Virus Research, 8 (1987) 33–41
Elsevier

VRR 00344

Nucleotide sequences of the PA and PB1 genes of B/Ann Arbor/1/66 virus: comparison with genes of B/Lee/40 and type A influenza viruses

Dan C. DeBorde [1], Clayton W. Naeve [2], M. Louise Herlocher [1] and Hunein F. Maassab [1]

[1] *Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, U.S.A. and* [2] *Department of Virology and Molecular Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, U.S.A.*

(Accepted for publication 30 March 1987)

Summary

The complete sequences of the PA and PB1 genome RNA segments of B/Ann Arbor/1/66 virus have been determined. The PA vRNA is 2308 bases long. Its complementary RNA has a single open reading frame of 2187 bases, capable of encoding a PA protein of 726 amino acids with a molecular weight of 83,175 Da. The predicted PA polypeptide has an overall net charge of −7.5 at pH 7.0. The PB1 vRNA is 2369 bases long. Its complementary RNA has a single open reading frame of 2277 bases, capable of encoding a PB1 protein of 752 amino acids with a molecular weight of 84,332 Da. The predicted PB1 polypeptide has an overall net charge of +18.5 at pH 7.0. Sequence homology comparisons of the PA and PB1 polypeptides from B/Ann Arbor/1/66 virus to the PA and PB1 polypeptides of type A influenza virus reveal respective homologies of approximately 38 and 60%. This high cross-type homology (61%) was previously reported for the PB1 protein of B/Lee/40 virus (Kemdirim et al., 1986). The cross-type homology for the PA protein is similar to that of other non-polymerase proteins, but is substantially lower than that seen for the PB1 protein. Thus, the high cross-type homology that exists for the PB1 gene does not appear to be a characteristic of all polymerase genes.

PA; PB1; Polymerase; Dideoxynucleotide sequencing; Influenza type A and B virus

Correspondence to: D.C. DeBorde, Dept. of Epidemiology, School of Public Health, University of Michigan, 109 Observatory, Ann Arbor, MI 48109, U.S.A.

Until 1986, complete sequence information for the RNA of type B influenza virus existed only for the non-polymerase genes of B/Lee/40 virus; then the B/Lee/40 PB1 gene sequence was published (Kemdirim et al., 1986). The polypeptide predicted from this PB1 gene showed a significantly higher level of sequence homology with the PB1 polypeptides of influenza type A viruses than that occurring for any of the other non-polymerase proteins, suggesting that a functional constraint might be operating as a significant selection mechanism, restricting sequence variation in the PB1 gene. In the absence of sequence information for the PB2 and PA genes of influenza type B virus, it was not known if this high level of cross-type homology would be typical of all three polymerase genes.

The PB1 and PA genes of influenza B/Ann Arbor/1/66 (B/AA/1/66) wild-type (wt) virus were sequenced as part of a project in which all six non-glycoprotein genes of both the cold-adapted (ca) and wt B/AA/1/66 viruses are being compared in order to catalogue the changes that occur during the process of cold-adaptation. With the sequencing of both wt and ca B/AA/1/66 viruses, we confirm the high cross-type homology for the PB1 polypeptide, but show that the PA polypeptide exhibits a level of cross-type homology similar to that of the NP (37%), HA_2 (39%), and NA (35%) proteins (Kemdirim et al., 1986).

The complete sequences of the PA and PB1 genes were determined by a combination of two RNA sequencing techniques. The first 60–70 nucleotides at the 3′-termini of both the PA and PB1 vRNA segments were sequenced by a direct chemical method described in Peattie (1979) using vRNA segments isolated on, and subsequently eluted from, a 3% polyacrylamide gel. The remaining sequences were determined by dideoxynucleotide chain termination sequencing procedures described previously (DeBorde et al., 1986). All ambiguities in these sequences were resolved using terminal deoxynucleotidyl transferase enzyme (DeBorde et al., 1986). The sequence that extended from each primer overlapped the position of the next primer by at least 20 nucleotides in every case. The complete nucleotide and predicted amino acid sequences for the PA and PB1 genes are presented in Figs. 1 and 4, respectively. The oligodeoxynucleotide primers were all 15 bases long, except for a 12-nucleotide-long primer beginning at residue 9 in PB1, and their positions are underlined in Figs. 1 and 4.

All sequence comparisons, manipulations, and calculations were performed using the programs developed by Queen and Korn (1984) and distributed by Beckman Instruments, Inc., Palo Alto, CA, as the Microgenie Sequence Software package.

The PA vRNA segment is 2308 nucleotides long (Fig. 1). The first protein initiation codon in the complementary RNA (cRNA) starts 30 nucleotides in from its 5′-end, and is encompassed by an open reading frame of 2187 nucleotides extending to the first termination codon beginning at nucleotide 2208. The PA polypeptide encoded by this open reading frame is 726 amino acids in length with a calculated molecular weight of 83,175 Da. A potential polyadenylation site composed of five consecutive adenine residues is present at nucleotides 2288–2292. The predicted mRNA would be 2292 nucleotides long, prior to capping and the addition of poly (A). No other major open reading frame exists. The next longest unterminated coding sequence in any of the reading frames would encode only 44 amino acids.

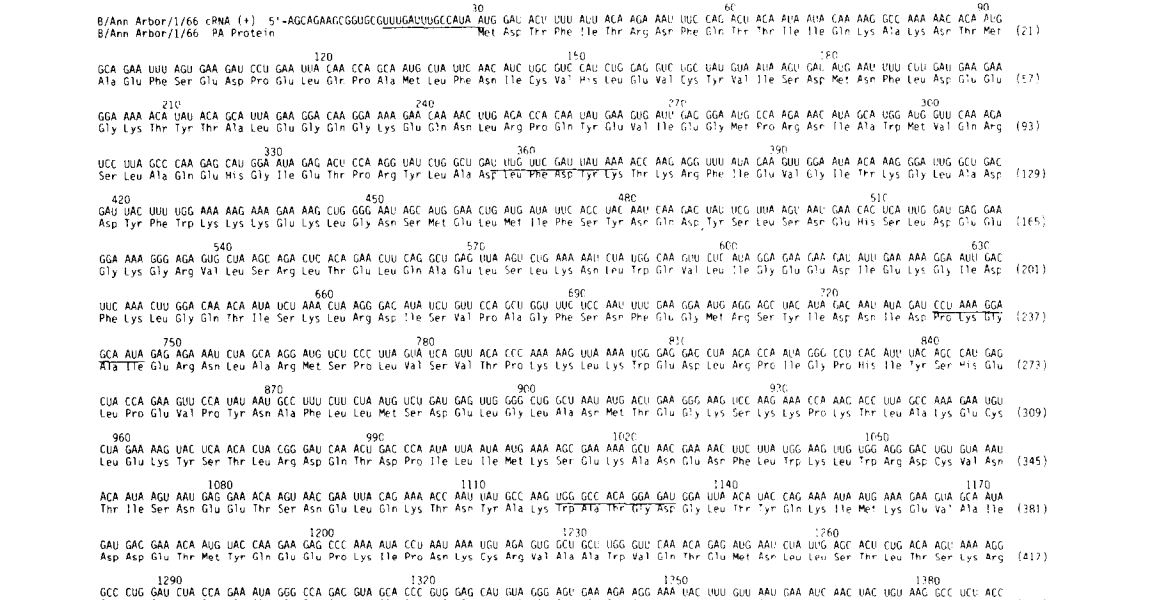

Fig. 1. Nucleotide and predicted polypeptide sequence of the PA gene of B/AA/1/66 virus. The sequence is presented 5′ → 3′ in the (+) messenger strand sense. The underlined areas indicate the position and sequence of the oligodeoxynucleotide primers that were used. The numbers in ( ) are the lengths of the nucleic acid and polypeptide at that point.

A comparison of the PA protein sequences of B/AA/1/66 and A/NT/60/68 using the Microgenie alignment program is shown in Fig. 2. Homologous amino acids are underlined. The overall charge of the B/AA/1/66 PA protein is predicted to be −7.5, based on a +1 charge for each arginine and lysine residue, a +0.5 charge for each histidine residue, and a −1.0 charge for each aspartic acid and glutamic acid residue calculated at pH 7.0. However, the net charge is asymmetrically distributed over the length of the protein as follows: the NH_2-terminal half of

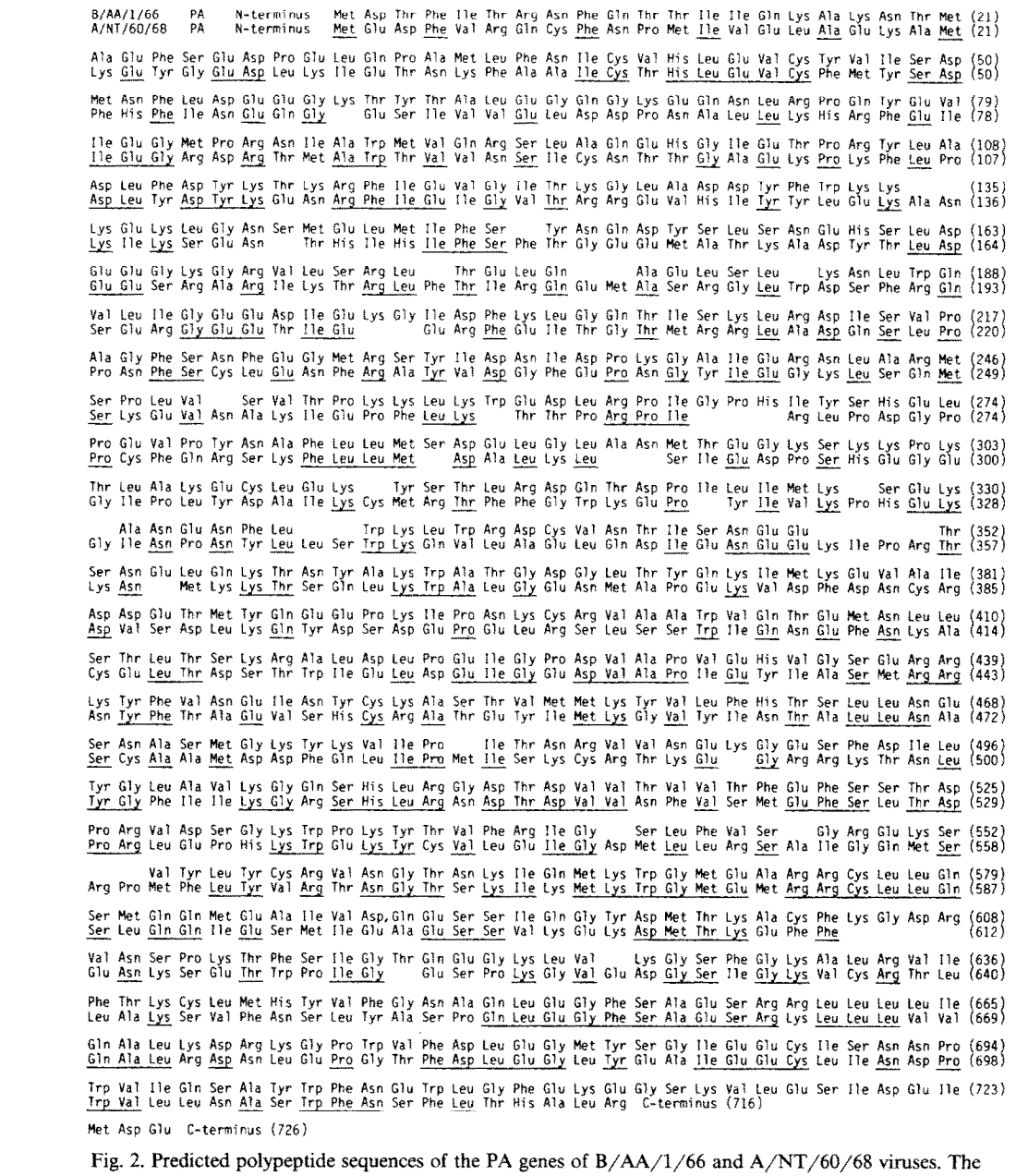

Fig. 2. Predicted polypeptide sequences of the PA genes of B/AA/1/66 and A/NT/60/68 viruses. The sequence is presented in the NH_2-terminal → COOH-terminal direction. The underlined areas indicate the amino acids that are in common for the two polypeptides. The numbers in ( ) are the lengths of the polypeptides at that point. Gaps were inserted in either sequence as necessary to provide the optimal alignment.

the B/AA/1/66 PA protein has a net charge of −8.5, while its COOH-terminal half has a net charge of +1.0. This unequal distribution of charge is not seen in the A/NT/60/68 PA protein. Its net charge is divided between its NH_2- and COOH-terminal ends as −8.0 and −11.0, respectively. Thus the overall net drop in negative charge between the influenza A and influenza B PA genes occurs in the COOH-terminal half of the protein.

While the two PA proteins exhibit a 38% overall level of homology, the amino acids range from a low of 22% conservation (histidine) to a high of 57% conservation (tryptophan). The charged amino acids are conserved as follows: histidine, 22%; arginine, 47%; lysine, 37%; aspartic acid, 43%; and glutamic acid, 48%. If the acidic or basic residues are considered as equivalent amino acids then the conservation for the acidic sites is 58% and for the basic sites, 47%. While these values are higher than the 38% overall homology, they do not imply rigid conservation of charged sites. Indeed, no amino acid is dramatically conserved between the sequences of these two viruses. The conserved regions are scattered throughout the protein, but the COOH-terminal half of the PA protein is slightly more conserved than the NH_2-terminal half (46 to 33%). Fig. 3 illustrates this homology difference over the length of the protein by a matrix comparison using a segment length of 40 amino acids in which 20 amino acids must match to give a positive result. It is obvious that the COOH-terminal end contains the region of highest homology overall, with a small region of high homology near the NH_2-terminal end. Thus, the COOH-terminal half of the polypeptide may be more important to the PA protein's function than the NH_2-terminal half. The importance of the difference in net charge between the B/AA/1/66 and A/NT/60/68 PA proteins in this region is not clear. Essentially the same asymmetric homology and net charge patterns were obtained with the PA gene of B/Singapore/222/79 (Dr. Debi Nayak, pers. commun.).

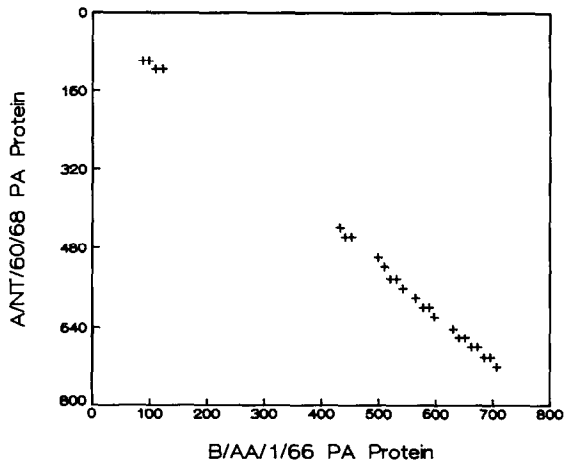

Fig. 3. Matrix comparison plot of B/Ann Arbor/1/66 and A/NT/60/68 PA polypeptides. Each symbol represents a site where at least 20 out of 40 amino acids were conserved. The numbers on the x and y axes represent the amino acid position in from the NH_2 terminus of each protein.

The PB1 vRNA segment is 2369 nucleotides long (Fig. 4). The first protein initiation codon in the cRNA starts 22 nucleotides in from the 5′-end, and an open reading frame of 2277 nucleotides extends from the first nucleotide at the 5′-end to the first termination codon beginning at nucleotide 2278. The PB1 polypeptide encoded in this open reading frame is 752 amino acids in length with a molecular weight of 84,332 Da. A polyadenylation site composed of six consecutive adenine residues is present at nucleotides 2348–2353. The predicted mRNA prior to capping and poly (A) addition would be 2353 nucleotides long. The PB1 RNA of B/AA/1/66 virus is one nucleotide longer than the B/Lee/40 PB1 RNA (Kemdirim et al., 1986). The difference in length is due to an addition of one cytidine residue in the B/AA/1/66 PB1 cRNA sequence at nucleotide 15 prior to the start codon for the PB1 polypeptide (see Fig. 4). The next longest stretch of nucleotides without a termination codon in either of the other reading frames can only code for 55 amino acids. The sequence predicts a protein with a net charge of +18.5 at pH 7.0, similar to the B/Lee/40 PA protein.

In addition to the one nucleotide insertion at position 15, there were 109 nucleotide mismatches resulting in 11 amino acid changes between the B/AA/1/66 and B/Lee/40 PB1 RNA and protein, respectively. Thus, the variation between these two viruses' predicted polypeptides is only 1.5%. Table 1 shows a compilation of the variation between five polypeptides from B/AA/1/66 and B/Lee/40 predicted by their respective nucleotide sequences. B/AA/1/66 sequences have been determined in this laboratory (manuscript in preparation) while sequence data for B/Lee/40 virus were derived from Kemdirim et al. (1986), Briedis and Tobin (1984), Briedis et al. (1982) and Briedis and Lamb (1982). The overall nucleotide variation was similar for all genes compared, but the amino acid variation ranged from a low of 1.5% (PB1) to a high of 7.8% (NS_1). These results are echoed in the percentage of possible non-silent and silent changes observed. The NS RNA was interesting in that the NS_1 gene had the highest percentage of non-silent changes, while NS_2 had the third lowest percentage of non-silent changes. NS_2 protein also had higher cross-type homology than did NS_1 protein (16.2% vs 9.7%) (Kemdirim et al., 1986) although these values are among the lowest for any of the viral proteins. Only the M_1 polypeptide shows as little variation as the PB1 polypeptide between these two viruses. Interestingly, the M_1 polypeptide does not show a correspondingly high cross-type homology with M_1 polypeptides of influenza A viruses as does the PB1 polypeptide. Its cross-type homology is only 25% (Kemdirim et al., 1986). This data may reflect a type-specific functional constraint existing for the M_1 gene of influenza B viruses as opposed to a cross-type constraint for the PB1 gene.

Because the PB1 gene of B/Lee/40 virus showed an extremely high cross-type homology with the PB1 genes of influenza A virus, approximately 60% (Kemdirim et al., 1986), we were interested in determining whether this relatedness would hold for polymerase genes in general. The B/AA/1/66 PA and PB1 polypeptides were compared to the PA and PB1 polypeptides of A/PR/8/34 and A/NT/60/68 viruses. The average cross-type homologies were 38 and 60%, respectively. Fig. 2 shows an optimal alignment of the predicted PA protein sequences of B/AA/1/66 and A/NT/60/68 viruses. Comparison with the A/PR/8/34 PA gene gave almost

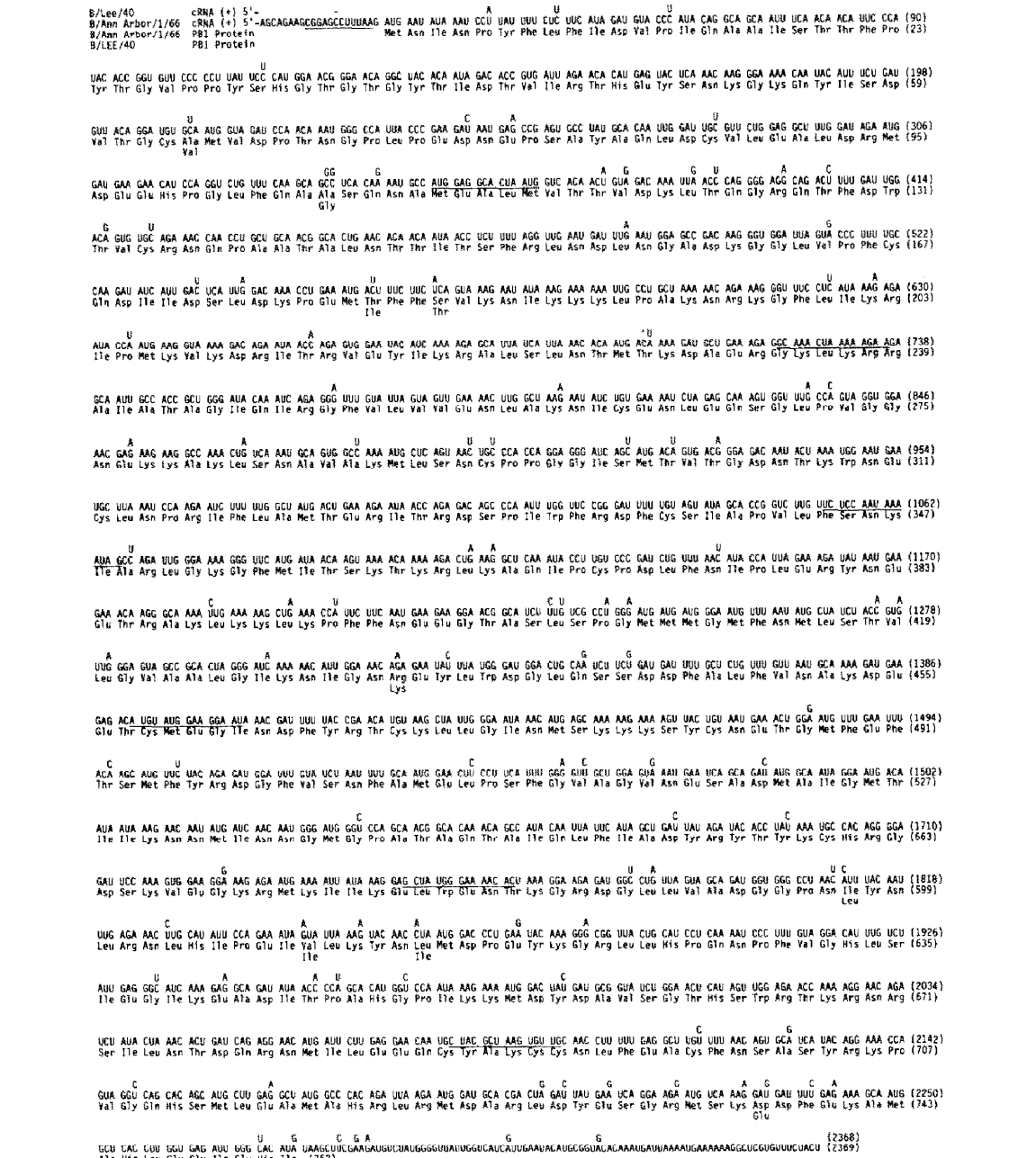

Fig. 4. Nucleotide and predicted polypeptide sequences of the PB1 genes of B/AA/1/66 and B/Lee/40 viruses. The sequence is presented 5′ → 3′ in the (+) messenger strand sense. The underlined areas indicate the position and sequence of the oligodeoxynucleotide primers that were used. The numbers in ( ) are the lengths of the nucleic acids and polypeptides at those points. Only sites of change are shown in the B/Lee/40 nucleotide and polypeptide sequences. Missing nucleotides are represented by −.

TABLE 1

VARIATION BETWEEN B/LEE/40 AND B/AA/1/66 VIRUSES.

Gene	No. nucleotide [a] mismatches/ total length	No. amino acid mismatches/ total length	No. non-silent mismatches/No. possible	No. silent mismatches/No. possible
PB1	109/2368 (4.6)	11/752 (1.5)	11/1922 (0.57)	93/800 (11.6)
NP	113/1841 (6.1)	26/560 (4.6)	28/1407 (1.9)	77/590 (13.0)
M_1	65/1191 (5.5)	4/248 (1.6)	4/ 624 (0.64)	31/271 (11.4)
N_1	61/1096 (5.6)	22/281 (7.8)	22/ 706 (3.1)	26/302 (8.6)
NS_2		3/122 (2.5)	3/ 315 (0.95)	9/131 (6.9)

[a] Values in this column are based on total RNA length, not on polypeptide encoding regions.
The total number of non-silent and silent positions possible in each coding region was calculated as previously described (Bishop et al., 1982).

identical homology results (data not shown). A similar figure for the PB1 gene is also not included, because we demonstrate that there is very little change between the PB1 proteins of B/Lee/40 and B/AA/1/66 viruses (see Fig. 4), and a comparison of B/Lee/40 and A/WSN/33 virus was previously published (Kemdirim et al., 1986). The cross-type homology data shows that the high level of relatedness seen for the B/Lee/40 PB1 gene extends to the PB1 gene of B/AA/1/66 virus, but that this high level of relatedness is not applicable to the PA gene, and hence to polymerase genes in general. In fact, this high cross-type homology appears to be unique for the PB1 gene since preliminary data comparing the first 435 amino acids of the B/AA/1/66 PB2 polypeptide with PB2 polypeptides of the same influenza A viruses used above, yields a homology of 38–40% (sequence not shown) similar to the PA proteins.

PB1 has been identified as the most likely polymerase protein to catalyze each successive nucleotide addition to the influenza RNA growing chains (Braam et al., 1983). This function should be essentially the same whether influenza A or B virus is involved, and thus little or no variation due to virus type would be expected. No divergence due to influenza virus type, coupled to a strong functional constraint, may explain why the PB1 protein, alone, has retained such a high sequence homology across influenza type A and B viruses.

Acknowledgments

We thank Susan Donabedian for her expert technical help in the preparation of viral RNA. The work in this paper was supported by Contract No. 1-AI-52564, National Institute of Allergy and Infectious Diseases, Development and Application Branch, Bethesda, MD 20892, and by a Biomedical Research Support Grant, RR 05584-21.

References

Bishop, D.H.L., Jones, K.L., Huddleston, J.A. and Brownlee, G.G. (1982) Influenza A virus evolution: Complete sequences of influenza A/NT/60/68 RNA segment 3 and its predicted acidic P polypeptide compared with those of influenza A/PR/8/34. Virology 120, 481–489.

Braam, J., Ulmanen, I. and Krug, R.M. (1983) Molecular model of a eucaryotic transcription complex: Functions and movements of influenza P proteins during capped RNA-primed transcription. Cell 34, 609–618.

Briedis, D.J. and Lamb, R.A. (1982) Influenza B virus genome: Sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS_1 and NS_2 proteins. J. Virol. 42, 186–193.

Briedis, D.J. and Tobin, M. (1984) Influenza B virus genome: Complete nucleotide sequence of the influenza B/Lee/40 virus genome RNA segment 5 encoding the nucleoprotein and comparison with the B/Singapore/222/79 nucleoprotein. Virology 133, 448–455.

Briedis, D.J., Lamb, R.A. and Choppin, P.J. (1982) Sequence of RNA segment 7 of the influenza B virus genome: partial amino acid homology between the membrane protein (M_1) of influenza A and B viruses and conservation of a second open reading frame. Virology 116, 581–588.

DeBorde, D.C., Naeve, C.W., Herlocher, M.L. and Maassab, H.F. (1986) Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotidyl transferase. Anal. Biochem. 157, 257–282.

Kemdirim, S., Palefsky, J. and Briedis, D.J. (1986) Influenza B virus PB1 protein: nucleotide sequence of the genome RNA segment predicts a degree of structural homology with the corresponding influenza A virus polymerase protein. Virology 152, 126–135.

Peattie, D.A. (1979) Direct chemical method for sequencing RNA. Proc. Natl. Acad. Sci. U.S.A. 76, 1760–1764.

Queen, C. and Korn, L.J. (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucl. Acids. Res. 12, 581–599.

(Manuscript received 9 February 1987; revised version received 30 March 1987)