

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<sup>1</sup>, Clayton W. Naeve<sup>2</sup>, M. Louise Herlocher<sup>1</sup>  
and Hunein F. Maassab<sup>1</sup>

<sup>1</sup> Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, U.S.A. and <sup>2</sup> 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<sub>2</sub> (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 un-terminated coding sequence in any of the reading frames would encode only 44 amino acids.

B/Ann Arbor/1/66 (+) 5'-AGCAGAAGCGGUGCGUUGUUGCCAU... AUG GAU UUU UUU ACA AGA AAU UUC CAC ACU ACA AUA UUA CAA AAG GCC AAA AAC ACA AUG  
 B/Ann Arbor/1/66 PA Protein Met Asp Trp Phe Ile Thr Arg Asp Phe Glu Thr Thr Ile Ile Glu Lys Ala Lys Asp Thr Met (21)

120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960 990 1020 1050 1080 1110 1140 1170 1200 1230 1260 1290 1320 1350 1380 1410 1440 1470 1500 1530 1560 1590 1620 1650 1680 1710 1740 1770 1800 1830 1860 1890 1920 1950 1980 2010 2040 2070 2100 2130 2160 2190 2220 2250 2280  
 GCA GAA UUU AGU GAA GAU CCU GAA UUA CCA CCA ACA AUG CUA UUC AAC UUC UGG GUC CAU UGC CUG GAC CUC UGC UAU GUA AUA ACU GAU AUG ACA AUG ACA GAA GAA  
 Ala Glu Phe Ser Glu Asp Pro Glu Leu Glr Pro Ala Met Leu Phe Asn Ile Lys Val His Leu Glu Val Lys Tyr Val Ile Ser Asp Met Asn Phe Leu Asp Glu Glu (57)  
 GGA AAA ACA GAU ACA GCA UUA GAA CCA GGA AAA GAA CAA ACA UUG ACA CCA CAA UAU GAA CUG AAU CAC GAG ACA ACA AAC AUA GAC UGG AUG SUU CCA ACA  
 Gly Lys Thr Tyr Thr Ala Leu Glu Gly Gln Gly Lys Glu Gln Asn Leu Arg Pro Gln Tyr Glu Val Ile Glu Gly Met Pro Arg Asp Ile Ala Trp Met Val Gln Arg (93)  
 UCC UUA GCC CAA GAG CAU GGA AUA GAC ACU CCA GGG UAU CUG CCU GAU UUG UUC GAU UAU AAA ACC AAG AGG UUU AUA CAA GGU GGA AUA ACA AAG GGA UNG GCU GAC  
 Ser Leu Ala Gln Glu His Gly Ile Glu Thr Pro Arg Tyr Leu Ala Asp Trp Phe Asp Tyr Lys Thr Lys Arg Phe Ile Glu Glu Gly Ile Thr Lys Gly Leu Ala Asp (129)  
 GAU UAC UUU UGG AAA AAG AAA GAA AAG CUG GCG AAU AGC AUG GAA CUG AUG AUA UUC ACC UAC AAU CAA CAG UAC UUG UUA ACU AAL CAA CAC UCA UUG GAU GAG GAA  
 Asp Tyr Phe Trp Lys Lys Lys Glu Lys Leu Gly Asn Ser Met Glu Leu Met Ile Phe Ser Tyr Asp Gln Asp Tyr Ser Leu Ser Asp Glu His Ser Leu Asp Glu Glu (165)  
 GGA AAA GGG AGA GUG CUA ACE AGA CUC ACA CAA UUU CAG GGU GAG UUA AGU CUG AAA AAU CUA UUG CCA GAU UUG AUA GGA GAA CAA CAU UAU GAA AAA GGA AUU GAC  
 Gly Lys Gly Arg Val Leu Ser Arg Leu Thr Glu Leu Asp Gln Ala Thr Ser Val Ala Asp Trp Phe Ser Leu Lys Asn Leu Trp Glr Val Leu Ile Glu Glu Asp Ile Glu Lys Ala Asp (201)  
 UUC AAA CUU GAA CAA ACA AUA CCU AAA AGG CAA AAU UCU GUU CCA GCU CGU UUC UCC AAU AAU UUA CAA GGA AUG AGG ACC UAC AAL AAU AUA AAU GAA CCU AAA GGA  
 Phe Lys Leu Gly Gln Thr Ile Ser Lys Leu Arg Asp Ile Ser Val Pro Ala Gly Phe Ser Asn Phe Glu Gly Met Arg Ser Tyr Ile Asp Asn Ile Asp Trp Lys Lys Ile Asp (237)  
 GCA AUA GAG AGA AAU CUA GCA AGG AUG UCU CCC UUA GUA UUA CUA GCU ACA CCC AAA AAG UUA AAA UGG CAG CAC CUA ACA CCA AUA GGG CCU CAC AAU UAC ACC CAU GAG  
 Ala Ile Glu Arg Asn Leu Ala Arg Thr Glu Leu Ser Val Thr Pro Lys Lys Leu Lys Tyr Glu Lys Leu Lys Tyr Glu Asp Lys Arg Phe Pro His Ile Tyr Ser His Asp (273)  
 CUA CCA GAA GGU CCA UAU AUC UUU CUU CUA AUG UCU GAU GAG UUG GGG CUG CCU AAU UUG ACU GAA GGG AAG UCC AAG AAA CCA AAG ACC UUA GCC AAA GAA UGU  
 Leu Pro Glu Val Pro Tyr Asn Ala Phe Leu Leu Met Ser Asp Glu Leu Gly Leu Ala Asn Met Thr Glu Gly Lys Ser Lys Lys Pro Lys Thr Leu Ala Lys Glu Cys (309)  
 CUA GAA AAG UAC UCA ACA CUA CGG GAU CAA AUC CAC CCA AUA UUA AUA AUG AAA AGC GAA AAA CCU AAG CAA AAG UUC UUA UGG AAG UUG UGG AGG GAC UGU GUA AAU  
 Leu Glu Lys Tyr Val Leu Ser Thr Ser Glu Thr Asp Gln Leu Met Lys Ser Glu Lys Ala Asn Asp Ser Glu Lys Tyr Lys Val Ile Pro Ile Thr Asp Ser Leu Leu Ser Thr Asp (345)  
 ACA AUA AGU AAU GAG GAA ACA AGU AAC GAA UUA CAG AAA ACC AAU UAU GCC AAG UCG GCC ACA GGA GAU GGA UUA ACA UAC CAG AAA AUA AUG AAA GAA GUA CCA AUA  
 Thr Ile Ser Asn Glu Glu Thr Ser Asn Glu Leu Gln Lys Thr Asn Tyr Ala Lys TDP ATA Thr Gly Asp Gly Leu Thr Tyr Gln Lys Ile Met Lys Glu Val Ala Ile (381)  
 GAU GAC GAA ACA AUA UAC CAA GAA GAC CCC AAA UUA CCU AAA UGU AUA GUG GCU GCU UGG UUU CAA ACA GAG AAU AAU CUA UUG ACC AAU CUA UUG ACC AAU CUA  
 Asp Asp Lys Thr Met Tyr Gln Glu Thr Ser Glu Pro Lys Thr Val Ile Pro Asn Glu Lys Ser Cys Arg Val Ala Ala Thr Trp Val Gln Thr Glu Asp Ser Phe Leu Ser Thr Lys Arg (417)  
 GCC CUG GAU CUA CAA GAA AUA GGG CCA GAG CUA GCA CCC CUG GAG CAU GUA GGG ACU GAA AGA AGG AAA UAC UUU UUU AAU GAA AUC AAC UAC UGU AAG CCC UCU ACC  
 Ala Leu Asp Leu Pro Glu Ile Gly Pro Asp Val Ala Pro Val Glu His Val Gly Ser Glu Arg Arg Lys Tyr Phe Val Asn Glu Ile Asn Tyr Cys Lys Lys Ala Ser Thr (453)  
 GUU AUG AUG AAG UAU GUA CUU UUU CAC ACU UCA UUA UUA AAU GAA ACC AAU GCC ACC AUG GAA AAA UUA AAA GUA AUA CCA AUA ACC AAC AGA GUA UTA AAU GAA AAP  
 Val Met Lys Tyr Val Leu Phe His Thr Ser Leu Leu Asn Glu Ser Asn Ala Ser Met Gly Lys Tyr Lys Val Ile Pro Ile Thr Asp Arg Val Val Asp Glu Lys (489)  
 GGA GAA AGU UUU GAC AUA UUU UUU GGU CUG GCG UUU AAA GGG CAA UCU CAA CUG AGG GGA GAG ACU GAU GGU GUA ACA GAU GUG ACU UUC GAA UUU AGU AGU ACA GAU  
 Gly Glu Ser Phe Asp Ile Leu Tyr Gly Leu Ala Val Lys Gly Glr Ser His Leu Arg Gly Asp Thr Asp Val Val Thr Val Val Thr Phe Glu Phe Ser Ser Thr Asp (525)  
 CCC AUG GUG CAC UCA GGA AAG UGG CCA AAA UUU ACU GUA UUU AGA AAU GGU UCC UUA UUU GUG AGU GGA AGG GAA AAA UCU GUG UAC CUA UUA UGC GGA GUG AAU GGU  
 Pro Arg Val Asp Ser Gly Lys Trp Pro Lys Tyr Thr Val Phe Asp Arg Ile Gly Ser Leu Phe Val Ser Gly Arg Glu Tyr Lys Val Ser Thr Asp Phe Thr Ser Lys Ala Leu (561)  
 ACA AAC AAG AUC CAA AUG AAA UGG GGA AUG GAA GCU AGA AGA UGU CUG CUC CAA UCA AUG CAA CAA AHS GAA CEA AAU GAU GAU CAA GAA UCA UCG AUA CAA GGA UAU  
 Thr Asn Lys Ile Gln Met Lys Trp Gly Met Glu Ala Arg Cys Leu Leu Gln Ser Met Glr Trp Met Glu Ala Ile Val Asp Glr Glu Ser Ser Ile Thr Glu Tyr (597)  
 GAC AUG ACC AAA CUC UHU UUC AAC GGA GAC AGA GUG AAU AGU CCC AAA AUU UUF AGU AAU UGG ACU CAA GAA GAA UUA CUA AAA GUA AAA UCC UUU GGG AAA CEA CUA  
 Asp Met Thr Lys Ala Cys Phe Lys Gly Asp Arg Tyr Val Asn Ser Phe Cys Lys Thr Phe Ser Ile Gly Thr Glr Glu Gly Lys Leu Val Lys Gly Ser Tyr Cys Arg Val Asn Glu (633)  
 AGA GUA AUA UUC ACC AAA UGU AUG CAC UAU GUA UUU GAA AAU GCC CAA UUG GAG GGG UUU AGU GCC GAA UCU AGC AGA CUC CUA CUG UUA AAU CAG GCA UUA ARG  
 Arg Val Ile Phe Thr Lys Lys Leu Met His Tyr Val Phe Gly Asp Ala Glr Leu Glu Gly Phe Ser Ala Glu Ser Arg Arg Leu Leu Leu Leu Ile Glr Ala Leu Lys (669)  
 GAC GAA AAG GGC CCU UGG GUA UUC GAC UUA GAG GGA AUG UAU UCU GGA AUA GAA GAA UGU AUC AAU AAC AAC CCU UUG GUA AUA CAG AGU GCA UAC UGG UUU AAU GAA  
 Asp Arg Lys Gly Pro Trp Val Phe Asp Leu Glu Gly Met Tyr Ser Cys Ile Glu Glu Cys Ile Ser Asp Asn Pro Trp Val Ile Glr Ser Ala Tyr TDP Phe Asp Glu (705)  
 UGG UUG GGC UUU GAA AAA GAG GGG AGU AAA GUA UUA GAA UCA AUA GAU GAA AUA AUG GAU CAA UGAAGAAGCGCAUGCGCCUACAAUUGUAGUACUAUUGUACUUAACACU  
 Trp Leu Gly Phe Glu Lys Glu Gly Ser Lys Val Leu Glu Ser Ile Asp Glu Ile Met Asp Glu (728)  
 CARUAAAAAGAAUUGAAGAUUAAARAAUUGCAGUGUUGUULUACU

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  $\text{NH}_2$ -terminal half of



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  $\text{NH}_2$ - 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  $\text{NH}_2$ -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  $\text{NH}_2$ -terminal end. Thus, the COOH-terminal half of the polypeptide may be more important to the PA protein's function than the  $\text{NH}_2$ -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.).

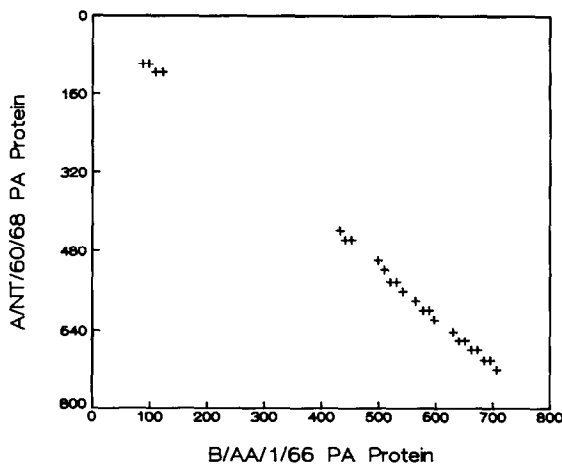


Fig. 3. Matrix comparison plot of B/Ann Arbor/1/66 and A/NT/60/68 PA polypeptides. Each symbol represents a site where at least 20 out of 40 amino acids were conserved. The numbers on the  $x$  and  $y$  axes represent the amino acid position in from the  $\text{NH}_2$  terminus of each protein.

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<sub>1</sub>). These results are echoed in the percentage of possible non-silent and silent changes observed. The NS RNA was interesting in that the NS<sub>1</sub> gene had the highest percentage of non-silent changes, while NS<sub>2</sub> had the third lowest percentage of non-silent changes. NS<sub>2</sub> protein also had higher cross-type homology than did NS<sub>1</sub> 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<sub>1</sub> polypeptide shows as little variation as the PB1 polypeptide between these two viruses. Interestingly, the M<sub>1</sub> polypeptide does not show a correspondingly high cross-type homology with M<sub>1</sub> 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<sub>1</sub> 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

B/Lee/40 cRNA (+) 5'-  
 B/Ann Arbor/1/66 cRNA (+) 5'-AGCAGAAAGCGGCCUUAAG AUG AAU AUA AAU CCU UAU UUU CUC UUC UUA GAU GUA CUA AUA CAG GCA GCA AUA UCA ACA ACA UUC CCA (90)  
 B/Ann Arbor/1/66 P81 Protein Met Asn Ile Asn Pro Tyr Phe Leu Phe Ile Asp Val Pro Ile Gln Ala Ala Ile Ser Thr Thr Phe Pro (23)  
 B/LEE/40 P81 Protein

UAC ACC GGU GUU CCC CCU UUU CCU CAA GGA ACG GCA ACA GGC UAC ACA AUA GAC ACC GUG AUU AGA ACA CAU GAG UAC UCA AAC AAG GAA AAA CAA UAC AUU UCU GAA (198)  
 Tyr Thr Gly Val Pro Pro Tyr Ser His Gly Thr Gly Thr Gly Tyr Thr Ile Asp Thr Val Ile Arg Thr His Glu Tyr Ser Asn Lys Gly Lys Gln Tyr Ile Ser Asp (59)

GUU ACA GGA UGU GCA AUG GUA GAA CCA ACA AAU GGG CCA UUA CCC GAA GAU AAU GAG CCG AGU GGC UUU GCA CAA UUG GAU UGU GUU CUG GAG GCU UUG GAU ACA AUG (306)  
 Val Thr Gly Cys Ala Met Val Asp Pro Thr Asn Gly Pro Leu Pro Glu Asp Asn Glu Pro Ser Ala Tyr Ala Gln Leu Asp Cys Val Leu Glu Ala Leu Asp Arg Met (95)

GAU GAA GAA CAU CCA CGU CUG UUU CAA GCA GGC UCA CAA AAU GCC AUG GAG GCA CUA AUG GUC ACA ACU GUA GAG AAA UUA ACC CAG GGG AGG CAG ACU UUU UAU UGG (414)  
 Asp Glu Glu His Pro Gly Leu Phe Gln Ala Thr Ala Ser Gln Asn Ala Met Glu Ala Leu Met Val Thr Thr Val Asp Lys Leu Thr Gln Gly Arg Gln Thr Phe Asp Trp (131)

ACA GUG UGC AGA AAC CAA CCU GCU GCA ACG GCA CUG AAC ACA ACA AUA ACC UCU UUU AGG UUG AAU GAU UUG AAU GGA GCC CAC AAG GGU GGA UUA GUA CCC UUU UUG (522)  
 Thr Val Cys Arg Asn Gln Pro Ala Ala Thr Ala Leu Asn Thr Thr Ile Thr Ser Phe Arg Leu Asn Asp Leu Asn Gly Ala Asp Lys Gly Gly Leu Val Pro Phe Cys (167)

CAA GAU AUC AUU GAC UCA UUG GAC AAA CCU GAA AUG ACU UUC UUC AUA GUA AAG AAU AUA AAG AAA PAA UUG CCU GCU AAA ACC AGA AAG GGU UUC CUC UUA AAG AGA (630)  
 Gln Asp Ile Ile Asp Ser Leu Asp Lys Pro Glu Met Thr Phe Ser Val Lys Asn Ile Lys Lys Lys Leu Pro Ala Lys Asn ACA AAG GAG GUC UUC UUC UUA AAG AGA (203)

AUA CCA AUG AAG GUA AAA GAG ACA AUA ACC AGA GGU GAA UAC ACA GCA UUA UCA UUA AAC ACA AUG ACA AAA CAU GCU CAA GAA GGC AAA CUA AAA ACA ACA (738)  
 Ile Pro Met Lys Val Lys Asp Arg Ile Thr Ala Thr Ala Leu Ser Leu Asn Thr Met Thr Lys Asp Ala Glu Arg Gly Lys Leu Lys Arg Arg (239)

SCA AUU GCC ACC GCU GGG AUA CAA AUG ASA GGG UUU GUA UUA GUA GUU GAA AAC UUG GCU AAG AAU AUC UGU GAA AAU CUA GAG CAA AGU GGU UUC CUA GGU GGA (846)  
 Ala Ile Ala Thr Ala Gly Ile Gln Ile Arg Gly Phe Val Leu Val Val Glu Asn Leu Ala Lys Asn Ile Cys Glu Asn Leu Glu Gln Ser Gly Leu Pro Val Gly Gly (275)

AAC GAG AAG AAG GCC AAA CUG UCA AAU GCA GUG GCU AAA AUG CUC AGU AAC UGC CCA CCA GGA GGG AUC AGC AUG ACU UUG AGC GGA CAC AAU ACU AAA UGG AAU GAA (954)  
 Asn Glu Lys Lys Ala Lys Leu Ser Asn Ala Val Ala Lys Met Leu Ser Asn Cys Pro Pro Gly Gly Ile Ser Met Thr Val Thr Gly Asp Asn Thr Lys Trp Asn Glu (311)

UGC UUA AAU CCA AGA AUC UUU UUG GCU AUG ACU GAA GAA AUA ACC AGA GAC AGC CCA AUU UGG UCG GAU UUU UGU AGU AUA GCA CCG GUC UUC UCC AAU AAA (1062)  
 Cys Leu Asn Pro Arg Ile Phe Leu Ala Met Thr Glu Arg Ile Thr Arg Asp Ser Pro Ile Trp Phe Arg Asp Phe Cys Ser Ile Ala Pro Val Leu Phe Ser Asn Lys (347)

AUA GCA AGA UUG GGA AAA GGG UUC AUG AUA ACA AGU AAA ACA AAA AGA CUG AAG GCU CAA AUA CCU UGU CCC GAU CUG UUA ACC AUA CUA UUA GAA AGA UAU AAU GAA (1170)  
 Ile Ala Arg Leu Gly Lys Gly Phe Met Ile Thr Ser Lys Thr Lys Arg Leu Lys Ala Gln Ile Pro Cys Pro Asp Leu Phe Asn Ile Pro Leu Glu Arg Tyr Asn Glu (383)

GAA ACA AGC GCA AAA UUG AAA AAG CUG AAA CCA UUC UUC AAU GAA GAA GCA ACG GCA UCU UUG UCG CCU GGG AUG AUG AUG GGA AUG UUU AAU AUG CUA UCU ACC GUC (1278)  
 Glu Thr Arg Ala Lys Leu Lys Lys Lys Leu Pro Phe Phe Asn Glu Gly Thr Ala Ser Leu Ser Pro Gly Met Met Gly Met Phe Asn Met Leu Ser Thr Val (419)

UUG GGA GUA GCC GCA CUA GGG AUC AAA AAC AUU GGA AAC AGA GAA UAU UUG GAU GGA CUG CAA UCU UCU Ser Asp AAU UUU GCU CUG UUU GUU AAU GCA AAA GAA GAA (1386)  
 Leu Gly Val Ala Ala Leu Gly Ile Lys Asn Ile Gly Asn Arg Glu Tyr Leu Trp Asp Gly Leu Gln Ser Ser GAA Asp Phe Ala Leu Phe Val Asn Ala Lys Asp Glu (455)

GAG ACA UGU AUG GAA GGA AUA AAC GAU UUU UAC CGA ACA UGU AAG CUA UUG GGA AUA AAC AUG AGC AAA AAG AAA AGU AUG AAU GAA ACU GAA UGU UUU GAA UUU (1494)  
 Thr Cys Thr Met Glu Gly Ile Arg Ser Phe Tyr Arg Thr Cys Lys Ser Thr Lys Lys Lys Ser Tyr Cys Asn Glu Thr Gly Met Phe Glu Phe (491)

ACA AGC AUG UUC UAC AGA CAU CGA UUU UUA UCU UAU UUU GCA AUG GAA CUA CUC UCA UUU GGG GUU GCU GUA CUA AAH GAA UCA GCA CUA UUC UCA AUA GGA AUG ACA (1502)  
 Thr Ser Met Phe Tyr Arg Asp Gly Phe Val Ser Asn Phe Ala Met Glu Leu Pro Ser Phe Gly Val Ala Gly Val Asn Glu Ser Ala Asp Met Ala Ile Gly Met Thr (527)

AUA AUA AAG AAC AAU AUG AUC AAC AAU GGG AUG GGU CCA GCA ACG GCA CAA GCA GCC AUA CAA UUA UUC AUA GCU GAU UAU ACA UAC ACC UAU AAA UGC CAC AGG GGA (1710)  
 Ile Ile Lys Asn Asn Met Ile Asn Asn Gly Met Gly Pro Ala Thr Ala Gln Thr Ala Ile Gln Leu Phe Ile Ala Asp Tyr Arg Tyr Thr Tyr Lys Cys His Arg Gly (663)

GAU UCC AAA GUG GAA GGA AAG AGA AUG AAA AUU AUA AAG GAG CUA UUG GAA AAC ACU AAA GGA AGA GAU GGC CUA UUA GUA GCA GAU GGU GGC CCU AAC AUU UAC AAU (1818)  
 Asp Ser Lys Val Val Glu Gly Lys Arg Met Lys Ile Ile Lys Glu Leu Trp Glu Asn Thr Lys Gly Arg Asp Gly Leu Leu Val Ala Asp Gly Gly Pro Asn Ile Tyr Asn (599)

UUG AGA AAC UGU CAU AUU CCA GAA AUA GUA UUA AAG UAC AAC CUA AUG GAC CCU GAA UAC AAA GGG CCG UUA CUG CAU CUC CAA AAU CCC UUU GUA GGA CAU UUG UCU (1926)  
 Leu Arg Asn Leu His Ile Pro Glu Ile Val Leu Lys Tyr Asn Leu Met Asp Pro Glu Tyr Lys Gly Arg Leu Leu His Pro Gln Asn Pro Phe Val Gly His Leu Ser (635)

AUU GAG GGC AUC AAA GAG GCA GAU AUA ACC CCA CCA GAU GGU CCA AUA AAG AAA AUG GAC UAU GAU GGG GUA UCU GGA ACU CAU AGU UGG AGA ACC AAA AGG ACA AGA (2034)  
 Ile Glu Gly Ile Lys Glu Ala Asp Ile Thr Pro Ala His Gly Pro Ile Lys Lys Met Asp Tyr Asp Ala Val Ser Gly Thr His Ser Trp Arg Thr Lys Arg Asn Arg (671)

UCU AUA CUA ACC ACU GAU CAG AGG AAC AUG AUU CUU GAG GAA CAA UGC UAC GCU AAG UGU UGC AAC CUC UUU GAG GCU UUU AAC AGU DCA UCA UAC AGG AAA CCA (2142)  
 Ser Ile Leu Asn Thr Asp Gln Arg Asn Met Ile Leu Glu Glu Gln Cys Tyr Ala Lys Cys Cys Asn Leu Phe Glu Ala Cys Phe Asn Ser Ala Ser Tyr Arg Lys Lys Pro (707)

YAL GGU CAG CAC AGC AUG CUU GAG CAG AUG GCC CAC AUA UGA AGA UAU GAA CCA CGA CUA GAU UAU GAA UCA GGA AGA AUG UCA AAG GAU GAU UUU GAG AUA GCA AUG (2250)  
 Val Gly Glu His Ser Met Leu Glu Ala Met Glu His Arg Leu Arg Met Asp Ala Arg Leu Asp Tyr Glu Ser Gly Arg Met Ser Lys Asp Asp Phe Glu Lys Ala Met (743)

GCN CAC UUU GUG GAG AUU UGG CAC AUA UAAGUUGAAGAUUGUUGGUGUUUUGUUGCAUUGAUAACUAGCGGUGUACAAUAUAUAAUAAGAAAGGUCUGUUCUUCU (2388)  
 Ala His Leu Gly Glu Ile Gly His Ile (752)  
 Tyr Met (752)

Fig. 4. Nucleotide and predicted polypeptide sequences of the P81 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 <sup>a</sup> 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 <sub>1</sub>	65/1191 (5.5)	4/248 (1.6)	4/624 (0.64)	31/271 (11.4)
N <sub>1</sub>	61/1096 (5.6)	22/281 (7.8)	22/706 (3.1)	26/302 (8.6)
NS <sub>2</sub>		3/122 (2.5)	3/315 (0.95)	9/131 (6.9)

<sup>a</sup> 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 (Kendirim 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<sub>1</sub> and NS<sub>2</sub> 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<sub>1</sub>) 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)