Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4527–4531, July 1983
Microbiology

Sequential mutations in hemagglutinins of influenza B virus isolates: Definition of antigenic domains

(antigenic variation/three-dimensional structure/evolution)

MARK KRYSTAL*, JAMES F. YOUNG*, PETER PALESE*, IAN A. WILSON†, JOHN J. SKEHEL‡, AND DON C. WILEY§

*Mount Sinai School of Medicine of the City University of New York, One Gustave L. Levy Place, New York, New York 10029; †Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037; ‡National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom; and §Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138

Communicated by Edwin D. Kilbourne, March 17, 1983

ABSTRACT **Comparative analysis of the amino acid sequences of hemagglutinins (HAs) of influenza B/Lee/40, B/Md/59, and B/HK/73 viruses has allowed examination of the molecular basis of antigenic variation in type B viruses. As seen with influenza type A viruses, antigenic drift in influenza B viruses proceeds mostly through the accumulation of amino acid substitutions within the HA1 portion of the HA molecule. However, the rate of variation observed among the influenza B virus HAs appears to be significantly lower than the observed rate of variation among influenza A virus HAs. The overall rate of amino acid change in the HA1s of the influenza B viruses studied is 2% per 10 years, whereas the HA1s of H3 influenza A viruses vary by 9.2% per 10 years. The sequences of the influenza B HAs were also examined in relation to the three-dimensional model for the A/Aichi/2/68 HA. When the primary amino acid sequences are compared, it appears that most of the important structural features of the type A HAs—such as the sialic acid binding site, the disulfide linkages, and the stem structure of the trimer—are conserved in the influenza B virus HAs. Regions are also identified where extensive amino acid substitutions have occurred among the three antigenically distinct influenza B virus HAs. The locations of these areas in the B HA structure correspond to antigenic regions proposed for the A virus HAs. In addition, modulation of antigenic regions in B virus HAs may also occur through amino acid deletions and variation in glycosylation sites.**

The influenza viruses have the unique capacity to undergo a high degree of antigenic variation within a short period of time. It is this property of the virus that has made it difficult to control the seasonal outbreaks of influenza throughout the human and animal populations (1–3). Through serologic and sequencing studies, two kinds of antigenic variation have been demonstrated in influenza A viruses. Antigenic shift occurs when the hemagglutinin (HA) is replaced in a new viral strain with an antigenically novel HA. This occurrence of new subtypes (H1, H2, H3) usually results in pandemics as illustrated in 1957 when H1 viruses were replaced by H2 viruses and in 1968 when H2 viruses were followed by H3 viruses. Antigenic drift occurs in influenza A viruses of a single subtype. Amino acid (4, 5) and nucleotide sequence analyses (6–11) suggest that antigenic drift occurs through a series of sequential mutations, resulting in amino acid changes in the polypeptide and differences in the antigenicity of the virus. Antigenic drift has also been observed to occur in influenza B viruses (12, 13), whereas antigenic shift has thus far been associated only with the influenza A viruses. Structural similarities of the HAs of influenza A and B viruses have been emphasized in a previous report (14), and comparison of sequences of NS (nonstructural), M (matrix), and NA (neuraminidase) genes has identified similarities in the structure of influenza A and B viral polypeptides (15).

In this report we examine the HA gene sequences from three influenza B field isolates, and we show that the rate of antigenic drift among influenza B viruses appears to be lower than that among influenza A viruses. In addition, similarities between type A and type B HAs allow us to propose possible antigenic sites on the influenza B HA molecule.

MATERIALS AND METHODS

Viruses. Influenza viruses B/Lee/40, B/Md/59, and B/HK/8/73 were grown in embryonated hens' eggs. Virus purification and RNA extractions were as described (14).

Priming and Sequencing of Ha Viral RNA. DNA fragments specific for the priming of the HA gene were isolated by using the full-length cDNA clone of the B/Lee/40 HA gene described previously (14). The various restriction fragments used as primers for the sequence analysis of the viral RNA HA segments encompassed bases 0–49, 105–261, 262–410, 411–495, 605–683, 600–752, 829–972, 973–1064, 1128–1235, 1235–1312, 1406–1495, 1496–1572, and 1573–1705.

Fifty to 200 ng of each primer was used for dideoxy sequencing of the influenza vRNAs (11, 16, 17). Because of ambiguities occasionally seen in the sequence when this procedure is used (11, 17), reactions were simultaneously carried out with B/Lee/40, B/Md/59, and B/HK/8/73 virus RNAs. Direct comparison of the gel patterns for the three RNAs facilitated the identification of nucleotide changes. The B/Lee/40 sequence was previously determined by analyzing the cloned DNA (14), and only three nucleotide differences were detected by using the two methods. These differences appeared at nucleotide positions 472, 477, and 478.

The nucleotide and amino acid sequences were stored, edited, and compared as described (14, 18–20).

RESULTS

Nucleotide Sequences. The nucleotide sequences of the RNA segments coding for the HA molecule of B/Lee/40, B/Md/59, and B/HK/8/73 viruses were determined by the dideoxy method, using restriction fragments derived from a cDNA clone of the B/Lee/40 HA as primers. A comparison of these HA sequences is shown in Fig. 1. Although these field isolates were obtained over a period of 33 years, their HA genes are closely related. The overall nucleotide change between the HA gene of the 1940 and the 1973 isolates is only 6.5%. Relative to B/Lee/40, 69 base changes are shared between the B/Md/59 and B/HK/8/73 genes. Only 18 changes are unique to the B/

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: HA, hemagglutinin.

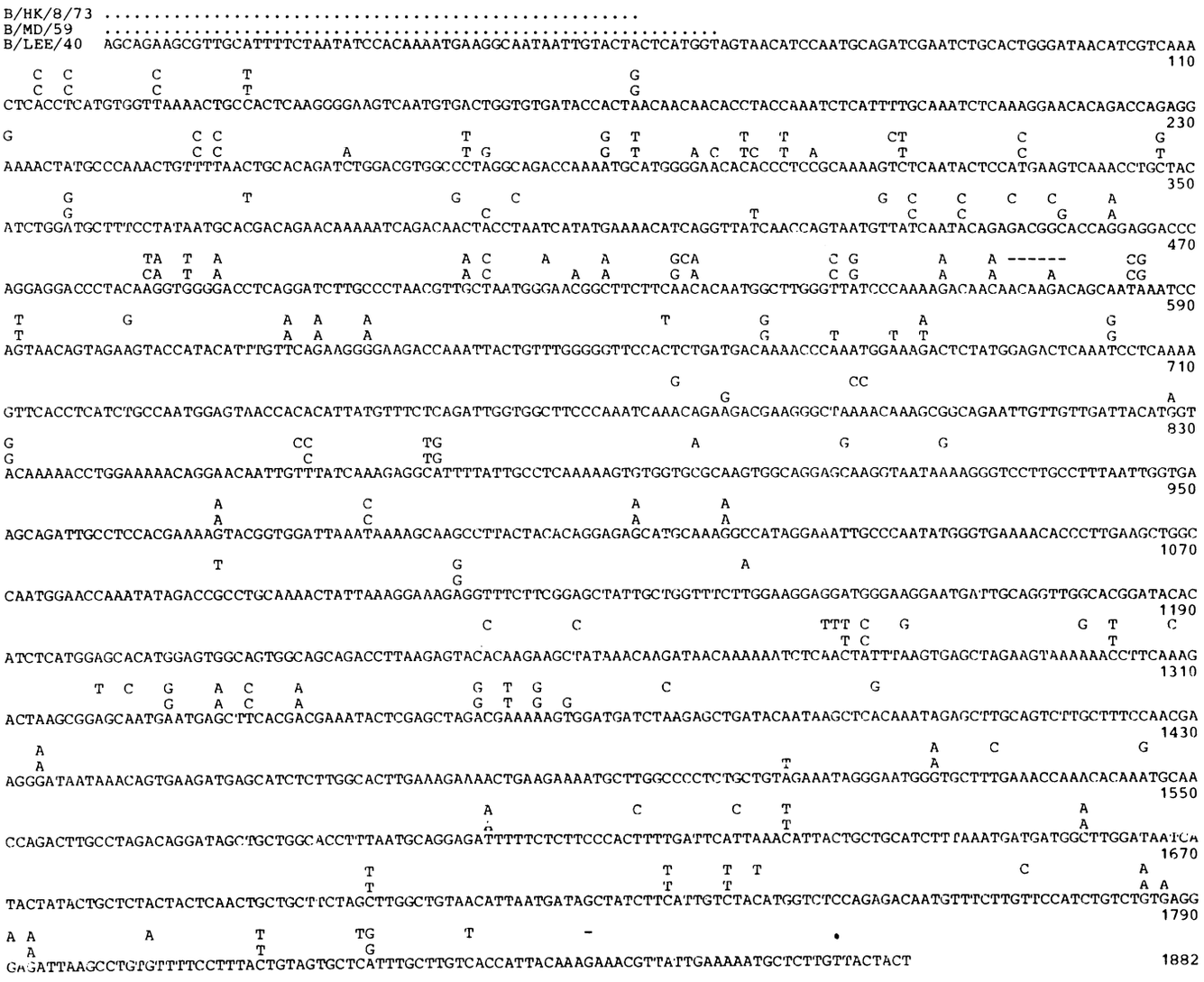


FIG. 1. Comparison of the nucleotide sequences from the influenza B/Lee/40, B/Md/59, and B/HK/8/73 virus HA genes. In the B/Md/59 and B/HK/8/73 sequences, only nucleotide changes relative to the B/Lee/40 sequence are indicated. The nucleotide sequences are complete except for 62 and 54 bases at the 5′ ends of the B/Md/59 and B/HK/8/73 genes, respectively (denoted by dots). Nucleotide deletions in the B/HK/8/73 gene are indicated by dashes.

Md/59 HA gene, whereas 48 base changes are seen in the B/HK/8/73 gene (Fig. 2). Thus, the sequential nature of the majority of these base changes is similar to that seen with influenza A virus HA (9, 11, 21) and NS genes (22).

Examination of the Influenza B HA Sequence in Relation to the Three-Dimensional Structure of the A/Aichi/68 HA. The three-dimensional structure of the crystallized HA from the A/Aichi/68 (H3N2) influenza virus provides an atomic model for interpreting results from HA sequence analyses (23, 24). Amino acid differences in antigenically distinct HAs from field and laboratory variants of H3 viruses have permitted the location of potential antigenic sites in this protein (regions A–E; ref. 24). Subsequent sequence studies of A/PR/8/34 (H1N1) antigenic variants have also demonstrated a close relationship between the HA structures of H1 and H3 viruses (25).

Because a close evolutionary relationship between influenza A and influenza B HAs is suggested through sequence comparison, we have examined the B virus HA sequences in relation to the three-dimensional structure of the A/Aichi/2/68 HA. The primary structures of the type A and type B HAs were aligned by using previously established criteria (14).

Most of the amino acids conserved among the HAs of different influenza A subtype viruses were also found to be conserved in the three B HA sequences, thus providing a struc-

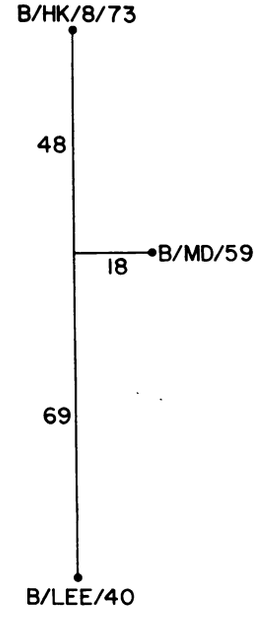


FIG. 2. Evolutionary tree of influenza B virus HAs based on nucleotide differences. Nucleotide positions 347, 485, 576, and 683 (positions where all three genes differ) were omitted from this analysis.

Correction. In the article "Sequential mutations in hemagglutinins of influenza B virus isolates: Definition of antigenic domains" by Mark Krystal, James F. Young, Peter Palese, Ian A. Wilson, John J. Skehel, and Don C. Wiley, which appeared in number 14, July 1983, of *Proc. Natl. Acad. Sci. USA* (**80**, 4527–4531), there were several computer errors in Fig. 1. The correct figure is shown below. This figure also contains nucleotide changes in the B/Md/59 sequence (position 302) and in the B/HK/8/73 sequence (positions 319, 457, and 672), which were detected subsequent to submission of the paper.

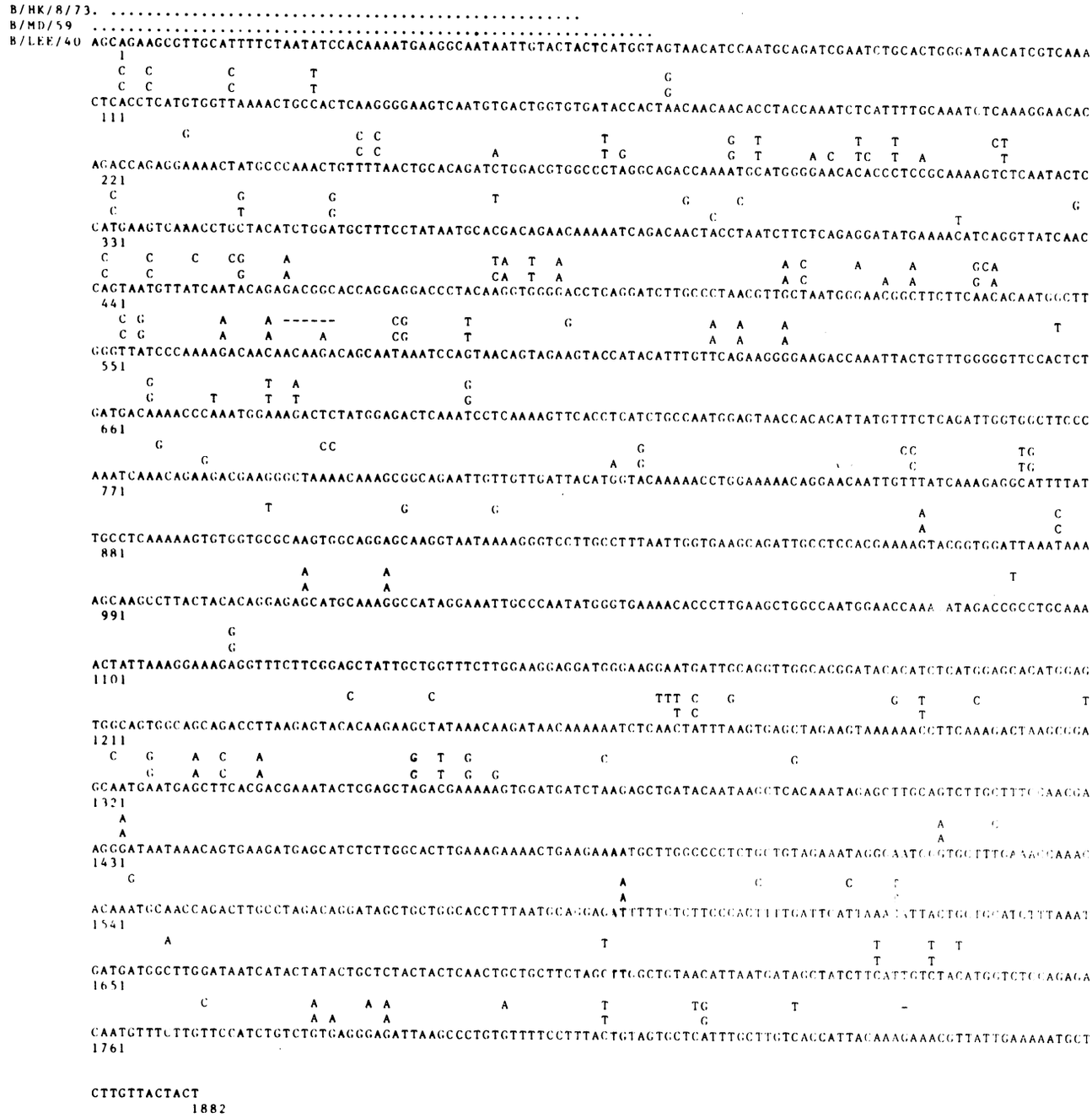


FIG. 1. Comparison of the nucleotide sequences from the influenza B/Lee/40, B/Md/59, and B/HK/8/73 virus HA genes. In the B/Md/59 and B/HK/8/73 sequences, only nucleotide changes relative to the B/Lee/40 sequence are indicated. The nucleotide sequences are complete except for 62 and 54 bases at the 5′ ends of the B/Md/59 and B/HK/8/73 genes, respectively (denoted by dots). Nucleotide deletions in the B/HK/8/73 gene are indicated by dashes.

tural framework for the B HA molecules. The results of this comparative analysis indicate that: (*i*) Five of the six disulfide bonds in the A/Aichi/2/68 HA were conserved in the B HAs (14-137HA2, 64-76, 97-139, 281-305, and 144HA2-148HA2). Note that the amino acids of the B HAs were numbered in accordance with the A/Aichi/2/68 amino acid sequence (ref. 10; Fig. 3). Computer-assisted model building also suggested that the 52-277 disulfide bond in the HA of the H3 subtype was probably replaced by a 52–61 disulfide bond in the B HAs. Due to a proposed deletion of amino acids 53–58, the B sequence in this region (residues 52–61) is Cys-Pro-Asn-Cys (Fig. 3). Two novel cysteines are located at positions 170 and 259 in the B HAs. These amino acid positions are opposite each other on two extended strands of β structure in the HA1 of the A/Aichi/2/68 HA. Therefore an extra interchain disulfide bond could link cysteines 170 and 259 in the B HAs. (*ii*) Another structurally conserved region is the proposed sialic acid binding site in the HA1 subunit of A and B HAs. In the A/Aichi/2/68 HA, this sialic acid binding site is supported by the Tyr-98, Trp-153, His-183, Tyr-195, Leu-194, Phe-147, Ile-252, Glu-190, Pro-99, Arg-229, and Gly-249 residues (23). These residues are conserved in the B HAs except for minor changes; the Tyr-98 and Glu-190 residues are changed to Phe and Gln/Leu, respectively (Fig. 3). (*iii*) The alternating hydrophobic residues along the long helix forming the center of the trimer's α-helical coiled coil in the A/Aichi/2/68 HA are preserved in the B HAs (23, 26). This region was shown to form the stalk of the HA molecule (23). (*iv*) Specific amino acids defining loops and helices that can be recognized by antibodies in the H3 structure are also found in the B HA sequences. For example, conserved Cys-139, Phe-147, and Phe-148 define an external loop corresponding to antigenic region A of the H3 HA; similarly Trp-153, Pro-162, Val/Leu-163, Leu-194, and Tyr-195 are conserved between A and B HAs and define a surface loop and helix (antigenic region B). All of these results suggest a close structural relationship between type A and type B HAs.

Antigenic Variation in the Influenza B HA. Fig. 4 *Left* and *Right* shows the locations of the amino acid differences between the HAs of B/Lee/40 and B/Md/59 and between B/Md/59 and B/HK/8/73 viruses, respectively. Some or all of these changes must be responsible for the different antigenic behavior exhibited by the viruses. In the study of antigenic variation in the H1 and H3 subtypes of influenza, antigenic variants selected by monoclonal antibodies have aided in defining the immunogenic areas of the HA (23, 25, 27). Because the sequences of antibody-selected influenza B virus variants are not available, the analysis here is based on the variations observed in the HA sequences of antigenically distinct field isolates of influenza B virus.

It has been shown that the HAs of the three B viruses used

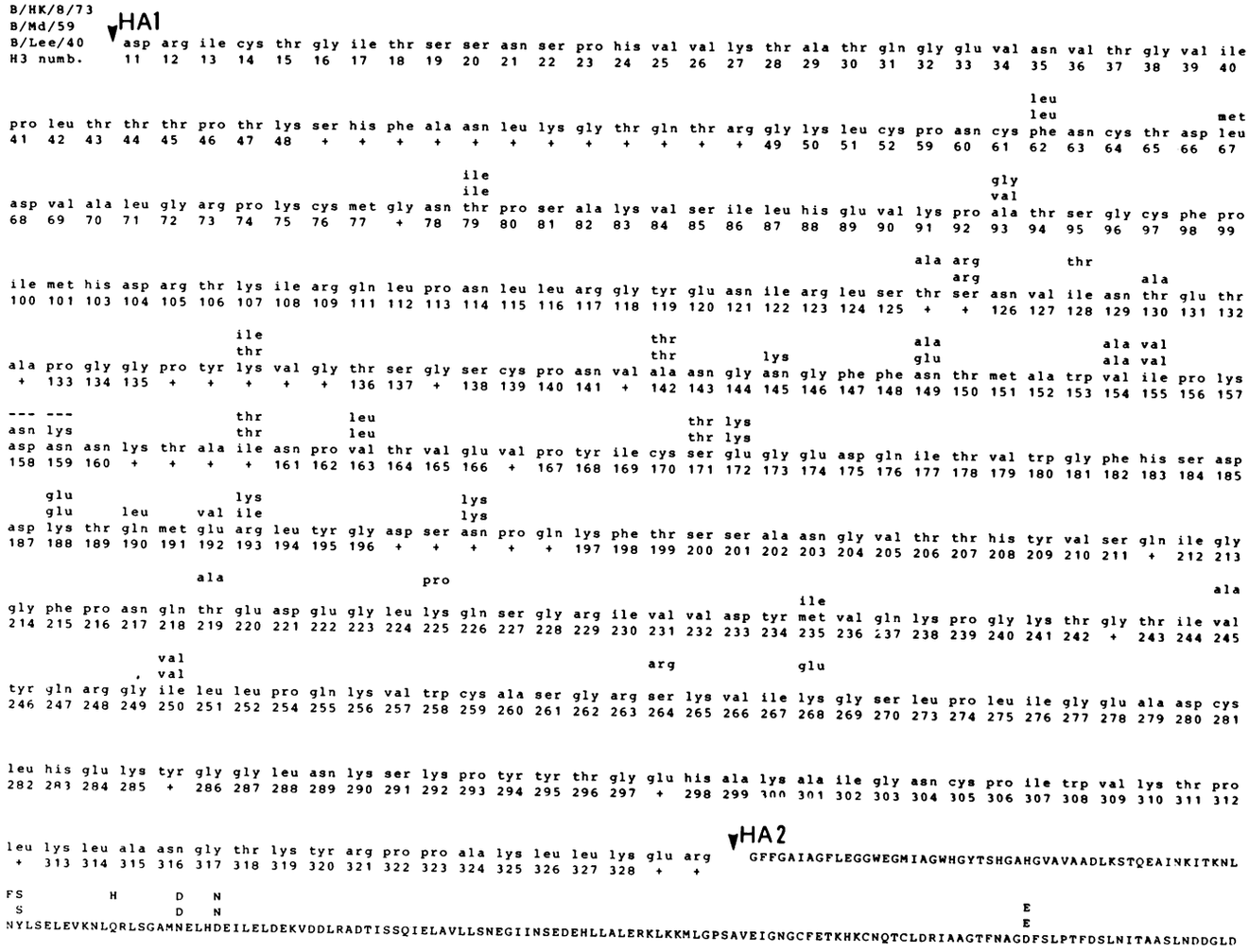


FIG. 3. Deduced amino acid sequences of the B/Lee/40, B/Md/59, and B/HK/8/73 virus HAs. Only changes in the B/Md/59 and B/HK/8/73 protein sequence relative to the B/Lee/40 sequence are indicated. The three-letter amino acid code is used for the HA1 polypeptide and the single-letter amino acid code is used for the HA2 subunit. The two-amino-acid deletion in B/HK/8/73 is indicated (---). As discussed in the text, the B virus sequences were aligned with the sequence of the A/Aichi/2/68 HA. The numbers under the B/Lee/40 HA1 reflect this alignment and correspond to the equivalent amino acids in the A/Aichi/2/68 HA model (24). Amino acids labeled + do not have corresponding residues in the A HA and reflect insertions in the B HA structure relative to the A/Aichi/2/68 HA model.

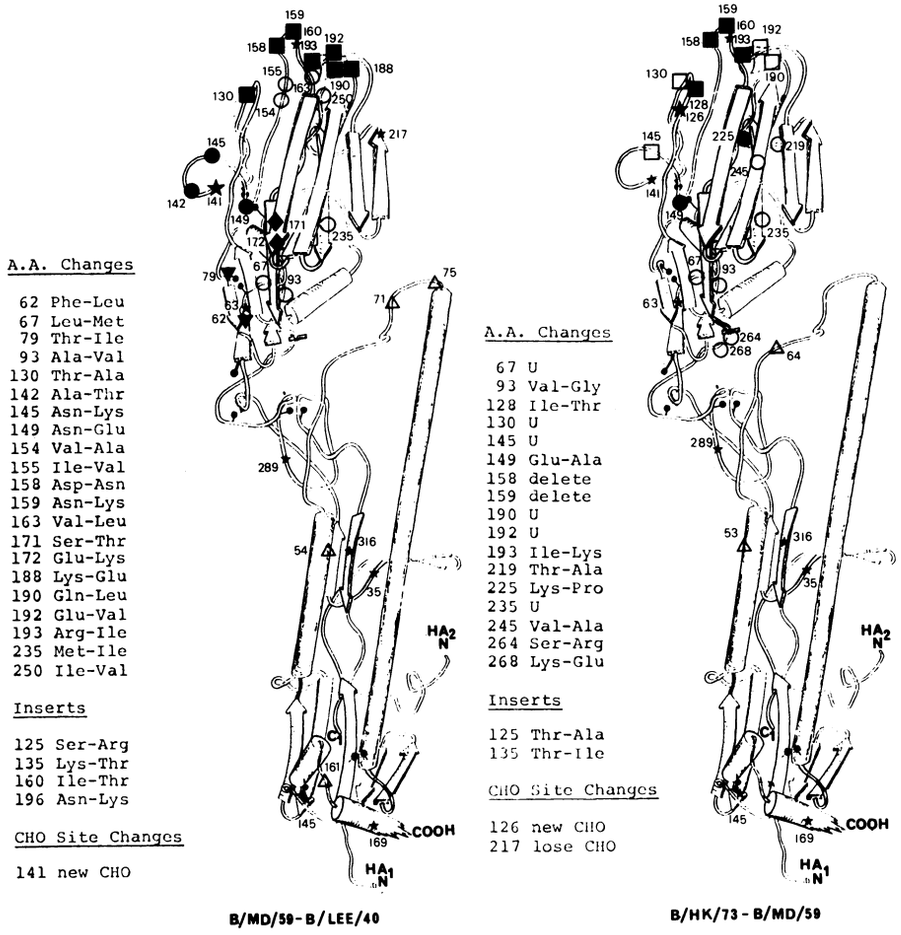


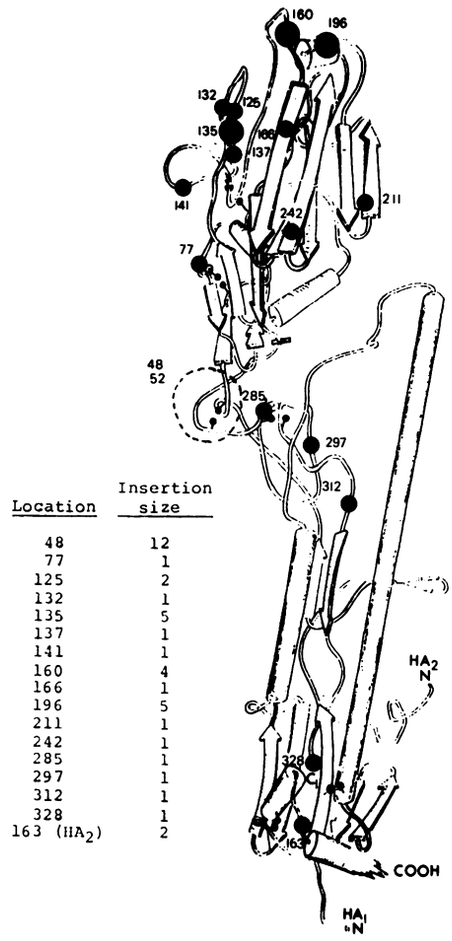
FIG. 4. Antigenic variation in influenza B virus HAs. The location of changed amino acids (A.A.) and the introduction of potential carbohydrate (CHO) attachment sites in the strain B/Md/59 relative to B/Lee/40 (*Left*) and in B/HK/73 relative to B/Md/59 (*Right*) are illustrated on schematic drawings of the A/Aichi/2/68 HA structure. Symbols denote the proposed antigenic region in the H3 subtype in which the B strain substitutions are found. ●, Region A (H3); ■, region B (H3); ◆, region D (H3); ▼, region E (H3); △, HA2 chain; ○, conservative or buried substitutions; ⋆, carbohydrate; ★, newly introduced carbohydrate site. A large number of amino acid changes (some of which must be responsible for the antigenic variation in the B strains), including the introduction of carbohydrate sites, are found in regions of the B HAs corresponding to areas of the A HA structure proposed to be important for antibody binding. In *Right*, U indicates that the amino acid found in that position in B/HK/73 is the same as that in B/Lee/40. The position is indicated in the figure by □.

in the analysis differ in their serological properties (12, 13, 28). Many nonconservative amino acid substitutions among the three HAs occur in regions of the molecules previously described as antigenically important in the HAs of H1 and H3 viruses (refs. 23 and 25; Fig. 4). (*i*) Substitutions at positions 145 and 225, a substitution in the insertion at position 135, and an acquired carbohydrate attachment site in the B/Md/59 and B/HK/8/73 HA at position 141 are all in the same location as antigenic region A of H3 viruses (designated antigenic site Ca2 in the H1 subtype). (*ii*) Changes at positions 128, 130, 158, 159, 188, 190, 192, and 193, mutations in the insertions after positions 125 and 160, and a novel carbohydrate site at position 126 in the B/HK/8/73 HA are all found in the region corresponding to antigenic region B of H3 viruses (designated antigenic site Sb in the H1 subtype). (*iii*) Substitutions at positions 171, 172, and 219 and the loss of a carbohydrate attachment site at position 217 in B/HK/8/73 fall in antigenic region D of the H3 HA (designated Ca1 in the H1 subtype). (*iv*) A change in position 79 and a conservative Phe/Leu change at position 62 fall in antigenic region E of the H3 HA (antigenic site Cb in the H1 subtype).

Thus, changes in the B HAs have occurred in all of the proposed antigenic regions except in the area corresponding to region C of the H3 subtype (23). This region of the influenza B HAs may be significantly altered with respect to the H3 HAs as a result of a 12-amino-acid insertion and a 6-amino-acid deletion (Fig. 5). Consequently, this area of influenza B HAs may not represent an antigenically important site. Note that the corresponding region in the H1 subtype is also not a proposed antigenic site, presumably due to blockage by a carbohydrate moiety (25). Structural analysis of the B HA indicates that in many regions insertions of amino acids have occurred relative to the A/Aichi/2/68 model (Fig. 5). The five large insertions (at positions 48, 125, 135, 160, and 196) and a number of small insertions (at positions 77, 132, 137, 141, 166, 211, and 242) all occur in areas to which antibodies are thought to bind in influenza A viruses (Fig. 5). The accommodation of extra residues in these areas demonstrates the plastic nature of these regions, which can readily accept amino acid substitutions and insertions. Also, within the B HAs analyzed, a deletion of two amino acids has occurred at positions 158 and 159 of the B/HK/8/73 HA1 subunit. This corresponds to antigenic region B of the H3 subtype (site Sb in the H1 subtype). It is interesting to note that the B/Md/59 nucleotide sequence contains a tandem repeat of A-C-A-A-A-A at nucleotides 566–577 and it is one of these six-nucleotide blocks that is deleted from the B/HK/8/73 sequence. The formation of this deletion might have been due to slippage of the RNA polymerase on the template during transcription.

DISCUSSION

Sequence comparison of three influenza B virus HAs has permitted an examination of the mechanism of antigenic drift in influenza B virus. As with influenza A viruses, antigenic drift in influenza B virus HAs appears to be occurring mostly through a series of single amino acid changes in the HA1 subunit. These changes occur along a common evolutionary lineage and presumably alter the antigenic structure of the HA molecules enough to allow the virus to escape neutralization by antibodies. One important difference between the influenza A and influenza B viruses may be the rates at which the HAs of these viruses vary with time. Influenza A viruses of the H3 subtype isolated 11 years apart (A/NT/60/68 and A/BK/1/79) exhibit a 10.1% amino acid difference in the HA1 portion of their HAs (11). In contrast, the HAs of influenza B viruses isolated 33 years apart (B/Lee/40 and B/HK/8/73) exhibit only a 6.6% amino acid difference. Although in any analysis of this kind it is unknown whether the strains analyzed are representative of the strains



Location	Insertion size
48	12
77	1
125	2
132	1
135	5
137	1
141	1
160	4
166	1
196	5
211	1
242	1
285	1
297	1
312	1
328	1
163 (HA_2)	2

FIG. 5. The location of amino acid insertions in the B HAs relative to the A/Aichi/2/68 HA illustrated on a schematic drawing of the Aichi HA structure. •, Insertions of one or two amino acids; ●, insertions of four or more amino acids. Most of the insertions occur in regions of the B HAs corresponding to antigenically important regions of the A HA structure. The dashed circle indicates insertion of 12 and deletion of 6 amino acids in the area corresponding to antigenic region C of the A virus HA.

circulating at the time, it appears that the amount of variation in the B HA1s is less than the amount seen among the H3 subtype viruses.

A detailed comparison of the influenza B HA with the 1968 influenza A HA atomic model indicates that the structures of the two proteins are very similar. Five of the six disulfide bonds are probably conserved in the influenza B virus, while two new disulfide bonds at positions 52–61 (instead of the 52–277 bond in influenza A virus) and 170–259 can be proposed. Also, the sialic acid binding site as well as the α-helical coiled coil structure that forms the stalk of the HA trimer are well conserved in the B viruses. In addition, most of the insertions and deletions that need to be proposed in the influenza B viruses relative to the A/Aichi/68 HA protein can be easily fitted into the H3 HA structural model (Fig. 3). These changes, as well as all potential glycosylation sites, are located on the surface of the A/Aichi/68 HA model (Figs. 4 and 5).

In the absence of information on amino acid positions defined by monoclonal-antibody-selected antigenic variants, the antigenic structure of the B strains cannot be definitively described. Nevertheless, sequence comparisons clearly indicate the existence in the influenza B HA of the architectural features associated with antibody binding in influenza A HAs. Furthermore, comparison of three antigenically distinct influenza B HAs (B/Lee/40, B/Md/59, and B/HK/8/73) reveals extensive nonconservative amino acid substitutions (involving polarity changes) in areas corresponding to antigenic regions A, B, D, and E of H3 viruses (ref. 23; Fig. 4). Also, in the B/HK/8/73 HA, a two-amino-acid deletion relative to the earlier B virus isolates has occurred at positions 158–159 (antigenic region B). This deletion, due to the loss of one block of a A-C-A-A-A-A tandem repeat, may represent an additional mechanism for causing antigenic variation in influenza field viruses. In this context, insertion of three nucleotides leading to an alteration of an antigenic site has been recently observed in a laboratory variant derived from influenza A/PR/8/34 virus (25). Change in the location of glycosylation sites has been proposed as another mechanism that modulates antigenic variation in influenza A H1 and H3 viruses (24, 25). The introduction of new oligosaccharide attachment sites at positions 126 (antigenic region B) and 141 (antigenic region A) between the B/Lee/40 and the B/HK/8/73 strains and the loss of a site at position 217 (antigenic region D) in B/HK/8/73 may be evidence that carbohydrate is utilized to expose or mask antigenic surfaces in influenza B viruses (25).

This work was supported in part by grants from the National Institutes of Health (AI 11823, AI 18998, and AI 13654 to D.C.W.) and the National Science Foundation (PC-77-11398 to D.C.W. for computing hardware). 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. Laver, W. G., Air, G. M., Dopheide, T. A. & Ward, C. W. (1980) *Nature (London)* **283**, 454–457.
5. Laver, W. G., Air, G. M. & Webster, R. G. (1981) *J. Mol. Biol.* **145**, 339–361.
6. Both, G. W. & Sleigh, M. J. (1980) *Nucleic Acids Res.* **8**, 2561–2575.
7. Gething, M. J., Bye, J., Skehel, J. J. & Waterfield, M. (1980) *Nature (London)* **287**, 301–306.
8. Min Jou, W., Verhoeyen, M., Devos, R., Saman, R., Fang, R., Huylebroeck, D. & Fiers, W. (1980) *Cell* **19**, 683–696.
9. Sleigh, M. J., Both, G. W., Brownlee, G. G., Bender, V. J. & Moss, B. A. (1980) *J. Virol.* **37**, 845–853.
10. Verhoeyen, M., Fang, R., MinJou, W., Devos, R., Huylebroeck, D., Saman, E. & Fiers, W. (1980) *Nature (London)* **286**, 771–776.
11. Both, G. W. & Sleigh, M. J. (1981) *J. Virol.* **39**, 663–672.
12. Chakraverty, P. (1972) *Bull. World Health Org.* **45**, 755–766.
13. Schild, G. C., Pereira, M. S., Chakraverty, P., Coleman, M. T., Dowdle, W. R. & Change, W. K. (1973) *Br. Med. J.* **4**, 127–131.
14. Krystal, M., Elliott, R. M., Benz, E. W., Young, J. F. & Palese, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4800–4804.
15. Lamb, R. A. (1983) in *The Genetics of Influenza Viruses*, ed. Palese, P. (Springer, Vienna), in press.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
17. Air, G. M. (1979) *Virology* **97**, 468–472.
18. Staden, R. (1977) *Nucleic Acids Res.* **4**, 4037–4051.
19. Staden, R. (1978) *Nucleic Acids Res.* **5**, 1013–1015.
20. Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610.
21. Sleigh, M. J. & Both, G. W. (1981) in *Genetic Variation among Influenza Viruses*, ed. Nayak, D. P. (Academic, New York), pp. 341–354.
22. Krystal, M., Buonagurio, D., Young, J. F. & Palese, P. (1983) *J. Virol.* **45**, 547–554.
23. Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981) *Nature (London)* **289**, 366–373.
24. Wiley, D. C., Wilson, I. A. & Skehel, J. J. (1981) *Nature (London)* **289**, 373–378.
25. Caton, A. J., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. (1982) *Cell* **31**, 417–427.
26. Ward, C. W. & Dopheide, T. A. (1980) *Aust. J. Biol. Sci.* **33**, 449–455.
27. Webster, R. G. & Laver, W. G. (1980) *Virology* **104**, 139–148.
28. Webster, R. G. & Becton, M. T. (1981) *J. Gen. Virol.* **54**, 243–251.