Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4800–4804, August 1982
Microbiology

Evolution of influenza A and B viruses: Conservation of structural features in the hemagglutinin genes

(nucleotide sequence homology/amino acid homology/functional domains of proteins/negative strand RNA virus)

MARK KRYSTAL*, RICHARD M. ELLIOTT*†, EDMUND W. BENZ, JR.‡, JAMES F. YOUNG*, AND PETER PALESE*

*Mount Sinai School of Medicine, New York, New York 10029; and ‡Albert Einstein College of Medicine, Bronx, New York 10461

Communicated by Edwin D. Kilbourne, May 5, 1982

ABSTRACT **The complete nucleotide sequence of the hemagglutinin (HA) gene of a type B influenza virus (B/Lee/40) was obtained by using cloned cDNA derived from the RNA segment. The gene is 1,882 nucleotides long and can code for a protein precursor of 584 amino acids. Structural features common to type A virus HAs are also conserved in the B virus HA. These include a hydrophobic signal peptide, hydrophobic NH_2 and COOH termini of the HA2 subunit, and a HA1/HA2 cleavage site involving an arginine residue. The sequence of the B HA gene and its deduced amino acid sequence were compared to those of a type A influenza virus (A/PR/8/34). When these two genes were aligned, it was found that 24% of the amino acids in the HA1 subunits and 39% of the amino acids in the HA2 subunits are conserved. This degree of relatedness between type B virus and type A virus HAs (intertypic comparison) is similar to the homologies observed among certain type A virus HAs (intratypic comparison). A close evolutionary relationship is therefore suggested between the HAs of type A and type B influenza viruses.**

Influenza is still a major disease in man, primarily because of the property of antigenic variation associated with influenza virus. There are three types of influenza viruses (A, B, and C) that are defined by the absence of serologic crossreactivity between their internal proteins (for review, see refs. 1–3). Influenza A viruses are further classified into subtypes based on antigenic differences of their glycoproteins, the hemagglutinin (HA) and neuraminidase (NA) (4).

The advent of recombinant DNA techniques has permitted the rapid cloning of RNA segments from different influenza A viruses and analysis of their sequences. This allowed for precise studies on the extent of variation among influenza A virus HAs belonging to the H1, H2, H3, and H7 subtypes (5–10). In addition, these studies have aided in our understanding of the structural features that determine the biological and antigenic properties of the HAs of influenza A viruses. In contrast, little is known about the HA molecule of influenza B viruses.

We report here the results of cloning studies designed to elucidate the sequence of an influenza B virus HA. Comparison of the B/Lee/40 virus HA with that of the A/PR/8/34 virus (5) reveals sequence and structural similarities suggesting a close evolutionary relationship.

MATERIALS AND METHODS

Virus and Bacterium. Influenza B virus B/Lee/40 was grown in embryonated hen's eggs. Purification of virus and extraction of RNA from the virus were as described (11). Plasmid pBR322 containing cloned virus-specific double-stranded DNA was propagated in *Escherichia coli* C600 cells as described (12).

Cloning and Sequence Analysis of the HA Gene. The synthesis of double-stranded cDNA molecules from influenza virus RNA by using synthetic dodecamer nucleotide primers has been described (12). The sequence of the HA gene of influenza B/Lee/40 virus was obtained from a pair of overlapping cDNA clones. A partial clone was first identified by end sequence analysis as deriving from the HA gene segment. It contained the terminal nucleotides used as primer and was found to possess sequences that could code for the NH_2-terminal amino acids previously identified for the HA of B/Lee/40 virus (13). From this 1,336-base pair (bp) clone, a restriction enzyme fragment (corresponding to nucleotides 973–1,236) was obtained and used as a hybridization probe to identify additional clones containing HA-specific sequences. In this way, another clone was identified; it was found by end sequence analysis to begin at nucleotide 483 and to extend to the 5′ end of the virion RNA, which was previously determined by direct sequence analysis of RNA (14). The sequence of each clone was determined by using the Maxam and Gilbert chemical modification procedure (15). DNA fragments suitable for sequence analysis were obtained either by strand separation or by secondary cleavage of 5′ end-labeled restriction enzyme fragments. The strategy for determining the nucleotide sequence of the HA gene is shown in Fig. 1. Sequence analyses were performed on both strands of the DNA except for the following short regions: nucleotides 47–164; nucleotides 1,456–1,494; and nucleotides 1,717–1,750. In all instances, the sequences across the boundaries of the restriction enzyme sites used to generate fragments were confirmed by using overlapping DNA fragments.

Computer Analysis of Sequences. Nucleotide sequence data were stored and edited in an IBM 370 computer at the University Computing Center of the City University of New York by using published programs (16–18). The nucleotide and amino acid sequences of the B/Lee/40 HA gene were graphically compared to those of the A/PR/8/34 virus by using a DEC 1170 computer at the Albert Einstein College of Medicine and were printed out via a Tektronix 4052 microcomputer coupled with a 4662 *x–y* plotter. For matrix comparison of the nucleotide sequences, a window of 10 nucleotides was used and a homology value of >75% was scored as positive (generating a solid dot). In comparing protein sequences a window of two amino acids was used and a dot was generated when both were identical. For the analysis of the relative hydrophilicity of the different hemagglutinins, a window of 10 amino acids was used to generate a relative hydrophilicity value based on published hydrophilicity values for each amino acid (19) (relative hydrophilicity = total hydrophilicity of decapeptide/10).

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.

Abbreviations: HA, hemagglutinin; bp, base pair(s).

† Present address: University of Glasgow, Institute of Virology, Glasgow, Scotland.

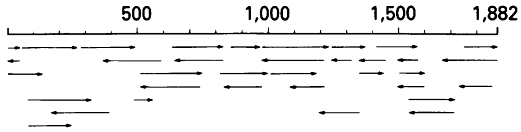

FIG. 1. Strategy for obtaining the nucleotide sequence of the influenza B/Lee/40 HA gene. Nucleotides are numbered from the start of the gene as in Fig. 2 (plus sense strand). The arrows indicate the direction of sequence analysis and the string of nucleotides that was determined in a particular restriction enzyme fragment.

RESULTS AND DISCUSSION

Nucleotide Sequence. The nucleotide sequence of the influenza B/Lee/40 HA gene is shown in Fig. 2. The entire segment is 1,882 bp long and contains an open reading frame beginning with an A-U-G triplet at position 34 and extending to a termination codon at nucleotide position 1,786. This open reading frame could code for a polypeptide of 584 amino acids. This leaves a stretch of 94 untranslated nucleotides following the termination codon. By analogy with influenza A and vesicular stomatitis virus genes, the poly(A) addition site is probably located in a region corresponding to the adenine-rich string of nucleotides at positions 1,862–1,866 (20). No additional open reading frames longer than 100 amino acids beginning with a methionine were detected on either the plus or minus sense strands.

Deduced Amino Acid Sequence. The only protein sequence information available on the influenza B/Lee/40 virus HA are partial sequences at the NH_2 termini of the HA1 and HA2 subunits (13). These data are compatible with the deduced amino acid sequences at positions 16–18 and 362–379, respectively (Fig. 2). Thus, our data suggest that the B virus HA precursor contains a signal peptide of 15 amino acids at its NH_2 terminus and an arginine cleavage site between the HA1 and HA2 subunits. The arrowheads in Fig. 2 indicate the postulated cleavage sites for the generation of the mature HA1 and HA2 polypeptide chains. The HA1 polypeptide would be composed of 346 amino acids with a M_r of 43,885 (not including any carbohydrate contribution). However, in the absence of COOH-terminal amino acid sequence data, the exact length of the HA1 polypeptide cannot be determined. In the six COOH-terminal amino acids, there are two lysine residues and one arginine residue, which may serve as secondary cleavage sites in a further processing of the HA1. Secondary trimming of the HA1 by a trypsin-like or carboxypeptidase B-like enzyme has been demonstrated with different HA subtypes of influenza A virus (10, 21, 22).

The NH_2 terminus of the HA2 polypeptide is the glycine res-

```
                                                                  ▼HA1
                                            MET LYS ALA ILE ILE VAL LEU LEU MET VAL VAL THR SER ASN ALA ASP ARG ILE CYS   19
AGC AGA AGC GTT GCA TTT TCT AAT ATC CAC AAA ATG AAG GCA ATA ATT GTA CTA CTC ATG GTA GTA ACA TCC AAT GCA GAT CGA ATC TGC   90

THR GLY ILE THR SER SER ASN SER PRO HIS VAL VAL LYS THR ALA THR GLN GLY GLU VAL ASN VAL THR GLY VAL ILE PRO LEU THR THR   49
ACT GGG ATA ACA TCG TCA AAC TCA CCT CAT GTG GTT AAA ACT GCC ACT CAA GGG GAA GTC AAT GTG ACT GGT GTG ATA CCA CTA ACA ACA  180

THR PRO THR LYS SER HIS PHE ALA ASN LEU LYS GLY THR GLN THR ARG GLY LYS LEU CYS PRO ASN CYS PHE ASN CYS THR ASP LEU ASP   79
ACA CCT ACC AAA TCT CAT TTT GCA AAT CTC AAA GGA ACA CAG ACC AGA GGA AAA CTA TGC CCA AAC TGT TTT AAC TGC ACA GAT CTG GAC  270

VAL ALA LEU GLY ARG PRO LYS CYS MET GLY ASN THR PRO SER ALA LYS VAL SER ILE LEU HIS GLU VAL LYS PRO ALA THR SER GLY CYS  109
GTG GCC CTA GGC AGA CCA AAA TGC ATG GGG AAC ACA CCC TCC GCA AAA GTC TCA ATA CTC CAT GAA GTC AAA CCT GCT ACA TCT GGA TGC  360

PHE PRO ILE MET HIS ASP ARG THR LYS ILE ARG GLN LEU PRO ASN LEU LEU ARG GLY TYR GLU ASN ILE ARG LEU SER THR SER ASN VAL  139
TTT CCT ATA ATG CAC GAC AGA ACA AAA ATC AGA CAA CTA CCT AAT CTT CTC AGA GGA TAT GAA AAC ATC AGG TTA TCA ACC AGT AAT GTT  450

ILE ASN THR GLU THR ALA PRO ARG GLY SER TYR LYS VAL GLY THR SER GLY SER CYS PRO ASN VAL ALA ASN GLY ASN GLY PHE PHE ASN  169
ATC AAT ACA GAG ACG GCA CCA AGA GGG TCC TAC AAG GTG GGG ACC TCA GGA TCT TGC CCT AAC GTT GCT AAT GGG AAC GGC TTC TTC AAC  540

THR MET ALA TRP VAL ILE PRO LYS ASP ASN ASN LYS THR ALA ILE ASN PRO VAL THR VAL GLU VAL PRO TYR ILE CYS SER GLU GLY GLU  199
ACA ATG GCT TGG GTT ATC CCA AAA GAC AAC AAC AAG ACA GCA ATA AAT CCA GTA ACA GTA GAA GTA CCA TAC ATT TGT TCA GAA GGG GAA  630

ASP GLN ILE THR VAL TRP GLY PHE HIS SER ASP ASP LYS THR GLN MET GLU ARG LEU TYR GLY ASP SER ASN PRO GLN LYS PHE THR SER  229
GAC CAA ATT ACT GTT TGG GGG TTC CAC TCT GAT GAC AAA ACC CAA ATG GAA AGA CTC TAT GGA GAC TCA AAT CCT CAA AAG TTC ACC TCA  720

SER ALA ASN GLY VAL THR THR HIS TYR VAL SER GLN ILE GLY GLY PHE PRO ASN GLN THR GLU ASP GLU GLY LEU LYS GLN SER GLY ARG  259
TCT GCC AAT GGA GTA ACC ACA CAT TAT GTT TCT CAG ATT GGT GGC TTC CCA AAT CAA ACA GAA GAC GAA GGG CTA AAA CAA AGC GGC AGA  810

ILE VAL VAL ASP TYR MET VAL GLN LYS PRO GLY LYS THR GLY THR ILE VAL TYR GLN ARG GLY ILE LEU LEU PRO GLN LYS VAL TRP CYS  289
ATT GTT GTT GAT TAC ATG GTA CAA AAA CCT GGA AAA ACA GGA ACA ATT GTT TAT CAA AGA GGC ATT TTA TTG CCT CAA AAA GTG TGG TGC  900

ALA SER GLY ARG SER LYS VAL ILE LYS GLY SER LEU PRO LEU ILE GLY GLU ALA ASP CYS LEU HIS GLU LYS TYR GLY GLY LEU ASN LYS  319
GCA AGT GGC AGG AGC AAG GTA ATA AAA GGG TCC TTG CCT TTA ATT GGT GAA GCA GAT TGC CTC CAC GAA AAG TAC GGT GGA TTA AAT AAA  990

SER LYS PRO TYR TYR THR GLY GLU HIS ALA LYS ALA ILE GLY ASN CYS PRO ILE TRP VAL LYS THR PRO LEU LYS LEU ALA ASN GLY THR  349
AGC AAG CCT TAC TAC ACA GGA GAG CAT GCA AAG GCC ATA GGA AAT TGC CCA ATA TGG GTG AAA ACA CCC TTG AAG CTG GCC AAT GGA ACC 1080
                                                ▼HA2
LYS TYR ARG PRO PRO ALA LYS LEU LEU LYS GLU ARG GLY PHE PHE GLY ALA ILE ALA GLY PHE LEU GLU GLY GLY TRP GLU GLY MET ILE  379
AAA TAT AGA CCG CCT GCA AAA CTA TTA AAG GAA AGA GGT TTC TTC GGA GCT ATT GCT GGT TTC TTG GAA GGA GGA TGG GAA GGA ATG ATT 1170

ALA GLY TRP HIS GLY TYR THR SER HIS GLY ALA HIS GLY VAL ALA VAL ALA ALA ASP LEU LYS SER THR GLN GLU ALA ILE ASN LYS ILE  409
GCA GGT TGG CAC GGA TAC ACA TCT CAT GGA GCA CAT GGA GTG GCA GTG GCA GCA GAC CTT AAG AGT ACA CAA GAA GCT ATA AAC AAG ATA 1260

THR LYS ASN LEU ASN TYR LEU SER GLU LEU GLU VAL LYS ASN LEU GLN ARG LEU SER GLY ALA MET ASN GLU LEU HIS ASP GLU ILE LEU  439
ACA AAA AAT CTC AAC TAT TTA AGT GAG CTA GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG AAT GAG CTT CAC GAC GAA ATA CTC 1350

GLU LEU ASP GLU LYS VAL ASP ASP LEU ARG ALA ASP THR ILE SER SER GLN ILE GLU LEU ALA VAL LEU LEU SER ASN GLU GLY ILE ILE  469
GAG CTA GAC GAA AAA GTG GAT GAT CTA AGA GCT GAT ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT TCC AAC GAA GGG ATA ATA 1440

ASN SER GLU ASP GLU HIS LEU LEU ALA LEU GLU ARG LYS LEU LYS LYS MET LEU GLY PRO SER ALA VAL GLU ILE GLY ASN GLY CYS PHE  499
AAC AGT GAA GAT GAG CAT CTC TTG GCA CTT GAA AGA AAA CTG AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATA GGG AAT GGG TGC TTT 1530

GLU THR LYS HIS LYS CYS ASN GLN THR CYS LEU ASP ARG ILE ALA ALA GLY THR PHE ASN ALA GLY ASP PHE SER LEU PRO THR PHE ASP  529
GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC AGG ATA GCT GCT GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT TTT GAT 1620

SER LEU ASN ILE THR ALA ALA SER LEU ASN ASP ASP GLY LEU ASP ASN HIS THR ILE LEU LEU TYR TYR SER THR ALA ALA SER SER LEU  559
TCA TTA AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG GAT AAT CAT ACT ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG 1710

ALA VAL THR LEU MET ILE ALA ILE PHE ILE VAL TYR MET VAL SER ARG ASP ASN VAL SER CYS SER ILE CYS LEU ***                  584
GCT GTA ACA TTA ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA GAC AAT GTT TCT TGT TCC ATC TGT CTG TGA GGG AGA TTA AGC 1800

CCT GTG TTT TCC TTT ACT GTA GTG CTC ATT TGC TTG TCA CCA TTA CAA AGA AAC GTT ATT GAA AAA TGC TCT TGT TAC TAC T          1882
```

FIG. 2. Complete nucleotide and deduced amino acid sequences of the influenza B/Lee/40 virus HA gene. The known cleavage sites at the NH_2 termini of the HA1 and HA2 subunits are indicated by arrowheads.

idue at amino acid position 362 (13). The HA2 would extend for 223 amino acids until the termination codon, resulting in a calculated M_r of 28,160 for the apoprotein. This number, and the values obtained for the HA1 subunit and the uncleaved HA, are in good agreement with published molecular weights determined by polyacrylamide gel electrophoresis (23–25).

Comparison of Type A and Type B HA Genes. The B/Lee/40 HA gene is ≈100 bases longer than are influenza A virus HA genes. The greater length of the influenza B virus HA gene was previously suggested by polyacrylamide gel electrophoretic examination of glyoxalated RNAs (26, 27). The deduced B/Lee/40 protein subunits are comparable in size to those of the corresponding influenza A viruses (Table 1). This suggests a possible overall structural similarity between type A and type B HAs.

This notion is further supported by a comparison of the relative hydrophilicity values of domains of the HAs from B/Lee/40 virus and various type A strains. Computer plots of relative hydrophilicity of polypeptides have proved useful in comparing the relatedness of structural properties of different proteins as well as in identifying possible antigenic sites in such proteins (28, 29). Fig. 3 shows the hydrophilicity plots of the B/Lee/40 HA and those of the type A virus HAs belonging to the H1, H2, and H3 subtypes. There are at least three conserved hydrophobic regions in each of the HA molecules. These regions correspond to areas of the signal peptide and the NH_2-terminal and COOH-terminal domains of the HA2 subunit. The domains are indicated in Fig. 3 and appear as areas with negative hydrophilicity values. In addition, in all four HAs, the hydrophobic COOH terminus of HA2 is preceded by a hydrophilic region of ≈100 amino acids. These data again imply a structural conservation of the influenza A and influenza B virus HAs.

A more precise comparison of A and B virus HAs was done by computer analysis of the specific nucleotide and amino acid sequences. The computer-generated graphic plot comparing nucleotide sequences of the B/Lee/40 and the A/PR/8/34 HAs is shown in Fig. 4. In this type of analysis, comparison of identical sequences will generate a continuous diagonal as well as a background of scattered dots caused by random homologies between the molecules. Therefore, when the B/Lee/40 and A/PR/8/34 sequences are compared, the extent of a diagonal line(s) is a measure of the homology between the two RNAs. A low degree of nucleotide homology is observed between the HA1 portions of the two genes, but several regions in the HA2 domain show significant nucleotide sequence homologies. The homology line observed in the HA2 domain does not cross the origin after extrapolation. This is due to a length difference of ≈60 bases within the HA1 coding region of the two genes.

A similar analysis comparing the amino acid sequences of the B/Lee/40 and A/PR/8/34 HAs is also shown in Fig. 4. The computer-generated plot shows that there are regions of homology between the HA2 domains of the two polypeptides.

Table 1. Comparison of human type A and type B influenza virus HA genes

Virus	Total length, nucleotides	HA1, amino acids	HA2, amino acids
H1*	1,778	326	222
H2*	1,773	324	222
H3*	1,765	328	221
B/Lee/40	1,882	346†	223†

* H1, A/PR/8/34; H2, A/Jap/305/57; H3, A/Aichi/2/68. H1, H2, and H3 are from refs. 5, 7, and 8, respectively.
† Deduced length of subunits without taking into account any secondary processing.

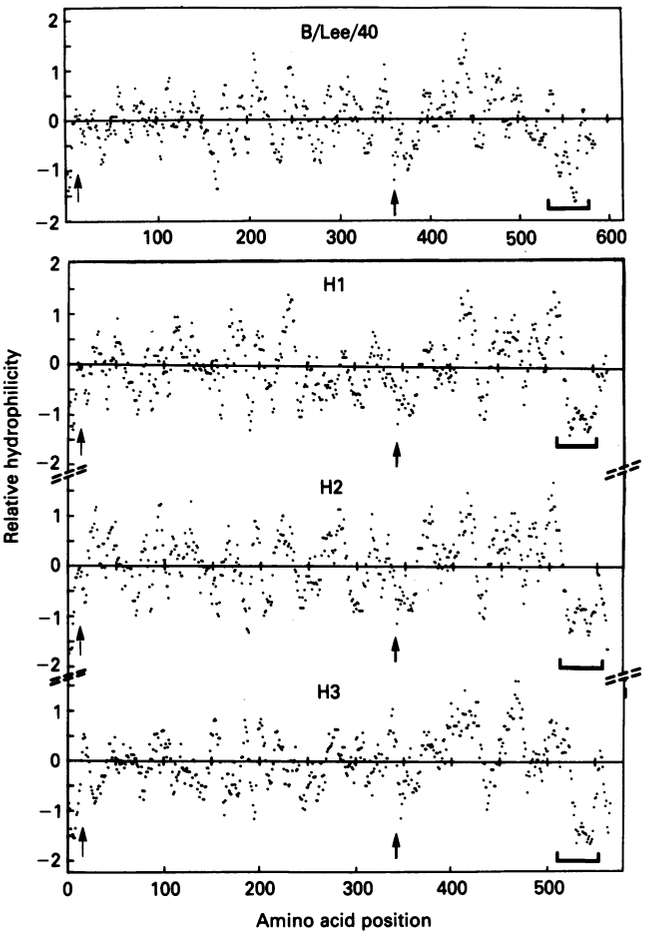

FIG. 3. Relative hydrophilicity plots of influenza A and B virus HAs. The plots are generated as described. B/Lee/40, B/Lee/40 virus HA; H1, A/PR/8/34 virus HA (5); H2, A/Jap/305/57 virus HA (7); H3, A/Aichi/2/68 virus HA (8). In every plot the arrows around amino acid positions 15–17 indicate the cleavage site of the signal peptides and the arrows at positions 340–365 designate the HA1/HA2 cleavage sites. Brackets near the end of each plot correspond to the hydrophobic membrane-bound portions of the HA2.

However, in this analysis, little homology between the HA1 polypeptides can be detected at this level of stringency.

Previously, the deduced amino acids from HAs of H1, H2, H3, and H7 subtypes were compared and a number of amino acids was found to be strictly conserved (5, 6). These residues are presumably important to structural and functional requirements of the HA. The preserved amino acids include most of the cysteine and some of the proline residues in the molecules. When these conserved amino acid residues are used as reference points, there is little difficulty in aligning the deduced amino acid sequences of the HA2 domains of the B/Lee/40 and A/PR/8/34 HAs. In addition, small insertions in either of the two sequences also allow for alignment of the HA1 regions (Fig. 5). (It should be noted that the alignment shown in Fig. 5 represents only one of several possible alignments of the two sequences.) In this way, 13 of the 15 cysteine residues found in the mature A/PR/8/34 HA can be matched with cysteine residues in the B/Lee/40 polypeptide and an overall amino acid conservation of 24% and 39% can be observed in the HA1 and HA2 subunits, respectively. Similar, but slightly lower, values were obtained when the B/Lee/40 HA was compared with those of the two other human influenza A subtype HAs (H2, A/Jap/305/57, and H3, A/Aichi/2/68) (data not shown).

The primary sequence homology between the B/Lee/40 and

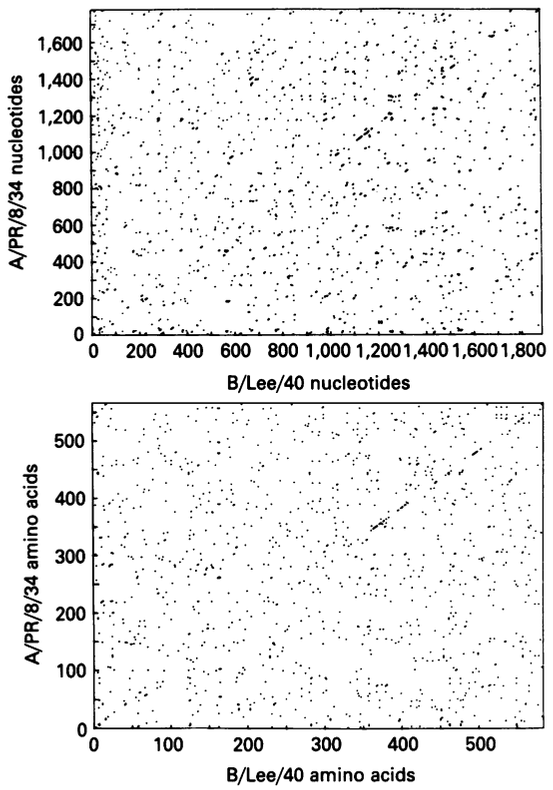

FIG. 4. Computer-generated graphic comparison of the nucleotide and deduced amino acid sequences of the influenza B/Lee/40 and A/PR/8/34 virus HA genes. The sequence of the A/PR/8/34 gene is from ref. 5. The parameters used for these matrix plots are as described.

the influenza A virus HAs is particularly striking because homologies of comparable values have been observed among different A type virus HAs. For example, in the H1 and H3 HAs only 35% and 53% of the amino acids are conserved in the HA1 and HA2 subunits, respectively (Table 2). Similar relationships are also observed at the nucleotide level (Table 2). It should also be noted that the M genes of influenza A and B viruses show a similar degree of homology. It was demonstrated that 25% of the amino acids are common in the M1 polypeptides of B/Lee/40 and A/Udorn/72 strains (30), confirming an evolutionary relationship of influenza A and B viruses.

Because HAs are glycoproteins, the sequence of the B/Lee/40 HA was analyzed for possible glycosylation sites. Ten potential sites (Asn-X-Thr or Asn-X-Ser) were found in the polypeptide sequence with 7 of these in the HA1 portion of the molecule. Only 3 of these 10 sites correspond to the 7 potential glycosylation sites found in the A/PR/8/34 virus HA sequence (Fig. 5). This is not surprising because the potential glycosylation sites are not even stringently conserved in the A virus HAs (6, 7).

To date, very few studies have shed light on the overall structural similarities between type A virus and type B virus HAs. RNA·RNA hybridization data suggested some homology between influenza A and B virus genes, although the distribution and the precise nature of this homology could not be assessed (31, 32). Similarly, partial protein sequence analysis of the NH_2 termini of the HA1 and HA2 subunits demonstrated conservation in this region of influenza A and B virus HAs (13). However, serological techniques failed to detect any significant crossreactivity between influenza A and B viruses (33), and comparisons of the RNAs of the two virus types by cDNA·RNA hybridization under stringent conditions showed no significant sequence homologies (34). In the present study it has been possible to compare a complete influenza B virus HA gene with the corresponding type A virus genes. When the predicted polypeptides are compared, many of the amino acids conserved among the influenza A virus HAs are also observed to be present in the B/Lee virus HA. In addition, many structural features—including a hydrophobic signal peptide, the HA1/HA2 cleavage site, and hydrophobic regions in the HA2 sub-

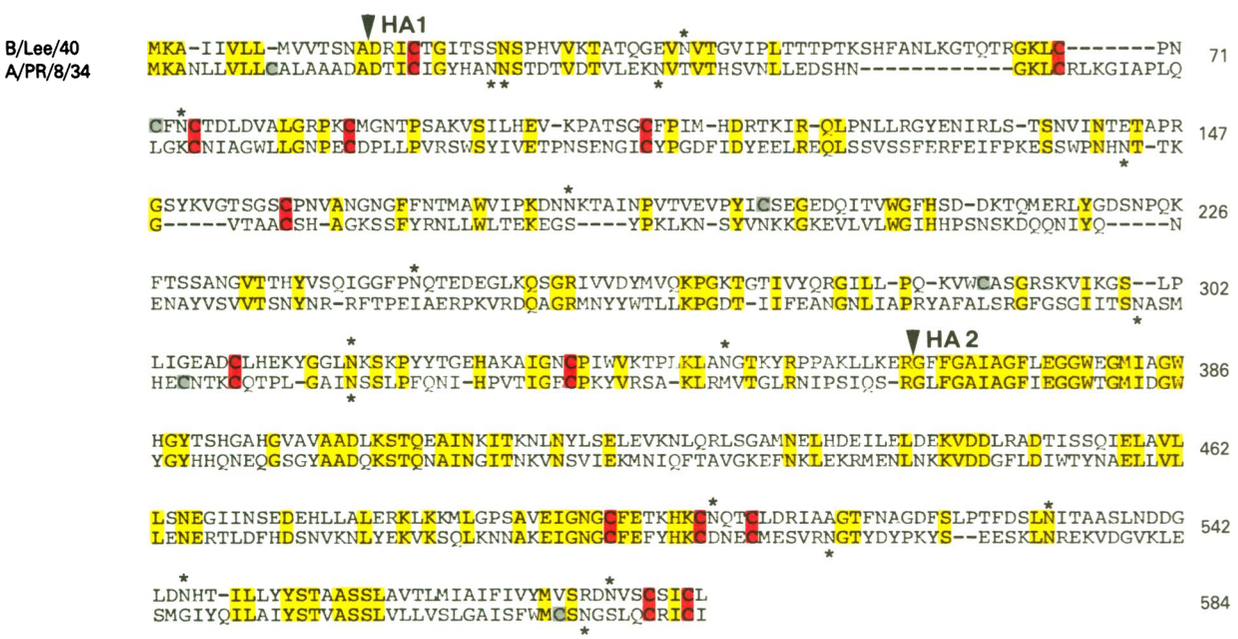

FIG. 5. Comparison of predicted amino acid sequences of the B/Lee/40 and A/PR/8/34 virus HAs. Alignment of the two sequences was accomplished by using (as reference points) amino acid residues that are conserved among influenza A virus HAs (see text). The aligned cysteine residues in the two sequences are colored red. All other aligned and paired residues are colored yellow. The unpaired cysteines are colored gray. The asterisks above the B/Lee/40 virus HA and below the A/PR/8/34 virus HA indicate potential glycosylation sites. The insertions that have been made to allow alignment of the two sequences are indicated by bars. Arrowheads indicate cleavage sites in the HA precursor polypeptide.

Table 2. Sequence homologies of HA genes from different influenza viruses

Comparison*	Nucleotide conservation, %†		Amino acid conservation, %†	
	HA1	HA2	HA1	HA2
H1 vs. H2	61	72	58	79
H1 vs. H3	45	58	35	53
H2 vs. H3	45	57	36	50
B vs. H1	36	48	24	39

* H1, A/PR/8/34; H2, A/Jap/305/57; H3, A/Aichi/2/68; and B, B/Lee/40. H1, H2, and H3 are from refs. 5, 7, and 8, respectively.

† Homologies were calculated as numbers of homologous residues/total number of residues. The total number of residues is the average of the length of the two sequences. The signal peptide sequences are not used for this comparison.

unit—are conserved in the B virus HA. In fact, the relationship of the B/Lee/40 HA to the A/PR/8/34 HA is comparable to that found among different influenza A virus subtype HAs (5–10, 35).

We are grateful to Dr. Robert P. Aaronson for helping us with the computer analysis of the data and we thank Ronald Taussig for many of the sequence analysis experiments and R. M. Medford for assistance in writing the computer programs. This work was supported in part by grants from the Reye's Syndrome Foundation, the National Science Foundation (PCM 07844), and the National Institutes of Health (AI 11823, NCI P30-CA 13330). J.F.Y. is a recipient of a Sinsheimer Scholar Award and P.P. is a recipient of an I. T. Hirschl Career Research Award. M.K. is the recipient of an American Lung Association postdoctoral fellowship.

1. Palese, P. & Young, J. F. (1982) *Science* **215**, 1468–1474.
2. Webster, R. G., Laver, W. G., Air, G. M. & Schild, G. C. (1982) *Nature (London)* **296**, 115–121.
3. Kilbourne, E. D. (1975) *The Influenza Viruses and Influenza* (Academic, New York).
4. W.H.O. Memorandum (1980) *Bull. World Health Organization* **58**, 585–591.
5. Winter, G., Fields, S. & Brownlee, G. G. (1981) *Nature (London)* **292**, 72–75.
6. Hiti, A. L., Davis, A. R. & Nayak, D. P. (1981) *Virology* **111**, 113–124.
7. Gething, M. J., Bye, J. & Waterfield, M. (1980) *Nature (London)* **287**, 301–306.
8. Verhoeyen, M., Fang, R., Min Jou, W., Devos, R., Huylebroeck, D., Saman, E. & Fiers, W. (1980) *Nature (London)* **286**, 771–776.
9. Fang, R., Min Jou, W., Huylebroeck, D., Devos, R. & Fiers, W. (1981) *Cell* **25**, 315–323.
10. Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. & Emtage, J. S. (1978) *Nature (London)* **282**, 471–477.
11. Ritchey, M. B., Palese, P. & Kilbourne, E. D. (1976) *J. Virol.* **18**, 738–744.
12. Baez, M., Taussig, R., Zazra, J. J., Young, J. F. & Palese, P. (1980) *Nucleic Acids Res.* **8**, 5845–5858.
13. Waterfield, M. D., Espelie, K., Elder, K. & Skehel, J. J. (1979) *Br. Med. Bull.* **35**, 57–63.
14. Desselberger, U., Racaniello, V. R., Zazra, J. J. & Palese, P. (1980) *Gene* **8**, 315–328.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–559.
16. Staden, R. (1977) *Nucleic Acids Res.* **4**, 4037–4051.
17. Staden, R. (1978) *Nucleic Acids Res.* **5**, 1013–1015.
18. Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610.
19. Levitt, M. (1976) *J. Mol. Biol.* **104**, 59–107.
20. Robertson, J. S., Schubert, M. & Lazzarini, R. A. (1981) *J. Virol.* **38**, 157–163.
21. Garten, W., Bosch, F. X., Linder, D., Rott, R. & Klenk, H.-D. (1981) *Virology* **115**, 361–374.
22. Bosch, F. X., Garten, W., Klenk, H.-D. & Rott, R. (1981) *Virology* **113**, 725–735.
23. Oxford, J. S. (1975) *J. Virol.* **12**, 827–835.
24. Tobita, K. & Kilbourne, E. D. (1975) *Arch. Virol.* **47**, 367–374.
25. Racaniello, V. R. & Palese, P. (1979) *J. Virol.* **29**, 361–373.
26. Desselberger, U. & Palese, P. (1978) *Virology* **88**, 394–399.
27. Palese, P., Racaniello, V. R., Desselberger, U., Young, J. F. & Baez, M. (1980) *Philos. Trans. R. Soc. London, Ser. B* **288**, 299–305.
28. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.
29. Both, G. W. & Sleigh, M. J. (1980) *Nucleic Acids Res.* **8**, 2561–2575.
30. Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1982) *Virology* **116**, 581–588.
31. Scholtissek, C. & Rott, R. (1969) *Virology* **39**, 400–407.
32. Scholtissek, C., Rohde, W. & Harms, E. (1977) *J. Gen. Virol.* **37**, 243–247.
33. Schild, G. C. & Dowdle, W. R. (1975) in *The Influenza Viruses and Influenza*, ed. Kilbourne, E. D. (Academic, New York), pp. 315–372.
34. Palese, P., Elliott, R. M., Baez, M., Zazra, J. J. & Young, J. F. (1981) in *Genetic Variation Among Influenza Viruses ICN-UCLA Symposia on Molecular and Cellular Biology*, ed. Nayak, D. P. (Academic, New York), Vol. 21, pp. 127–140.
35. Air, G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7639–7643.