The small hydrophobic protein of human respiratory syncytial virus: comparison between antigenic subgroups A and B

Peter L. Collins,* Robert A. Olmsted† and Philip R. Johnson†

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

The nucleotide and amino acid sequences of the mRNA and predicted polypeptide of the integral membrane small hydrophobic (SH) protein of human respiratory syncytial virus strain 18537 (a prototype strain of antigenic subgroup B) were determined from cloned cDNA. At the nucleotide and amino acid levels there was 78% and 76% identity, respectively, with the previously described SH mRNA and protein of strain A2 (a prototype strain of subgroup A). Most of the amino acid substitutions occurred in the predicted ectodomain (50% identity). The pattern of post-translational processing of the strain 18537 SH protein was very similar to that of strain A2, yielding a non-glycosylated form and two glycosylated forms. Analysis of released virions of strain A2 by immunoprecipitation with SH-specific antibodies suggested that the major non-glycosylated species and one of the glycosylated species are virion structural components.

Human respiratory syncytial virus (RSV), a member of the pneumovirus genus of the paramyxovirus family (Kingsbury et al., 1978), is an enveloped virus that is an important aetiological agent of respiratory tract disease (McIntosh & Chanock, 1989). The genome of RSV is a single negative strand of RNA of 15222 nucleotides (Collins et al., 1987; Collins, 1990 and references therein; D. S. Stec, M. A. Mink & P. L. Collins, unpublished data) that is transcribed into 10 major species of mRNA, each of which encodes a major viral protein (Collins et al., 1984, 1987; Collins, 1990).

One of the 10 RSV proteins is a small (64 amino acids for strain A2) hydrophobic species that is called 1A or SH (Collins et al., 1984; Collins & Wertz, 1985; Huang et al., 1985; Olmsted & Collins, 1989). The nucleotide and amino acid sequence of the SH mRNA and protein of RSV strain A2 of RSV antigenic subgroup A were determined by cDNA sequencing (Collins & Wertz, 1985). Biochemical studies (Olmsted & Collins, 1989) showed that the SH protein is an integral membrane protein that, at least in the case of strain A2, accumulates intracellularly in at least two glycosylated and two nonglycosylated forms that are described later. The C terminus of the SH protein appeared to be oriented extracellularly based on the finding that antibodies raised against a synthetic peptide representing the C-terminal 12 amino acids bound to the SH protein when reacted with intact RSV-infected cells (Olmsted & Collins, 1989).

On the basis of its size, gene map location, and integral membrane status, the RSV SH protein appears to be analogous to the 44 amino acid SH protein of simian virus 5 (Hiebert et al., 1988 and references therein). Also, nucleotide sequence analysis identified an mRNA for mumps virus that would encode an SH-like protein, although direct identification of the putative mumps virus SH protein has not yet been reported (Elango et al., 1989). Interestingly, a number of paramyxoviruses including Sendai virus, parainfluenza virus type 3, measles virus and Newcastle disease virus do not contain a separate SH gene or encode an SH-like protein. On the other hand, possible counterparts exist for influenza A and B viruses, which encode small integral membrane proteins called, respectively, M₂ and NB (Williams & Lamb, 1988; Zebedee & Lamb, 1988 and references therein). These small membrane proteins might have similar functional roles in these paramyxoviruses and orthomyxoviruses.

To characterize the RSV SH protein further, we have determined the nucleotide sequence of the SH mRNA of RSV strain 18537, representing RSV antigenic subgroup B, for comparison with that for the subgroup A strain A2 (Collins & Wertz, 1985). As shown in Fig. 1, the 18537 SH mRNA is approximately the same size as its strain A2 counterpart [408 compared to 405 nucleotides, exclusive of poly(A)], and shares 78% sequence identity. As noted previously (Johnson & Collins, 1988b), the...
gene-start signal of the SH gene of strain 18537 contains a single nucleotide difference (3′ CCCGCUUUUA) compared with its counterpart in strain A2, whereas the gene-start signal for each of the other nine genes was exactly conserved between the subgroups. The gene-end signal also contains a single nucleotide substitution (3′ UCAUCAUAABAA) relative to its strain A2 counterpart, but nucleotide differences in the semi-conserved gene-end signals are common (Johnson & Collins, 1988b).

The predicted strain 18537 SH protein is one amino acid longer than its strain A2 counterpart (65 compared to 64 amino acids) and shares 76% sequence identity. Both of the SH proteins have a central hydrophobic core (amino acids 23 to 41) that presumably is involved in its integral membrane association, and both SH proteins have two potential acceptor sites for N-linked carbohydrate (Asn 3 and Asn 52) whose positions are exactly conserved. The amino acid differences between the two strains are not evenly distributed within the SH protein sequence: the N-terminal region (amino acids 1 to 22) and the putative membrane-spanning domain (amino acids 23 to 41) were relatively well conserved (91% and 84% identity, respectively), whereas the C-terminal putative ectodomain (nucleotides 42 to 64) was much more divergent (50% identity).

In previous work (Johnson et al., 1987; Johnson & Collins, 1988a, b, 1989, 1990; and references therein), comparison of the mRNA and predicted protein sequences for six other genes of strains A2 and 18537 showed that these two prototype subgroup strains have a high degree of sequence identity for most of the mRNA and protein sequences: for example, the 1B, 1C, N, P and F proteins have 92, 87, 96, 90 and 91% sequence identity between subgroups. On the other hand, several nucleotide and amino acid sequence domains were poorly conserved between subgroups: for example, the ectodomain of the G protein (apart from a single circumscribed conserved region), the signal peptide of the F protein, and a small domain of the P protein each have 53% or less amino acid sequence identity between subgroups. For these different proteins and polypeptide domains, a high degree of sequence conservation is suggestive of functional or structural importance, whereas a low degree of conservation implies that the specific sequence of that region is not important for RSV replication. Following this reasoning, the relatively well conserved N-terminal and membrane-spanning regions of the SH protein might be important functional or structural domains, whereas the poorly conserved C-terminal region does not appear to contain a specific circumscribed sequence that is important for RSV replication (although the possibility exists that the three-dimensional structure might contain conserved features not apparent in the amino acid sequences).

The SH proteins in extracts of cells infected with strain A2 or 18537 were compared by immunoprecipitation with rabbit antiserum raised against a synthetic peptide representing either the C-terminal 12 or N-terminal eight amino acids of the SH protein of strain A2. As described previously (Olmsted & Collins, 1989), the C-terminal sequence-specific antiserum precipitated four species of SH protein that comigrated with SH protein synthesized in vitro (Collins et al., 1984) and which is the appropriate size to represent either the C-terminal 12 or N-terminal eight amino acids of the SH protein of strain A2. As described previously (Olmsted & Collins, 1989), the C-terminal sequence-specific antiserum precipitated four species of strain A2 SH protein (Fig. 2): (i) a 7·5K species (SH0) that comigrates with SH protein synthesized in vitro (Collins et al., 1984) and which is the appropriate size to be the full-length unmodified SH protein; (ii) a 4·8K species (SH1) that is carboxy-coterminal with SH0 and might be generated by translational initiation at the second AUG (Met 23) in the SH open reading frame;
(iii) a 13K to 15K species (SHg) that contains one or more N-linked carbohydrate side-chains of the non-complex type; (iv) a 21K to 30K species (SHp) that is generated by the post-translational addition of polylactosaminoglycan to the N-linked carbohydrate of SHg (H. Sheshberadaran, M. A. Williams, R. A. Olmsted, P. L. Collins & R. A. Lamb, unpublished data). Previously, the SHo, SHg and SHp species were shown to be expressed abundantly at the cell surface whereas the SHt species appeared to be strictly intracellular (Olmsted & Collins, 1989).

In immunoprecipitations involving strain A2 SH proteins, the N terminal-specific antiserum (which had not been previously described) precipitated the SHo, SHg and SHp proteins but not the SHt protein (Fig. 2). The lack of detectable reactivity with the SHt protein was confirmed in multiple experiments with long film exposures and supports the interpretation that SHt is an N-terminally truncated form of SHo.

The two antipeptide sera raised against peptides representing strain A2 also were immunoreactive, albeit weakly, with the SH protein of strain 18537 despite the sequence differences between the two strains. As shown in Fig. 2, three forms of the SH protein of strain 18537 were precipitated by both antisera and were identified on the basis of shared characteristics with their strain A2 counterparts. These were (i) the unmodified 18537 SHo protein, which was identified by its comigration with 18537 SH protein synthesized in reticulocyte lysates in response to intracellular 18537 mRNA (not shown), (ii) the strain 18537 SHg protein, which was identified as a glycoprotein because it was labelled with [3H]glucosamine during a short labelling period (not shown) whereas in longer labelling periods its accumulation was markedly decreased in favour of the SHp protein (Fig. 3), a result that would be consistent with a role as precursor to the SHp protein as described for strain A2 (Olmsted & Collins, 1989; H. Sheshberadaran et al., unpublished data) and (iii) the strain 18537 SHp protein, which was identified by labelling with [3H]glucosamine, by its diffuse appearance and decreased electrophoretic mobility during SDS-PAGE, and by the slow rate of its intracellular appearance during radiolabelling compared with SHo and SHg. The strain 18537 SHo protein migrated more rapidly during SDS-PAGE than did its strain A2 counterpart, with an apparent Mr of 4.4K compared to 7.5K, although the calculated Mr values of the proteins are almost identical (7536 for strain A2 and 7529). However, this difference probably is not due to post-translational proteolysis of the strain 18537 protein because it was reactive with both the N- and C-terminal-specific antisera (Fig. 2), and because it was identical in electrophoretic mobility to the in vitro synthesized strain 18537 SH protein (not shown). Presumably, the difference is an artefact of SDS-PAGE. A strain 18537 counterpart to the SHt protein of strain A2 was not detected, but this could be due to low abundance of this species and reduced immunoreactivity with the strain A2-specific antipeptide antibodies. Thus, two aspects of
the complex post-translational processing of the SH protein were conserved between the two subgroups: first, processing can include the addition of both N-linked sugars and polylactosamine, and secondly the addition of both types of sugar is inefficient, resulting in the accumulation of the unglycosylated (SH0), N-glycosylated (SHg) and dually glycosylated (SHp) forms.

We have not investigated whether the relative amounts of the different forms of the SH protein are the same for both strains. Examination of Fig. 2 would suggest that a greater proportion of the SH protein of strain 18537 is glycosylated. The most straightforward approach to determining the molar amounts of the different SH species would be quantification of SDS-PAGE profiles of SH protein immunoprecipitated with antipeptide antibodies, such as shown in Fig. 2. However, the sequences represented by peptides used to prepare the antibodies either contain a potential glycosylation site, as in the case of the N-terminal peptide, or are immediately adjacent to a potential glycosylation site, as in the case of the C-terminal peptide. A carbohydrate side-chain at either site might reduce immunoreactivity, resulting in an underestimation of the relative abundance of the glycosylated forms. The previous finding that the C terminus appeared to be extracellular (Olmsted & Collins, 1989) would suggest that the C-terminal site is utilized. On the other hand, the N-terminal-specific antibodies bound the SHp protein less efficiently relative to the other species than did the C-
terminal-specific antibodies (Fig. 2), indicating that it would be premature to make conclusions on the sites of carbohydrate attachment.

Previously, RSV strain A2 virions that had been purified on sucrose gradients were shown by SDS-PAGE to contain SH₀ protein, albeit in amounts that were low relative to the other RSV structural proteins and could be detected only with long film exposures (Huang et al., 1985). Factors that complicate the identification of RSV structural proteins include the low yield of released virus in tissue culture, the lability of the virion and the tendency of released virions to be contaminated with intracellular proteins and cellular membranes. This experiment was repeated using similar procedures under conditions that, from our experience, routinely result in the recovery of 50% to 100% of input infectivity (Fig. 3). Analysis of [³⁵S]methionine-labelled infected cell lysates and RSV virions (Fig. 3a) revealed levels of virion-associated SH₀ protein that were readily detectable with (lane SH) or without (lane V) immunoprecipitation with SH-specific antibodies. In comparison, the non-structural 1B protein was readily detected in lysates of infected cells (Fig. 3, lane I), but was greatly reduced or absent in the virion preparation (lane V). In addition, analysis of [³H]glucosamine-labelled infected cell lysates and RSV virions (Fig. 3b) showed that the SHp protein was a prominent radiolabelled species both in infected cell lysates (lane I) and in virions with (lane SH) or without (lane V) immunoprecipitation.

In summary, comparison of the predicted sequences of the SH proteins of strains A2 and 18537 showed that the C-terminal third of the molecule, which was predicted in earlier work (Olmsted & Collins, 1989) to be extracellular, is relatively divergent between the two subgroups. This suggests that RSV replication does not have a requirement for a specific sequence in this polypeptide domain. Despite the sequence diversity, the pattern of inefficient glycosylation involving both N-linked carbohydrate and polylactosaminoglycan was well conserved, suggesting that the multiple forms of SH have importance in RSV replication.

The amino acid sequence of the proposed ectodomain of the SH protein, like that of the G protein, was poorly conserved between the antigenic subgroups. As discussed previously for the G protein (Johnson et al., 1987; Johnson & Collins, 1990), we hypothesize that polypeptide domains that are extracellular antigens are subject to selective pressure from host immunity. This would result in a greater amount of amino acid substitution except in situations where there are structural or functional constraints. This latter caveat could explain the high degree of amino acid conservation for the F protein both between the antigenic subgroups (Johnson & Collins, 1988a) and among different paramyxoviruses (Spriggs et al., 1986). In this context, it is interesting to note that synthetic peptides representing amino acids 45 to 62 and 51 to 60 of the strain A2 SH protein were shown to contain epitopes for RSV-specific T and B lymphocytes, respectively (Nicholas et al., 1988). In addition to being divergent, both of these epitopes include a potential site for attachment of N-linked carbohydrate (Asn 52). The presence of sugar at this site might alter or block immune recognition of SHg and SHp and thereby might be another mechanism for reducing the efficiency of host immunity.

Finally, preparations of RSV strain A2 virions contained SH₀ and SHp in amounts that equalled or exceeded their relative amounts in infected cell lysates. We cannot presently exclude the possibility that the Sho and SHp proteins in the virus preparation were derived from contaminating cellular membranes bearing cell surface SH protein. However, the relative abundance of the SH₀ and SHp proteins in the virion preparations suggests that these two forms of the SH protein are virion structural components. This point might be resolved by immunoelectron microscopy using the antipeptide sera.

This work was performed in the laboratories of Drs Robert M. Chanock and Brian R. Murphy. We thank them for their interest and support. We thank Dr W. Lee Maloy of the National Institute of Allergy and Infectious Diseases for the synthesis of peptides, and thank Ena Camargo for assistance in the preparation of cells and virus.

References


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 20 December 1989; Accepted 6 February 1990)