatively similar to the profile seen for MVM (Storgaard et al., 1997).
in ORF 1 prior to the first small intron (NS1 is encoded in ORF 3 in this region) and would be predicted to encode a unique small non-structural protein (NS3; Alexandersen et al., 1988). In contrast to AA V5 and B19 (see below), unspliced RNAs from P3 and P36 are apparently not polyadenylated at the proximal site, even though this polyadenylation signal is consensus and is used efficiently for the R2/H11032 and RX RNAs.

Parvoviruses of Humans and Non-Human Primates

Dependoviruses

AAVs have been isolated from multiple species including humans and other primates (Parks et al., 1967; Blacklow et al., 1968; Bantel-Schaal and zur Hausen, 1984; Georg-Fries et al., 1984; Gao et al., 2002), birds (Yates et al., 1973), bovine (Coria and Lehmkuhl, 1978), sheep (Clarke et al., 1979), and snakes (Farkas et al., 2004). Survey of the transcriptional profiles of human AAV serotypes 1–6, has shown that, with respect to their transcription profiles, these viruses fall into two main groups (Qu and Pintel, unpublished). The first group, for which AAV2 can be taken as the prototype, includes AAV1, AAV2, AAV3, AAV4, and AAV6. They all possess similar large (approximately 320 nucleotides) introns and a non-consensus donor and acceptor site. The RNAs generated from these viruses all polyadenylate only at a site at the right-hand end of the genome (although they all contain a cryptic AAUAAA site within their intron), and they show a variable dependence on helper viruses. AAV2 is very dependent on both helper virus and Rep for proper RNA processing and high levels of

Figure 18.4 Transcription map of Aleutian mink disease parvovirus (AMDV). The AMDV genome is depicted on the top of the figure. The promoters (P3 and P36), splice donors (D) and acceptors (A), polyadenylation sites ([pA]p and [pA]d), translation initiation sites and terminal palindromes (TR) are shown. The individual RNAs that have been detected, and the proteins they are predicted to encode are also shown. The different open reading frames (ORFs) that are used are shown in different shading patterns.
transcription initiation and AA V3 and AA V4 are similar. AA V1 and AA V6, however, generate relatively low levels of spliced RNA products and are significantly less responsive to Ad5 helper functions for enhancement of RNA processing. The second main group, comprising AA V5, Avian-AA V and goose parvovirus (GPV), use an internal polyadenylation site within the central viral intron. Plasmid clones of these viruses express constitutively higher levels of spliced RNAs and so are less dependent on Ad5 or HSV helper virus, or Rep, for these aspects of gene expression. They will be described in detail below. Although portions of the genomes of AA V7 and AA V8 have been isolated from monkey tissue (Gao et al., 2002), infectious virus has not yet been characterized.

AAV2-LIKE ADENO-ASSOCIATED VIRUSES

General transcription organization of AA V2. The transcriptional organization of AA V2 has been previously reviewed (Carter et al., 1990; Muzyczka and Berns, 2001). The AA V2 genome has three different promoters identified by their map positions of P5, P19, and P40 (Green and Roeder, 1980; Lusby and Berns, 1982; see Figure 18.5). All of the AA V2 transcripts contain a single intron, which uses a 5’ splice donor at nucleotide 1906, and one of either two splice acceptors at nucleotide 2201 (A1) or nucleotide 2228 (A2) (see Carter et al., 1990). Unspliced P5- and P19-generated RNAs encode Rep78 and Rep52, respectively, while spliced P5- and P19-generated RNAs encode Rep68 and Rep40. Alternative splicing to either A1 or A2 generates two isoforms each for Rep68 and Rep40; however, individual functions for the different isoforms generated from RNAs using these different acceptors has not yet been reported. Spliced P40-generated RNAs using the A1 acceptor encode the VP1 capsid protein, while spliced RNAs using the A2 acceptor encode VP2 and VP3 (Trempe and Carter, 1988a). Unspliced P40 mRNAs express very little capsid proteins even though they contain the entire VP1 ORF (Becerra et al., 1988). This may be because there are a number of short ORFs within the unspliced P40 transcripts that are available and may be used instead (Carter et al., 1990). Recombinant AA V vectors have been shown to express transcripts generated from the AA V2 ITR (Flotte et al., 1992, 1993; Haberman et al., 2000) and a similar set of ITR-initiated transcripts has been detected at low levels during AA V2 infection (Nayak, Tullis and Pintel, unpublished). The biological function of these ITR-initiated transcripts remains to be defined. AA V2 gene expression is tightly controlled by both the AAV-encoded Rep proteins (Qiu and Pintel, 2002) and co-infecting helper virus (see Carter et al., 1990).

Similar to the autonomous parvovirus MVM, there is a temporal order to the appearance of AA V2 RNAs during infection (Labow et al., 1986; Trempe and Carter, 1988a; Mouw and Pintel, 2000). Unspliced P5-generated transcripts are detectable prior to the significant accumulation of other AA V2 RNAs. Ultimately, P19-generated RNAs accumulate to levels greater than those generated from P5, and P40-generated transcripts come to predominate in the total RNA pool. In addition, the percentage of each class of AA V2 RNA that is spliced increases during infection (Mouw and Pintel, 2000). The degree of this increase is different for the P5 and P19 RNA compared with those generated by P40. At late times post infection, 90 percent of P40 RNAs, but only 50 percent of RNA generated by P19, and <10 percent of RNAs generated by P5, are spliced. Surprisingly, therefore, the same AA V2 intron is excised to different final levels from those different RNA species. In addition, it has been observed that the ratio of splicing to A2 relative to A1 also increases as AA V2 infection progresses; however, as noted above, the functional relevance of this is not clear (Mouw and Pintel, 2000). All AA V2 RNAs have been shown to be quite stable during infection (Mouw and Pintel, 2000).

AAV2 RNA SPlicing

Pre-mRNA splicing is a critical determinant for productive AA V infection. Although unspliced RNAs migrate to the cytoplasm and unspliced P5- and P19-generated RNAs are

Figure 18.5 Transcription map of AA V2. The AA V2 genome is depicted at the bottom of the figure, and includes the location of the viral promoters (P5, P19, P40), the small intron donor (D) and acceptors (A1 and A2), the termination site for the Rep proteins, and the inverted terminal repeats (ITRs). The major transcript classes, and the proteins that they encode are shown above. The different open reading frames (ORFs) that are used are shown in different shading patterns.
translated into essential viral proteins (see below), the ratio of the viral capsid proteins is determined by alternative splicing and the ratio of Rep78 and Rep52 versus Rep68 and Rep40 proteins depends upon the level of splicing of the P5 and P19-generated RNAs (Carter et al., 1990; Myuzyczka and Berns, 2001). Splicing of AAV2 pre-mRNAs requires the participation of both the helper virus (e.g. adenovirus; Myez et al., 1980; Trempe and Carter, 1988b; Berns and Giraud, 1996; Mouw and Pintel, 2000), or herpes virus (Bullert et al., 1981; Mishra and Rose, 1990; Mouw and Pirone, 2000). Adenovirus gene products in enhancement of AAV2 RNA splicing has not be achieved using those herpes gene products suggests that Rep enhancement of splicing is not clear. The role of the large Rep proteins in AAV2 RNA splicing. AAV2 large Rep proteins (Rep78/68) can act to increase the ratio of spliced to unspliced AAV RNA when they are targeted to the transcription template via a Rep-binding element (RBE; Qiu and Pintel, 2002). An extended RBE provided either by the P5 promoter or the viral ITR can function to target Rep in a way that sustains full enhancement of the splicing of P40-generated RNA. This effect requires the presence of adenovirus gene products, and if the transcription template lacks an extended Rep-binding element, neither Ad5 nor Rep have significant detectable effects on AAV2 RNA splicing. It has so far not been possible to separate Rep enhancement of splicing from Rep activation of transcription initiation. In addition, the required RBE is both location and orientation independent, similar to the classically defined transcription enhancer-factor binding sites, which suggested that Rep might be acting co-transcriptionally to effect RNA processing.

Although the Rep effect on AAV2 RNA processing appears to be linked to its transcriptional transactivation activity, transactivation of transcription initiation is not sufficient for enhancement of RNA processing (Qiu and Pintel, 2002). Activation of the P40 transcription unit by other activators targeted to transcription template (such as Gal-VP16 and MVM N51) does not enhance AAV2 RNA splicing, although transcription of the P40 promoter is strongly transactivated by these proteins.

Rep enhancement of the relative levels of spliced RNA decreases as the distance between the promoter and the intron of the affected transcription unit increases (Qiu and Pintel, 2002). This is consistent with a co-transcriptional model of enhancement of RNA processing by Rep, and explains why the same AAV2 intron is excised to a much greater extent from the proximal P40-generated transcripts than from upstream P5 and P19 transcripts.

It is not yet clear how Rep and adenovirus affect AAV RNA processing; indeed, it is not yet clear whether the combined effect of adenovirus and Rep is directly on splicing or other RNA processing mechanisms that ultimately determine the steady-state levels of spliced AAV RNA. Together these factors may alter the composition of the elongating RNA polymerase II complex, specifically, by altering levels or types of RNA processing factors associated with the RNA polymerase II carboxy terminal domain (CTD). Alternatively, a direct interaction of Rep with cellular splicing factors could
be possible. The recent demonstration of Rep co-localization in replication centers with helper adenovirus, or HSV, single-strand DNA-binding proteins (ssDBP and ICP8, respectively), following co-transfection (Stracker et al., 2004), suggests it is also possible that Rep functions to enhance AAV2 RNA splicing by targeting AAV2 to these replication centers, which normally are active for splicing (Bridge et al., 1993; Pombo et al., 1994; Bridge et al., 1996).

As described above, the ratio of spliced to unspliced P5- and P19-generated RNAs increases over the course of infection. The mechanism underlying this increase is not known; conceivably, this might be due to either an increase in Rep production, or perhaps, differences in viral localization as infection proceeds.

Export of AAV2 RNA. An unusual feature of AAV2 gene expression is that unspliced RNAs from P5, P19, and P40 promoter accumulate in the cytoplasm at high levels. The P5- and P19-generated unspliced RNAs are translated to generate the non-structural proteins, Rep78 and Rep50, respectively. In general, unspliced cellular and viral RNAs remain in the nucleus. Their export typically requires a regulated pathway including a specific cis-acting export signal, such as the CTE on the HIV rev RNA (Malim et al., 1989). The export and translation of unspliced P5- and P19-generated RNAs may compensate for the poor splicing of P5- and P19-generated RNAs, impaired because of their distance from the intron (see above).

During adenovirus infection, the adenovirus E1B55K-E4orf6 complex both inhibits the export of cellular mRNAs and is also required for efficient cytoplasmic accumulation of adenovirus mRNAs. The E1B55K-E4orf6 complex has also been suggested to be necessary for the timely cytoplasmic accumulation of AAV RNAs during adenovirus co-infection (Samulski and Shenk, 1988). These experiments were performed using adenovirus mutants that individually did not generate either the E1B 55K or E4orf6 proteins. Because of recent experiments highlighting the critical importance of these proteins in multiple steps during the replication and expression of AAV2, their role in export of AAV2 RNA probably needs to be re-evaluated. Following transfection of an infectious clone of AAV2 into 293 cells in the absence of Ad5, unspliced AAV2 RNAs can be efficiently exported to the cytoplasm (Mouw and Pintel, 2000), suggesting that their export can be achieved in the absence of additional adenovirus gene products.

Whether the export of unspliced AAV2 RNAs requires a particular cis-acting signal that directs it through a regulated export pathway, or, alternatively, whether these RNAs are exported via the TAP/Aly1 (Stutz et al., 2000; Zhou et al., 2000) cellular pathway will be interesting to determine.

RNA SPlicing OF OTHER AA V2-LIKE SEROTYPES OF AAV

AAV1, 2, 3, 4, and 6 share a substantial homology within the intron region and exhibit identical donor sites and highly similar acceptor sites (see Figure 18.6). Splicing of AAV3 and AAV4 is basically similar to that of AAV2 with regards to their dependence on helper virus; however, the splicing of AAV1 and AAV6 appears to be less dependent on helper virus (Qiu, Farris and Pintel, unpublished). Most of the P41-generated transcripts generated by AAV1 and AAV6 RepCap plasmids following transfection of 293 cells remain unspliced even in the presence of Ad5. Lower levels of splicing of P40-generated products are also seen during AAV1 and AAV6 viral infection. The levels of splicing seem to be governed by intron sequences. The AAV2 intron, when placed in an AAV1 background, splices at high efficiency, while the AAV1 intron still splices poorly when inserted into AAV2 (Qiu, Farris and Pintel, unpublished). These results suggest that it may be possible to identify cis-acting sequences within the AAV2 intron that confer splicing responsiveness to adenovirus and Rep.

AAV5-LIKE ADENO-ASSOCIATED VIRUSES

GENERAL FEATURES OF AAV5 RNA TRANSCRIPTION

While the basic transcription profile of AAV5 has similarities to that of AAV2, there are also significant differences (see Figure 18.7). Most surprisingly, RNAs generated from both the AAV5 P7 and P19 promoters are efficiently polyadenylated at a site ([pA]p, at nucleotide 2193) lying within the intronic region in the center of the genome (Qiu et al., 2002; nucleotide numbers refer to GenBank accession number AF085716). Because P7- and P19-generated transcripts are polyadenylated at this site and not spliced, Rep78 and Rep52 are the only Rep proteins detected during AAV5 infection of 293 cells. Four different species of P41-generated RNAs have been detected in cells infected with AAV5 or transfected with full-length AAV5 clones. Two unspliced species are detected, one polyadenylated at ([pA]p and one polyadenylated at the distal site ([pA]d) at nucleotide 4434. P41-generated RNAs polyadenylated at ([pA]d are also spliced using the A1 or A2 acceptor and generate VP1 and VP2/VP3, respectively. The function of the unspliced P41-generated RNAs is unknown.

Mapping of the AAV5 transcripts also demonstrated a transcription initiation site within the AAV5 inverted terminal repeat (ITR; Qiu et al., 2002). A similar transcript generated from the ITR of recombinant AAV2 has also been identified (Flotte et al., 1992, 1993). The initiation site of the AAV5 ITR-generated transcripts has been mapped to the terminal resolution site in the hairpin, and are unspliced 4.7 kb RNAs that polyadenylate at ([pA]d. The function of the ITR-initiated RNA remains unknown; however, these molecules are relatively more abundant during viral infection, or following transfection of a replicating, compared with non-replicating, plasmid clones of AAV5 (Qiu et al., 2002).

SPLICING AND POLYADENYLATION OF AAV5 RNA

The AAV5 intron, which is 240 nucleotides, is considerably smaller than the 322 nucleotides AAV2 intron. It has a non-consensus 5’ splice site (AAG/GTATGA) at nucleotide 1990, and the 3’ acceptor site A1 is at nucleotide 2204, and site A2 is at 2231 (Qiu et al., 2002). The polypyrimidine tract in the
AAV DONOR SITES:

AAV1 (1982)

AAV2 (1985)

AAV3 (1982)

AAV4 (1936)

AAV6 (1987)

AAV5 (1968)

GPV (2178)

[Sequence alignment and comparison between AAV and GPV donor sites]

AAV ACCEPTOR AND POLYADENYLATION SITES:

AAV1 (2178)

AAV2 (2158)

AAV3 (2153)

AAV4 (2207)

AAV5 (2163)

GPV (2383)

[Sequence alignment and comparison between AAV and GPV acceptor sites]

Figure 18.6 Nucleotide sequence of the AAV and GPV 5' donor and 3' acceptor sites. The nucleotide sequence of the 5' splice donor region, and the 3' splice acceptor and polyadenylation site region are aligned for AAV1, AAV2, AAV3, AAV4, AAV5, and compared with AAV6 and GPV. A consensus sequence is also presented. The intron bordering nucleotides and potential polyadenylation signals are shown in red. The downstream element (DSE) required for AAV5 polyadenylation at (pA)p is also shown.
Humans and non-human primates 263

region is pyrimidine-rich and splicing of AAV5 intron is constitutively relatively high even in the absence of adenovirus (Qiu et al., 2002). Interestingly, splicing at the A1 site is dependent on the A2 site; mutation of the A2 AG motif also debilitates splicing to A1 (Qiu, Ye and Pintel, unpublished). In contrast, for AAV2, splicing using A1 is independent of A2 (Qiu, Ye and Pintel, unpublished). Also in contrast to AAV2, neither the Rep protein nor additional adenovirus gene products are required to achieve efficient AAV5 promoter activity and pre-mRNA splicing following transfection of a RepCap plasmid clone lacking ITRs into 293 cells (Qiu et al., 2002). The independence of AAV5 splicing is not due merely to its smaller size, but rather to the nature of cis-sequences within the context of the intron (Qiu et al., 2004).

AAV5 RNAs generated from P7 and P19, which encode the large and small Rep proteins, respectively, from the large ORF in the 5′/H11032 half of the genome, use (pA)p at high efficiency (Qiu et al., 2002, 2004). RNAs generated from P41 promoter, however, which encode the virus capsid proteins from the large ORF in the 3′/H11032 half of the genome, use this nearby site with significantly reduced efficiency; they primarily read through to the distal site (pA)d (Qiu et al., 2004). The intron region of AAV5 contains two relatively consensus AAUAAA signals at nucleotides 2177 and 2191, which are immediately upstream of the first intron acceptor A1 at nucleotide 2204 (Qiu et al., 2004). RNA cleavage and polyadenylation occurs 11–14 nucleotides downstream of the first AAUAAA motif (Qiu et al., 2002). AAV2 also contains a consensus AUAAA polyadenylation signal in this region; however, its use can only be detected by sensitive PCR techniques (Tullis and Pintel, unpublished).

Sequences within the AAV5 intron are both sufficient and necessary to direct cleavage and polyadenylation of P7-generated transcripts. Efficient polyadenylation of both P7- or P19-driven transcripts and P41-driven transcripts at (pA)p requires a U-rich downstream element (DSE) that overlaps with the A2 3′ splice acceptor site (Qiu et al., 2004). In addition, efficient polyadenylation of P7-driven transcripts at (pA)p involves an upstream element (USE) that lies prior to the initiation of the P41 transcripts (between nucleotides 1266 and 1637; Qiu et al., 2004). Because the USE lies upstream of the start of the P41-generated RNAs, these RNAs are not subject to the additional level of control that this element imposes on P7-driven RNAs. The USE is not required if the DSE is intact; however, it can compensate for the loss of DSE function (Qiu et al., 2004). Although the relationship between these two elements is not yet clear, the USE, which is lacking in P41-generated RNAs, probably functions to ensure that the AAV5 P7- and P19-generated RNAs are polyadenylated at (pA)p at high efficiency.

Polyadenylation of both P7- or P19-driven transcripts and P41-driven transcripts at (pA)p is significantly increased when it is produced from a replication-competent molecule such as viral infection or an infectious clone (Qiu et al., 2004). This is not merely due to the presence of the ITRs in cis, because ITR-containing templates that cannot replicate due to Rep mutation do not show this increase. Although most P41-generated RNAs read-through (pA)p, 25–28 percent of P41-generated RNA during AAV5 co-infection with adenovirus are polyadenylated at (pA)p. Whether these RNAs have any role during AAV5 infection is not known.

As mentioned above, the AAV5 DSE-containing A2 5′ splice site plays a dual role in both polyadenylation and splicing, and there is clearly a complex regulation between these two processes. The DSE overlaps, but is not congruent with, the 3′ acceptor of the AAV5 intron. The relative strengths of these two signals must allow both for

Figure 18.7  Transcription map of AAV5. The AAV5 genome is depicted at the bottom of the figure, and includes the location of the viral promoters (Inr, P7, P19, P41), the small intron donor (D) and acceptors (A1 and A2), the termination site for the Rep proteins, the internal polyadenylation site (pAp), and the inverted terminal repeats (ITRs). The major transcript classes, and the proteins that they encode are shown above. The different open reading frames (ORFs) that are used are shown in different shading patterns. It is not clear whether a portion of Inr-initiated RNA is spliced or not; this is indicated by a dashed line.
polyadenylation of the upstream promoter-generated RNAs that generate the essential Rep protein and for efficient splicing to generate the correct ratio of spliced P41-generated RNAs to produce the individual capsid proteins at the appropriate stoichiometry. The relative efficiency of these processes is controlled to a significant degree by the strength of the cis-acting signals governing them, and they probably have evolved to program the optimal amounts of viral gene products required for the AAV5 life cycle (Qiu et al., 2004).

In a similar vein, when the (pA)p region of AAV5 is replaced with the internal AAUAAA signal and AAV2, which is normally cryptic in AAV2, or when the internal region of the AAV2 intron is inserted into AAV5 and linked to the AAV5 DSE-containing A2 3′ splice site, polyadenylation at the AAV2 AAUAAA site becomes highly efficient (Qiu and Pintel, unpublished). This suggests that at least one element that differentiates the use of this site between the two viruses is the presence of the DSE in AAV5.

Polyadenylation at (pA)p increases as the distance between the RNA initiation site and the intron and (pA)p site is increased (Qiu and Pintel, 2004). The steady-state level of RNAs polyadenylated at (pA)p is independent of the promoter used or of the intervening sequence but is dependent upon competition with splicing, inhibition by U1 snRNP binding to the intron donor, and the intrinsic efficiency of the cleavage/polyadenylation reaction (Qiu and Pintel, 2004). Each of these determinants shows a marked dependence on the distance between the RNA initiation site and the intron and (pA)p.

The simplest model to explain the interdependence of splicing and polyadenylation in the AAV5 system is that strong binding of U1 to the donor facilitates both splicing and inhibition of (pA)p and that, perhaps as a consequence of 5′ exon definition, U1 binding is stronger when the promoter is close (e.g. P41-generated RNAs, which are inhibited for polyadenylation at (pA)p and splice) and is weaker when the distance is large (e.g. P7-generated transcripts, which splice poorly, escape inhibition by U1 snRNP, and are efficiently polyadenylated at (pA)p) (Figure 18.8). 5′-Terminal exon definition typically requires that factors binding at or near the cap site interact with and stabilize the binding of factors (usually U1 snRNP) to the proximal 5′ splice site, and this interaction may be more critical when the proximal 5′ donor is non-consensus (Lewis et al., 1996). The cap structure, via binding to the cap-binding complex (CBC; Izaurealde et al., 1994, 1995) has been suggested to affect both splicing and polyadenylation. A nuclear CBC (Inoue et al., 1989; Colot et al., 1996) facilitates association of U1 snRNP with the cap-proximal 5′ splice site, and this interaction is not dependent on a strict spacing between the cap and the affected 5′ splice site (Lewis et al., 1996). It seems feasible that for AAV5, CBC binding to the cap site may stabilize factors binding to the non-consensus AAV5 donor, but only when the cap site is very close (78 nucleotides for P41), but not at a distance (1668 nucleotides for P7, 1088 nucleotides for P19). If interactions between components of the CBC and U1 snRNP mediate the distance-related donor inhibition seen for AAV5, it is due to the strengthening of the association of U1 snRNP to the non-consensus donor site, which would lead to the enhancement of U1’s inhibitory activity.

Such a model is also consistent with the distance-related splicing of AAV2 RNA. In that case, the AAV2 Rep protein enhances splicing of proximal P40-generated transcripts to a much greater degree than it enhances splicing of RNAs generated far upstream from P5. It may be that for AAV2, the Rep protein, in the presence of adenovirus, facilitates the stabilization of U1 snRNP to the AAV2 donor site when the RNA initiation site is close to the intron. A potential mechanism that might explain the distance-related processing of AAV RNA, using the AAV5 system as a model, is depicted in Figure 18.8.

**GOOSE PARVOVIRUS (GPV) AND MUSCOVY DUCK PARVOVIRUS (MDPV)**

GPV and MDPV are autonomously replicating viruses that were originally classified into the genus *Parvovirus*, but they are genetically more related to AAVs and accordingly have now been reclassified into the *Dependovirus* genus (Brown et al., 1995; Zadori et al., 1995). The genome of GPV shares greater than 65 percent identity at the nucleotide level with the AAVs (Brown et al., 1995; Zadori et al., 1995) and the Rep protein of GPV shares functional homology with the Rep protein of AAV2 (Smith et al., 1999). The gene expression profile of GPV shares some similarity to that of AAV5; however, it also retains some characteristics of the *Parvovirus* genus (Qiu and Pintel, unpublished). GPV contains an efficiently spliced small intron (247 nucleotides) that is similar in size to AAV5. The 5′ splice site is located at nucleotide 2208, and the 3′ splice sites are located at nucleotide 2436 (A1) and 2454 (A2; nucleotide numbers refer to GenBank accession number NC_001701).

Also similar to AAV5, the Rep-encoding RNAs generated by the GPV P9 promoter efficiently uses an internal polyadenylation site that has a DSE including a U-rich sequence located within its A2 polypyrimidine tract (Qiu and Pintel, unpublished). However, mutational analysis has shown that an additional U-rich stretch downstream of the intron forms part of the GPV DSE signal (Qiu and Pintel, unpublished). GPV has only one AAUAAA signal, which is located at nucleotide 2416, upstream of the A1 acceptor site. Interestingly, the sequence around that signal is identical to the corresponding region of AAV2, although it is not efficiently used in AAV2 for polyadenylation of AAV2 pre-mRNA (Figure 18.6). The transcription profile of MDPV is similar to that of GPV (Qiu and Pintel, unpublished).

**AVIAN-ADENO-ASSOCIATED VIRUS (A-AAV)**

An avian AAV (A-AAV) was first isolated from the Olson strain of quail bronchitis adenovirus (Yates et al., 1973).
The genome of A-AAV is 4694 nucleotides in length and has organization similar to that of other AAVs (Bossis and Chiorini, 2003). The entire genome of A-AAV displays 56–65 percent identity at the nucleotide level with the other known AAVs (Bossis and Chiorini, 2003).

The intron of A-AAV is larger than that of AAV2. It is 453 nucleotides in length, with the 5' splice site at nucleotide 1824, and 3' splice acceptor sites at nucleotides 2247 (A1) and 2277 (A2; nucleotide numbers refer to GenBank accession number AY186198). Like AAV5, splicing of A-AAV RNAs shows less dependence in 293 cells on Ad5. However, in contrast to AAV5, approximately half of the A-AAV RNAs generated from the upstream P5 and P19 promoters are polyadenylated at a (pA)p site (nucleotide 2228) that lies within the intron (Qiu and Pintel, unpublished).

**Erythroviruses**

**HUMAN PARVOVIRUS B19**

**GENERAL TRANSCRIPTION ORGANIZATION**

A total of nine transcripts have been identified from the B19 genome both in viral infection of permissive cells (Oxawa et al., 1987) and following transfection of non-permissive cells (St Amand., 1991; Figure 18.9, left panel). Only one promoter for B19 has been identified, at map unit...
6 (the TATA box is at nucleotide 499; Blundell et al., 1987; Doerig et al., 1987; Ozawa et al., 1987, 1988a; Liu et al., 1991) and so the processing of P6-generated pre-mRNAs is critical in determining the relative levels of B19 gene products. All transcripts driven by P6 have a short leader sequence of 56 nucleotides (nucleotides 531 to 586; nucleotide numbers refer to GenBank accession number AY386330) at their 5′ end. Three transcripts (designated R1, R2, and R3 in Figure 18.9, left panel) terminate at a proximal polyadenylation site [(pA)p] at nucleotide 2819 in the middle of the genome. RNA R1 (Figure 18.9, left panel) is unspliced and encodes the large non-structural protein NS1, which initiates at an AUG codon at nucleotide 616. RNA R2 (Figure 18.9, left panel) is spliced between a donor site at nucleotide 587 (D1) and an acceptor at nucleotide 2089 (A1–1) and encodes a 7.5 kDa protein, the function of which is currently unknown. RNA R3 (Figure 18.9, left panel) is spliced between D1 and an acceptor at nucleotide 2209 (A1–2). No protein product of the (pA)p-terminated R3 RNA has yet been identified.

Six additional transcripts that use a polyadenylation site at nucleotide 5170 at the far right side of the genome (pA)d are also generated. All of these RNAs splice the first intron, using D1 and either A1–1 or A1–2. Two of these, R4 and R5, are not spliced further and encode the VP1 protein. For R5, which is spliced between D1 and A1–2, the VP1-initiating AUG at nucleotide 2624 is the eighth AUG present in the mRNA, and it shows reduced translation of VP1 protein in vitro (Ozawa et al., 1988b). For R4, and all molecules splicing between D1 and A1–1 (including R6 and R8), the full 7.5 kDa ORF is retained. Whether multiple ORFs are translated from any of these potentially polycistronic mRNAs is not known. RNAs that splice the first intron and also splice a second intron, between a donor site at nucleotide 2363 (D2) and an acceptor at nucleotide 3224 (A2–1; R6 and R7), encode the capsid protein VP2, which, in R7: initiates at the first available AUG at nucleotide 3305. VP2 is translated in the same ORF as VP1. Finally, a set of RNAs that splice the first intron and are also spliced between D2 and a second acceptor, much further downstream at nucleotide 4883 (at A2–2; the R8 and R9 mRNAs), encode the 11 kDa proteins. Although it is produced at high abundance, the function of this protein has not yet been characterized.

Figure 18.9  Transcription maps of B19 and SPV. The map of B19 is shown on the left, and the map of SPV is shown on the right. In each case, the viral genome is depicted, and includes the location of the viral promoter (P6), the donors and acceptors, the polyadenylation sites (pA)p and (pA)d, and the terminal repeats (TR). The major transcript classes, their size in nucleotides, and the proteins that they encode are shown below. Open reading frames (ORFs) used to encode the viral proteins are shown above the genome.
Because B19 has a single promoter, temporal regulation of B19 gene expression is also governed to a great extent by processing of the primary P6-generated transcripts (Blundell et al., 1987; Doerig et al., 1987; Ozawa et al., 1987, 1988a; Liu et al., 1991). Accumulation of the NS1 mRNAs that are polyadenylated at (pA)p and not spliced, precedes accumulation of the mRNAs for the structural proteins in non-permissive cells (Yoshimoto et al., 1991).

Small RNAs (<1 kb; R2, R3, R8, R9) are the most abundant RNAs generated during B19 infection (Ozawa et al., 1987; St Amand, 1993). Among them, the 0.8 kb small RNA, R2, which encodes a 7.5 kDa protein, is the most abundant species in non-permissive cells as the ratio of A1–1 usage is greater than for A1–2. (Yoto and Pintel, unpublished). R2 is more abundant than R3 in permissive bone marrow cells from a patient with sickle cell anemia (Ozawa et al., 1987). Likewise, R4 is more abundant than R5 in non-permissive cells. The combination of P1 transcripts (R4 + R5) are more abundant than the NS1 transcript (R1) in viral infection (Ozawa et al., 1987; Liu et al., 1992); however, the VP2 transcripts R6 + R7 are even more abundant. Splicing to the downstream A2 acceptor A2–2 is greater than to the upstream A2 acceptor (A2–1) resulting in more transcripts encoding the 11 kDa protein (R8 + R9), than transcripts encoding VP2 in non-permissive cells (St Amand et al., 1991). As discussed below, there are differences in the relative ratios of the various RNAs generated during viral infection of permissive cells compared with plasmid transfection in non-permissive COS cells, especially in the relative abundance of the NS1-encoding R1 transcripts, which is governed both by a choice between splicing at D1 and polyadenylation at the (pA)p site (Liu et al., 1992).

It has recently been reported that following infection of semipermissive MB02 cells, B19 RNA, rather than splicing between D1 and A1–1 and A1–2, is spliced predominantly between A1–2 and a non-consensus (CT rather than GT) donor site 35 nucleotides downstream of D1 (Brunstein et al., 2000). Such splicing would be predicted to generate a protein that initiates at an AUG codon in the NS1 reading frame, two amino acids prior to the non-consensus donor, and would then be fused, after the splice, into the 7.5 kDa ORF. All B19 spliced mRNAs spliced in this way would be predicted to encode a short NS1/7.5 kDa fusion ORF at their 5’ ends. As mentioned below, a similar donor site has also been detected at low levels in RNA generated by SPV infection of monkey liver (Vashisht et al., 2004). The generality of this splicing event and how these alternative splice sites are related to the tropism of erythroviruses is not yet clear.

DIFFERENCES BETWEEN PERMISSIVE AND NON-PERMISSIVE CELLS, AND REGULATION OF POLYADENYLATION IN THE MIDDLE OF THE GENOME

Replication of B19 is restricted to the erythroid lineage from BFU-E to erythroblasts, with susceptibility to virus increasing with differentiation (Shimomura et al., 1992; Morita et al., 2001). Human bone marrow is permissive and cell lines UT-7/Epo-S1, derived from a megakaryoblastic leukemia cell line (Shimomura et al., 1992; Gallinella et al., 2000; Morita et al., 2001; Morita et al., 2003) and KUB812Ep (Miyagawa et al., 1999), derived from a chronic myeloid leukemia cell line, are at least semipermissive for viral replication.

In non-permissive cells, the majority of B19 transcripts are spliced at the first donor; however, unspliced transcripts, which polyadenylate at (pA)p and encode NS1, are relatively more abundant than those for VP1 (Liu et al., 1992). In permissive cells, the abundance of RNAs spliced at D1 is increased even further relative to unspliced R1 mRNAs. Furthermore, viral infection in non-permissive cells is accompanied by increased read-through of the (pA)p site and so the ratio of RNAs polyadenylated distally at the (pA)d site (R4–R9), relative to RNAs, which polyadenylate proximally at (pA)p (R1 + R2 + R3), is also greater in permissive infections (Liu et al., 1992). These two factors significantly increase the capsid-encoding transcripts (R4–R7), as well as RNAs R8 and R9 relative to the level of the NS1-encoding (R1) transcripts in permissive cells. Whether this difference in the RNA profile is directly related to the inability of virus to replicate in non-permissive cells, or whether a cell-type difference in RNA processing prevents replication has not been firmly established. It has been shown that a B19 plasmid, driven to replicate in normally non-permissive COS cells by the SV40 origin of replication, generates a greater relative abundance of RNAs that read-through the (pA)p site, suggesting a model in which replication facilitates the generation of the read-through capsid-encoding transcripts (Beard et al., 1989; St Amand et al., 1991; Liu et al., 1992; Luo and Astell, 1993; St Amand and Astell, 1993). There are additional differences seen in B19 expression in permissive compared with non-permissive cells that have not been fully characterized. In addition, in non-permissive COS cells, B19 RNAs that are polyadenylated at (pA)p are the spliced R2 RNAs, which encode the 7.5 kDa protein. Therefore, regulation of splicing at the first donor site can be considered to be a critical determinant that governs the relative level of NS1, which is encoded by R1. D1 usage appears to be greater in permissive compared with non-permissive cells.

The B19 (pA)p signal in the middle of the genome, ATTAAA, is unusual. Mutational studies have shown that this non-consensus ‘ATTAAA’ motif is used as a polyadenylation signal, while the similar motif, ‘AATAAC’, which is immediately adjacent downstream, is not (Liu et al., 1992; Yoto and Pintel, unpublished). Mutation of the second motif, however, results in enhanced polyadenylation at (pA)p suggesting that the second motif can compete with the first motif for efficient polyadenylation at this site (Yoto and Pintel). There are a number of examples in which non-consensus polyadenylation signals are used to regulate polyadenylation (St Amand, 1991; Tabaska and Zhang, 1999). Non-consensus sequences are typically weaker binding sites for CPSF than consensus sites and this may allow the pre-mRNAs that eventually encode B19 VP1, VP2, and the 11 kDa proteins to read-through the internal polyadenylation site more efficiently and for this signal to remain unused within the VP1-encoding mRNA.
Alternative polyadenylation of B19 RNA is governed by many factors, including competition with splicing of the pre-miRNAs, as has been demonstrated for AAVS (Qu and Pintel, 2004). Characterization of the relationship between B19 RNA processing and permissivity to infection will yield important insights into B19 pathogenesis.

SIMIAN PARVOVIRUS (SPV)

Although there have been a number of non-human primate erythroviruses identified, including simian parvovirus (O’Sullivan et al., 1994), pig-tailed macaque parvovirus and Rhesus parvovirus (Green et al., 2000), only the simian parvovirus (SPV) has a single promoter (P6) and two polyadenylation sites ([pA]p and [pA]d). However, the [pA]p signal in the middle of genome is consensus (AAUAAA) and has a putative GU-rich DSE (Liu et al., 2004). The NS and Cap regions of SPV and B19 are more than 50 percent of homologous (Brown et al., 1995a). Like B19, SPV has a single polyadenylation donor (predicted to be D3 (Vashisht et al., 2004)) and two polyadenylation sites ([pA]p and [pA]d). However, the [pA]p signal in the middle of genome is consensus (AAUAAA) and has a putative GU-rich DSE (Liu et al., 2004).

At least 13 SPV transcripts are generated by alternative splicing and polyadenylation following transfection of non-permissive COS cells (Liu et al., 2004) and viral infection of monkey liver (Vashisht et al., 2004). Although the transcription maps of SPV and B19 are generally similar, they differ in a number of important ways. SPV pre-mRNAs contains three introns, instead of the two introns found for B19. The third intron is within the capsid gene region and RNAs in which this intron is spliced are abundant (there is approximately three times more spliced RNA than unspliced RNA across this region). These RNAs encode two SPV 14 kDa proteins (analogous to the B19 11 kDa protein), which initiates at an AUG in the exon preceding the third intron. Also, SPV RNAs were found that were spliced between the first donor of the first intron and the acceptor of the second intron (D1 to A2, in Figure 18.9, right panel). A portion of these molecules were not additionally spliced and encode VP2; those from which the third intron was also removed encode the 14 kDa protein. Similar to B19, SPV has a polyadenylation signal ([pA]p) in the middle of genome that directs efficient polyadenylation; however, unlike B19, the SPV signal is a consensus AAUAAA motif. Efficient polyadenylation at [pA]p of both spliced and unspliced miRNAs (encoding a putative 10 kDa protein, analogous to the B19 7.5 kDa protein and SPV NS1, respectively) was detected. The 14 kDa protein, which is approximately 3-fold more abundant than the VP2 protein, has an unknown function and was localized to both in the nucleus and cytoplasm (Liu et al., 2004).

The first intron donor site D1 is at nucleotide 279 and it is joined to either the A1–1 (nucleotide 1792) or A1–2 (nucleotide 1792) acceptor site (nucleotide numbers refer to GenBank accession number U26342). All SPV RNAs that are spliced between D1 and A1–1 retain an open reading frame for a protein analogous to the B19 7.5 kDa protein. In SPV this proteins, which has an apparent molecular weight of 10 kDa, is translated from the R2 molecule that polyadenylates at [pA]p. As is the case for B19, whether this ORF, which is retained in all RNAs splicing between D1 and A1–1, is translated bicistronically from these RNAs is not known. The second intron donor site D2 (nucleotide 2063) is joined to A2 at nucleotide 3087. Splicing between D1-A2 and D2-A2 is found at approximately equal levels. The novel third intron in SPV RNA uses a donor D3 (nucleotide 3208) and is joined to either of two polyadenylation sites ([pA]p and [pA]d). However, the [pA]p signal in the middle of genome is consensus (AAUAAA) and has a putative GU-rich DSE (Liu et al., 2004).

The small RNAs generated by SPV (the 0.9 kb R2 and R3 RNAs, the 0.8 kb R12 and 0.5 kb R12 + R13 RNAs) are abundant both in viral infection of monkey liver tissue (Vashisht et al., 2004) and following transfection of COS cells (Liu et al., 2004), similar to B19 (Ozawa et al., 1987).
PARVOVIRUSES OF INVERTEBRATES
(DENSOVIRINAE)

The densoviruses (DNVs) are a large group of viruses that infect a number of invertebrate species, including insects and shrimp. Although their capsids and genomes are structurally similar to paroviruses infecting vertebrates, they share little sequence homology with them, and so they have been classified as a subfamily – Densovirinae – within the Parovirinae family. They replicate autonomously and have both (+) and (−) sense single-strand genomes that are packaged into separate viral particles. Numerous DNVs have been isolated, but only a subset of these have been characterized at the molecular level. Three groups have so far been identified, and most remain unclassified. One genus, the densoviruses, which contain the densoviruses from Junonia coenia (Jc DNV), Galleria mellonella (Gm DNV), and Mythimna loreyi (Ml DNV), have a genome of approximately 6 kb in length, relatively long ITR structures, and an ambisense genome. A second genus, the brevidensoviruses, which include densoviruses of Aedes (Ae DNV and Aa DNV), and of shrimp, have smaller 4 kb mono- (as opposed to ambi-) sense genomes flanked by different termini. A third genus, the iteraviruses, which include densoviruses of Caspalia extranea (Ce DNV) and Bombyx mori (Bm DNV), also are not ambisense, and have genomes of approximately 5 kb in length, which are flanked by short ITRs.

Both the brevidensoviruses and iteraviruses have significant ORFs in each half of the molecule (Boublik et al., 1994; Li et al., 2001; Fediere et al., 2002). The ORFs in the left half of the genome encode the non-structural proteins NS1 and NS2 from unspliced RNAs generated from a promoter at the left-hand end. These proteins are initiated by different AUG codons in two differencelucleotide ORFs. It is unclear where the 3′ terminus is for the RNAs that encode the NS proteins. The capsid coding ORF in the right-hand end of the molecule is translated from an unspliced mRNA generated from a promoter in the center of the genome. The different capsid proteins are apparently generated from this mRNA by a leaky scanning mechanism.

The transcription pattern of the Densovirus genus has been more fully characterized (Tijssen et al., 2003; Fediere et al., 2004). Maps of Gm DNV and Ml DNV have been determined in detail and are quite similar (see Figure 18.10; the GenBank accession number for Gm DNV is L32896). These viruses have ambisense genomes. The non-structural proteins are expressed from an ORF in the left half of the genome. Pre-mRNA is generated from a single promoter near the left-hand end and is polyadenylated in the center of the genome. An unspliced version of this RNA encodes the NS3 protein from the first available ORF (ORF 1) at the 5′ end of the molecule. This pre-mRNA is also spliced to generate a second mRNA, from which an intron that encompasses the NS3 coding region is removed. The resulting spliced RNA encodes the NS1 and NS2 proteins. NS1 initiates at an AUG immediately after the splice donor and continues...
Parvovirus RNA processing strategies

in ORF 2 to the 3' end of the mRNA. The smaller NS2 is initiated a few nucleotides downstream of NS1, and continues in ORF3, until its termination within the NS1 gene. The capsid proteins are translated from an ORF on the right-hand side of the genome; however, from a RNA generated in the opposite orientation and from a template of the opposite sense as those encoding the non-structural proteins. A single major transcript is made that is driven in the opposite sense as those encoding the non-structural proteins. Alternative splicing is a key to determining the size of the 5' terminal exon increases, excision of the intron becomes less efficient.

The various AA V serotypes show a wide spectrum of dependence on helper virus for the expression and processing of their RNA. This ranges from AA V2, which is dependent on both helper virus and Rep for efficient transcription and processing of the P40-generated RNAs, to AA V5, which is less responsive to helper virus and the role of the Rep protein, GPV, which shares significant homology with the AA Vs, is helper independent. Further characterization of their expression profiles may shed light on the evolutionary relationship between these viruses.

Splicing of AA V2 RNA requires both the Rep protein targeted to the transcription template and helper virus co-infection. Rep's role in RNA processing has not been separated from its role as a transactivator of transcription initiation. The distance-dependence of AA V2 RNA splicing is most likely related to the strength of interaction of the U1snRNP with the intron donor, and this interaction is governed by the mechanism that controls 5' exon definition. Therefore, as the size of the 5' terminal exon increases, excision of the intron becomes less efficient. Whether Rep facilitates splicing of P40 RNA by directly influencing RNA processing factors associated with the elongating RNA polymerase, or whether Rep facilitates splicing by localizing the transcription template to helper virus-induced nuclear sites that are rich in RNA processing factors has not been clearly determined. However, most AA V serotypes are much less dependent on helper virus for the transcription and processing of their RNA. Perhaps this is related to a longer co-evolution of AA V2 and adenovirus in a tissue culture system.

Because the transcription profiles of numerous parvoviruses have now been determined, some generalities can be drawn. All the parvoviruses use a wide variety of alternative RNA processing strategies to maximize the coding capacity from their small genomes. Different processing strategies result in the generation of multiple RNAs that typically encode one large and 2–5 smaller non-structural proteins that share at least some amino acid sequence, and 2–3 overlapping capsid proteins. A single major transcript is made that is driven in the opposite sense as those encoding the non-structural proteins, which is a critical feature of infection.

Most parvovirus genomes are transcribed in overlapping transcription units from multiple promoters and all contain small introns in the center of the genome. All transcripts generated by the multiple promoters of the autonomous parvoviruses MVM, H1, CPV, and PPV and PPV and the adeno-associated viruses type 1, 2, 3, 4, and 6 polyadenylate at the far right-hand end of the genome.

For MVM the small central intron appears to be the focal point for entry of the spliceosome onto MVM RNA. Splicing of the overlapping small intron is very efficient, and the small intron is excised at similar levels from all classes of MVM RNA. Subsequent to small intron recognition and splicing, the upstream NS2-specific exon is defined and excision of the upstream large intron occurs. This second step must be slowed (likely by the large intron non-consensus donors and acceptors), so that singly spliced RNA can exit to the cytoplasm to encode NS1.

AA V1, 2, 3, 4, and 6 have a single intron that is excised to greater levels from RNAs generated by the intron proximal P40 promoter than from RNAs generated by P19 or P5. Therefore, these viruses generate relatively more Rep78 and Rep52 from unspliced RNAs generated by upstream promoters and more spliced RNAs from P40, which encode the capsid proteins. The level of spliced AA V2-generated RNA increases as infection progresses. Expression of Rep78 and Rep52 requires that unspliced AA V RNAs be transported efficiently to the cytoplasm. Therefore, translation of Rep78 and Rep52 from unspliced RNAs may compensate for the inefficient splicing of P5- and P19-generated RNAs.

SUMMARY

The capsid proteins are translated from an ORF near the right-hand hairpin and also polyadenylate in the center of the genome. This transcript encodes the multiple major capsid proteins, presumably by a leaky-scanning type mechanism.

Because the transcription profiles of numerous paroviruses have now been determined, some generalities can be drawn. All the paroviruses use a wide variety of alternative RNA processing strategies to maximize the coding capacity from their small genomes. Different processing strategies result in the generation of multiple RNAs that typically encode one large and 2–5 smaller non-structural proteins that share at least some amino acid sequence, and 2–3 overlapping capsid proteins. Alternative splicing is a key to determining the relative accumulated levels of parovirus proteins, which is a critical feature of infection.

Most parovirus genomes are transcribed in overlapping transcription units from multiple promoters and all contain small introns in the center of the genome. All transcripts generated by the multiple promoters of the autonomous paroviruses MVM, H1, CPV, and PPV and PPV and the adeno-associated viruses type 1, 2, 3, 4, and 6 polyadenylate at the far right-hand end of the genome.

For MVM the small central intron appears to be the focal point for entry of the spliceosome onto MVM RNA. Splicing of the overlapping small intron is very efficient, and the small intron is excised at similar levels from all classes of MVM RNA. Subsequent to small intron recognition and splicing, the upstream NS2-specific exon is defined and excision of the upstream large intron occurs. This second step must be slowed (likely by the large intron non-consensus donors and acceptors), so that singly spliced RNA can exit to the cytoplasm to encode NS1.

AA V1, 2, 3, 4, and 6 have a single intron that is excised to greater levels from RNAs generated by the intron proximal P40 promoter than from RNAs generated by P19 or P5. Therefore, these viruses generate relatively more Rep78 and Rep52 from unspliced RNAs generated by upstream promoters and more spliced RNAs from P40, which encode the capsid proteins. The level of spliced AA V2-generated RNA increases as infection progresses. Expression of Rep78 and Rep52 requires that unspliced AA V RNAs be transported efficiently to the cytoplasm. Therefore, translation of Rep78 and Rep52 from unspliced RNAs may compensate for the inefficient splicing of P5- and P19-generated RNAs.
The erythroviruses B19 and SPV have a single promoter and so alternative processing of a single pre-mRNA molecule generates the full spectrum of RNAs that are expressed by these viruses. Polyadenylation at an internal site is also a critical event for erythroviruses. Use of (pA)p prevents the accumulation of RNAs required to generate the capsid proteins and the regulation of this internal polyadenylation event has been suggested to be a critical determinant of B19 tropism and pathogenesis. In contrast to AAV5, for both B19 and SPV, the internal polyadenylation site remains unused in the capsid encoding VP1 mRNA. It is interesting that B19, SPV, AMDV, GPD, DPV, and AAV5, all of which use an internal polyadenylation site, are clinically pathogenic in their hosts.

In addition to the complex alternative RNA processing strategies that are used to generate multiple mRNA molecules from compact genomes, the expression of the parvovirus genome is further governed, in certain cases, by translational control of these RNAs. For example, the AAV2 capsid proteins VP2 and VP3 are encoded by the same mRNA, and the relative levels at which these two proteins accumulate is controlled by differences in translation initiation. Perhaps the most striking example is found in the densoviruses. This group uses a novel ambisense strategy to express its genome, and multiple capsid proteins are generated from a single mRNA molecule, presumably by a leaky-scanning type mechanism.

The small parvovirus genomes use complex patterns of initiation, processing, export, and translation of their RNA to express their genomes. Characterization of these features will further our understanding of the biology of these viruses, and help unravel their evolutionary relationships. In addition these viruses provide a rich system, amenable to study, in which to examine basic cellular mechanisms of gene expression.

REFERENCES


Miyaogawa E., Yoshida T., Takahashi H. et al. 1999. Infection of the erythroid cell line, KUB182Ep6 with human parovirus B19 and...


Qiu J. and Pintel D. J. 2004. Alternative polyadenylation of adeno-associated virus type 5 RNA within an internal intron is governed by the distance between the promoter and the intron and is inhibited by U1 small nuclear RNP binding to the intervening donor. Journal of Biological Chemistry 279: 14889–98.


Analysis of the complete nucleotide sequences of goose and muscovy duck paroviruses indicates common ancestral origin with adeno-associated virus 2. Virology 212: 562–73.

Alternative splicing of pre-mRNAs encoding the nonstructural proteins of minute virus of mice is facilitated by sequences within the downstream intron. Journal of Virology 68: 2849–59.

Efficient excision of the upstream large intron from P4-generated pre-mRNA of the parovirus minute virus of mice requires at least one donor and the 3′ splice site of the small downstream intron. Journal of Virology 69: 6170–9.

Sequences within the parovirus minute virus of mice NS2-specific exon are required for inclusion of this exon into spliced steady-state RNA. Journal of Virology 69: 5864–8.