Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6817–6821, November 1982
Biochemistry

Complete nucleotide sequence of the neuraminidase gene of influenza B virus

(myxoviruses/evolution of viral proteins/membrane glycoprotein structure/nucleotide sequence comparison by metric analysis)

MICHAEL W. SHAW, ROBERT A. LAMB, BRUCE W. ERICKSON, DALIUS J. BRIEDIS, AND PURNELL W. CHOPPIN

The Rockefeller University, New York, New York 10021

Contributed by Purnell W. Choppin, August 2, 1982

ABSTRACT **The complete nucleotide sequence of the neuraminidase gene of influenza virus B/Lee/40 was derived from a cloned cDNA copy of virion RNA segment 6 and its corresponding mRNA. The RNA segment contains 1,557 virus-specific nucleotides, and the protein encoded by the longest open reading frame has a total of 466 amino acids with a molecular weight of 51,721. As is the case with the influenza A virus neuraminidases, the deduced amino acid sequence of the influenza B protein includes a single hydrophobic region near the amino terminus which would be capable of spanning the lipid bilayer of the viral or cell membrane. There are four potential glycosylation sites in the protein, two of which are near the amino-terminal hydrophobic region. Comparisons of the nucleotide and amino acid sequences with those of influenza A virus neuraminidases revealed seven regions of extensive homology within the central portion of the molecules, including 12 conserved cysteine residues. Five other cysteine residues in the terminal portions were also conserved.**

Both influenza A and influenza B viruses have a genome consisting of eight RNA segments which encode similar proteins (1). However, the two viruses cannot participate in reassortment of RNA segments to yield stable genotypically mixed virions (2, 3), and there is no immunological crossreactivity between the viral proteins. Similarities have been noted between the sequences of the two smallest genome segments of influenza A and B viruses (4–6), suggesting an evolutionary relationship. Unlike influenza A viruses, type B viruses have not undergone periodic major antigenic shifts in their surface glycoproteins—hemagglutinin and neuraminidase (NA)—presumably due to the lack of genetic reassortment in nature (7), which may be related to the absence of an animal reservoir for influenza B viruses.

The NA is encoded by a RNA segment that is the sixth largest in most strains (1, 8, 9). It projects from the surface of virions as a tetramer of identical molecules with a monomer molecular weight of 58,000–68,000, depending on the virus (10–13). Protease treatment releases the tetrameric head from the membrane-associated portion of the spike; this head has enzymatic and antigenic activities and a monomer molecular weight of 48,000–58,000 (10, 12). The complete nucleotide sequences of the NAs from several influenza A virus strains (14–16) suggest that the glycoprotein is bound to viral and cellular membranes by its amino terminus. Studies involving protease treatment of the virion and analyses of the intact and cleaved protein have supported this concept (13). Comparisons of the sequences have also revealed the presence of several conserved regions, of which some may be involved in the active site of the enzyme (13–17).

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.

In the present study, the NA gene of influenza B/Lee/40 virus was cloned and its sequence was determined and examined for homologies with the influenza A virus sequences. The results indicate significant structural similarities between the immunologically dissimilar NAs of influenza A and B viruses.

MATERIALS AND METHODS

Virus and Virus-Specific RNAs. The Lee/40 strain of influenza B virus was grown in embryonated eggs. Genome RNA segments were isolated from virions, and cytoplasmic mRNAs were isolated from infected HeLa cells and purified as described (4).

Cloning, Restriction Endonuclease Analysis, and Nucleotide Sequence Analysis of Virus-Specific DNA. The cloning and restriction endonuclease mapping of influenza B/Lee/40 virus genome segment 6 (B-NA DNA) was performed as described (4–6) for segments 7 and 8. DNA sequence analysis was performed by the procedure of Maxam and Gilbert (18) with minor modifications (19).

Hybrid-Arrested Translation. Hybrid-arrested translation of viral mRNAs in a wheat germ system to confirm the genome RNA segment that had been cloned and separation of the translation products on a $NaDodSO_4$/polyacrylamide gel were carried out as described (20).

Sequence Comparison. The nucleotide sequences of NA genes were compared by the mathematical method of metric analyses (21). The Sellers algorithms SS (22) and UU (23) were used to calculate the distance between two sequences, which is defined (21) as the minimum of the number of base changes plus twice the number of deletions needed to convert these sequences into a common sequence. The probability that the sequence similarity of two regions is not due to base composition alone was measured by a Monte Carlo method (21) involving the calculation of the distances between 3,600 permuted pairs of sequences having the same base compositions as the initial regions.

RESULTS

Selection of Cloned B-NA DNA. Double-stranded DNA duplexes were formed by hybridizing cDNA strands that had been synthesized separately from viral genome RNA and from virus-specific mRNA and cloned into plasmid pBR322 at the single *Pst* I site, as described (5). Plasmid DNA from positive colonies was purified, the inserted sequences were excised with *Pst* I, and the lengths of the inserts were determined on 4% polyacrylamide gels. The longest DNA insert was hybridized to total mRNAs from influenza B virus-infected cells; in hybrid-ar-

Abbreviations: NA, influenza virus neuraminidase; B-NA DNA, cloned DNA of influenza B virus neuraminidase gene.

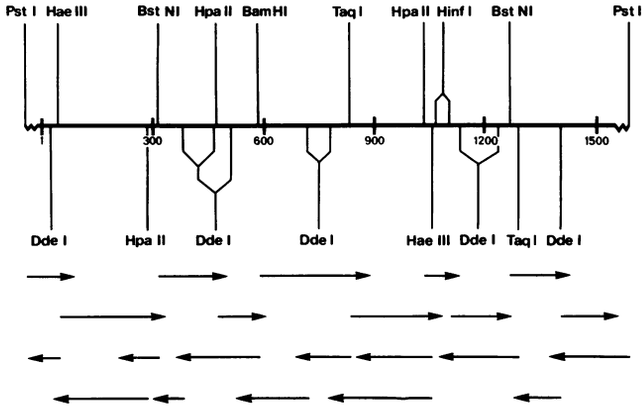

FIG. 1. Restriction endonuclease cleavage map and sequence determination strategy used for cloned B-NA DNA. The direction of mRNA transcription is from left to right. The zigzag line at each end represents the G-C linker used for plasmid construction. The origin of each arrow represents the restriction enzyme site labeled for sequence determination, and the body of each arrow denotes the direction and extent of analysis performed.

rested translation experiments, it specifically prevented the *in vitro* synthesis of the NA protein, confirming that the clone was derived from the gene for this protein.

Restriction Enzyme Mapping and Complete Nucleotide Sequence Determination of Influenza B Virus Genome RNA Segment 6. A partial restriction endonuclease cleavage map of the B-NA DNA was obtained and used as the basis for the sequence determination strategy as shown in Fig. 1. All restriction sites used for sequence analysis were overlapped, and the sequence of both DNA strands was determined for >99% of the full length of the B-NA DNA.

The nucleotide sequence of the mRNA sense strand of the cloned B-NA DNA is shown in Fig. 2. The B-NA DNA was shown to represent a full-length copy of the NA gene, as indicated by the finding of close agreement between the sequences at both the 5′ and 3′ termini of the insert and those of the viral genome segment determined by direct RNA sequence determination (24, 25). The virus-specific portion of the cloned B NA-DNA is 1,557 nucleotides long, excluding 10 nucleotides derived from cellular RNA by the "cap-transfer" process in influenza virus mRNA transcription (26). The first AUG codon oc-

```
non-viral  1                  20                  40                  60                  80                  100
AGACACACGC  AGCAGAAGCAGAGCATATTCTTAGAACTGAAGTGAACAGGCCAAAAATGAACA ATG CTA CCT TCA ACT GTA CAA ACA TTA ACC CTA TTA CTC ACA TCA GGG GGA GTA
                                                                  (N-terminus) Met-Leu-Pro-Ser-Thr-Val-Gln-Thr-Leu-Thr-Leu-Leu-Leu-Thr-Ser-Gly-Gly-Val-(18)

          120                 140                 160                 180                 200
TTA TTA TCA CTA TAT GTG TCA GCC TCA TTG TCA TAC TTA TTG TAT TCG GAT GTA TTG CTA AAA TTT TCA TCA ACA AAA ACA ACT GCA CCA ACA ATG TCA TTA GAG
Leu-Leu-Ser-Leu-Tyr-Val-Ser-Ala-Ser-Leu-Ser-Tyr-Leu-Leu-Tyr-Ser-Asp-Val-Leu-Leu-Lys-Phe-Ser-Ser-Thr-Lys-Thr-Thr-Ala-Pro-Thr-Met-Ser-Leu-Glu-(53)

      220                 240                 260                 280                 300
TGC ACA AAC GCA TCA AAT GCC CAG ACT GTG AAC CAT TCT GCA ACA AAA GAG ATG ACA TTT CCA CCC CCA GAG CCG GAG TGG ACA TAC CCT CGT TTA TCT TGC CAG
Cys-Thr-[Asn-Ala-Ser]-Asn-Ala-Gln-Thr-Val-[Asn-His-Ser]-Ala-Thr-Lys-Glu-Met-Thr-Phe-Pro-Pro-Pro-Glu-Pro-Glu-Trp-Thr-Tyr-Pro-Arg-Leu-Ser-Cys-Gln-(88)

320                 340                 360                 380                 400                 420
GGC TCA ACC TTT CAG AAG GCA CTC CTA ATT AGC CCT CAT AGG TTC GGA GAG ATC AAA GGA AAC TCA GCT CCC TTG ATA ATA AGA GAA CCT TTT GTT GCT TGT GGA
Gly-Ser-Thr-Phe-Gln-Lys-Ala-Leu-Leu-Ile-Ser-Pro-His-Arg-Phe-Gly-Glu-Ile-Lys-Gly-Asn-Ser-Ala-Pro-Leu-Ile-Ile-Arg-Glu-Pro-Phe-Val-Ala-Cys-Gly-(123)

          440                 460                 480                 500                 520
CCA AAA GAA TGC AGA CAC TTT GCT CTG ACC CAT TAT GCA GCT CAG CCG GGG GGA TAC TAC AAT GGA ACA AGA AAG GAC AGA AAC AAG CTG AGG CAT CTA GTA TCA
Pro-Lys-Glu-Cys-Arg-His-Phe-Ala-Leu-Thr-His-Tyr-Ala-Ala-Gln-Pro-Gly-Gly-Tyr-Tyr-[Asn-Gly-Thr]-Arg-Lys-Asp-Arg-Asn-Lys-Leu-Arg-His-Leu-Val-Ser-(158)

          540                 560                 580                 600                 620
GTC AAA TTG GGA AAA ATC CCA ACT GTG GAA AAC TCC ATT TTC CAC ATG GCA GCT TGG AGC GGA TCC GCA TGC CAT GAT GGT AGA GAA TGG ACA TAT ATC GGA GTT
Val-Lys-Leu-Gly-Lys-Ile-Pro-Thr-Val-Glu-Asn-Ser-Ile-Phe-His-Met-Ala-Ala-Trp-Ser-Gly-Ser-Ala-Cys-His-Asp-Gly-Arg-Glu-Trp-Thr-Tyr-Ile-Gly-Val-(193)

      640                 660                 680                 700                 720
GAT GGT CCT GAC AAT GAT GCA TTG GTC AAA ATA AAA TAT GGA GAA GCA TAT ACT GAC ACA TAT CAT TCC TAT GCA CAC AAC ATC CTA AGA ACA CAA GAA AGT GCC
Asp-Gly-Pro-Asp-Asn-Asp-Ala-Leu-Val-Lys-Ile-Lys-Tyr-Gly-Glu-Ala-Tyr-Thr-Asp-Thr-Tyr-His-Ser-Tyr-Ala-His-Asn-Ile-Leu-Arg-Thr-Gln-Glu-Ser-Ala-(228)

740                 760                 780                 800                 820                 840
TGC AAT TGC ATC GGG GGA GAT TGT TAT CTT ATG ATA ACA GAC GGC TCA GCT TCA GGA ATT AGT AAA TGC AGA TTT CTT AAA ATT AGA GAG GGT CGA ATA ATA AAA
Cys-Asn-Cys-Ile-Gly-Gly-Asp-Cys-Tyr-Leu-Met-Ile-Thr-Asp-Gly-Ser-Ala-Ser-Gly-Ile-Ser-Lys-Cys-Arg-Phe-Leu-Lys-Ile-Arg-Glu-Gly-Arg-Ile-Ile-Lys-(263)

          860                 880                 900                 920                 940
GAA ATA CTT CCA ACA GGA AGA GTG GAG CAC ACT GAA GAG TGC ACA TGC GGG TTC GCC AGC AAT AAA ACC ATA GAA TGT GCC TGT AGA GAC AAC AGT TAC ACA GCA
Glu-Ile-Leu-Pro-Thr-Gly-Arg-Val-Glu-His-Thr-Glu-Glu-Cys-Thr-Cys-Gly-Phe-Ala-Ser-[Asn-Lys-Thr]-Ile-Glu-Cys-Ala-Cys-Arg-Asp-Asn-Ser-Tyr-Thr-Ala-(298)

          960                 980                 1000                1020                1040
AAA AGA CCC TTT GTC AAA TTA AAT GTG GAA ACT GAT ACA GCT GAA ATA AGA TTG ATG TGC ACA AAG ACT TAT CTA GAC ACT CCC AGA CCG GAT GAT GGA AGC ATA
Lys-Arg-Pro-Phe-Val-Lys-Leu-Asn-Val-Glu-Thr-Asp-Thr-Ala-Glu-Ile-Arg-Leu-Met-Cys-Thr-Lys-Thr-Tyr-Leu-Asp-Thr-Pro-Arg-Pro-Asp-Asp-Gly-Ser-Ile-(333)

      1060                1080                1100                1120                1140
GCA GGG CCT TGC GAA TCT AAT GGA GAC AAG TGG CTT GGA GGC ATC AAA GGA GGA TTC GTC CAT CAA AGA ATG GCA TCT AAG ATT GGA AGA TGG TAC TCC CGA ACG
Ala-Gly-Pro-Cys-Glu-Ser-Asn-Gly-Asp-Lys-Trp-Leu-Gly-Gly-Ile-Lys-Gly-Gly-Phe-Val-His-Gln-Arg-Met-Ala-Ser-Lys-Ile-Gly-Arg-Trp-Tyr-Ser-Arg-Thr-(368)

1160                1180                1200                1220                1240                1260
ATG TCT AAA ACT AAC AGA ATG GGG ATG GAA CTG TAT GTA AAG TAT GAT GGT GAC CCA TGG ACT GAC AGT GAT GCT CTT ACT CTT AGT GGA GTA ATG GTT TCC ATA
Met-Ser-Lys-Thr-Asn-Arg-Met-Gly-Met-Glu-Leu-Tyr-Val-Lys-Tyr-Asp-Gly-Asp-Pro-Trp-Thr-Asp-Ser-Asp-Ala-Leu-Thr-Leu-Ser-Gly-Val-Met-Val-Ser-Ile-(403)

          1280                1300                1320                1340                1360
GAA GAA CCT GGT TGG TAT TCT TTT GGC TTC GAA ATA AAG GAC AAG AAA TGT GAT GTC CCT TGT ATT GGG ATA GAG ATG GTA CAC GAT GGT GGA AAA GAT ACT TGG
Glu-Glu-Pro-Gly-Trp-Tyr-Ser-Phe-Gly-Phe-Glu-Ile-Lys-Asp-Lys-Lys-Cys-Asp-Val-Pro-Cys-Ile-Gly-Ile-Glu-Met-Val-His-Asp-Gly-Gly-Lys-Asp-Thr-Trp-(438)

          1380                1400                1420                1440                1460
CAT TCA GCT GCA ACA GCC ATT TAC TGT TTG ATG GGC TCA GGA CAA TTG CTA TGG GAC ACT GTC ACA GGC GTT GAT ATG GCT TTA TAA TAG AGGAATGGTTGGATCTGTT
His-Ser-Ala-Ala-Thr-Ala-Ile-Tyr-Cys-Leu-Met-Gly-Ser-Gly-Gln-Leu-Leu-Trp-Asp-Thr-Val-Thr-Gly-Val-Asp-Met-Ala-Leu (C-terminus) 466 amino acids

1480                1500                1520                1540                1557
CTAAACCCTTTGTTCCTATTTTATTTGAACAGTTGTTCTTACTAGATTTAATTGTTTCTGAAAAATGCTCTTGTTACTACT 3′
```

FIG. 2. Nucleotide sequence of the mRNA sense strand of the cloned B-NA DNA. The deduced amino acid sequence of the open reading frame following the second ATG is shown with the potential glycosylation sites in boxes. The first ATG and the first termination codon in the reading frame beginning at this codon are underlined.

curs at nucleotides 47–49, but the open reading frame for translation that begins at this codon terminates at nucleotide 346. This would yield a protein with a molecular weight of only 11,242, which is incompatible with the size of the NA protein. The second AUG at nucleotides 54–56 is followed by an open reading frame extending to two consecutive termination signals at nucleotides 1,452–1,457, suggesting that this second AUG is the initiation site for the translation of the NA. Translation beginning at this codon would yield a protein of 466 amino acids, molecular weight 51,721, which is the appropriate size for the NA. This protein would contain four potential glycosylation sites (i.e., as asparagine followed by an unspecified amino acid, followed by serine or threonine) at amino acids 56, 64, 144, and 284. As shown in Table 1, the amino acid composition predicted from the nucleotide sequence is in close agreement with results of an amino analysis of the B/Lee/40 NA (27).

Comparison of Nucleotide and Amino Acid Sequences of Influenza A and B Virus NAs. The complete nucleotide sequence of the influenza B/Lee/40 NA was compared to sequences of the PR/8/34 (14), WSN/33 (15), and Udorn/72 (16) strains of influenza A virus by using computer-assisted metric analysis (21–23). Significant areas of homology at both the nucleotide and amino acid levels were detected; the most extensive homology was in the central portion of the genes. In the example in Fig. 3, the central 744 nucleotides of the B/Lee/40 NA gene are compared to the central 741 nucleotides from A/Udorn/72. This area encodes the middle 248 and 247 amino acids of the NAs from B/Lee/40 and A/Udorn/72, respectively, and includes >50% of the sequence of each protein. A total of 87 amino acids are conserved, with the significant inclusion of 12 cysteine residues. A potential glycosylation site is also conserved in this region, at amino acids 144 (B/Lee) and 146 (A/Udorn).

The alignment of the central portions of B/Lee and A/Udorn in Fig. 3 shows seven regions of significant homology (Table 2). These regions, ranging from 45 to 126 nucleotides, were obtained by omitting those portions that either are out of phase or can be aligned in more than one equivalent way. The similarity of these regions is much too great to be due to chance alone. This is illustrated by region 5, in which the probability that two nucleotide sequences having the base compositions of B/Lee region 807–890 and A/Udorn region 776–859 would differ by only 43 of 84 bases is <0.001. The other regions of homology are even more similar than region 5. The similarity of these seven regions of homology is thus not due to their base compositions but results from another cause—e.g., their evolution from a common ancestral gene.

Table 1. Amino acid composition of influenza virus B/Lee/40 NA

Residue	Residues per 100 residues	
	Predicted	Determined*
Asx	8.6	10.0
Thr	8.6	8.2
Ser	8.2	7.6
Glx	7.1	7.9
Pro	4.7	4.7
Gly	8.8	9.6
Ala	6.4	6.6
Cys	3.9	2.7
Val	4.7	4.9
Met	3.0	2.5
Ile	5.6	5.6
Leu	8.2	8.8
Tyr	4.3	3.9
Phe	2.8	2.8
His	2.8	2.6
Lys	6.0	6.8
Arg	4.5	4.8
Trp	1.9	†

* From Laver and Baker (27).
† Not determined.

DISCUSSION

The NA gene of influenza B/Lee/40 virus found here is approximately 10% longer than that of A/PR/8/34 NA (14). The B/Lee/40 protein encoded by the longest open reading frame would contain 466 amino acids (molecular weight, 51,721) with four potential glycosylation sites. Previous estimates of the molecular weight of the influenza B virus NA based on electrophoretic mobility or amino acid composition have varied from 60,000 to 68,000 for the glycoprotein monomer, including an estimated 12–20% carbohydrate content (1, 10–13, 27, 28), giving a range for the polypeptide chain of 48,000 to 59,000. The amino acid composition deduced from the nucleotide sequence is in agreement with results obtained several years ago from analysis of the purified glycoprotein (27).

The deduced amino acid sequence of influenza B/Lee/40 NA supports the hypothesis that the amino-terminal portion is inserted into the viral membrane, an orientation suggested by Fields *et al.* (14) for influenza A virus on the basis of the nucleotide sequence of the A/PR/8/34 strain. In the influenza A NA, the only hydrophobic region sufficiently long to span a membrane is near the amino terminus before amino acid 36. Similarly, in the B/Lee/40 protein the only uncharged potential membrane-spanning sequence extends from residues 4 to 34. Similar results have been found in the amino acid sequences deduced for the NAs of A/WSN/33 (15) and A/Udorn/72 (16). The positions of the four potential glycosylation sites in the influenza B/Lee/40 protein also support the concept of an amino-terminal membrane insertion. Removal of the portion of the protein possessing NA activity from intact B/Lee/40 virions with trypsin is associated with the loss of ≈50% of the carbohydrate content of the protein, leaving 50% associated with the stalk that is attached to the viral membrane and is isoleucine-poor (28). In the amino acid sequence deduced in the present studies, two of the four potential glycosylation sites are found within the first 70 residues from the amino terminus, which is an area remarkable for its lack of isoleucine residues in spite of the long hydrophobic region.

The size of influenza B virus NA, the amino-terminal hydrophobic region, and the homology between the A and B viruses all suggest that translation of B/Lee NA begins at the second AUG codon from the 5′ end of the mRNA. Although in the majority of eukaryotic mRNAs, including those of the other influenza virus genes whose sequences have been determined, the first AUG codon is utilized as the initiation site for translation, there are cases in which the first AUG codon is not used (reviewed in ref. 29). To exclude a sequence error in the immediate 5′-terminal region encompassing the first two AUG codons of the B-NA DNA, this region was subjected to sequence determination (both strands) several times. To eliminate the possibility of a copy error introduced during the cloning procedures, the nucleotide sequences at the 5′ end of a second full-length B-NA clone was determined, and the sequence was confirmed. In the influenza B virus NA mRNA, the first and second AUG codons are separated only by four nucleotides, and thus the first AUG codon may be bypassed by the hypothesized "scanning mechanism" (29) of initiation of protein synthesis. However, it is unusual that there is a long open reading frame,

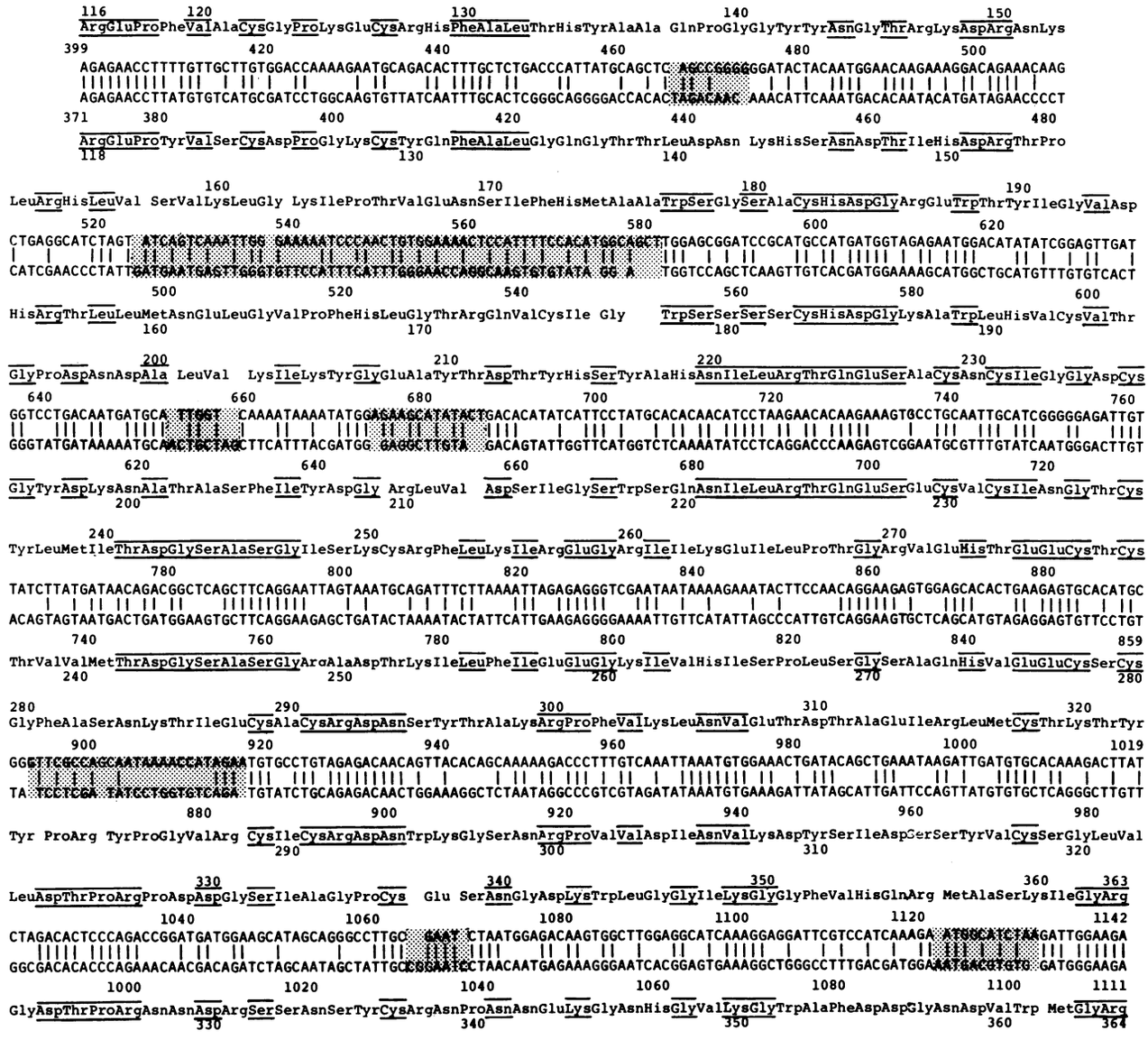

FIG. 3. Metric analysis, using algorithms SS and UU, of the NA genes from influenza B/Lee/40 nucleotides 399–1,142 (upper rows) and influenza A/Udorn/72 nucleotides 371–1,111 (lower rows) (16). Common amino acids are indicated by horizontal lines; common nucleotides are indicated by vertical lines. Stippling identifies regions in which the reading frames for the two sequences are out of phase.

which could code for 100 amino acids, following the first AUG codon. We have not observed synthesis of this protein, but its possible existence (and thus of a bicistronic mRNA) has not been rigorously excluded. Such a protein would be rich in isoleucine and have an uncharged amino terminus of 40 amino acids, except for histidine and arginine residues at positions 16 and 18, and two potential glycosylation sites immediately at its amino terminus.

Table 2. Characterization of seven regions of homology in the center of NA genes of influenza A and B viruses

Region	Bases, no.	Positions of bases in sequences: B/Lee	Positions of bases in sequences: A/Udorn	Same bases, %	Same amino acids, %	bc/b*
1	66	399–464	371–436	53	46	0.470
2	45	477–521	449–493	53	40	0.467
3	42	582–653	551–622	60	50	0.403
4	108	687–794	656–763	58	61	0.417
5	84	807–890	776–859	49	39	0.512
6	126	918–1,043	884–1,009	58	38	0.421
7	48	1,071–1,118	1,040–1,087	54	31	0.458

* The normalized distance—the distance in base changes (bc) between the aligned regions divided by the number of bases (b) in each region. In each case the probability that the regions of the two viruses would be this similar by chance is <0.001.

Computer-assisted analyses of sequence homologies at the nucleotide and amino acid levels were undertaken with algorithms designed to identify deletion/insertion events without reading frame bias (21–23). The example shown in Fig. 3 encompasses the central half of the proteins encoded by B/Lee/40 and A/Udorn/72, where homology is greatest. Although the three-dimensional structure of the NA is not yet known in detail, the extensive conservation of this central region of the influenza A and B proteins, which are immunologically distinct, suggests that this region may encompass the active site of the enzyme. The active site would be expected to have stringent requirements with regard to spatial arrangement and, therefore, amino acid sequence. The conservation of cysteine residues provides striking evidence for homology among the NAs of the different viruses. In the central portion of the molecule, 12 cysteine residues are conserved. As shown in Fig. 4, these

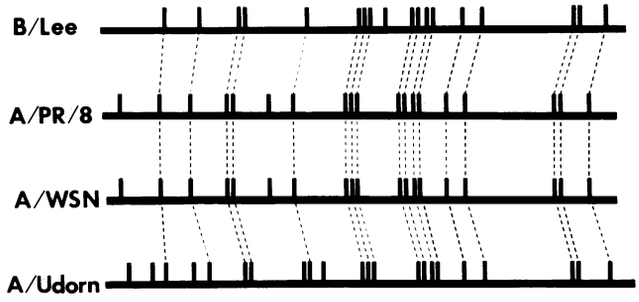

FIG. 4. Schematic representation of the conserved cysteine residues in the B/Lee strain of influenza B virus and three influenza A virus strains. The vertical solid lines indicate the placement of the residues in the amino acid sequence of the protein, and the dotted lines connect homologous residues in the different proteins.

residues are conserved not only in B/Lee and A/Udorn but also in the A/PR/8 and A/WSN strains. In addition, five other cysteine residues in the terminal portions of the protein are conserved in all four strains. It would be expected that some, if not all, of these cysteine residues would be involved in disulfide bonds important in determining the structure of the protein and its active site.

A striking feature of the amino acid sequences predicted from the nucleotide sequences of eight different subtypes of influenza A NAs is that the first six amino acids at the amino terminus are identical (13). The amino acid sequence in this region of the influenza B/Lee NA is different; however, it is noteworthy that the third residue, proline, is the same in the A and B proteins.

Electron microscopic studies of detergent-released and protease-released NAs suggest that although still possessing antigenic and enzymatic activities, the protease-released polypeptides have lost the stalk by which the subunits of the tetrameric head are attached to the membrane (10, 12). The amino acid and nucleotide compositions of the NAs analyzed thus far would therefore suggest that the highly conserved central portion of the molecule is in the globular head where enzymatic activity resides. The lack of detectable antigenic crossreactivity between the influenza A and B NAs (30) suggests that this conserved region is relatively inaccessible to antibody. This concept is also supported by the conclusion that the inhibition of NA activity by antibodies was due to steric hindrance rather than binding of the antibody to the active site (31).

We thank Mary A. Conners for excellent technical assistance and Janet M. Sekulski for programming the Sellers algorithms. This work was supported by Research Grants AI-05600 and AI-18746 from the National Institute of Allergy and Infectious Diseases and CA-28571 from the National Cancer Institute. M.W.S. is a Postdoctoral Fellow of the American Cancer Society; D.J.B. was a Postdoctoral Fellow of the Medical Research Council of Canada; and R.A.L. is an Irma T. Hirschl Awardee and an Established Investigator of the American Heart Association.

1. Racaniello, V. R. & Palese, P. (1979) *J. Virol.* **29,** 361–373.
2. Compans, R. W. & Choppin, P. W. (1975) in *Comprehensive Virology,* eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum, New York), Vol. 4, pp. 179–252.
3. Kilbourne, E. D. (1975) in *The Influenza Viruses and Influenza,* ed. Kilbourne, E. D. (Academic, New York), pp. 1–14.
4. Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1981) *Virology* **112,** 417–425.
5. Briedis, D. J. & Lamb, R. A. (1982) *J. Virol.* **42,** 186–193.
6. Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1982) *Virology* **116,** 581–588.
7. Webster, R. G. & Laver, W. G. (1971) *Prog. Med. Virol.* **13,** 271–338.
8. Palese, P. & Schulman, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 2142–2146.
9. Ueda, M., Tobita, K., Sugiura, A. & Enomoto, C. (1978) *J. Virol.* **25,** 685–686.
10. Laver, W. G. & Valentine, R. C. (1969) *Virology* **38,** 105–119.
11. Lazdins, I., Haslam, E. A. & White, D. O. (1972) *Virology* **49,** 758–765.
12. Wrigley, N. G., Skehel, J. J., Charlewood, P. A. & Brand, C. M. (1973) *Virology* **51,** 525–529.
13. Blok, J., Air, G. M., Laver, W. G., Ward, C. W., Lilley, G. G., Woods, E. F., Roxburgh, C. M. & Inglis, A. S. (1982) *Virology* **119,** 109–121.
14. Fields, J., Winter, G. & Brownlee, G. G. (1981) *Nature (London)* **290,** 213–217.
15. Hiti, A. L. & Nayak, D. P. (1982) *J. Virol.* **41,** 730–734.
16. Markoff, L. & Lai, C.-J. (1982) *Virology* **119,** 288–297.
17. Blok, J. & Air, G. M. (1980) *Virology* **107,** 50–60.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
19. Lamb, R. A. & Lai, C.-J. (1980) *Cell* **21,** 475–485.
20. Lamb, R. A., Choppin, P. W., Chanock, R. M. & Lai, C.-J. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1857–1861.
21. Erickson, B. W. & Sellers, P. H. (1982) in *Time Warps, String Edits, and Macromolecules,* eds. Sankoff, D. & Kruskal, J. B. (Addison-Wesley, Reading, MA), in press.
22. Sellers, P. H. (1974) *SIAM J. Appl. Math.* **26,** 787–793.
23. Sellers, P. H. (1980) *J. Algorithms* **1,** 359–373.
24. Skehel, J. J. & Hay, A. J. (1978) *Nucleic Acids Res.* **5,** 1207–1219.
25. Desselberger, U., Racaniello, V. R., Zazra, J. J. & Palese, P. (1980) *Gene* **8,** 315–328.
26. Krug, R. M., Broni, B. A. & Bouloy, M. (1979) *Cell* **18,** 329–334.
27. Laver, W. G. & Baker, N. (1972) *J. Gen. Virol.* **17,** 61–67.
28. Allen, A. K., Skehel, J. J. & Yuferov, V. (1977) *J. Gen. Virol.* **37,** 625–628.
29. Kozak, M. (1981) in *Current Topics in Microbiol. and Immunol.,* ed. Shatkin, A. J. (Springer, Berlin), Vol. 93, pp. 81–123.
30. Paniker, C. K. J. (1968) *J. Gen. Virol.* **2,** 385–394.
31. Rafelson, M. E., Schneir, M. & Wilson, V. M. (1963) *Arch. Biochem. Biophys.* **103,** 424–430.