Evolutionary changes in influenza B are not primarily governed by antibody selection

(nucleotide sequence/hemagglutinin/neuraminidase)

GILLIAN M. AIR†, ADRIAN J. GIBBS‡, W. GRAEME LAVER§, AND ROBERT G. WEBSTER¶

*Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294; ‡Research School of Biological Sciences and §John Curtin School of Medical Research, Australian National University, Canberra City, ACT 2601 Australia; and ¶Saint Jude Children’s Research Hospital, P.O. Box 318, Memphis, TN 38101

Communicated by Frank Fenner, March 12, 1990 (received for review February 9, 1990)

ABSTRACT Influenza B viruses evolve more slowly than human influenza A, but no reasons for the difference have been established. We have analyzed sequence changes in the hemagglutinin and neuraminidase of influenza B viruses (and have determined four hemagglutinin sequences, of B/Bonn/43, B/USSR/100/83, B/Victoria/3/85, and B/Memphis/6/86) in relation to antigenic properties and compared these with similar analyses of variation in influenza A antigens. Independent of the slower rate of change in influenza B antigens, only approximately 30% of nucleotide changes in either the hemagglutinin or neuraminidase gene sequence result in amino acid changes in the protein, whereas in influenza A 50% of nucleotide changes result in altered amino acids. Thus, there is less selection for change, or less tolerance to change, in the influenza B antigens. This is similar to findings with influenza C and findings with influenza A viruses that replicate in lower animals and birds and is closer to the type of variation found in other RNA viruses. We propose that human influenza A is unique in that it is the only virus group in which antibody selection dominates evolutionary change.

Sequence analysis of the HA1-coding portion of the hemagglutinin (HA) gene of influenza B viruses showed that the rate of change is less than in influenza A viruses and that there are multiple lineages of influenza B (1). However, there has been little attempt to correlate sequence changes with antigenic change in influenza B surface antigens or to explain why there is less change.

In influenza A, there are two kinds of antigenic variation, known as antigenic drift and antigenic shift. During drift there is a progressive accumulation of antigenic changes that result from accumulating sequence changes in the HA and neuraminidase (NA) genes. The antigenic changes can often be correlated with specific amino acid sequence changes in known antigenic sites (2–5). In contrast, when antigenic shift occurs, there is a sudden and dramatic change in the HA, with or without an equally startling change in the NA. There is evidence, in some instances, that the new virus resulted from a reassortment of genes, possibly involving the extensive reservoir of influenza viruses that exist in human, horse, pig, and particularly bird populations (6).

In influenza B viruses, no antigenic shifts have ever been detected, and there are no subtype divisions of the surface antigens as in influenza A viruses. This difference may arise because influenza B occurs in only one host species (humans). Antigenic variation occurs in influenza B but, when measured with polyclonal antisera, the extent of variation is considerably less than in influenza A (7). Isolates can be distinguished by monoclonal antibodies, and these analyses have indicated that there is no clear progressive antigenic drift in influenza B; the pattern of variation is very erratic. Several pairs of viruses isolated many years apart showed remarkable similarity, whereas viruses isolated from a single epidemic showed extensive antigenic differences (8, 9).

This paper reports the sequence of HA genes and antigenic properties of four more influenza B viruses§ isolated from 1943 to 1986 and combines these data with previously determined sequences to extend previous work (1) into an analysis of the mode and rate of genetic change in influenza B viruses compared with that of influenza A.

METHODS

Viruses were propagated and the genomic RNA was extracted and sequenced using the dideoxynucleotide method as described (10). The cDNA sequences of the HA genes of influenza viruses B/Bonn/43, B/USSR/100/83, B/Victoria/3/85, and B/Mem/6/86 were determined.

The sequence and antigenic differences among influenza A and B viruses were compared using almost all the complete HA and NA sequences of each type or subtype in the GenBank (release 60.0) and EMBL (release 19.0) data bases. The sequences were analyzed using the University of Wisconsin program package, version 6.1 (11), and their relationships were assessed using the neighbor-joining method (12). The sequences (references are given in the data bases) examined were as follows. For N2 NA: X-7(F1) (=A/Ri/5/57), A/Tokyo/3/67, A/NT/60/68, A/Udorn/72, A/Victoria/3/75, and A/Bangkok/1/79. For H3 NA: X-31 (A/Aichi/2/68), A/Victoria/3/75, A/Memphis/1/71, A/Memphis/102/72, A/England/321/77, A/Bangkok/1/79, A/Philippines/2/82, A/Michigan/1/85, and A/Memphis/6/86. The Phil/82 and Mich/85 sequences (5) are not in the data bases and were entered manually. For B NA: B/ Lee/40, B/Maryland/59, B/Hong Kong/8/73, B/Singapore/222/79, B/Oregon/5/80, B/USSR/100/83, B/Victoria/3/85, B/Leningrad/179/86, B/Memphis/6/86, and B/Memphis/3/89. For B HA: B/Lee/40, B/Maryland/59, B/HK/8/73, B/Singapore/222/79, B/Oregon/5/80, B/England/222/82, and also B/Bonn/43, B/USSR/100/83, B/Victoria/3/85, and B/Memphis/6/86 (this paper).

RESULTS AND DISCUSSION

Four nucleotide sequences of influenza B HA genes were determined. Copies of the nucleotide and translated amino acid sequences are available from the authors. The total numbers of pairwise differences among the gene and protein

Abbreviations: HA, hemagglutinin; NA, neuraminidase.
†To whom reprint requests should be addressed.
§The sequences reported in this paper have been deposited in the EMBL data base (accession numbers: X13552 for B/USSR/100/83, X13551 for B/Memphis/6/86, X13550 for B/Bonn/43, and X13553 for B/Victoria/3/85.)
Evolution: Air et al.


Table 1. Differences between influenza B HA sequences

<table>
<thead>
<tr>
<th>Residue differences, no.</th>
<th>B/Bonn/43</th>
<th>B/Md/59</th>
<th>B/HK/73</th>
<th>B/Sing/79</th>
<th>B/Ore/80</th>
<th>B/Eng/82</th>
<th>B/USSR/83</th>
<th>B/Vic/85</th>
<th>B/Mem/86</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/Lee/40</td>
<td>88, 37</td>
<td>93, 33</td>
<td>131, 36</td>
<td>134, 37</td>
<td>137, 38</td>
<td>146, 40</td>
<td>142, 40</td>
<td>145, 38</td>
<td>149, 40</td>
</tr>
<tr>
<td>B/Bonn/43</td>
<td>0</td>
<td>94, 32</td>
<td>123, 36</td>
<td>123, 34</td>
<td>125, 35</td>
<td>138, 38</td>
<td>133, 37</td>
<td>143, 39</td>
<td>149, 42</td>
</tr>
<tr>
<td>B/Md/59</td>
<td>0</td>
<td>0</td>
<td>66, 17</td>
<td>72, 18</td>
<td>73, 19</td>
<td>87, 22</td>
<td>81, 24</td>
<td>87, 25</td>
<td>90, 27</td>
</tr>
<tr>
<td>B/HK/73</td>
<td>0</td>
<td>0</td>
<td>34, 12</td>
<td>35, 13</td>
<td>52, 16</td>
<td>43, 17</td>
<td>49, 16</td>
<td>54, 18</td>
<td></td>
</tr>
<tr>
<td>B/Sing/79</td>
<td>0</td>
<td>16, 4</td>
<td>29, 7</td>
<td>26, 8</td>
<td>51, 15</td>
<td>58, 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/Ore/80</td>
<td>0</td>
<td>34, 6</td>
<td>27, 7</td>
<td>53, 16</td>
<td>58, 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/Eng/82</td>
<td>0</td>
<td>43, 9</td>
<td>71, 19</td>
<td>74, 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/USSR/83</td>
<td>0</td>
<td>60, 20</td>
<td>64, 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/Vic/85</td>
<td>0</td>
<td>10, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First number is nucleotide differences; the second number is amino acid differences.

sequences of all the influenza B Has where the complete sequences is known is shown in Table 1.

Influenza B HA and NA antigenic changes can be clearly distinguished using monoclonal antibodies (10, 13), and Table 2 shows the analyses of influenza B Has in hemagglutination inhibition tests. The cross-reactivities among the viruses tend to be greater the shorter the time interval between their isolations, but there are some exceptions——e.g., antibodies 419/2 and 232/1. The closest relationships found with monoclonal antibodies are between B/USSR/83 and Mem/86. With polyclonal chicken serum, Lee/40, HK/73, Sing/79, Vic/85, and Mem/86 are clearly distinguished from the other type B viruses.

Rate and Mode of Change of Influenza B HA and NA Genes

Differences in the nucleotide and amino acid sequences of the influenza B HA gene and protein (Table 1) were converted to differences per 100 sites and used to calculate a similarity network by the neighbor-joining method (12). The branch lengths in the network provided a corrected distance matrix. Similar matrices were calculated for the NA of influenza B (10) and for human influenza A H3 HA and N2 NA gene and protein sequences. All eight distance matrices significantly and positively correlated with the differences in times of isolation of the viruses. As there was a clear time-dependent drift, each network was represented as a dendrogram (Fig. 1), the root of which was arbitrarily placed at the mid-point of the branch to the earliest isolate. The dendrograms confirm that both surface antigens of a single subtype of influenza A viruses have a single major evolutionary line with short side branches, whereas influenza B trees have longer branches, in essence several coexisting lines, confirming the antigenic data (8) and HA1 and NS1 gene sequence data (1). The influenza B dendrograms also give some indication of reassortment among the genes of the most recent isolates (between 1979 and 1989).

The antigenic analyses (Table 2) also show multiple lineages, but the antigenic relationships are not the same as the sequence relationships. For example, no amino acid substitutions can be found that would explain the reactivity patterns of monoclonal antibodies 419/2 and 232/1.

The time dependence of amino acid changes shows clearly when the position of each isolate in the horizontal axis of the dendrogram (i.e., difference from the "root") is plotted against the year of isolation (Fig. 2). The slopes of the linear regressions for these points are measures of the average rates of change of the antigens and their genes (Table 3). Although we calculated the linear regression to obtain average rates of change, it is noticeable in Fig. 2 that the rate of change of the influenza A antigens appears to diminish in the most recent isolates. There are insufficient data to be sure that this is significant.

Table 3 shows that the Has and also NAs of influenza B do not change as rapidly as those of influenza A; nucleotide sequence changes in influenza B HA and NA occur at 20–30% of the rates in influenza A, whereas the amino acid rate in influenza B is 10–20% of the rate in influenza A. The rate of nucleotide change for the whole HA gene (Table 3) is almost identical to that obtained for the HA1 coding portion and NS genes (0.103 and 0.11 changes per 100 nucleotides per year, respectively) (1).

Why Do the Influenza B Surface Antigens Evolve More Slowly Than Those of Influenza A? The rate of evolution of an organism depends primarily on the basic mutation rate, and the ability of those mutants to survive various selection barriers, including their ability to compete successfully with other individuals of the population. We have sought to identify which parts of the mutation/selection process could explain the differences between influenza A and B.

Is the Influenza B Polymerase More Accurate? RNA polymerases lack the editing mechanisms of DNA polymerases (14), but not all RNA viruses evolve rapidly, due partly to different error rates (15) and partly to differences in types or extents of selection pressures. The only comparative data available on error rates is the frequency of selecting escape mutants. For influenza B HA, the frequency was indeed low, 10⁻⁷ to 10⁻⁸, compared with 10⁻³ for influenza A HA or NA

Table 2. Cross-reactivity between the Has of influenza B viruses

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Number</th>
<th>log₁₀ HI titer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td></td>
<td>Lee/40</td>
</tr>
<tr>
<td>Monoclonal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HK/8/73</td>
<td>174/1</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>313/2</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>419/2</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>163/5</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>162/1</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>232/1</td>
<td>2.3</td>
</tr>
<tr>
<td>Mem/6/86</td>
<td>BM15A</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>1A/1/86</td>
<td>1/2</td>
</tr>
<tr>
<td></td>
<td>HK/8/73</td>
<td>125</td>
</tr>
</tbody>
</table>

HI, hemagglutination inhibition; <, less than 1.0. Polyclonal antiserum against HK/8/73 was from a chicken.
Influenza A NA, however, the frequency was $10^{-5}$ to $10^{-2}$ (10), which is similar to that seen in influenza A and also in vesicular stomatitis virus and Sendai virus, which do not show antigenic drift (17). This is in accord with the measured polymerase error rate in influenza A of $1.5 \times 10^{-5}$ mutations per nucleotide per infectious cycle (15). By assuming that the epitope contains 20 amino acids (18) and that one-third of changes are silent, the number of changes per epitope is $6 \times 10^{-4}$ per infectious cycle. We conclude that polymerase activity is unlikely to be a factor in the different rates of evolution of influenza A and B viruses, and either positive or negative selection must be the driving force.

**Is There Selection at the Nucleotide Level?** The average base composition of all influenza A sequences in GenBank release 60.0 (adenine = 33.0%, guanine = 23.7%, cytosine = 19.7%, and uracil = 23.6%) is not very different from the composition of all influenza B sequences (adenine = 35.7%, guanine = 22.0%, cytosine = 18.4%, and uracil = 23.8%). We also determined the compositions of individual segments of A/PR/8/34 and B/Ann Arbor/1/86 and found all segments have the same relative composition (adenine > uracil > guanine > cytosine). The composition of human coding sequences (19) is very different (adenine = 24.5%, guanine = 26.8%, cytosine = 27.4%, and uracil = 21.3%). The G+C content of the human coding sequences is 54%, in influenza mRNAs it is only 42%, and it is the same in the negative (viral) RNA.

The reason for the high G+C content of human coding sequences, while "junk" sequences are higher in A+T, is not understood. Adenine and uracil require less energy of synthesis and thus might be preferred for rapid transcription/replication of viral RNA. It might also be important to influenza virus to more easily unwind the RNA:RNA duplex or to minimize secondary structure in the separate strands.

**Codon Use in Influenza B HA and NA.** We looked for differences in codon use between influenza A and B viruses but did not find any significant differences.

It is thought that the nucleotide sequences of virus genomes are selected so that their codon use mimics that of their hosts and thus optimally uses the translation system of the host. A striking difference between influenza A and B is that the latter viruses have only been isolated from humans whereas influenza A viruses are widely distributed in pigs, horses, and particularly, birds. We, therefore, examined the codon frequencies to see if the type B viruses appeared more human-like or the type A more chicken-like, but no correla-

**FIG. 1.** Dendrograms showing the relationships and time-dependent sequence changes in the HA and NA genes and proteins of influenza A and influenza B viruses. Calculated using neighbor-joining method (12) as described in text.

**FIG. 2.** Time dependence of nucleotide and amino acid changes (differences per 100 sites from the root) in type A and type B influenza virus HA and NA. The position of each isolate in the horizontal axis of the dendrogram is plotted against year of isolation. $\triangle$, Nucleotides; $\circ$, amino acids. The earliest isolate is omitted as it was used to provide the external root to the dendrogram.
Table 3. Rates of change of HA and NA of influenza A and B

<table>
<thead>
<tr>
<th>Virus</th>
<th>Protein</th>
<th>Nucleotides</th>
<th>Amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influenza A</td>
<td>HA</td>
<td>0.323 ± 0.037</td>
<td>0.509 ± 0.049</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.311 ± 0.049*</td>
<td>0.468 ± 0.072</td>
</tr>
<tr>
<td>Influenza B</td>
<td>HA</td>
<td>0.103 ± 0.011*</td>
<td>0.051 ± 0.011*</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.063 ± 0.009</td>
<td>0.098 ± 0.016</td>
</tr>
</tbody>
</table>

Influenza A is from humans. Values are mean ± SEM of slopes of linear regressions of data in Fig. 2.

*Position of the HA of B/Donn/43 was anomalous, probably reflecting a more complex relationship with B/Lee/40 than appears in the dendrograms, so it was omitted from the calculations. If included, the rates are 0.082 (nucleotides) and 0.022 (amino acids).

The lack of selection of influenza B is similar to the situation when influenza A viruses replicate in birds. Sequences of H3 HA in viruses isolated from ducks in Japan from 1977 to 1985 show few differences (23) and, when the percent coding changes in all pairwise comparisons are calculated, the values range from 9% to 21%; a sharp contrast to the 47% coding changes in H3 HAs of human viruses. When two N9 NAs were compared, one from a tern, the other from a whale, the percentage of coding changes between the two was 24%. None of the amino acid substitutions occurs near known neutralizing epitopes (24). It seems that if there is any selection of influenza B or avian influenza A viruses at the protein level, it is not by antibodies. Since humans mount an immune response to influenza B that is not obviously different to that against influenza A, the reason why those antibodies do not select is not clear. It is possible that the bulk of antibodies that bind influenza B antigens are not neutralizing. The difficulty in obtaining escape mutants may be due to lower affinities of monoclonal antibodies, although binding constants of serum antibodies directed against influenza B were not significantly different to those of influenza A (25). Efforts to crystallize the HA of influenza B have not so far been successful but crystals have been grown of influenza B NA and of antibody Fab fragments complexed with influenza B NA (27, 28).

We have been unable to identify any trends in the nucleotide or amino acid sequences of influenza B HA and NA that would identify a major selection pressure, either positive or negative. It is possible that the evolution is random, although the rates observed are higher than would be expected to be randomly fixed in the population from the estimated error rate of influenza RNA polymerase (15).

Although our focus has been on the question of why influenza A evolves differently from influenza A, the question should probably be turned around. Sequence data are not so extensive for other groups of RNA viruses, but the emerging data show no evidence of progressive antigenic drift. A new epidemic virus is selected from a pool of circulating viruses.
that differ in antigenic properties, but there is little evidence that evolution in response to antibody pressure occurs in paramyxoviruses, poliovirus, or influenza type C (1, 26) as well as type B. Thus the data discussed in this paper support the hypothesis that, although influenza A viruses in humans evolve by antibody selection of variants of the two surface antigens, there is no equivalent driving force in influenza B evolution or when influenza A viruses replicate in ducks.

It may be that human influenza A is unique in that it is able to produce a series of antigenically selected mutants that are as fit as the parental population and is the only virus that undergoes true antigenic drift.

We thank many colleagues for fruitful discussions, and Lynn Ritchie, Anna Browne, Tim Thomas, Lili Wong, and James Whethead for excellent technical assistance. This collaborative project was greatly helped by international dialing facilities donated by the Australian Overseas Telecommunications Commission. This work was supported in part by Grants AI-18203 and AI-08831 from the National Institutes of Health.