Mutational analysis of the avian pneumovirus conserved transcriptional gene start sequence identifying critical residues

Nicole L. Edworthy and Andrew J. Easton

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Seven of the eight genes in the avian pneumovirus (APV) genome contain a conserved 9 nt transcriptional start sequence with the virus large (L) polymerase gene differing from the consensus at three positions. The sequence requirements of the APV transcriptional gene start sequence were investigated by generating a series of mutations in which each of the nine conserved bases was mutated to each of the other three possible nucleotides in a minigenome containing two reporter genes. The effect of each mutation was assessed by measuring the relative levels of expression from the altered and unaltered gene start sequences. Mutations at positions 2, 7 and 9 significantly reduced transcription levels while alterations to position 5 had little effect. The L gene start sequence directed transcription at levels approximately 50 % below that of the consensus gene start sequence. These data suggest that there are common features in pneumovirus transcriptional control sequences.

Avian pneumovirus (APV) causes acute respiratory infection in domestic poultry throughout most of the world (Jones, 1996) and it has been detected in several wild bird species, although the disease impact in these species is not yet clear (Bennett et al., 2002, 2004; Lwamba et al., 2002; Shin et al., 2000, 2002). APV is classified as a member of the family Paramyxoviridae, the subfamily Pneumovirinae and the genus Metapneumovirus (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988). It is distinguished from members of the genus Pneumovirus by the order of its genes – 3’-N-P-M-F-M2-SH-G-L-5’ (Ling et al., 1992; Randhawa et al., 1996, 1997; Yu et al., 1992a, b) and the absence of non-structural protein genes (Randhawa et al., 1997).

RNA synthesis in pneumoviruses requires a ribonucleoprotein complex comprising the nucleocapsid (N) protein bound to the genomic RNA together with the phosphoprotein (P) and large (L) polymerase protein. Efficient, progressive, transcription is enhanced by the M2-1 protein, a feature that is unique to the members of the subfamily Pneumovirinae (Collins et al., 1995; Naylor et al., 2004). Transcription of virus mRNA is carried out by the virion-associated polymerase complex in a progressive stop–start manner from the 3’ to the 5’ end of the genome (Dickens et al., 1984; Pringle & Easton, 1997), and is directed by gene start and gene end sequences that flank the transcription units. The gene order in non-segmented negative-strand RNA viruses is significant, as their transcription strategy results in more mRNA transcripts from genes at the 3’ end of the genome, with levels of mRNA progressively decreasing in a step-wise manner to the 5’ end. In Human respiratory syncytial virus (HRSV) and APV, the gene start sequence is conserved in all of the genes except for the L gene (Li et al., 1996; Ling et al., 1992, 1995; Randhawa et al., 1996). The APV consensus is GGGACAAGU in mRNA sense (Li et al., 1996; Ling et al., 1992, 1995; Randhawa et al., 1996; Yu et al., 1991, 1992a) and the L gene start sequence has three differences (underlined), giving a sequence of AGGACAAU (Randhawa et al., 1996). A consensus gene end sequence at the 3’ end of the transcription unit is thought to be involved in termination and polyadenylation (Bukreyev et al., 1996; Jacobs et al., 2003; Ling et al., 1992; Randhawa et al., 1996). Termination and polyadenylation of the mRNA occur at the gene junction, followed either by progression of the polymerase across the non-transcribed intergenic region to the next gene or relocation of the polymerase complex to the genome 3’ terminus to begin the process again. The reinitiation of transcription of the next gene then occurs (Collins et al., 1986).

Here, we report the effect of mutation of the APV gene start sequence on gene expression, and identify the key residues in the gene start sequence that control transcription. Each of the nine bases of the conserved APV gene start sequence was mutated to the three other possible nucleotides. A dicistronic minigenome for APV, similar to the dicistronic HRSV minigenome (Kuo et al., 1996), was kindly provided by Dr J. Smith, University of Warwick, UK. The minigenome was constructed using the APV leader and
The conserved gene start sequence was located at the beginning of both reporter genes. The CAT gene end transcription termination sequence was taken from the APV P gene and the Luc gene was terminated with the L gene end sequence. The intergenic region between the two reporter genes was UCGAU. Placing this next to the last A residue of the polyadenylation site of the P gene end sequence generated a ClaI site that was used for cloning. The newly created intergenic region contains a pyrimidine as the first base, which is common with APV intergenic regions, but does not otherwise resemble any of the wild-type regions. The consensus gene start sequence at the beginning of the Luc gene was followed with a 4 nt AACC sequence (in mRNA sense) to preserve the Kozak sequence (Kozak, 1986) around the Luc AUG start codon. Mutagenic primers of 22–29 nt in length were designed to create each of the 27 desired mutants of the Luc gene start sequence in the APV dicistronic minigenome. PCRs were carried out using a mutagenic primer and a standard T7 promoter primer that annealed with the T7 promoter sequence in the vector. In addition, a non-consensus APV L gene start sequence and an ‘intermediate’ mutant with two of the three bases unique to the non-consensus L gene start sequence were also created using mutagenic primers. All mutations were confirmed by sequencing. Details of the primer sequences are available upon request from the authors.

Fig. 1. Diagram of negative-strand RNA dicistronic minigenome. The virus genome sense RNA transcribed from the dicistronic minigenome plasmid used in the study contained CAT and Luc genes each flanked by conserved APV gene start and gene end transcriptional regulation sequences. The construct contained the APV leader and trailer terminal sequences, as indicated.

The levels of expression of the Luc protein, relative to the level of the CAT protein, compared with the wild type for each of the mutations were calculated and are presented in Fig. 2. Although the levels of reporter gene expression from the mutants covered a wide range, all but one mutation (C5U) decreased activity to below 60% of the control. Consideration of all of the mutations showed that changes from a G to an A, in which the purine nature of the base is conserved, were better tolerated than alteration of a G residue to either a U or C. Mutations were more detrimental (any change resulting in a value below 25% of control) when present at positions 2, 7 and 9. At position 1 the change from a G to an A was the best tolerated mutant with the other alternatives giving very low (G to U) or no (G to C) gene expression. At positions 2 and 3, the change of the conserved G residue to any other nucleotide reduced expression significantly, with little difference between the three mutants at each position, although the introduction of a U residue at position 2 reduced expression more than the A or C substitutions. The nucleotide that tolerated mutation with the least effect on transcription was at position 5, where the introduction of a U residue gave little reduction in reporter gene expression. Overall, levels of gene expression of the mutants ranged from 0% for G1C to almost wild-type levels for C5U.

Of particular interest for APV was the set of mutations that, individually and collectively, form the L gene start sequence. It was found that these three mutations caused a similarly decreased level of expression (40–50%) when present individually or when combined in the L gene start sequence, and also that a change of two of the three bases (G1A/G8A) that make up the L gene start sequence similarly reduced gene expression (Fig. 2). It is interesting to note that there was no additive effect of the multiple mutations. The conserved APV gene start sequence appears to be very important for efficient transcription and the differences in the L gene start sequence from the consensus maintained in the other genes may, coupled with the effect of the position of the L gene in the APV genome, be an important factor in transcriptional regulation of the L gene.

It should be noted that the A7U mutant introduces a translational start codon, AUG, which may alter the protein reading frame from the naturally occurring one. This will affect the observed activity of the Luc enzyme produced from these transcripts, and hence the level of expression for this mutant cannot be considered to be truly representative of transcriptional capacity. Results of CAT ELISA and Luc enzyme assays have previously been shown to closely match levels of mRNA transcription from minigenomes measured by Northern blot analysis (Marriott et al., 1999). To clarify the effect of the introduction of this start codon, and to confirm that the mutations were affecting mRNA transcription rather than having an indirect effect on another aspect of gene expression, Northern blot analysis was carried out.
using mRNA isolated from HEp2 cells previously infected with recombinant vaccinia virus expressing T7 RNA polymerase in a plasmid-based rescue system as described by Marriott et al. (1999), with plasmid amounts optimized for APV (0.4 μg of the N plasmid and dicistronic minigenome, 0.2 μg of the P and L plasmids and 0.02 μg of the M2-1 plasmid) transfected into approximately 2 × 10⁶ cells using Lipofectin (Invitrogen). A range of mutants were chosen to represent the different levels of Luc activity seen in Fig. 2 and this included the A7U mutant (Fig. 3). A riboprobe containing sequences from both CAT and Luc genes was generated by transcription from an APV dicistronic minigenome and the blots were processed as described by Marriott et al. (1999). It can be seen that the levels of Luc mRNA relative to those of CAT mRNA are consistent with the protein levels determined for the mutants. The mRNA of the mutant found to have the least effect, C5U, had both the CAT and Luc band at a similar relative intensity to that of the wild-type minigenome. The mutations with the most detrimental effect in the protein assay were G1C where no Luc protein was detected and mutant G1U which expressed Luc protein at levels of approximately 5% of the wild-type control. From Fig. 3 it can be seen that no Luc mRNA can be detected for either of these mutants. In contrast, mutant G1A had an intermediate effect on Luc protein production and it can be seen that the Luc mRNA band is of moderate intensity compared with the wild-type Luc mRNA and their CAT mRNA levels are of similar intensity. It can be seen that mutant A7U did not direct the synthesis of detectable levels of Luc mRNA, indicating that this mutation had a significant effect on transcription.

The APV gene start sequence is similar to that of HRV, the representative virus for the subfamily Pneumovirinae. Mutational analysis of the gene start sequence has been done for HRV (Kuo et al., 1997) and features similar to the data obtained here were found. At position 1, the APV and the HRV sequences were intolerant of change, with no mutation achieving above 40% of wild-type sequence activity and a change from a G to an A was the best tolerated mutation. The favouring of the A at position 1 was unexpected because this nucleotide is not found naturally in any HRV gene start sequence, though it is seen in the APV L gene start sequence. At positions 2 and 3 of the HRV consensus sequence, any change significantly reduced gene expression, but a change from a G to an A had the least effect.
on gene expression. Position 5 for both HRSV and APV was more tolerant of mutation and mutation to a U gave almost 100% activity in both systems. A change at position 6 for HRSV significantly reduced gene expression, but the effect of mutation at this position was less dramatic for APV. For both APV and HRSV a mutation at position 5 to a C residue gave the highest level of expression. At position 7, for both viruses, any change decreased the expression levels to below 20% of controls. Position 9, for both viruses, was least tolerant of mutation of any kind, with any change decreasing expression levels to below 10% of the wild-type level. Overall, both viruses were relatively sensitive to change at positions 7 and 9, and position 5 was relatively insensitive to change.

The APV gene start sequence mutation data presented here differ from those of HRSV in that the APV gene start sequence showed a trend to more readily accept changes from a G to an A. However, the APV gene start sequence was also more sensitive to mutation overall than that of HRSV and none of the mutations resulted in enhanced activity of the gene start sequence. The APV gene start sequence also tolerated best the changes at the positions found to differ from the consensus in the L gene start sequence over any other possible nucleotide. This trend was also found at positions 4 and 10 of the HRSV gene start sequence (mRNA sense) in the gene start sequence for the L gene if the A at position 9 of the HRSV gene start sequence is considered to be an insertion rather than a substitution (Kuo et al., 1997). The data presented here, together with those of Kuo et al. (1997) for HRSV, indicate that there is a consistent pattern for the nucleotides important for initiating efficient transcription by pneumoviruses. In addition, the expression of the L gene is regulated not only by genome location but also by its unique gene start sequence, which is less efficient than the consensus present in the other genes.

Acknowledgements

We would like to thank Dr Joanne Smith for providing the APV dicistronic minigenome used in this work and Dr Tony Marriott for constructive comments. This work was supported by EC contract no. QLK2-CT-2002-01699.

References


