Genes 1 and 2 of pneumonia virus of mice encode proteins which have little homology with the 1C and 1B proteins of human respiratory syncytial virus

Philip Chambers, Craig R. Pringle and Andrew J. Easton*

Department of Biological Sciences, Warwick University, Coventry CV4 7AL, U.K.

Genes 1 and 2 of pneumonia virus of mice (PVM) consist of 410 and 571 nucleotides and encode proteins of 113 and 156 amino acids respectively. The proteins show no extensive (gene 1 analogous to 1C) or low (gene 2 analogous to 1B) homology to their presumed counterparts in human respiratory syncytial virus (HRSV). The strongest homology is between regions of approximately 35 amino acids located near the carboxy termini of the gene 2 product and the 1B protein with 29% identity, although a lower level of homology can be detected throughout much of these proteins (18% identity overall). These observations contrast with the conservation of 1C and 1B proteins between subgroups of HRSV and with the conservation of nucleocapsid proteins between HRSV and PVM.

Pneumoviruses differ from the other paramyxoviruses in morphological and genetic criteria such as the diameter of the viral nucleocapsid and the number and arrangement of the viral genes (Pringle, 1987). Pneumonia virus of mice (PVM) is a pneumovirus first isolated from apparently healthy mice which is capable of causing lethal pneumonia on passage (Horsfall & Hahn, 1939, 1940). Serological evidence suggests that infection with PVM or a closely related virus is widespread in many animals including man (Pringle & Eglin, 1986).

A notable difference between the genome of the prototype pneumovirus, human respiratory syncytial virus (HRSV), and those of the other paramyxoviruses is the presence of two additional small genes, 1C and 1B, located closer to the 3' terminus of the negative strand genome than the gene encoding the major nucleocapsid protein (Collins & Wertz, 1983: Dickens et al., 1984). This arrangement is maintained in the genome of PVM, where the two small genes are currently designated genes 1 and 2 (Chambers et al., 1990a, b). Given a single entry for transcription by the pneumovirus RNA polymerase at the 3' end of the genome, and attenuation of transcription at successive intergenic regions (Collins et al., 1986; Dickens et al., 1984), this location would ensure high levels of transcription of the mRNAs corresponding to PVM genes 1 and 2, and probably thereby ensure high levels of synthesis of the relevant gene products. The high proportion of gene 1- and 2-specific cDNA clones made from mRNAs, and the high level of proteins subsequently identified as the gene 1 and 2 products, are evidence that genes 1 and 2 are indeed transcribed and translated at high levels in PVM-infected cells (Chambers et al., 1990a; Cash et al., 1979; Ling & Pringle, 1989).

Little is known about the HRSV 1C and 1B proteins beyond their deduced amino acid sequences (Collins & Wertz, 1985; Elango et al., 1985), which are well conserved between subgroups A and B of HRSV (Johnson & Collins, 1989). The high levels of synthesis of all of these proteins and the conservation of amino acid sequences of the HRSV proteins suggest that these proteins have some important, although still unknown, role in virus-infected cells. The 1C and 1B proteins are present in only trace quantities in virions (Huang et al., 1985) and are thus considered to be non-structural proteins. Little is known of the function of the non-structural proteins which are generated from the gene encoding the phosphoprotein (P) of the morbilliviruses and paramyxoviruses (Lamb & Paterson, 1991; Kolakofsky et al., 1991). There is no published evidence of the antigenicity of the two non-structural proteins of HRSV (Kimman & Westerbrink, 1990), and no murine helper T cell response to 1B was elicited by a recombinant vaccinia/1B virus (Openshaw et al., 1988).

Molecular cloning of PVM mRNAs for genes 1 and 2 and identification of the polypeptide products has been reported (Chambers et al., 1990a). Several full-length (with rearranged 5' termini) and partial clones of PVM...
Fig. 1. Nucleotide sequence of PVM genes 1 and 2. The nucleotide sequence is shown as positive-sense cDNA and includes the intergenic region (Chambers et al., 1990b). Numbering starts from the first base of gene 1. The deduced amino acid sequences of open reading frames are given using the single-letter code beneath the nucleotide sequence. Sequences in full-length gene 1-specific cDNA clones, possibly via DNA polymerase I-mediated resolution of hairpin loops and priming of second-strand cDNA synthesis (Barr et al., 1991).

Verification of the order of genes 1 and 2 and determination of the intergenic sequence have been described (Chambers et al., 1990b). Nucleotide sequence determination (Sanger et al., 1977) was performed after subcloning into bacteriophage M13mp18 and M13mp19 (Messing & Vieira, 1982), and sequences were analysed by PVM genes 1 and 2 (Chambers et al., 1990b). Poly(A) tails on the mRNAs correspond to oligo(A) tracts at the 3' termini of most of the PVM mRNAs we have sequenced (Barr et al., 1991; our unpublished results) and is associated with the generation of inverted sequences in full-length gene 1-specific cDNA clones, possibly via DNA polymerase I-mediated resolution of hairpin loops and priming of second-strand cDNA synthesis (Barr et al., 1991).

Gene 1 encodes a single long open reading frame of 113 amino acids from the first AUG triplet, which is in a good context for initiation of protein synthesis (Kozak, 1981; PuXXAUGG). The predicted Mr of the polypeptide (12-9K) is in fair agreement with the size found on SDS-PAGE (14K to 16K depending on the gel system; Cash et al., 1979; Ling & Pringle, 1989; Chambers et al., 1990a). The net charge on the polypeptide at pH 7, assuming a charge of +1 on lysine and arginine, +0.5 on histidine and −1 on aspartate and glutamate, would be +0.5, giving an almost neutral pI (6.1). In fact, two polypeptides with approximately the correct apparent Mr, and almost neutral pIs (they are described as slightly acidic by Ling & Pringle, 1989) are resolved by two-dimensional PAGE, but the nature of the difference between them is not known.

The HRSV 1C protein has a low predicted net charge (−1). It may also appear heterogeneous on two-dimensional PAGE as it undergoes intracellular processing with a loss of 1.5K in apparent Mr, and a phosphorylated protein of approximately the right apparent Mr, to be either 1C or 1B has been detected in HRSV-infected cells (Huang et al., 1985; Lambert et al., 1988). Heterogeneity of Mr, charge and the level of phosphorylation has also been described for the non-structural proteins of paramyxoviruses such as Newcastle disease virus (Chambers & Samson, 1980, 1982).
Gene 2 extends from nucleotides 419 to 989 (571 nucleotides). The 5'-terminal 41 bases of gene 2 contain three short overlapping series of complementary sequences (Fig. 1). Two sets of these complementary sequences [indicated as (a) and (b) in Fig. 1] are associated with cDNA rearrangements similar to those described by Barr et al. (1991) in the two cDNA clones that reached the 5' terminus of gene 2 mRNA.

Gene 2 encodes a polypeptide of 156 amino acids from the first AUG triplet. Again, the predicted Mr of the protein product (18-2K) is in agreement with estimates from SDS-PAGE (17K to 20K; Cash et al., 1979; Ling & Pringle, 1989; Chambers et al., 1990a). The first AUG is separated from the second in-frame AUG triplet by nine bases. It is interesting to note that the non-structural protein C/Y of Sendai virus has two pairs of initiation codons (Curran & Kolakofsky, 1988). Neither of the first two AUGs in gene 2 of PVM is in a completely favourable context for initiation (Kozak, 1981), but it is possible that both are used because the gene 2 product in PVM-infected cells has two forms which differ only slightly in Mr, (the difference is estimated at 1K) and not at all in pi (Ling & Pringle, 1989). The size and charge differences contributed by the first four amino acids in the open reading frame would be 0-5K and zero respectively, which is consistent with the differences seen in PVM-infected cells.

The net charge of the polypeptides encoded by gene 2 would be -6 at pH 7 based on the values quoted above (with a pi of 5-9), which is consistent with the observation that the polypeptides subsequently identified as gene 2 products are more acidic than the gene 1 products when PVM-infected cell extracts are resolved by two-dimensional PAGE (Ling & Pringle, 1989; Chambers et al., 1990a). The predicted PVM gene 2 product(s) and the HRSV 1B protein differ substantially in net charge (-6 or +6 respectively). The net negative charge of the PVM gene 2 product(s) is due to the unusual C-terminal sequence of six consecutive acidic amino acids.

The nucleotide sequences at the 5' termini of PVM genes 1 and 2 are probably capable of generating hairpin loops during cDNA synthesis, and may also be important for mRNA secondary structure formation. It may be significant that the 5'-terminal region of the gene 2 mRNA has potential to form stem-loop structures between bases 1 to 11 and 14 to 41 (corresponding to bases 419 to 420 and 423 to 450 in Fig. 1), because this region contains the two potential AUG initiation codons for the gene 2 product discussed above.

If stem-loop structures in some way affect initiation of protein synthesis at the first two AUGs in PVM gene 2 (neither of which is in the best context for initiation), this may allow access to the third AUG triplet which starts a second, overlapping 69 residue open reading frame (Fig. 1). There is no evidence for the existence of such a protein and there is no corresponding second open reading frame in the 1B gene of HRSV (Collins & Wertz, 1985), but low Mr, proteins in PVM-infected cells have not been fully characterized (Cash et al., 1979; Ling & Pringle, 1989). Surprisingly, translation of the negative-sense sequence of PVM gene 2 also generated two substantial overlapping reading frames (not shown), which emphasizes the caution needed when considering the second positive-sense open reading frame.

The nucleotide sequences of PVM genes 1 and 2 are not detectably homologous with those of 1C, 1B or any other HRSV gene. Databank searches failed to locate any strong matches with the nucleotide or deduced amino acid sequences of genes 1 and 2. Pairwise dot-matrix comparisons (Queen & Korn, 1984; Staden, 1982) of the amino acid sequences were more successful, in that some homology between the carboxy termini of the PVM gene 2 product and the 1B protein of HRSV was found (not shown). An alignment of the complete amino acid sequences of the PVM gene 2 product and the HRSV 1B protein obtained with the program CLUSTAL (Higgins & Sharp, 1988) is shown in Fig. 2. Homology is strongest between approximately 35 residues at the carboxy termini of each of the proteins, but alignment of the carboxy-terminal 80 to 90 residues is strengthened because there is similarity between the hydrophathy (Hopp & Woods, 1981) and secondary structure predictions (Garnier et al., 1978) for the relevant parts of the PVM gene 2 product and the HRSV 1B protein (data not shown).

No strong overall similarities were found in any pairwise comparisons involving the PVM gene 1 product or the HRSV 1C protein, but the 1C protein of the B subgroup HRSV (Johnson & Collins, 1989) and the PVM gene 1 product have four identical amino acids at their amino termini.

In the absence of data on the function of the pneumovirus non-structural proteins, discussion of se-
quence relationships is bound to be speculative. The low level of amino acid sequence homology we have found between the PVM gene 1 and 2 products and the 1C and 1B proteins of HRSV is surprising because these proteins presumably share some common functions, which is reflected in the inter-subgroup conservation of these proteins in HRSV (Johnson & Collins, 1989). Furthermore, there are close similarities between other PVM and HRSV proteins, exemplified by the high homology between the nucleocapsid proteins (60% identity; Barr et al., 1991). Low homology between the corresponding polyproteins from the subgroups of HRSV is most striking for the virus attachment protein (G), where it correlates with unusual amino acid compositions which are rich in proline, serine and threonine (Wertz et al., 1985; Johnson et al., 1987). The 1C and 1B proteins of HRSV have no such gross anomalies in their amino acid compositions (Table 1 in Collins & Wertz, 1985), but the PVM gene 1 and 2 products and the HRSV 1C and 1B proteins each contain two to five cysteines, which seems rather high for such small proteins that presumably function entirely within the cytoplasm of virus-infected cells. The two cysteines of the HRSV 1B protein align with two of the cysteines of the PVM gene 2 product (Fig. 2), perhaps suggesting they have some important function.

At the moment, the amino acid sequences of PVM gene 1 and 2 products are difficult to align with those of the HRSV 1C and 1B proteins, and better comparisons must await multiple alignments which include the relevant protein sequences from additional pneumoviruses such as turkey rhinotracheitis virus (Collins & Gough, 1988; Ling & Pringle, 1988).

We thank John Barr and David Matthews for technical assistance. This work was supported by grants from the Medical Research Council and the Wellcome Trust.

References


CHAMBERS, P., MATTHEWS, D. A., PRINGLE, C. R. & EASTON, A. J. (1990b). The nucleotide sequences of intergenic regions between nine genes of pneumonia virus of mice establish the physical order of these genes in the viral genome. Virus Research 18, 263-270.


JOHNSON, P. R. & COLLINS, P. L. (1989). The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial virus (RSV) of antigenic subgroups A and B; sequence conservation and divergence within RSV genomic RNA. Journal of General Virology 70, 1539-1547.


(Received 29 April 1991; Accepted 21 June 1991)