

A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminidase

(myxoviruses/mRNA structure/gene structure/overlapping reading frames)

MICHAEL W. SHAW, PURNELL W. CHOPPIN, AND ROBERT A. LAMB*

The Rockefeller University, New York, New York 10021

Contributed by Purnell W. Choppin, April 25, 1983

ABSTRACT RNA segment 6 of the influenza B virus genome codes for a previously unidentified polypeptide designated NB. The reading frame for this polypeptide begins with the first AUG codon on the mRNA and overlaps the reading frame for the viral neuraminidase by 292 nucleotides. The amino acid sequence of polypeptide NB deduced from the nucleotide sequence of the B/Lee/40 strain consists of 100 amino acids with a molecular weight of 11,242. The sequence contains four potential glycosylation sites, and the protein has been found to be glycosylated in infected cells. NB has not been found in virions. Sucrose gradient sedimentation and analysis of the structure of the mRNA by nuclease S1 mapping and sequence analysis by the primer extension method indicated that polypeptide NB and the neuraminidase are translated from a single bicistronic mRNA. A protein analogous to NB has not been found with influenza A virus, and this represents a major difference between the two virus types.

The sequences of the RNA segments 4, 6, 7, and 8 of influenza A and B viruses indicate significant similarities for each segment between the virus types at both the nucleotide and amino acid levels (1–5). The sixth RNA segment in both influenza A and B viruses encodes the neuraminidase (NA) protein (6–8). The nucleotide sequence of the influenza B/Lee/40 NA gene differs from the other influenza virus RNA segments in that the NA protein is translated from an open reading frame initiating at the second AUG codon from the 5' end of the mRNA (5). The first AUG codon, which is separated from the second by four nucleotides, is followed by an open reading frame of 100 amino acids (5). This open reading frame, now designated NB (see Fig. 1), overlaps the NA reading frame by 292 nucleotides and has not been found in influenza A NA genes (9–12). Influenza viruses use overlapping reading frames on RNA segments 7 and 8 for the synthesis of polypeptides as part of their coding strategy (1, 2, 13–15). We therefore searched for a polypeptide product (NB) translated from the NB reading frame. The nucleotide sequence, which predicts that polypeptide NB would contain seven cysteine and 18 isoleucine residues and four potential glycosylation sites, has been used to select appropriate labeled precursors. We report here evidence for the synthesis of this glycoprotein. Analysis of the mRNAs derived from RNA segment 6 indicated the presence of only a single nucleotide species, which appears to be bicistronic.

MATERIALS AND METHODS

Viruses, Cells, mRNA Isolation, and Fractionation. Influenza B virus strains Lee/40 and MB/50 were grown in embryonated eggs, and the GL/1760/54 strain was grown in HKCC cells (16). Virus-specific mRNAs were isolated from HeLa cells 8 hr after infection (1, 17, 18). Sucrose gradient fractionation

of mRNAs and *in vitro* translation was done as described (18).

Cloned DNA of RNA Segment 6 (B NA DNA). The clone of B NA DNA used in these studies was constructed (2) and its sequence was determined (5) as described.

Hybrid-Arrest Translation. This procedure was done as described (13) using cloned B NA DNA.

Nuclease S1 Mapping and Sequence Analysis of NA mRNA by the Primer Extension Method. These procedures were done as described (14, 19). 5' Labeling of DNA fragments was done using T4 polynucleotide kinase and 3' labeling by using *Escherichia coli* DNA polymerase I (Klenow fragment).

Isotopic Labeling of Polypeptides in Infected Cells and Peptide Mapping. These procedures were done as described (20). Digestion of extracts with *endo-β-N*-acetylglucosaminidase H [Endoglycosidase H] was done as described (21) with modifications as indicated below.

RESULTS

Identification of Polypeptide NB in Infected Cells. The nucleotide sequence of RNA segment 6 of influenza B virus (Fig. 1) predicts that a polypeptide derived from the NB open reading frame should be detectable by using either cysteine or isoleucine as radioisotopic precursors. Such a polypeptide was not detected by using [³⁵S]methionine (1, 5) but NB would contain only one methionine residue, and this is the initiator methionine that is normally cleaved from mature polypeptides. HeLa cells infected with the Lee/40 and MB/50 strains of influenza B virus were labeled with [³⁵S]cysteine or [³⁵S]methionine and analyzed on polyacrylamide gels. As shown in Fig. 2, in addition to the known polypeptides of influenza B virus, a previously undetected polypeptide (NB) of $M_r \approx 17,700$ was found. NB migrates differently in the two strains, reflecting its virus specificity. As shown in Fig. 3A, NB could also be detected by using [³H]isoleucine. Infection of the HKCC line of hamster kidney cells with B/GL/1760/54 virus also showed the synthesis of NB (data not shown); this cell system produces high titers of infectious virus (16). The NB open reading frame predicts a translation product that would have a M_r of 11,242 before post-translational modification; however, there are four possible glycosylation sites (Asn-X-Thr or Ser) in NB, and the attachment of carbohydrate to any of these sites would cause an increase in molecular weight. To investigate the glycosylation of NB, B/Lee-infected HeLa cells were labeled with [³H]-glucosamine and the polypeptides were compared with those of infected cells labeled with [³H]isoleucine. As shown in Fig. 3A, in addition to the viral glycosylated polypeptides, the hem-

Abbreviations: NA, influenza virus neuraminidase; B NA DNA, cloned DNA of influenza B virus RNA segment 6; NB, previously unrecognized influenza B virus protein; endo H, *endo-β-N*-acetylglucosaminidase H; NS₂, influenza virus nonstructural protein ($M_r \approx 14,000$).

* Present address: Dept. of Biochemistry, Molecular and Cell Biology, Northwestern Univ., Evanston, IL 60201.

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.

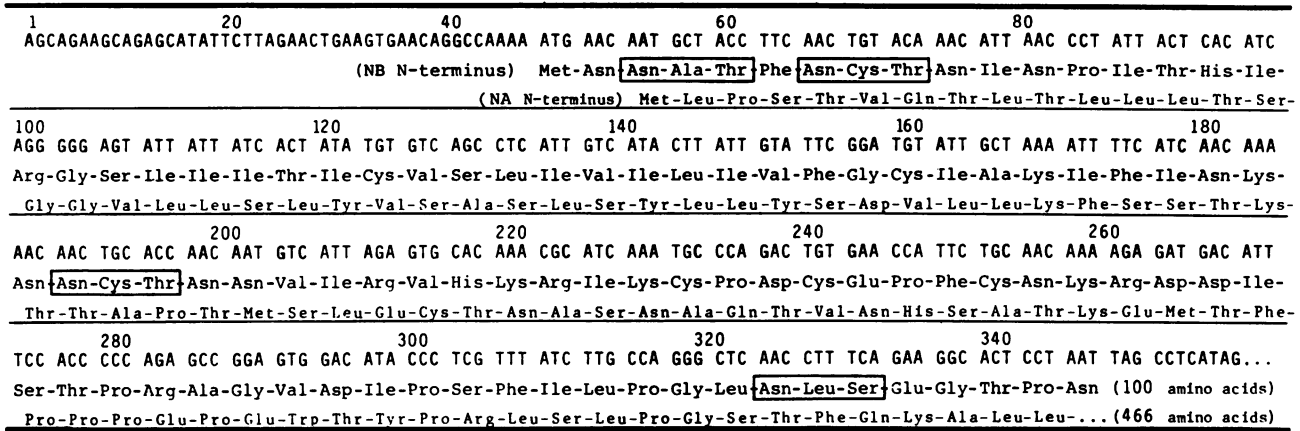


FIG. 1. Nucleotide sequence of the 5' region of B NA DNA, in the mRNA sense, showing the deduced amino acid sequences of the overlapping NB and NA reading frames (derived from ref. 5). Potential glycosylation sites in NB are boxed.

agglutinin and NA, a glycosylated polypeptide was detected (lane B/Lee GlcNH) that comigrated with NB labeled with [³H]isoleucine (lane B/Lee, Ile). Treatment of [³⁵S]cysteine-labeled infected cell extracts with endo H to remove carbohydrate chains resulted in the disappearance of NB from its normal position in the gel and the concomitant appearance of a polypeptide of M_r 11,000–12,000, the expected size of unglycosylated NB (NB₀) (Fig. 3B).

In Vitro Synthesis of NB₀ and Hybrid-Arrest Translation. Influenza B/Lee virus-specific mRNAs were translated in wheat germ extracts using [³H]isoleucine as labeled precursor. As shown in Fig. 4 (lane C), a polypeptide (NB₀) of M_r 11,000–12,000, migrating slightly slower than the viral nonstructural protein NS₂, was observed. Hybridization of the total B NA DNA to the mRNAs and subsequent translation specifically prevented

the translation of unglycosylated NA (NA₀) and NB₀ (Fig. 4, lane T). Hybridization using the smaller 5' BamHI fragment of B NA DNA specifically prevented the translation of NA₀ and NB₀, whereas hybridization with the larger 3' BamHI fragment of B NA DNA inhibited translation of NA₀ but not of NB₀ (Fig. 4, lanes 5' and 3'). Thus, these results map the NB coding region in the mRNA sense within the 5' 0.38 map units of the B NA DNA, confirming the above predictions.

Tryptic Peptide Mapping of NA, NB, NA₀, and NB₀. The specific labeling of polypeptide NB with isoleucine and cysteine but not with methionine, the changes in mobility of NB due to glycosylation, and the hybrid-arrest translation experi-

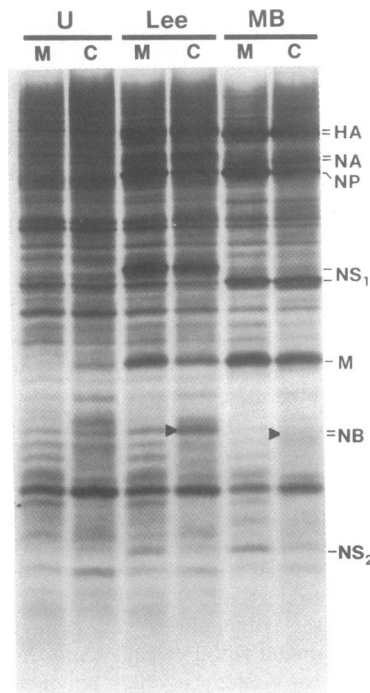


FIG. 2. Synthesis of polypeptide NB in HeLa cells infected with two strains of influenza B virus. Eight hours after infection, cells were labeled for 1 hr with [³⁵S]methionine (lanes M) or [³⁵S]cysteine (lanes C) and subjected to electrophoresis on a 20% polyacrylamide gel. Lanes: Lee, B/Lee/40; MB, B/MB/50; U, uninfected cells. Polypeptide NB is indicated by arrowheads.

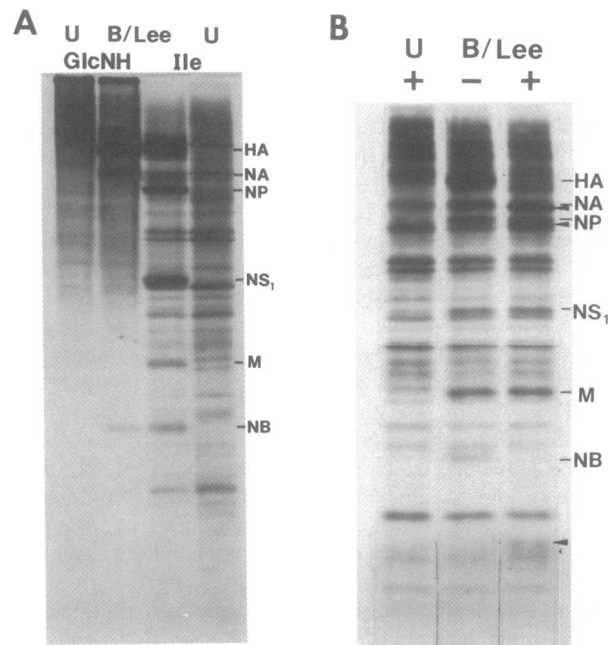


FIG. 3. (A) Glycosylation of polypeptide NB in influenza B virus-infected HeLa cells. Lanes: GlcNH, [³H]glucosamine-labeled cells; Ile, [³H]isoleucine-labeled cells; B/Lee, influenza B/Lee/40 virus-infected cells; U, uninfected cells. Samples were analyzed on a 20% polyacrylamide gel. (B) Treatment of influenza B/Lee/40 virus-infected cells with endo H. Arrows indicate position of HA, NA, and NB after endo H treatment. Eight hours after infection, cells were labeled for 1 hr with [³⁵S]cysteine and solubilized with 0.5% NaDodSO₄/50 mM Tris-HCl, pH 7.5/100 mM NaCl/2 mM EDTA/Trasyol (20 units/ml). An equal volume of 0.3 M sodium citrate (pH 5.5) was added (21). Extracts were incubated for 16 hr at 37°C with (+) or without (-) endo H at 0.033 unit/ml.

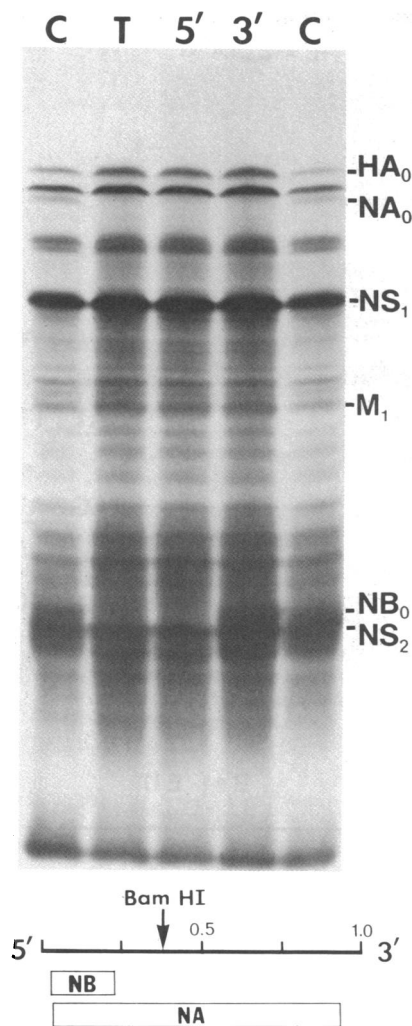


FIG. 4. Hybrid-arrested translation of NA and NB mRNA with restriction fragments of B NA DNA. Poly(A)-containing mRNAs from B/Lee/40-infected HeLa cells were hybridized with total B NA DNA (lane T), with the small (lane 5') or large (lane 3') *Bam*HI fragments of B NA DNA, or no added DNA (lanes C). mRNAs were translated in wheat germ extracts using [³H]isoleucine as labeled precursor and analyzed on a 20% polyacrylamide gel. A schematic representation of the B NA DNA with the *Bam*HI cleavage site and the NB and NA open reading frames is shown below.

ment strongly suggest that NB is the polypeptide encoded by the first open reading frame (designated NB) in the influenza B virus RNA segment 6. To provide further evidence and to eliminate the unlikely possibility that NB is a fragment of NA, [³⁵S]cysteine-labeled polypeptides synthesized both *in vitro* and *in vivo* were analyzed by tryptic peptide mapping (Fig. 5). NB and NA synthesized *in vivo* have different tryptic peptides (Fig. 5, top row). Similarly, NB₀ and NA₀ synthesized *in vitro* have tryptic peptides that are distinct (Fig. 5, middle row). Comparison of NB₀ synthesized *in vitro* with NB synthesized *in vivo* revealed that they are closely related (Fig. 5, left column), as shown by the mixture of equal amounts of NB₀ and NB; however, some differences could be due to glycosylation of peptides. A comparison of NA₀ and NA is also shown (Fig. 5, middle column). From the nucleotide sequence, it can be predicted that polypeptide NB includes four cysteine-containing tryptic peptides and that one of these is large and hydrophobic and unlikely to move from the origin in electrophoresis or chromatography. Therefore, the extra peptides in NB synthesized *in vivo* are probably host-cell contaminants.

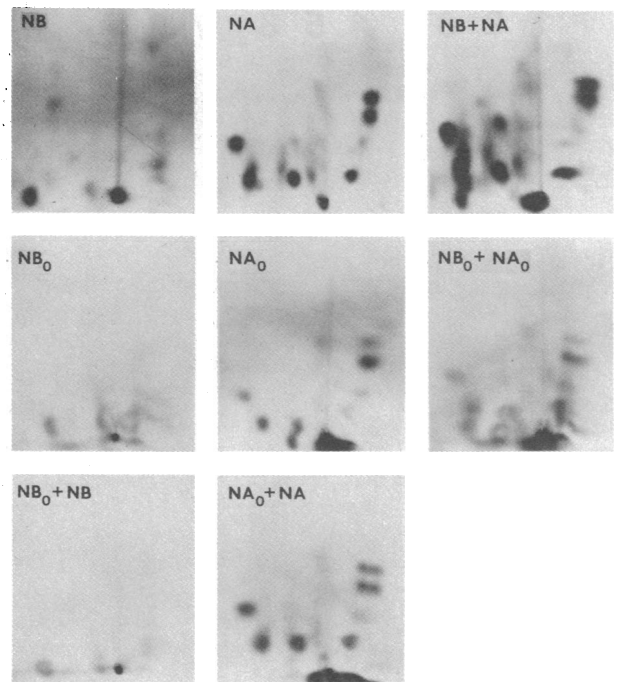


FIG. 5. Tryptic peptide maps of [³⁵S]cysteine-labeled NB and NA synthesized in HeLa cells and NB₀ and NA₀ synthesized *in vitro* in wheat germ extracts. Appropriate mixtures of these polypeptides are also shown.

Analysis of the mRNA for NB and NA. Sucrose gradient fractionation of influenza B virus mRNAs and translation of the mRNA in each fraction using [³⁵S]cysteine as a labeled precursor indicated that the mRNA for NA₀ and NB₀ sediments in the same position (Fig. 6). NS₂ labels poorly with [³⁵S]cysteine (Fig. 2) and therefore is not seen in Fig. 6; however, as shown previously (1, 2), the NS₂ mRNA (≈400 nucleotides) sediments more slowly than the NB₀ mRNA although the polypeptides are of similar size (122 versus 100 amino acids). This sedimentation experiment suggested that a single mRNA coded for both NA and NB. Further experiments were done to exclude the pos-

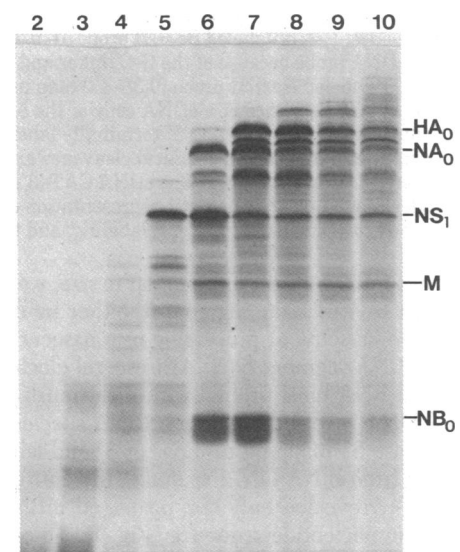


FIG. 6. Translation *in vitro* of influenza B/Lee/40 virus mRNAs separated on a 15–30% (wt/vol) sucrose gradient. Fractions were collected from the top of the gradient and translated in wheat germ extracts using [³⁵S]cysteine as labeled precursor.

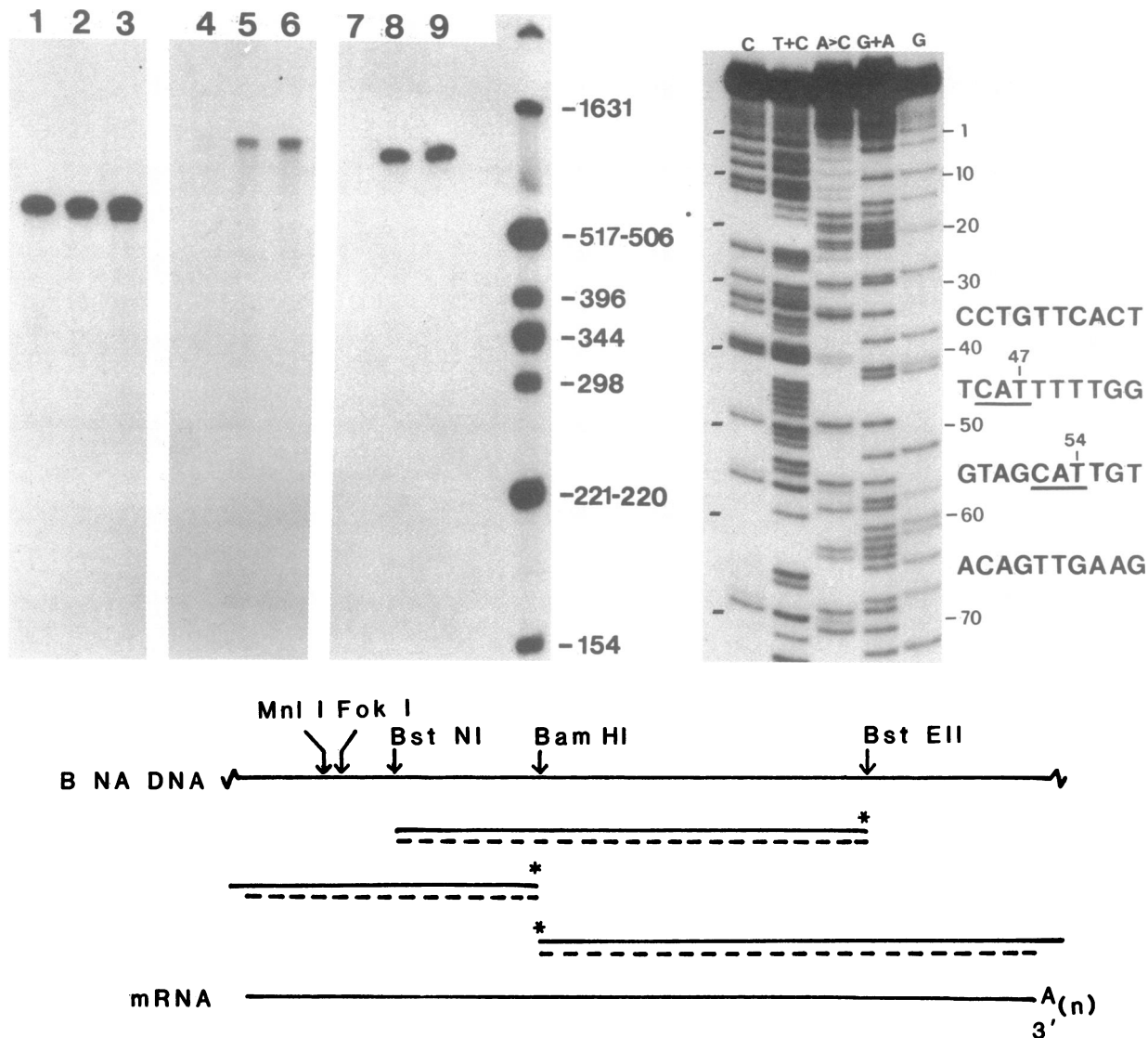


FIG. 7. (Upper Left) Nuclease S1 mapping of influenza B/Lee/RNA segment 6-derived mRNA. Poly(A)-containing mRNAs from influenza B/Lee/40 virus-infected HeLa cells were hybridized with 5'- or 3'-labeled B NA DNA fragments, digested with nuclease S1, and analyzed on 4% polyacrylamide gels containing 9 M urea. Lanes: 1-3, 4-6, and 7-9: no mRNA, 1 unit of mRNA, and 2 units of mRNA, respectively, hybridized with various probes. Lanes: 1-3, 5'-labeled *Bam*HI probe (0-0.38 map unit) (the protected DNA fragment would be expected to be slightly smaller than the reassociated DNA probe because of the G-C linker and nonviral sequences present in the probe); 4-6, *Bst*NI/*Bst*EII fragment, 5' labeled at the *Bst*EII site; 7-9, 3'-labeled *Bam*HI probe (0.38-1.0 map unit). mRNA units are arbitrary and DNA markers are from 32 P-labeled *Hinf*I-digested pBR322. (Upper Right) Sequence of a cDNA copy of the 5'-terminal region of RNA segment 6-derived mRNA. A small primer from B NA DNA [Mnl I-Fok I, nucleotides 140-172] and 5'-terminally labeled at nucleotide 172 was hybridized to influenza B/Lee/40 mRNAs and extended with reverse transcriptase. Base-specific chemical cleavages are indicated over each lane. The sequence of the cDNA copy is interpreted with the nucleotides numbered as shown in Fig. 1. The cDNA CAT (3'-5') complement of the two AUG codons for NB (nucleotides 47-49) and NA (nucleotides 54-56) are underlined. (Lower) Schematic representation of the B NA DNA and the restriction endonuclease fragments. The continuous line represents DNA; the asterisks, the sites of 32 P labeling; and the dotted line, the fragments protected by mRNA after nuclease S1 digestion.

sibility that two mRNA transcripts, similar in size, were derived from influenza B virus RNA segment 6, either by differential transcription or subsequent processing of a nascent transcript as found with RNA segment 8 (2). Agarose gel electrophoresis of mRNAs and RNA blots probed with nick-translated 32 P-labeled B NA DNA revealed only one mRNA species (data not shown). Nuclease S1 mapping using 5' or 3' end-labeled fragments derived from B NA DNA, spanning the entire length of the B NA DNA, revealed only one protected mRNA species homogeneous in nucleotide sequence (Fig. 7). To confirm the nuclease S1 mapping data, the sequence of the 5' end of the mRNA was determined by the primer extension method. A small B NA DNA fragment (Mnl I-Fok I, nucleotides 140-172] 5' uniquely labeled at nucleotide 172 was hybridized to virus-specific mRNA. The primer was extended with reverse transcrip-

tase and the nucleotide sequence of the cDNA was determined (Fig. 7). The sequence is homogeneous through the virus-specific portion of the mRNA, including the AUG codons beginning both the NB and the NA reading frames. Thus, the results of these experiments indicate that the mRNA for NB and NA have the same nucleotide sequence and that NA and NB are synthesized from a bifunctional mRNA derived from RNA segment 6.

DISCUSSION

The data presented here indicate that influenza B virus RNA segment 6 encodes two distinct glycoproteins, NA and NB, using overlapping reading frames. The synthesis of NB has been found with three different strains of influenza B virus, indi-

cating that it is not a nucleotide sequence peculiarity of one strain.

The finding of two initiation codons separated by four nucleotides giving rise to separate polypeptides, NA and NB, is another exception to the "scanning model" (22) for initiation of protein synthesis at the AUG codon closest to the 5' terminus of a mRNA. In other cases where the first AUG is not used it is usually followed by in-phase termination codons (23). However, there are a few cases in which two AUG codons in the same eukaryotic mRNA are used, and most of these occur in viral systems. A mRNA from the Elb region of adenovirus contains two initiator AUG codons from which two tumor antigens ($M_r \approx 21,000$ and $54,000$) are translated from different reading frames (24). The agnoprotein of simian virus 40 is probably translated from the same mRNA as viral protein 1, using separate AUG codons and nonoverlapping reading frames (25, 26), and viral proteins 2 and 3 of simian virus may be translated from the same mRNA using different initiator AUG codons. Some simian virus 40 recombinant DNA vectors contain both the agnoprotein AUG and a foreign gene AUG, and either can initiate protein synthesis (19, 27, 28). A mRNA of infectious pancreatic necrosis virus is thought to be polycistronic (29) and two bunyaviruses, snow shoe hare and La Crosse, have a minus-strand RNA segment encoding two overlapping reading frames (30) that probably code for two separate polypeptides using different AUG codons, but it is not known whether there is more than one mRNA species.

The S1 nuclease mapping, the primer extension sequence analysis, and the gradient fractionation of mRNAs all indicate that only one mRNA population is derived from influenza B virus RNA segment 6 and the data therefore suggest that the mRNA is bifunctional. However, a subtle modification giving rise to separate mRNA pools such as form 1 or form 2 cap structures or internal methylation has not been completely eliminated. Further indirect evidence for a single mRNA species is provided by finding that the synthesis of NA and NB are simultaneous during the course of infection (unpublished data). *In vivo*, NB and NA are made in roughly equivalent amounts whereas, *in vitro*, considerably more NB₀ is synthesized than NA₀ [cf. Fig. 3 and Fig. 4]. A similar finding was made with the two tumor antigens for adenovirus Elb region (24). This may reflect the ability of the wheat germ extracts to translate smaller polypeptides better than larger ones because of the "early quitting" phenomenon or a preference for initiation at the first AUG because of the *in vitro* conditions.

The open reading frame encoding the NB glycoprotein of influenza B virus RNA segment 6 has not been observed with influenza A viruses and it is a clear example of a structural difference between the two virus types. NB has not been detected in released virions (data not shown). It remains to be determined whether NB is inserted into membranes and, if so, whether this is in an NH₂- or COOH-terminal orientation or whether NB is secreted. The NB protein has stretches of hydrophobic amino acids at both its NH₂-terminus (residues 1–15, 19–40) and at its COOH terminus (residues 84–95) and both hydrophobic terminal regions contain potential glycosylation sites (Fig. 1). If NB is an integral membrane protein, then the region inserted in the membrane would not be expected to be glycosylated.

The function of NB in influenza B virus is at present unknown, but it is likely to be of significance because otherwise the overlapping reading frame would probably have been lost by the natural mutation rate. There are several biological differences between influenza A and B viruses. Genetic reas-

sortment does not occur between the two viruses, and influenza B virus does not undergo periodic major antigenic shifts as does influenza A virus. Influenza B virus has not been found in lower animals, and influenza B virus may be a more "toxic" virus; i.e., reactions to influenza B vaccines have at times been higher and the incidence of Reye syndrome is greater after influenza B than influenza A virus infection. There is as yet no evidence to implicate NB in any of these properties, but it is intriguing that it is the only protein not shared by the two viruses.

We thank Andrew Koopman for excellent technical assistance. This research was supported by Research Grants AI-05600 and AI-20201 from the National Institute of Allergy and Infectious Diseases and PCM 80-13464 from the National Science Foundation. M.W.S. was a Postdoctoral Fellow of the American Cancer Society and R.A.L. is an Established Investigator of the American Heart Association.

- Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1981) *Virology* **112**, 417–425.
- Briedis, D. J. & Lamb, R. A. (1982) *J. Virol.* **42**, 186–193.
- Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1982) *Virology* **116**, 581–588.
- Krystal, M., Elliot, R. M., Benz, E. W., Jr., Young, J. F. & Palese, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4800–4804.
- Shaw, M. W., Lamb, R. A., Erickson, B. W., Briedis, D. J. & Choppin, P. W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6817–6821.
- Palese, P. & Schulman, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2142–2146.
- Ueda, M., Tobita, K., Sugiura, A. & Enomoto, C. (1978) *J. Virol.* **25**, 685–686.
- Raccaniello, V. R. & Palese, P. (1979) *J. Virol.* **29**, 361–373.
- Fields, S., Winter, G. & Brownlee, G. G. (1981) *Nature (London)* **290**, 213–217.
- Hiti, A. L. & Nayak, D. P. (1982) *J. Virol.* **41**, 730–734.
- Markoff, L. & Lai, C.-J. (1982) *Virology* **119**, 288–297.
- Blok, J. & Air, G. M. (1982) *Virology* **121**, 211–229.
- Lamb, R. A., Choppin, P. W., Chanock, R. M. & Lai, C.-J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1857–1861.
- Lamb, R. A. & Lai, C.-J. (1980) *Cell* **21**, 475–485.
- Lamb, R. A., Lai, C.-J. & Choppin, P. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4170–4174.
- Choppin, P. W., Lazarowitz, S. G. & Goldberg, A. R. (1975) in *Negative Strand Viruses*, eds. Mahy, B. W. J. & Barry, R. D. (Academic, London), Vol. 1, pp. 105–119.
- Collins, P. L., Hightower, L. E. & Ball, L. A. (1978) *J. Virol.* **28**, 324–336.
- Lamb, R. A. & Choppin, P. W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4908–4912.
- Lamb, R. A. & Lai, C.-J. (1982) *Virology* **123**, 237–256.
- Lamb, R. A., Etkind, P. R. & Choppin, P. W. (1978) *Virology* **91**, 60–78.
- Erickson, A. H., Conner, G. E. & Blobel, G. (1981) *J. Biol. Chem.* **256**, 11224–11231.
- Kozak, M. (1978) *Cell* **15**, 1109–1123.
- Kozak, M. (1981) in *Current Topics in Microbiology and Immunology*, ed. Shatkin, A. J. (Springer, Berlin), Vol. 93, pp. 81–123.
- Bos, J. L., Polder, L., Bernards, R., Schrier, P., van den Elsen, P., van der Eb, A. & van Ormondt, H. (1981) *Cell* **27**, 121–131.
- Jay, G., Nomura, S., Anderson, C. W. & Khoury, G. (1981) *Nature (London)* **291**, 346–349.
- Mertz, J. E., Murphy, A. & Barkan, A. (1983) *J. Virol.* **45**, 36–46.
- Gruss, P., Efstratiadis, A., Karathanasis, S., Konig, M. & Khoury, G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6091–6095.
- Gruss, P., Rosenthal, N., Konig, M., Ellis, R. W., Shih, T. Y., Scolnick, E. M. & Khoury, G. (1982) in *Eukaryotic Viral Vectors*, ed. Gluzman, Y. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 13–17.
- Mertens, P. P. C. & Dobos, P. (1982) *Nature (London)* **297**, 243–246.
- Akashi, H. & Bishop, D. H. L. (1983) *J. Virol.* **45**, 1155–1158.