Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 4075–4079, June 1985
Biochemistry

Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein

(mRNA sequence/glycoprotein structure/paramyxovirus/*O*-linked oligosaccharides)

GAIL W. WERTZ*, PETER L. COLLINS*†, YUNG HUANG*‡, CHRIS GRUBER§, SEYMOUR LEVINE§, AND L. ANDREW BALL¶‖

*Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC 27514; §Department of Immunology and Microbiology, Wayne State University, Detroit, MI 40201; and ¶Department of Biochemistry and ‖Biophysics Laboratory, University of Wisconsin, Madison, WI 53706

Communicated by Carl W. Gottschalk, March 8, 1985

ABSTRACT **The major surface glycoprotein (G) of human respiratory syncytial (RS) virus has an estimated mature M_r of 84,000–90,000. Among a library of cDNA clones prepared from RS virus mRNAs, we identified clones that hybridized to a message that encoded a M_r 36,000 polypeptide that was specifically immunoprecipitated with anti-G antiserum. The amino acid sequence of the G protein backbone was determined by nucleotide sequence analysis of several of the cDNA clones. It contains a combination of structural features that make it unique among the known viral glycoproteins. The G mRNA is 918 nucleotides long and contains a single major open reading frame that encodes a polypeptide having 298 amino acid residues with a M_r of 32,587, a finding consistent with the M_r 36,000 estimate for the *in vitro* translation product of the G mRNA. This suggests that >50% of the molecular weight of the mature glycoprotein may be contributed by carbohydrate. Glycosylation of G is largely resistant to tunicamycin, an inhibitor of the attachment of *N*-linked oligosaccharides, suggesting that the majority of the carbohydrate residues are attached via *O*-glycosidic bonds. In accordance with this, serine and threonine residues, the acceptor sites for *O*-linked oligosaccharides, comprise 30.6% of the total amino acid composition. There are also four potential acceptor sites for *N*-linked oligosaccharides. The amino acid sequence lacks both an NH_2-terminal hydrophobic signal sequence and a COOH-terminal hydrophobic region. Instead, a strongly hydrophobic region is located between amino acid residues 38 and 66. This region may serve as both the signal to insert the nascent polypeptide through the membrane and as the membrane anchor site.**

Human respiratory syncytial (RS) virus is the major cause of severe lower respiratory tract infection in infants and children throughout the world. The virus has an unusual pathogenesis: infection can take place in the presence of circulating antibody and reinfection is common (1). Since no effective vaccine is available, annual epidemics occur during which the virus is a major cause of hospital admissions for bronchiolitis and pneumonia in children (1). Clearly, more information about the virus is needed to help combat the disease; therefore, we undertook a study of its molecular biology, with an emphasis on the major surface antigens.

RS virus is classified as a paramyxovirus (2). The prototypical paramyxoviruses contain two membrane glycoproteins: F, which is responsible for cell fusion, and HN, which possesses hemagglutinating and neuraminidase activities (3). RS virus also has two membrane glycoproteins (4). Its F protein mediates cell fusion (5), resulting in the formation of syncytia that constitute the characteristic cytopathology of RS virus infection. However, its larger glycoprotein G, which has a mature glycosylated form with an estimated M_r of 84,000–90,000, has neither hemagglutinating nor neuraminidase activities (6, 7).

By the use of cDNA cloning, we recently established that the negative-stranded genome of human RS virus contains information for 10 unique mRNAs (8, 9). The corresponding 10 proteins were identified by *in vitro* translation of individual mRNAs after purification by hybridization to individual cDNA clones (9). All but 2 of the protein products synthesized *in vitro* had counterparts of the same electrophoretic mobility and related peptide maps among the authentic RS virus-specific proteins labeled in infected cells. However, 2 of the polypeptides made *in vitro*, a M_r 59,000 and a M_r 36,000 species, did not have counterparts in infected cells and were tentatively identified as the nonprocessed precursors of the two viral glycoproteins, F and G, respectively. The identification of the M_r 59,000 polypeptide as the uncleaved and nonglycosylated precursor of the F protein was confirmed by peptide mapping, immunoprecipitation, and sequence determination (refs. 10 and 11; unpublished observations). However, initial attempts to relate the M_r 36,000 polypeptide to the remaining unidentified RS virus-specific protein, the M_r 84,000–90,000 major surface glycoprotein G, by peptide mapping were unsuccessful because of the extensive glycosylation of the authentic G protein. Here we report (*i*) the use of antisera prepared against electrophoretically purified G protein to identify G mRNA and its corresponding cDNA clones; (*ii*) the construction and nucleotide sequence determination of complete cDNA clones of the G mRNA; and (*iii*) the deduced amino acid sequence of the encoded G protein. The data show that the RS virus G protein contains a combination of structural features that make it unique among the known viral glycoproteins.

MATERIALS AND METHODS

cDNA Clones. cDNAs were synthesized by using intracellular viral mRNA as template, inserted into the *Pst* I site of pBR322 by homopolymer tailing, and cloned in *Escherichia coli* HB101 (8, 9). Five independently derived cDNA clones to the G mRNA were used for nucleotide sequence analysis; cDNAs 2B-4 and 2B-8 were isolated from a library constructed in previous work (8, 9), and cDNAs 2B-6b, 2B-16, and 2B-17 were isolated from a second library constructed essentially as described by Land *et al.* (12).

DNA Sequence Analysis. DNA sequences were determined by the chemical method of Maxam and Gilbert (13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: RS virus, respiratory syncytial virus.

†Present address: Laboratory of Infectious Disease, National Institutes of Health, Bethesda, MD 20205.
‡Present address: Department of Pathology, Case Western Reserve University, Cleveland, OH 44106.

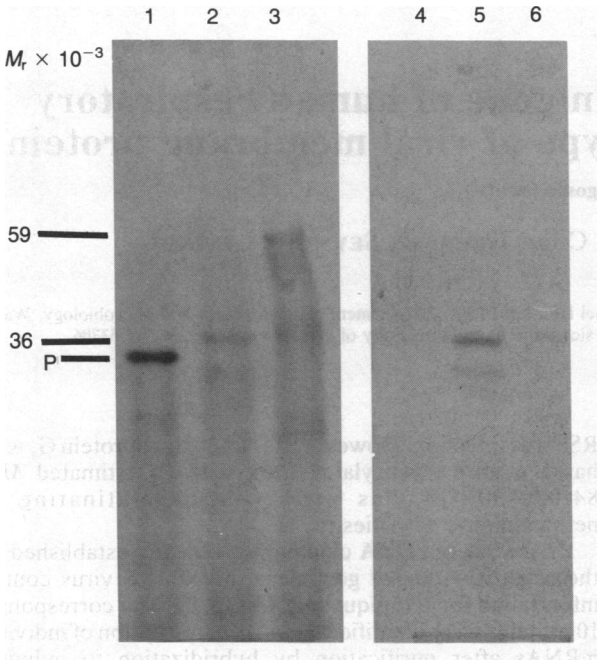

FIG. 1. Immunoprecipitation of RS virus G-specific polypeptide. RS virus mRNAs coding for the P, M_r 36,000, and M_r 59,000 proteins were isolated by hybridization selection with individual cDNAs. After translation *in vitro* in a wheat germ extract, the products of each reaction were analyzed by electrophoresis on a 12.5% $NaDodSO_4$/polyacrylamide gel either before (lanes 1–3) or after (lanes 4–6) immunoprecipitation with antiserum specific for the RS virus G protein. Lanes 1 and 4, P protein; lanes 2 and 5, M_r 36,000 protein; lanes 3 and 6, M_r 59,000 protein.

Primer Extension. G clone 2B-16 was digested with restriction enzymes *Dde* I and *Bst*NI. The 17-nucleotide antimessage sense strand from positions 48–64 in the complete sequence was labeled with ^{32}P at the 5′ end and isolated by electrophoresis on a 10% sequencing gel where it was resolved from its complementary strand that was two nucleotides longer. The 17-nucleotide end-labeled primer was coprecipitated in ethanol with 5 μg of mRNA isolated from RS virus-infected cells. The pellet was washed, dried, and resuspended in 10 μl of 4× reverse transcriptase buffer containing 120 mM 2-mercaptoethanol, 440 mM KCl, and each of the four dNTPs at 1.5 mM. The mixture was incubated for 10 min at 42°C. Reverse transcriptase (100 units) and 26 μl of H_2O were then added to a final vol of 40 μl and the reaction mixture was incubated for 2 hr at 42°C. The extended primer was isolated by electrophoresis on a 10% sequencing gel, eluted from the gel, recovered by ethanol precipitation, sequenced by the chemical method (13), and analyzed on an 8% sequencing gel.

Cell-free Translation and Immunoprecipitation. *In vitro* translations were carried out as described (9). Products of *in vitro* translation were denatured by boiling for 2 min in the presence of 2% $NaDodSO_4$ and immunoprecipitated in buffer A (0.02 M Tris·HCl, pH 7.4/0.15 M NaCl/1% deoxycholate/1% Triton X-100/0.1% $NaDodSO_4$/bovine serum albumin) (1 mg/ml), by incubation for 4 hr at 4°C with antibody followed by precipitation with IgGsorb (The Enzyme Center, Malden, MA). Precipitates were washed three times in buffer (0.15 M NaCl/0.02 M Tris·HCl, pH 7.4/1% Triton X-100/0.5% deoxycholate/0.1% $NaDodSO_4$) and analyzed by electrophoresis on 12.5% $NaDodSO_4$/polyacrylamide gels. Antibody to the G protein was prepared by electrophoresis of purified viral proteins on $NaDodSO_4$/polyacrylamide gels. Slices containing the G protein were excised, emulsified in Freund's complete or incomplete adjuvant, and injected into New Zealand White rabbits as described by Gruber and Levine (14). This antiserum is directed primarily against the G protein, but it also reacts with denatured bovine serum albumin (14).

RESULTS AND DISCUSSION

Identification of the G Protein mRNA. cDNA clones corresponding to the mRNAs that coded for the RS virus M_r 34,000 (P) protein, M_r 36,000 protein, and M_r 59,000 (F) protein (9) were used to hybrid-select their respective mRNAs, which were then translated *in vitro*. The labeled protein products of each translation were immunoprecipitated with G-specific antiserum (14) and subjected to electrophoresis on a 12.5% $NaDodSO_4$/polyacrylamide gel (Fig. 1). The M_r 36,000 protein was specifically precipitated by the G-specific antiserum, thereby identifying it as G specific and identifying the mRNA previously designated 2B [M_r, 0.39×10^6 (9)] as the G mRNA. The cell-free translation system used here lacked glycosylation activity; therefore, these data indicated that the M_r 36,000 polypeptide was the nonprocessed precursor of the major glycoprotein G. The M_r 84,000–90,000 estimate for the molecular weight of G was based on mobility in $NaDodSO_4$/polyacrylamide gels run under reducing conditions. Because there is no evidence for multimeric forms of G (ref. 7; unpublished results), these data indicate that, within the limits of accuracy for determining molecular weights of glycoproteins in $NaDodSO_4$/poly-

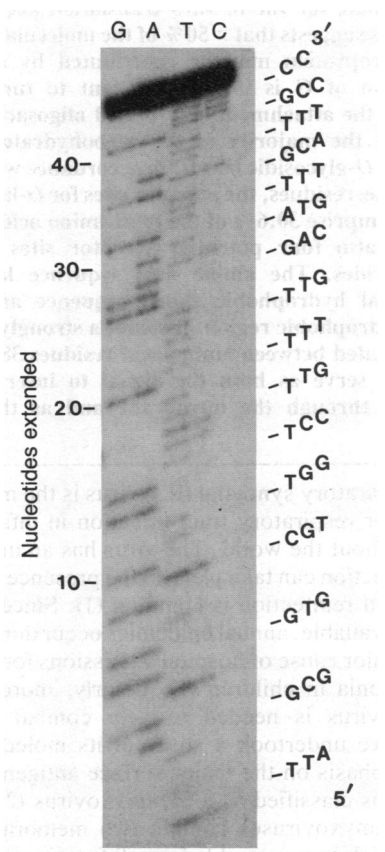

FIG. 2. Sequence of the 5′ end of the G mRNA. A 17-nucleotide primer from the G cDNA clone was labeled at the 5′ end, annealed to the G mRNA, and extended to the 5′ end of the message by using reverse transcriptase. The extended primer was isolated by electrophoresis on a 10% sequencing gel, sequenced by the chemical method (13), and analyzed on an 8% sequencing gel. Lane headings designate the reaction specificities as follows: G (G), A (G and A), T (C and T), C (C). Nucleotide sequence (antimessage sense) is shown.

acrylamide gels, at least 50% of the molecular weight of the mature G protein is contributed by carbohydrate.

cDNA Sequencing. To examine the structure of the RS virus G protein in detail, we sequenced cDNA clones of its mRNA. Chemical sequencing was carried out on five independently derived cDNA clones, three of which represented complete copies of the G mRNA. Each position in the sequence was determined at least three times from at least two independently derived clones and >90% of the sequence was confirmed in three independently derived clones. The end of the cDNA that corresponded to the 3′ end of the mRNA was identified by the presence of a terminal poly(dA) tract. The sequence of the 5′ end of the G mRNA was determined by sequencing the three complete cDNA clones and was confirmed by primer-extension experiments. In the primer-extension experiments, a 17-nucleotide primer was annealed to mRNA and extended to the 5′ terminus using reverse transcriptase. Chemical sequencing of the extended primer was carried out. The 5′ sequence of the mRNA (Fig. 2) was 5′ G-G-G-G-C-A-A-A-U . . ., notable because in the cDNA clones, the terminal run of G residues was indistinguishable from the adjoining poly(dG) tails added during cloning. These data showed that the sequence of the first nine nucleotides of the G mRNA was the same as that of the eight other RS virus mRNAs examined to date (11).

Nucleotide Sequence of the G mRNA. The complete sequence of the G mRNA exclusive of poly(A) was 918 nucleotides long (Fig. 3). The sequence contained a single major open reading frame that occupied 98.7% of the coding capacity of the message. All other reading frames were blocked by numerous termination codons. The major open reading frame coded for a polypeptide of 298 amino acid residues (Fig. 3), having a predicted M_r of 32,587. This finding is consistent with our estimate by gel electrophoresis of 36,000 for the molecular weight of the major translation product of G mRNA (Fig. 1; ref. 9). Somewhat unusually, the major open reading frame initiated at the second AUG in the sequence rather than the first, which initiated a reading frame of only 15 codons. It has been noted that in 95% of eukaryotic mRNAs examined, the 5′ proximal AUG served as the initiator codon (15). However, according to a compilation of sequences surrounding favored initiation codons, the utilized second AUG in the G mRNA sequence had a more favorable sequence environment than the first (15).

Comparison of the G mRNA sequence with that of other known viral glycoprotein genes showed no significant homology. Comparison of the G mRNA sequence with seven of the other RS virus genes sequenced to date showed no significant homology other than that noted previously (11) at both mRNA termini: the 5′ terminus as described above and the 3′ conserved sequence . . . 5′ A-G-U-$\frac{A}{U}$-A-(N)1–4-poly(A). However, it was observed that the 3′ end of the G mRNA shown in Fig. 3 was identical to the previously published sequence of the 3′ end of the RS virus N mRNA (16) for 286 nucleotides. The published sequence for the N mRNA was based on sequencing a single cDNA clone, and we have found that the sequence reported (16) for the 3′ end of N is incorrect

```
  1  GGG GCA AAT GCA AAC ATG TCC AAA AAC AAG GAC CAA CGC ACC GCT AAG ACA TTA
                         Met Ser Lys Asn Lys Asp Gln Arg Thr Ala Lys Thr Leu

 55  GAA AGG ACC TGG GAC ACT CTC AAT CAT TTA TTA TTC ATA TCA TCG TGC TTA TAT
     Glu Arg Thr Trp Asp Thr Leu Asn His Leu Leu Phe Ile Ser Ser Cys Leu Tyr

109  AAG TTA AAT CTT AAA TCT GTA GCA CAA ATC ACA TTA TCC ATT CTG GCA ATG ATA
     Lys Leu Asn Leu Lys Ser Val Ala Gln Ile Thr Leu Ser Ile Leu Ala Met Ile

163  ATC TCA ACT TCA CTT ATA ATT GCA GCC ATC ATA TTC ATA GCC TCG GCA AAC CAC
     Ile Ser Thr Ser Leu Ile Ile Ala Ala Ile Ile Phe Ile Ala Ser Ala Asn His

217  AAA GTC ACA CCA ACA ACT GCA ATC ATA CAA GAT GCA ACA AGC CAG ATC AAG AAC
     Lys Val Thr Pro Thr Thr Ala Ile Ile Gln Asp Ala Thr Ser Gln Ile Lys Asn

271  ACA ACC CCA ACA TAC CTC ACC CAG AAT CCT CAG CTT GGA ATC AGT CCC TCT AAT
     Thr Thr Pro Thr Tyr Leu Thr Gln Asn Pro Gln Leu Gly Ile Ser Pro Ser Asn

325  CCG TCT GAA ATT ACA TCA CAA ATC ACC ACC ATA CTA GCT TCA ACA ACA CCA GGA
     Pro Ser Glu Ile Thr Ser Gln Ile Thr Thr Ile Leu Ala Ser Thr Thr Pro Gly

379  GTC AAG TCA ACC CTG CAA TCC ACA ACA GTC AAG ACC AAA AAC ACA ACA ACA ACT
     Val Lys Ser Thr Leu Gln Ser Thr Thr Val Lys Thr Lys Asn Thr Thr Thr Thr

433  CAA ACA CAA CCC AGC AAG CCC ACC ACA AAA CAA CGC CAA AAC AAA CCA CCA AGC
     Gln Thr Gln Pro Ser Lys Pro Thr Thr Lys Gln Arg Gln Asn Lys Pro Pro Ser

487  AAA CCC AAT AAT GAT TTT CAC TTT GAA GTG TTC AAC TTT GTA CCC TGC AGC ATA
     Lys Pro Asn Asn Asp Phe His Phe Glu Val Phe Asn Phe Val Pro Cys Ser Ile

541  TGC AGC AAC AAT CCA ACC TGC TGG GCT ATC TGC AAA AGA ATA CCA AAC AAA AAA
     Cys Ser Asn Asn Pro Thr Cys Trp Ala Ile Cys Lys Arg Ile Pro Asn Lys Lys

595  CCA GGA AAG AAA ACC ACT ACC AAG CCC ACA AAA AAA CCA ACC CTC AAG ACA ACC
     Pro Gly Lys Lys Thr Thr Thr Lys Pro Thr Lys Lys Pro Thr Leu Lys Thr Thr

649  AAA AAA GAT CCC AAA CCT CAA ACC ACT AAA TCA AAG GAA GTA CCC ACC ACC AAG
     Lys Lys Asp Pro Lys Pro Gln Thr Thr Lys Ser Lys Glu Val Pro Thr Thr Lys

703  CCC ACA GAA GAG CCA ACC ATC AAC ACC ACC AAA ACA AAC ATC ATA ACT ACA CTA
     Pro Thr Glu Glu Pro Thr Ile Asn Thr Thr Lys Thr Asn Ile Ile Thr Thr Leu

757  CTC ACC TCC AAC ACC ACA GGA AAT CCA GAA CTC ACA AGT CAA ATG GAA ACC TTC
     Leu Thr Ser Asn Thr Thr Gly Asn Pro Glu Leu Thr Ser Gln Met Glu Thr Phe

811  CAC TCA ACT TCC TCC GAA GGC AAT CCA AGC CCT TCT CAA GTC TCT ACA ACA TCC
     His Ser Thr Ser Ser Glu Gly Asn Pro Ser Pro Ser Gln Val Ser Thr Thr Ser

865  GAG TAC CCA TCA CAA CCT TCA TCT CCA CCC AAC ACA CCA CGC CAG TAG TTA CTT
     Glu Tyr Pro Ser Gln Pro Ser Ser Pro Pro Asn Thr Pro Arg Gln End

919  AAA AAA AAA AAA AAA AA 935
```

FIG. 3. Nucleotide sequence of the RS virus G mRNA and the predicted protein sequence. The serine and threonine residues, potential sites for attachment of *O*-linked carbohydrate chains, are underlined; the potential acceptor sites for *N*-linked carbohydrate residues (Asn-X-Ser/Thr; X is not Pro) are indicated by a closed circle; and the proline residues are marked by a triangle beneath the amino acid sequence. Nucleotide variability was observed at four positions: position 76 (shown as A, was G in one of four cDNAs), position 85 (shown as T, was G in one of three cDNAs), position 94 (shown as T, was C in one of four cDNAs), and position 553 (shown as C, was G in one of three cDNAs).

and suggest that it probably represents that of an aberrant transcript (unpublished observations).

Glycosylation of the G Protein. The amino acid sequence of G presented in Fig. 3 shows that serine and threonine residues together account for a remarkable 30.6% of the total amino acid composition. This high content of serine and threonine is a characteristic feature of certain glycoproteins, which contain extensive *O*-linked oligosaccharide chains that are attached exclusively via these two amino acids (17). To test the possibility that the carbohydrate of the RS virus G protein might be linked via *O*-glycosidic bonds, the effect of tunicamycin on glycosylation was examined. As shown in Fig. 4, only the RS virus G and F proteins were labeled by incorporation of [^{3}H]glucosamine. The drug tunicamycin, which inhibits the addition of *N*- but not *O*-linked oligosaccharides, did not block addition of >90% of the carbohydrate of G, whereas it completely inhibited glycosylation of the RS virus F protein (Fig. 4). Appropriate controls showed that the inhibition of glycosylation of F was not merely due to inhibition of protein synthesis (data not shown). These data indicate that the majority of the carbohydrate of G is resistant to tunicamycin and, therefore, presumably *O*-linked, whereas that of F is completely *N*-linked. As described above (Fig. 3), the protein backbone of G has a M_r of 32,587, consistent with the estimate by gel electrophoresis of 36,000 (Fig. 1; ref. 9) for the nonglycosylated product of translation *in vitro* of the G mRNA. These data taken together with the data in Fig. 4 and the lack of evidence for multimeric forms of G suggest that more than one-half of the M_r of G is contributed by carbohydrate, the majority of which is *O*-linked. Independent pulse-labeling experiments using inhibitors of *N*- and *O*-linked glycosylation have identified a M_r 33,000 G-specific precursor in infected cells (14), providing further evidence that more than one-half of the M_r of G is contributed by *O*-linked carbohydrate. The amino acid sequence of G also revealed four potential sites for the attachment of *N*-linked oligosaccharides (Fig. 3) and, indeed, the slightly increased electrophoretic mobility of G protein made in the presence of tunicamycin (Fig. 4) confirmed that the mature protein contained some *N*-linked sugars.

If the oligosaccharide chains of G resemble *O*-linked chains in other glycoproteins, which usually contain 2–5 sugar residues and have M_r values of 600–1500 (17), the RS virus G protein may carry 40–80 separate *O*-linked carbohydrate chains distributed among the 91 serine and threonine residues identified here. In this respect, the G protein would resemble the erythrocyte membrane protein glycophorin A, in which 60% of the mass is contributed by *O*-linked oligosaccharides that occupy about one-half of the potential attachment sites (18, 19), the rat thymocyte W3/13 antigen (20), and submaxillary mucin in which one-third of the amino acid residues are substituted serine or threonine (19). By contrast, RS virus F protein (11) and other known paramyxovirus glycoproteins contain only 10–20% of their mass as carbohydrate residues, and these are attached exclusively via *N*-glycosidic bonds (Fig. 4; ref. 21).

Structural Features of the G Protein. Two other distinctive features of the amino acid sequence of G were revealed by hydrophobicity analyses. A plot of local hydrophobicity versus amino acid position revealed the lack of both a hydrophobic signal sequence at the NH_2 terminus and a hydrophobic transmembrane anchor region near the COOH terminus (Fig. 5). Both features are prominent in the structure of the RS virus F protein (11) and in most other viral glycoproteins (23). However, local hydrophobicity analysis of the G protein sequence showed that the NH_2 terminus was distinctly hydrophilic. The most hydrophobic region lay between residues 38 to 66 (Fig. 5). Of these 28 amino acids, 15 were hydrophobic, 12 were neutral, and none were charged. The next longest hydrophobic area extended from residues 166 to 186, but of these 20 uncharged residues, only 6 were hydrophobic, 8 were neutral, and 3 had polar groups. The other hydrophobic areas—residues 21–31, 83–100, 110–120, and 243–260—were either too short to be considered good potential membrane spanning sites or were interrupted by charged residues. We consider the region between residues 38 and 66 to be the best candidate for membrane interaction, perhaps serving both to direct the transfer of the protein across the membrane during its synthesis or processing and to anchor the mature protein in the lipid bilayer. A

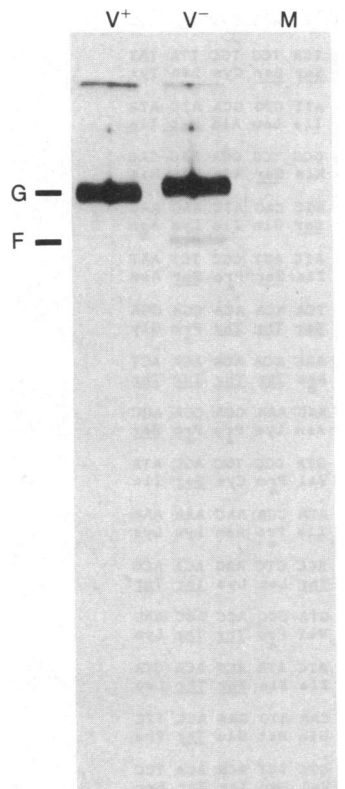

FIG. 4. Effect of tunicamycin on glycosylation of RS virus G and F proteins. Virus was purified from infected cells grown in the absence (V$^-$) or presence (V$^+$) of 2 μg of tunicamycin per ml and the effect of the drug on glycosylation was assayed by incorporation of [^{3}H]glucosamine from 15 to 30 hr p.i. An autoradiograph of a 15% polyacrylamide gel used to resolve the labeled proteins is shown. Lane M, mock-infected cells labeled with [^{3}H]glucosamine in the presence of tunicamycin.

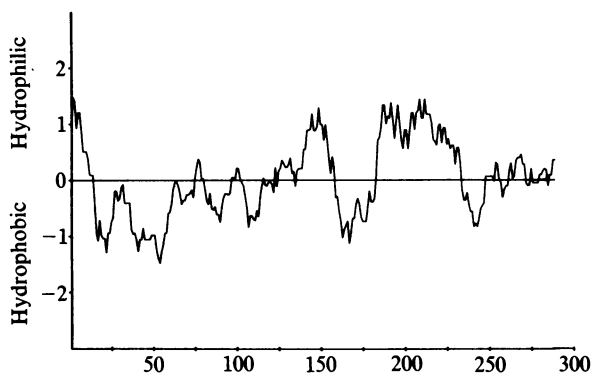

FIG. 5. Plot of hydrophobic and hydrophilic regions in the G protein. A window of 10 amino acids was used to calculate local hydrophobicity of each position using the procedure of Hopp and Woods (22). Horizontal axis denotes amino acid sequence position.

similar role has been postulated for an analogous internal hydrophobic domain of the influenza neuraminidase (24) and of some other membrane proteins (25, 26). Further analogy with influenza neuraminidase (24, 27) would suggest that the COOH-terminal portion of the RS virus G protein, from residues 67 to 298, might be exposed on the external surface of the viral envelope. While this orientation remains to be directly established, it is consistent with the fact that this portion of the protein contains 77 of the 91 potential sites for the attachment of *O*-linked oligosaccharides, and sugar residues are located exclusively on the external surface of plasma membranes. Several proteins of cellular origin are also known to adopt this orientation with respect to the plasma membrane (28).

Also present in this COOH-terminal 231-residue portion of the molecule are the 30 proline residues, which contribute to an unusually high (10.1%) proline content for the protein, and the four cysteine residues, which are clustered in a 14-residue stretch. The profound influence of proline on protein structure suggests that the unusual abundance of proline, which also has been observed in glycoproteins of mucinous origin having abundant *O*-linked oligosaccharides, constitutes a major determinant of the three-dimensional structure of these proteins.

The sequence of the RS virus G protein reported here shows that it has few characteristics in common with previously described viral membrane proteins (3, 11, 21, 24, 27, 29). Instead, its high content of serine, threonine, and proline, and extensive *O*-linked glycosylation are features held in common with various mucinous glycoproteins (17). This observation and the lack of detectable homology with available sequences of any other viral glycoprotein including the coronavirus E1 (*O*-linked) glycoprotein (26) and the HN protein of simian virus 5 (31) suggest an evolutionary origin for the RS virus G protein distinct from the glycoproteins of other enveloped viruses. Furthermore, the RS virus G protein lacks an NH_2-terminal hydrophobic signal sequence and contains instead an internal hydrophobic region whose function remains to be determined. Little is known about the biosynthesis of such membrane proteins or of *O*-linked oligosaccharides, whose assembly has been reported to occur at a subcellular location different from that of the well-studied *N*-linked glycoproteins (30). The RS virus G protein possesses both of these features and its availability as a cloned gene affords the opportunity to study the intracellular maturation, sorting, and membrane anchoring of such a protein. Furthermore, the construction of eukaryotic and prokaryotic expression vectors designed to produce glycosylated or nonglycosylated G protein will allow us to examine the relationships between G protein, its glycosylation, and the development of immunity to RS virus disease.

This work was supported by grants from the National Institute of Allergy and Infectious Diseases AI12464 and AI20181 (to G.W.W.) and AI18270 (to L.A.B.). S.L. was supported by the Thrasher Research Fund.

1. Chanock, R. M., Kim, H. W., Brandt, C. D. & Parrott, R. H. (1982) in *Viral Infections of Humans: Epidemiology and Control*, ed. Evans, A. S. (Plenum, New York), pp. 471–489.
2. Kingsbury, D. W., Bratt, M. A., Choppin, P. W., Hanson, R. P., Hosaka, Y., ter Meulen, V., Norrby, E., Plowright, W., Rott, R. & Wunner, W. H. (1978) *Intervirology* **10**, 137–152.
3. Choppin, P. W. & Scheid, A. (1980) *Rev. Infect. Dis.* **2**, 40–61.
4. Peeples, M. & Levine, S. (1979) *Virology* **95**, 137–145.
5. Walsh, E. E. & Hruska, J. (1983) *J. Virol.* **47**, 171–177.
6. Richman, A. V., Pedreira, F. A. & Tauraso, N. M. (1971) *Appl. Microbiol.* **21**, 1099–1100.
7. Gruber, C. & Levine, S. (1983) *J. Gen. Virol.* **64**, 825–832.
8. Collins, P. L. & Wertz, G. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3208–3212.
9. Collins, P. L., Huang, Y. T. & Wertz, G. W. (1984) *J. Virol.* **49**, 572–578.
10. Huang, Y. T. (1983) Dissertation (University of North Carolina, Chapel Hill).
11. Collins, P. L., Huang, Y. T. & Wertz, G. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7683–7687.
12. Land, H., Gretz, M., Hauser, H., Lindenmaier, W. & Schutz, G. (1981) *Nucleic Acids Res.* **9**, 2251–2266.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
14. Gruber, C. & Levine, S. (1985) *J. Gen. Virol.*, in press.
15. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
16. Elango, N. & Venkatesan, S. (1983) *Nucleic Acids Res.* **11**, 5941–5951.
17. Kornfeld, R. & Kornfeld, S. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans*, ed. Lennarz, W. J. (Plenum, New York), pp. 1–34.
18. Thomas, D. B. & Winzler, R. J. (1969) *J. Biol. Chem.* **244**, 5943–5946.
19. Sharon, N. & Lis, H. (1981) *Chem. Eng. News*, 21–44.
20. Brown, W. R. A., Barclay, A. N., Sunderland, C. A. & Williams, A. F. (1981) *Nature (London)* **289**, 456–460.
21. Klenk, H. D. & Rott, R. (1980) *Curr. Top. Microbiol. Immunol.* **90**, 19–48.
22. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.
23. Sabatini, D. D., Kreibich, G., Murimoto, T. & Adesnik, J. (1982) *Cell Biol.* **92**, 1–22.
24. Fields, S., Winter, G. & Brownlee, G. G. (1981) *Nature (London)* **290**, 213–217.
25. Frank, G., Brunner, J., Hauser, H., Wacker, H., Semenza, G. & Zuber, H. (1978) *FEBS Lett.* **96**, 183–188.
26. Armstrong, J., Niemann, H., Smeekens, S., Rottier, P. & Warren, G. (1984) *Nature (London)* **308**, 751–752.
27. Blok, J., Air, G., Laver, W., Ward, G., Lilley, G., Woods, E. F., Roxburgh, G. & Inglis, A. (1982) *Virology* **119**, 109–121.
28. Kriel, G. (1981) *Annu. Rev. Biochem.* **50**, 317–348.
29. Rose, J. & Gallione, C. (1981) *J. Virol.* **39**, 519–528.
30. Hanover, J. A., Lennarz, W. J. & Young, J. D. (1980) *J. Biol. Chem.* **255**, 6713–6716.
31. Hiebert, S. W., Paterson, R. G. & Lamb, R. A. (1985) *J. Virol.* **53**, 1–6.