

Sequence Comparison of Wild-Type and Cold-Adapted B/Ann Arbor/1/66 Influenza Virus Genes

DAN C. DEBORDE,¹ ARMEN M. DONABEDIAN, M. LOUISE HERLOCHER, CLAYTON W. NAEVE,* AND HUNEIN F. MAASSAB²

Department of Epidemiology, School of Public Health, University of Michigan, 109 Observatory St., Ann Arbor, Michigan, 48109, and

**Department of Virology and Molecular Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, 38101*

Received August 10, 1987; accepted November 24, 1987

Consensus sequences for both wt and ca B/Ann Arbor/1/66 viral PB2, PB1, PA, NP, M, and NS genes were directly determined from vRNA using a combination of chemical and chain-termination sequencing methods. There were 105 sites of difference between the wt and ca sets of these six RNA genes. The differences resulted in 26 amino acid substitutions distributed over the six proteins. The sequence changes were compared to the sequences of other known influenza type B wt viruses to pinpoint those changes that were unique to the ca B/Ann Arbor/1/66 virus. Of the 26 amino acid differences, only 11 were unique to the cold-adapted virus. These unique sites were distributed among five of the six genes. The NS protein had no amino acid substitutions. The sequence changes are discussed in terms of their probable mode of origin and selection, and in terms of their importance to the cold-adapted, temperature-sensitive, and attenuation phenotypes of ca B/AA/1/66 virus. The sequence and organization of the PB2 gene and predicted protein are also given. The PB2 gene was 2396 nucleotides long, and it encoded a predicted protein of 770 amino acids with a molecular weight of 88,035 Da for the wt virus and 88,072 Da for the ca virus. Both proteins were predominantly hydrophilic, and each had an overall charge of +24.5 at pH 7.0. © 1988 Academic Press, Inc.

INTRODUCTION

The cold-adapted (ca) B/Ann Arbor/1/66 (B/AA/1/66) influenza virus line was recently chosen for use in the production of ca reassortant live virus vaccines for testing in man (Maassab *et al.*, 1985). The rationale behind the production of these ca reassortant vaccines is based on the experience amassed from successful trials using influenza type A ca reassortant vaccines. The optimal vaccine reassortant derives the viral genes encoding the two surface antigens, the hemagglutinin (HA) and the neuraminidase (NA), from the wild-type (wt) parent. Its other six genes are derived from the ca donor virus, and it is these genes which carry the characteristic ca, temperature-sensitive (ts), and attenuated (att) phenotypes. Thus, for the purposes of a safe and effective influenza A virus vaccine, all that is required is the assured presence of all six nonsurface genes from the ca A/Ann Arbor/60 donor virus (for review, Maassab *et al.*, 1985). Laboratory studies have confirmed this phenotypic transfer to influenza B reassortant viruses by the six nonsurface genes of ca B/AA/1/66 in primary chick kidney (PCK) tissue culture and in ferrets (Maassab *et al.*, 1986).

Human trials have been done and are being done to extend the validity of this approach for ca influenza B reassortant vaccine viruses (Davenport *et al.*, 1977; Monto *et al.*, 1982; Reeve *et al.*, 1982; Keitel *et al.*, 1986). While the transfer of these phenotypes is guaranteed by the transfer of all six nonsurface genes, the total number of mutations which brings about the ca, ts, and att phenotypes and their locations in the genome are not known. It is also not clear whether these phenotypes are engendered only by the mutations that were selected by the cold-adaptation process, or by other mutations selected in an unknown manner.

Initial studies used various methods to reveal that each of the eight genes of the ca A/Ann Arbor/6/60 (A/AA/6/60) donor virus displayed some difference from the corresponding gene of the wt A/AA/6/60 virus from which it was derived (Cox *et al.*, 1986). Such genomic-wide changes are also probable with the influenza B virus system since just the electrophoretic comparison of the wt and ca B/AA/1/66 viral RNA segments demonstrated that changes occur in at least six of the eight genes (Maassab *et al.*, 1985). Only the NA and NS ca and wt gene pairs did not migrate differently under the various conditions tested. Other studies for both influenza A and B viruses (Snyder *et al.*, 1987; Donabedian *et al.*, 1987) have used a genetic approach in an attempt to analyze the particular gene(s) associated with the three phenotypes. They

¹ Present address: Department of Microbiology, University of Montana, Missoula, MT 59812.

² To whom requests for reprints should be addressed.

isolated reassortants containing all wt influenza genes except for one gene from the ca parent virus. These "single-gene" reassortants were then analyzed for expression of the ca, ts, and att phenotypes. Single-gene analysis works reasonably well for the ts and att phenotypes, but less well for the ca phenotype since the ability to replicate efficiently at 25°, not a normal wt virus characteristic, can be inhibited by the presence of any wt gene that does not function well at this sub-optimal temperature. In addition, some ca viral genes may act only in concert with one or more of the other ca donor virus genes to express these phenotypes, and hence, not be identified by single-gene studies.

In an effort to determine the total extent of the sequence changes that were selected in the six non-surface genes during the cold-adaptation of B/AA/1/66 wt virus, we sequenced both the wt and ca B/AA/1/66 strains' PB2, PB1, PA, NP, M, and NS genes. These sequence data provide a complete catalog of the nucleotide changes that existed between these two virus lines, and these changed sites were compared to other known wt influenza B virus sequences to pinpoint changes unique to ca B/AA/1/66 viral genes. These data coupled with information derived from the single-gene reassortment studies mentioned above will provide a base of knowledge for future experiments which will attempt to link particular mutations to specific phenotypes. (See following article, Donabedian *et al.*, 1988).

MATERIALS AND METHODS

Chemicals and enzymes

Oligodeoxynucleotide primers were purchased from the University of Michigan DNA Synthesis Facility or were made by one of the authors, C. W. Naeve, with an Applied Biosystems Model 380A DNA Synthesizer. The deoxynucleoside triphosphates, dimethyldichlorosilane, γ -(methacryloxy)propyltrimethoxysilane, diethyl pyrocarbonate (DEP), and sodium borohydride were purchased from Sigma Chemical Co., St. Louis, Missouri. The dideoxynucleoside 5'-triphosphates were obtained from Boehringer-Mannheim, Indianapolis, Indiana. Reverse transcriptase was the standard preparation derived from avian myeloblastosis virus by Life Sciences, Inc., St. Petersburg, Florida. Terminal deoxynucleotidyl transferase (TdT) was either from Life Sciences, Inc., or from Bethesda Research Laboratories, Gaithersburg, Maryland. T4 RNA ligase and T4 polynucleotide kinase were also from Bethesda Research Laboratories. Dimethyl sulfate and aniline were from Aldrich Chemical Co., Milwaukee, Wisconsin. Anhydrous hydrazine was purchased from Fisher Sci-

entific Co., Fair Lawn, New Jersey. [α - 32 P]deoxyadenosine 5'-triphosphate, tetra-(triethylammonium) salt was purchased from Dupont, NEN Research Products, Boston, Massachusetts, while [γ - 32 P]adenosine 5'-triphosphate was from ICN Radiochemicals, Irvine, California.

Viruses and tissue culture

The viruses studied in these experiments were wt and ca B/AA/1/66 strains, and wt B/Melbourne/43 and B/AA/1/86 viruses. The cold-adaptation of the wt B/AA/1/66 virus together with the passage histories for both wt and ca viruses are given in Fig. 1. The original isolate was cultured in primary chick kidney cells from a human nasal wash sample (Maassab *et al.*, 1986). It was not cloned, and therefore the starting virus was a population of individual viruses probably of related but varying individual sequences. The production of viral stocks for the isolation of vRNA for sequencing was performed exactly as presented previously (DeBorde *et al.*, 1986).

Preparation of vRNA

The isolation of vRNA was performed as presented previously (DeBorde *et al.*, 1986), except that the RNA from virus grown in 60–100 embryonated eggs was

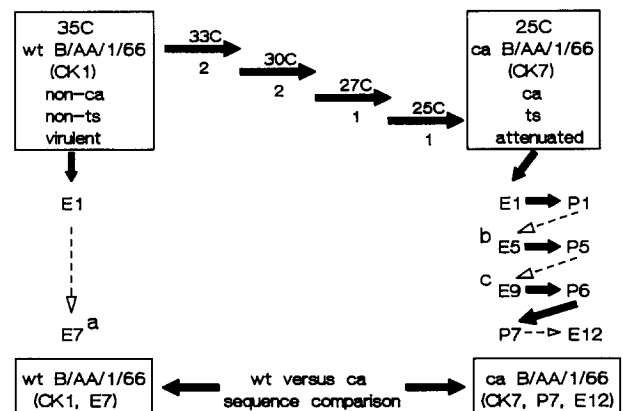


FIG. 1. Cold-adaptation and passage history of wt and ca B/AA/1/66 viruses. The cold-adaptation process is represented by the staggered arrows connecting the top two boxes. The passages were done in PCK cells at the temperatures indicated above each arrow. The number below the arrow represents number of passages at each temperature. The dashed arrows represent a number of replicative steps as follows: ^aConsecutive embryonated egg passages of the wt virus pool at high dilution. ^bConsecutive plaquing and egg amplification of the ca virus pool. ^cConsecutive embryonated egg passages of the ca virus pool at high dilution. The bottom two boxes represent the wt and ca virus pools that were sequenced in this paper.

resuspended in a final volume of 50 μ l, not 50 ml as erroneously stated in that paper.

End-labeling of vRNA

Purified vRNA (30 to 50 μ g) and 100 μ Ci of [32 P]cytidine bis-phosphate (pCp) were dried down together in a vacuum concentrator and resuspended in 5 μ l of 2 \times RNA ligase buffer, (100 mM HEPES, pH 7.5, 20 μ g/ml BSA, 6.6 μ M ATP, 30 mM MgCl₂), 1 μ l dimethylsulfoxide, and 1 μ l 0.5 mM ATP. Thirty to thirty-five pCp ligation units of T4 RNA ligase was added, and the total volume was adjusted to 10 μ l with H₂O if necessary. The reaction was incubated at 4° overnight. Two hundred microliters of 0.3 M sodium acetate was added to the ligation mixture, followed by 600 μ l of 95% ethanol, and the RNA was precipitated at -20° for at least 4 hr. The RNA was pelleted for 5 min in a microfuge, washed once with 70% ice-cold ethanol, pelleted once more, and dried in a vacuum concentrator. The RNA was resuspended in H₂O and mixed with an equal volume of loading buffer (0.2 \times TBE, 0.2% SDS, 20% sucrose, and 0.01% bromophenol blue).

Isolation of individual vRNA segments

For the chemical sequencing reactions, the purified and [32 P]pCP 3' end-labeled vRNA was separated into its eight segments on a 3% polyacrylamide gel in 1 \times TBE, (0.09 M Trizma base, 0.09 M boric acid, and 2 mM Na₂EDTA) (Peacock and Dingman, 1968) with 0.01% sodium dodecyl sulfate (SDS). This level of SDS still protected the RNA from any contaminating RNases during the long electrophoretic run, but did not preclude visualization of the RNA with ethidium bromide. These conditions separated all segments except RNAs 4 and 5 (NP and HA). To separate these two RNAs, 6 M urea was added to the above gel. The RNA was electrophoresed at 37° for 16 hr at 200 V, constant voltage for the nondenaturing gel, and at 37° for 21 hr at 240 V, constant voltage for the 6 M urea gel. After electrophoresis the vRNA segments were visualized by immersing the gel in 1 liter of 1 \times TBE buffer containing 0.01% SDS and 1 μ g/ml ethidium bromide for 25 min at room temperature. The gel was then removed from the ethidium bromide solution, rinsed with 1 \times TBE buffer, and placed on a sheet of plastic wrap on a uv transilluminator. The desired RNA was excised from the gel with little or no excess gel material. The end-labeled vRNA segments were then eluted from the gel slices using an International Biotechnologies, Inc. (New Haven, CT), unidirectional electroeluter. The running buffer was 0.5 \times TBE buffer plus 0.01% sarcosine. The high salt re-

tion buffer was 3.0 M sodium acetate with 0.03% sarcosine and 0.01% bromophenol blue. Elution proceeded at 100 V, constant voltage, for 1-hr intervals. The smaller segments, NS and M, were completely eluted in 1 hr; NP, HA, and NA segments took 2 to 3 hr; and the three polymerase segments, PB2, PB1 and PA took 3 to 4 hr. Completion of elution was monitored by Cerenkov counting of the original gel slice and hourly samples. Recovery was 70 to 90% of the original sample. The isolated vRNA segments were precipitated at -20° overnight with 3 vol of 95% ethanol.

Chain-termination sequencing reactions and gels

These procedures, including the use of TdT to eliminate ambiguities, were performed exactly as described previously (DeBorde *et al.*, 1986). Only one compression area was encountered. It was in the PB2 gene and was resolved by running the standard sequence reaction products on a gel that was a 7 M urea, 8% polyacrylamide gel modified by the addition of 40% deionized formamide (Martin, 1987). To determine the wt and ca sequences, the sequencing reaction products were run side by side on the same electrophoretic gel. Several independent runs, with and without TdT enzyme treatment, were performed for each primer.

Chemical sequencing reactions and gels

These procedures were taken from Peattie (1979) and followed exactly, including the electrophoresis and autoradiography procedures. Recommendations given in Peattie (1979) for the safe use and disposal of sodium borohydride, aniline, DEP, and hydrazine were followed.

Sequence analysis programs

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

RESULTS

Figures 2-7 present the cRNA (plus strand) sequences and the predicted polypeptide sequences for the PB2, PB1, PA, NP, M, and NS genes. Each gene sequence was determined by a combination of chemical and chain-termination sequencing procedures (see Materials and Methods for details and/or

B/AA/1/66 ca PB2 cRNA (+) 5'.....A (50).....
 B/AA/1/66 wt PB2 cRNA (+) 5'-AGCAGAAGCGAGCGUUUCAAG AUG ACA UUG GCC AAA AUU GAA UUG UUG AAA CAA CUG UUA AGG GAC AAU GAA GCC AAA ACG GUA UUG AAA (92)
 B/AA/1/66 wt PB2 Protein N-terminus Met Thr Leu Ala Lys Ile Glu Leu Leu Lys Thr Lys Leu Leu Arg Asp Asn Lys Thr Val Leu Lys (23)
 B/AA/1/66 ca PB2 Protein N-terminus

CAA ACA ACG GUA GAC CAA UAU AAC AUA UUA AGA AAA UUC AAU ACA UCA AGA AUU GAA AGG AAC CCU UCA UUA AGG AUG AAG UGG GCC AUG UGU UCU AAU UUU CCC UUG GCU (203)
 Gln Thr Thr Val Asp Gln Tyr Asn Ile Ile Arg Lys Phe Asn Thr Ser Arg Ile Glu Arg Asn Pro Ser Leu Arg Met Lys Trp Ala Met Cys Ser Asn Phe Pro Leu Ala (60)

.....A (256).....
 CUG ACC AAG GGU GAU AUG GCA AAU AGA AUC CCU UUG GAA UAC AAG GGA AUA CGA CUU AAA ACA AAU GCU GAA GAC AUA GGA ACC AAA GGC CAA AUG UGC UCA AUA GCA GCA (314)
 Leu Thr Lys Gly Asp Met Ala Asn Arg Ile Pro Leu Glu Tyr Lys Gly Ile Arg Leu Lys Thr Asn Ala Glu Asp Ile Gly Thr Lys Gly Gln Met Cys Ser Ile Ala Ala (97)
Gln (78).....

.....A (341).....U (359).....
 GUU ACC UGG UGG AAU ACA UAU GGA CCC AUA GGA GAU ACU GAA GGG UUC GAA AAG GUC UAC GAA AGC UUU UUU CUC AGA AAG AUG AGA CUU GAC AAU GCC ACU UGG GCC CGA (425)
 Val Thr Trp Trp Asn Thr Tyr Gly Pro Ile Gly Asp Thr Glu Gly Phe Glu Lys Val Tyr Glu Ser Phe Phe Leu Arg Lys Met Arg Leu Asp Asn Ala Thr Trp Gly Arg (134)

.....U (476).....
 AUA ACU UUU GGC CCA GUU GAA AGA GUG AGA AAA AGG GUA CUG CUA AAC CCC CUC ACC AAG GAA AUG CCU CCA GAU GAA GCG AGC AAU GUG AUA AUG GAA AUA UUG UUC CCU (536)
 Ile Thr Phe Gly Pro Val Glu Arg Val Arg Lys Arg Val Leu Leu Asn Pro Leu Thr Lys Glu Met Pro Pro Asp Glu Ala Ser Asn Val Ile Met Glu Ile Leu Phe Pro (171)

.....A (572).....U (626).....
 AAA GAA GCA GGA AUA CCA AGA GAA UCU ACU UGG AUG CAU AGG GAA CUG AAA GAA AAA AGA GAA AAA UUG AAA GGA ACG AUG AUA ACC CCC AUU GUA CUG GCA UAC AUG (647)
 Lys Glu Ala Gly Ile Pro Arg Glu Ser Thr Trp Met His Arg Glu Leu Ile Lys Glu Lys Arg Glu Lys Leu Lys Gly Thr Met Ile Thr Val Ala Ser Asn Pro Leu Glu Leu Ala (208)
Ile (183).....

.....C (728).....
 CUU GAG AGA GAA CUG GUU GCC CGA AGA AGG UUC CUG CCA GUG GCA GGA GCA ACA UCA GCC GAG UUC AUA GAA AUG CUA CAU UGC UUA CAA GGU GAA AAU UGG AGA CAA AUA (758)
 Leu Glu Arg Glu Leu Val Ala Arg Arg Arg Phe Leu Pro Val Ala Gly Ala Thr Ser Ala Glu Phe Ile Glu Met Leu His Cys Leu Gln Gly Glu Asn Trp Arg Gln Ile (245)

.....A (828).....
 UAU CAC CCA GGA GGG AAU AAA CUA ACU GAA UCU AGG UCU CAA UCA AUG AUU GUA GCU UGU AGA AAA AUA GUC AGA AGA UCA AUA GUC GCA UCA AAC CCA CUA GAG CUA GCU (869)
 Tyr His Pro Gly Gly Asn Lys Leu Thr Glu Ser Arg Ser Gln Ser Met Ile Val Ala Cys Arg Lys Ile Val Arg Arg Ser Ile Val Ala Ser Asn Pro Leu Glu Leu Ala (282)
Ile (269).....

.....U (899).....
 GUA GAA AAU GCA AAC AAG ACU GUG AUA GAC ACU GAA CCU UUA AAA UCA UGU CUG GCA GCC AUA GAC GGA GGU GAU GUA GCC UGU GAC AUA AUA AGA GCU GCA UUA GGA CUA (980)
 Val Glu Ile Ala Asn Lys Thr Val Ile Asp Thr Glu Pro Leu Lys Ser Cys Leu Ala Ala Ile Asp Gly Gly Asp Val Ala Cys Asp Ile Ile Arg Ala Ala Leu Gly Leu (319)

.....G (983).....C (1008).....(1052).....U.....A (1058).....(1088).....A.....
 AAA AUC AGA CAA AGA CAA AGA UUU GGA AGG CUU GAA CUA AAG AGA AUA UCA GGA AGA GGA UUC AAA AUA GAC GAA GAG AUA UUA AUC GGG AAC GGA ACA AUA CAG AAG AUU (1091)
 Lys Ile Arg Gln Arg Gln Arg Phe Gly Arg Leu Glu Leu Lys Arg Ile Ser Gly Arg Gly Phe Lys Asn Asp Glu Glu Ile Leu Ile Gly Asn Gly Thr Ile Gln Lys Ile (356)

.....C (1160).....A (1184).....
 GGA AUA UGG GAC GGA GAA GAG GAG UUC CAU GUA AGA UGU GGU GAA UGC AGG GGA AUA UUA AAA AAG AGU AAA AUG AGA AUG GAA AAA CUA CUG AUA AAU UCA GCC AAA AAG (1202)
 Gly Ile Trp Asp Gly Glu Glu Glu Phe His Val Arg Cys Gly Glu Cys Arg Gly Ile Leu Lys Lys Ser Lys Met Arg Met Glu Lys Leu Ile Asn Ser Ala Lys Lys (393)

.....A (1313).....
 GAG GAC AUG AAA GAU UUA AUA AUC UUG UGC AUG GUA UUU UCU CAA GAC ACU AGG AUG UUC CAA GGA GUG AGA GGA GAA AUA AAU UUU CUU AAU CGA GCA GGC CAA CUU UUA (1313)
 Glu Asp Met Lys Asp Leu Ile Ile Leu Cys Met Val Phe Ser Gln Asp Thr Arg Met Phe Gln Gly Val Arg Gly Glu Ile Asn Phe Leu Asn Arg Ala Gly Gln Leu Leu (430)

.....G (1334).....
 UCU CCA AUG UAC CAA CUC CAA CGA UAU UUU UUG AAU AGG AGC AAC GAC CUU UUU GAU CAA UGG GGG UAU GAG GAA CCA CCC AAA GCA AGU GAA CUA CAU GGG AUA AAU GAA (1424)
 Ser Pro Met Tyr Gln Leu Gln Arg Tyr Phe Leu Asn Arg Ser Asn Asp Leu Phe Asp Gln Trp Gly Tyr Glu Pro Pro Lys Ala Ser Glu Leu His Gly Ile Asn Glu (467)

.....A (1535).....
 UUA AUG AAU GCA UCU GAC UAU ACG UUG AAA GGG GUU GUA GUA ACA AAA AAU GUG AUU GAU GAC UUU AGU UCU ACU GAA ACA GAA AAA GUA UCU AUA ACA AAA AAU CUA AGU (1535)
 Leu Met Asn Ala Ser Asp Tyr Thr Leu Lys Gly Val Val Val Thr Lys Asn Val Ile Asp Asp Phe Ser Ser Thr Glu Thr Glu Lys Val Ser Ile Thr Lys Asn Leu Ser (504)

.....A (1646).....
 UUA AUA AAA AGG ACU GGG GAA GUC AUA AUG GGG GCU AAU GAC GUA AGU GAA UUA GAA UCA CAA GCA CAG CUA AUG AUA ACA UAU GAU ACA CCU AAG AUG UGG GAG AUG GGA (1646)
 Leu Ile Lys Arg Thr Gly Glu Val Ile Met Gly Ala Asn Asp Val Ser Glu Leu Glu Ser Gln Ala Gln Leu Met Ile Thr Tyr Asp Thr Pro Lys Met Trp Glu Met Gly (541)

.....A (1757).....
 ACA ACC AAA GAA CUG GUG CAA AAC ACC UAC CAA UGG GUG CUA AAA AAU UUG GUA ACA CUG AAG GCU CAG UUU CUU CUG GGA AAA GAA GAC AUG UUC CAA UGG GAU GCA UUU (1757)
 Thr Thr Lys Glu Leu Val Gln Asn Thr Tyr Gln Trp Val Leu Lys Asn Leu Val Thr Leu Lys Ala Gln Thr Leu Leu Gly Lys Glu Asp Met Phe Gln Trp Asp Ala Phe (578)

.....A (1868).....
 GAA GCA UUU GAA AGC AUA AUC CCC CAG AAG AUG GCU GGC CAG UAC AGU GGA UUU GCA AGA GCA GUG CUC AAA CAA AUG AGA GAC CAA GAG GUU AUG AAA ACU GAC CAG UUC (1868)
 Glu Ala Phe Glu Ser Ile Ile Pro Gln Lys Met Ala Gly Gln Tyr Ser Gly Phe Ala Arg Ala Val Leu Lys Gln Met Arg Asp Gln Glu Val Met Lys Thr Asp Gln Phe (615)

.....A (1913).....
 AUA AAG UUG UUG CCU UUC UGU UUC UCA CCA CCA AAA UUA AGG AGC AAU GGG GAG CCU UAU CAA UUC UUG AGG CUU AUG UUG AAG GGA GGA GGG GAA AAU UUC UUC GAA GUA (1979)
 Ile Lys Leu Leu Pro Phe Cys Phe Ser Pro Pro Lys Leu Arg Ser Asn Gly Glu Pro Tyr Gln Phe Leu Arg Leu Met Leu Lys Gly Gly Glu Asn Phe Ile Glu Val (652)
Arg (630).....

.....C (2003).....G (2081).....
 AGG AAA GGG UCC CCU CUA UUC UCU UAC AAU CCA CAA ACA GAA GUC CUA ACU AUA UGC GGC AGA AUG AUG UCA UUA AAA GGA AAA AUU GAA GAU GAA GAA AGA AAU AGA UCA (2090)
 Arg Lys Gly Ser Pro Leu Phe Ser Tyr Asn Pro Gln Thr Glu Val Leu Thr Ile Cys Gly Arg Met Met Ser Leu Lys Gly Lys Ile Glu Asp Glu Glu Arg Asn Arg Ser (689)

.....(2138).....C.....A (2141).....
 AUG GGG AAU GCA GUA UUG GCA GGC UUU CUC GUU AGU GGC AAG UAU GAU CCG GAU CUU GGA GAU UUC AAA ACU AUU GAA GAA CUU GAA AAG CUA AAA CCG GGG GAA AAA GCA (2201)
 Met Gly Asn Ala Val Leu Ala Gly Phe Leu Val Ser Gly Lys Tyr Asp Pro Asp Leu Gly Asp Phe Lys Thr Ile Glu Glu Leu Glu Lys Leu Lys Pro Gly Glu Lys Ala (726)

.....C (2273).....
 AAC AUC UUA CUU UAU CAA GGA AAG CCC GUU AAA GUA GUU AAA AGG AAA AGA UAU AGU GCU UUA UCC AAU GAU AUU UCA CAA GGA AUU AAG AGA CAA AGA AUG ACA GUU GAG (2312)
 Asn Ile Leu Leu Tyr Gln Gly Lys Pro Val Lys Val Val Lys Arg Lys Arg Tyr Ser Ala Leu Ser Asn Asp Ile Ser Gln Gly Ile Lys Arg Gln Arg Met Thr Val Glu (763)

.....(2362).....G.C.(2364).....
 UCC AUG GGG UGG GCC UUG AGC UAAUUAUUUUUCCAUUAUUCAUUAUUCAUUAUUUGAGUGAAAAAUGCUCGUGUUUCUCAU (2396)
 Ser Met Gly Trp Ala Leu Ser C-terminus (770)

Fig. 2. PB2 gene. The ca nucleotide sequence differences are presented above the wt sequence and the amino acid changes are presented below the wt sequence. Each site is followed or preceded by its position number in parentheses and, again, at the end of each line. The position (and unless otherwise noted) the sequence of each oligodeoxynucleotide primer is indicated by underlining of the wt sequence. A (—) in the sequence represents the deletion of the nucleotide in the corresponding sequence.

B/AA/1/66 ca cRNA (+) 5'-.....C (53).....
 B/AA/1/66 wt cRNA (+) 5'-AGCAGAAGCGGUGCGUUGAUUUGCCAU AUG GAU ACU UUU AUU ACA AGA AAU UUC CAG ACU ACA UUA UUA CAA AAG GCC AAA AAC ACA AUG GCA (95)
 B/AA/1/66 wt PA Protein N-terminus Met Asp Thr Phe Ile Thr Arg Asn Phe Gln Thr Thr Ile Ile Gln Lys Ala Lys Asn Thr Met Ala (22)
 B/AA/1/66 ca PA Protein

...GAA UUU AGU GAA GAU CCU GAA UUA CAA CCA GCA AUG CUA UUC AAC AUC UGC GUC CAU CUG GAG GUC UGC UAU GUA UUA AGU GAU AUG AAU UUU CUU GAU GAA GAA GGA (203)
 Glu Phe Ser Glu Asp Pro Glu Leu Gln Pro Ala Met Leu Phe Asn Ile Cys Val His Leu Glu Val Cys Tyr Val Ile Ser Asp Met Asn Phe Leu Asp Glu Glu Gly (58)

...AAA ACA UAU ACA GCA UUA GAA GGA CAA GGA AAA GAA CAA AAU UUG AGA CCA CAA UAU GAA GUG AUU GAG GGA AUG CCA AGA AAC AUA GCA UGG AUG GUU CAA AGA UCC (311)
 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 Asn Ile Ala Trp Met Val Gln Arg Ser (94)

...UUA GCC CAA GAG CAU GGA UUA GAG ACU CCA AGG UAU CUG GCU GAU UUG UUC GAU UAU AAA ACC AAG AGG UUU UUA GAA GUU GGA UUA ACA AAG GGA UUG GCU GAC GAU (419)
 Leu Ala Gln Arg His His Gly Ile Glu Thr Pro Arg Tyr Leu Ala Asp Leu Phe Asp Tyr Lys Thr Lys Arg Phe Ile Glu Val Gly Ile Thr Lys Gly Leu Ala Asp Asp (130)

...UAC UUU UGG AAA AAG AAA GAA AAG CUG GGG AAU AGC AUG GAA CUG AUG UUA UUC AGC UAC AAU CAA GAC UAU UCG UUA AGU AAU GAA CAC UCA UUG GAU GAG GAA GGA (527)
 Tyr Phe Trp Lys Lys Lys Glu Lys Leu Gly Asn Ser Met Glu Leu Met Ile Phe Ser Tyr Asn Gln Asp Tyr Ser Leu Ser Asn Glu His Ser Leu Asp Glu Glu Gly (166)
 Ser (160)..... UC. (507-508).....

...AAA GGG AGA GUG CUA AGC AGA CUC ACA GAA CUU CAG GCU GAG UUA AGU CUG AAA AAU CUA UGG CAA GUU CUC UUA GGA GAA GAA GAU AUU GAA AAA GGA AUU GAC UUC (635)
 Lys Gly Arg Val Leu Ser Arg Leu Thr Glu Leu Ser Leu Lys Asn Leu Trp Gln Val Leu Ile Gly Glu Asp Ile Glu Lys Asp Ile Glu Lys Cys Leu Phe (202)

...AAA CUU GGA CAA ACA UUA UCU AAA CUA AGG GAC AUA UCU GUU CCA GCU GGU UUC UCC AAU UUU GAA GGA AUG AGG AGC UAC AUA GAC AAU AUA GAU CCU AAA GGA GCA (743)
 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 Pro Lys Gly Ala (238)

...AUA GAG AGA AAU CUA GCA AGG AUG UCU CCC UUA GUA UCA GUU ACA CCC AAA AAG UUA AAA UGG GAG GAC CUA AGA CCA AUA GGG CCU CAC AUU UAC AGC CAU GAG CUA (851)
 Ile Glu Arg Asn Leu Ala Arg Met Ser Pro Leu Val Ser Val Thr Pro Lys Lys Leu Lys Trp Glu Asp Leu Arg Pro Ile Gly Pro His Ile Tyr Ser His Glu Leu (274)
 Asn (271)..... A. (841).....

...CCA GAA GUU CCA UAU AAU GCC UUU CUU CUA AUG UCU GAU GAG UUG GGG CUG GCU AAU AUG ACU GAA GGG AAG UCC AAG AAA CCA AAG ACC UUA GCC AAA GAA UGU CUA (959)
 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 Leu (310)

...GAA AAG UAC UCA ACA CUA CGG GAU CAA ACU GAC CCA AUA UUA AUG AAA AGC GAA AAA GCU AAC GAA AAC UUC UUA UGG AAG UUG UGG AGG GAG UGU AUA AAU ACA (1067)
 Glu Lys Tyr Ser Thr Leu Arg Asp Gln Thr Asp Pro Ile Leu Ile Met Lys Ser Glu Lys Ala Asn Glu Asn Phe Leu Trp Lys Leu Trp Arg Asp Cys Val Asn Thr (346)
 C. (1044).....

...AUA AGU AAU GAG GAA ACA AGU AAC GAA UUA CAG AAA ACC AAU UAU GCC AAG UGG GCC ACA GGA GAU GGA UUA ACA UAC CAG AAA AUA AUG AAA GAA GUA GCA AUA GAU (1175)
 Ile Ser Asn Glu Glu Thr Ser Asn Glu Leu Gln Lys Thr Asn Tyr Ala Lys Trp Ala Thr Gly Asp Gly Leu Thr Tyr Gln Lys Ile Met Lys Glu Val Ala Ile Asp (382)

...GAC GAA ACA AUG UAC CAA GAA GAG CCC AAA UUA CCU AAU AAA UGU AGA UGU GCU GCU UGG GUU CAA ACA GAG AUG AAU CUA UUG AGC ACU CUG ACA AGU AAA AGG GCC (1283)
 Asp Glu Thr Met Tyr Gln Glu Glu Pro Lys Ile Pro Asn Lys Cys Arg Val Ala Ala Trp Val Gln Thr Glu Met Asn Leu Leu Ser Thr Leu Thr Ser Lys Arg Ala (418)

...CUG GAU CUA CCA GAA UUA GGG CCA GAC GUA GCA CCC GUG GAG CAU GUA GGG AGU GAA AGA AGG AAA UAC UUU GUU AAU GAA AUC AAC UAC UGU AAG GCC UCU ACC GUU (1391)
 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 Ala Ser Thr Val (454)
 Met (431).....

...AUG AUG AAG UAU GUA CUU UUU CAC ACU UCA UUA UUA AAU GAA AGC AAU GCC AGC AUG GGA AAA UAU AAA GUA UUA CCA UUA ACC AAC AGA GUA GUA AAU GAA AAA GGA (1499)
 Met 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 Asn Arg Val Val Asn Glu Lys Gly (490)

...GAA AGU UUU GAC AUA CUU UAU GGU CUG GCG GUU AAA GGG CAA UCU CAU CUG AGG GGA GAU ACU GAU GUU GUA ACA GUU GUG ACU UUC GAA UUU AGU AGU ACA GAU CCC (1607)
 Glu Ser Phe Asp Ile Leu Tyr Gly Leu Ala Val Lys Gly Gln Ser His Leu Ser His Leu Arg Gly Asp Thr Asp Val Val Thr Val Val Thr Phe Glu Phe Ser Ser Thr Asp Pro (526)
 (1514).G C. (1518)..... (495) Met ... His (497).....

...AGA GUG GAC UCA GGA AAG UGG CCA AAA UAU ACU GUA UUU AGA AUU GGU UCC UUA UUU GUG AGU GGA AGG GAA AAA UCU GUG UAC CUA UAU UGC CGA GUG AAU GGU ACA (1715)
 Arg Val Asp Ser Gly Lys Trp Pro Lys Tyr Thr Val Phe Arg Ile Gly Ser Leu Phe Val Ser Gly Arg Glu Lys Ser Val Tyr Leu Tyr Cys Arg Val Asn Gly Thr (562)
 ..U (1718).....

...AAC AAG AUC CAA AUG AAA UGG GGA AUG GAA GCU AGA AGA UGU CUG CUU CAA UCA AUG CAA CAA AUG GAA GCA AUA GUU GAU CAA GAA UCA UCG UUA CAA GGA UAU GAC (1823)
 Asn Lys Ile Gln Met Lys Trp Gly Met Glu Ala Arg Arg Cys Leu Leu Gln Ser Met Gln Gln Met Glu Ala Ile Val Asp Gln Glu Ser Ser Ile Gln Gly Tyr Asp (598)
 Glu (590).....

...AUG ACC AAA GCU UGU UUC AAG GGA GAC AGA GUG AAU AGU CCC AAA ACU UUC AGU AUU GGG ACU CAA GAA GGA AAA CUA GUA AAA GGA UCC UUU GGG AAA GCA CUA AGA (1931)
 Met Thr Lys Ala Cys Phe Lys Gly Asp Arg Val Asn Ser Pro Lys Thr Phe Ser Ile Gly Thr Gln Glu Gly Lys Leu Val Lys Gly Ser Phe Gly Lys Ala Leu Arg (634)

...GUA AUA UUC ACC AAA UGU UUG AUG CAC UAU GUA UUU GGA AAU GCC CAA UUG GAG GGG UUU AGU GCC GAA UCU AGG AGA CUU CUA CUG UUA AUA CAG GCA UUA AAG GAC (2039)
 Val Ile Phe Thr Lys Cys Leu Met His Tyr Val Phe Gly Asn Ala Gln Leu Glu Gly Phe Ser Ala Glu Ser Arg Arg Leu Leu Leu Leu Ile Gln Ala Leu Lys Asp (670)

...AGA AAG GGC CCU UGG GUA UUC GAC UUA GAG GGA AUG UAU UCU GGA UUA GAA GAA UGU AUU AGU AAC AAC CCU UGG GUA UUA CAG AGU GCA UAC UGG UUU AAU GAA UGG (2147)
 Arg Lys Gly Pro Trp Val Phe Asp Leu Glu Gly Met Tyr Ser Gly Ile Glu Glu Cys Ile Ser Asn Asn Pro Trp Val Ile Gln Ser Ala Tyr Trp Phe Asn Glu Trp (706)

...UUG GGC UUU GAA AAA GAG GGG AGU AAA GUA UUA GAA UCA AUA GAU GAA AUA AUG GAU GAA UGA AAG AAG GGC CAU AGC GCU CAU UUU UGU UUA UGU AUC UUA AUC AUC AUC (2270)
 Leu Gly Phe Glu Lys Glu Gly Ser Lys Val Leu Glu Ser Ile Asp Glu Ile Met Asp Glu C-terminus (726)
 G. (2235).....

...AAAAGAAUUGAGAAUAAAAUUGCAGGUGUUCUACU-3' (2308)

Fig. 4. PA gene. The ca and wt sequences are presented as in Fig. 2, with identical conventions for numbering and primer location and sequences.

B/AA/1/66 ca NP cRNA (+) 5'-----
 B/AA/1/66 wt NP cRNA (+) 5'-AGCAGAAGCAGCAAGUUUUUUUGUGAACUUAAGUACCAACAAAACUUAAGAAUCAA AUG UCC AAC AUG GAU AUU GAC GGC AUC AAC ACU GGA ACA AUU GAC AAA (106)
 B/AA/1/66 wt NP Protein N-terminus Met Ser Asn Met Asp Ile Asp Gly Ile Asn Thr Gly Thr Ile Asp Lys (16)
 B/AA/1/66 ca NP Protein

ACA CCA GAA GAA AUA ACU UCC GGA ACC AGU GGG GCA ACC AGA CCA AUC AUC AAG CCA GCA ACC CUU GCC CCA CCA AGC AAU AAA C (187) ..C (211) ..G (211) ..
 Thr Pro Glu Glu Ile Thr Ser Gly Thr Ser Gly Ala Thr Arg Pro Ile Ile Lys Pro Ala Thr Leu Ala Pro Pro Ser Asn Lys Arg Thr Arg Asn Pro Ser Pro Glu Arg Ala (220)

G(221)U(232) ..U GA.(238-240)
 ACC ACA AGC AGC GAA GCG AUU GUC GGA AGG AGA ACC CAA AAG AAA CAA ACC CCG ACA GAG AUA AAG AGC GUC UAC AAU AUG GUA GUG AAA CUG GGU GAA UUC UAC AAC CAG (334)
 Thr Thr Ser Ser Glu Ala Ile Val Gly Arg Arg Thr Gln Lys Lys Gln Thr Pro Thr Glu Ile Lys Lys Ser Val Tyr Asn Met Val Val Lys Leu Gly Glu Phe Tyr Asn Gln (92)
 Ala (55) Asp (61)

AUG AUG GUC AAA GCU GGA CUC AAC GAU GAC AUG GAG AGA AAC CUA AUC CAA AAU GCA CAU GCU GUG GAA AGA AUU CUA UUG GCU GCU ACU GAU GAC AAG AAA ACU GAA UAC CAA (448)
 Met Met Val Lys Ala Gly Leu Asn Asp Asp Met Glu Arg Asn Leu Ile Gln Asn Ala His Ala Val Glu Arg Ile Leu Leu Ala Ala Thr Asp Asp Lys Lys Thr Glu Tyr Gln (130)

AAG AAA AAG AAU GCC AGA GAU GUC AAA GAA GGG AAA GAA GAA AUA GAC CAC AAC AAA ACA GGA GGC ACC UUU UAU AAG AUG GUA AGA GAU GAU AAA ACC AUC UAC UUC AGC CCU (562)
 Lys Lys Lys Asn Ala Arg Asp Val Lys Lys Lys Lys Lys Ile Asp His Asn Lys Thr Gly Gly Thr Phe Tyr Lys Met Val Arg Asp Asp Lys Lys Thr Ile Tyr Phe Ser Pro (168)

AUA AGA AUU ACC UUU UUA AAA GAA GAG GUG AAA ACA AUG UAC AAG ACC ACC AUG GGG AGU GAU GGU UUC AGU GGA CUA AAU CAC AUC AUG AUU GGG CAU UCA CAG AUG AAC GAU (676)
 Ile Arg Ile Thr Phe Leu Lys Glu Glu Val Lys Thr Met Tyr Lys Thr Thr Met Gly Ser Asp Gly Phe Ser Gly Leu Asn His Ile Met Ile Gly His Ser Gln Met Asn Asp (206)

GUC UGU UUC CAA AGA UCA AAG GCA CUA AAA AGA GAU GGA CUU GAC CCU UCA UUA AUC AGU ACU UUU GCA GGA AGC ACA CUC CCC AGA AGA UCA GGU GCA ACU GGU GUU GCG AUC (790)
 Val Cys Phe Glu Arg Ser Lys Ala Leu Lys Arg Val Gly Leu Asp Pro Ser Leu Ile Ser Thr Phe Ala Lys Asp Lys Thr Thr Pro Arg Arg Ser Thr Leu Thr Thr Thr Ala Ile (244)

AAA GGA GGU GGA ACU UUA GUG GCA GAA GCC AAU CGA UUU AUA GGA AGA GCA AUG GCA GAC AGA GGG CUA UUG AGA GAC AUC AGA GCC AAG ACG GCC UAU GAA AAG AUU CUU CUG (904)
 Lys Gly Gly Gly Thr Leu Val Ala Glu Ala Ile Arg Phe Ile Gly Arg Ala Met Ala Asp Arg Gly Leu Leu Arg Asp Ile Arg Ala Lys Thr Ala Tyr Glu Lys Ile Leu Leu (282)

AAU CUG AAA AAC AAG UGC UCU GCG CCC CAA CAA AAG GCU CUA GUU GAU CAA GUG AUC GGA AGU AGA AAC CCA GGG AAU GCA GAC AUC AUA GAA GAC CUA ACC CUG CUU GCC CGA AGC (1018)
 Asn Leu Lys Asn Lys Cys Ser Ala Pro Gln Gln Lys Lys Lys Ala Leu Val Ile Gln Gln Ser Arg Asn Pro Gly Ile Ala Asp Ile Ala Asp Ile Glu Asp Leu Thr Leu Leu Ala Arg Ser (320)

AUG GUC GUU GUC AGG CCC UCU GUA GCG AGC AAA GUG GUG CUU CCC AUA AGC AUU AAU GCU AAA AUA CCU CAA CUA GGG UUC AAU GUU GAA GAA UAC UCU AUG GUU GGG UAU GAA (1132)
 Met Val Val Val Arg Pro Ser Val Ala Ser Lys Val Val Leu Pro Ile Ser Ile Asn Ala Lys Ile Pro Gln Leu Gly Phe Asn Val Glu Glu Tyr Ser Met Val Gly Tyr Glu (358)

GCC AUG GCU CUU UAU AAU AUG GCA ACA CCU GUU UCC AUA UUA AGA AUG GGA GAC GAU GCA AAA GAU AAA UCA CAA UUA UUC UUC AUG UCU UGC UUU GGA GCU GCC UAU GAA GAC (1246)
 Ala Met Ala Leu Tyr Asn Met Ala Thr Pro Val Ser Ile Leu Ser Phe Met Gly Asp Asp Ala Lys Asp Lys Ser Gln Leu Phe Phe Met Ser Cys Phe Gly Ala Ala Tyr Glu Asp (396)

CAA AGA GUU UUG UCU GCA CUA ACC GGC ACA GAA UUC AAG CCU AGG UCA GCA UUA AAG UGC AAG GGU UUC CAC GUU CCA GCA AAG GAG CAA GUG GAA GGA AUG GGG GCA GCU CUG (1360)
 Gln Arg Val Leu Ser Ala Leu Thr Gly Thr Glu Phe Lys Pro Arg Ser Ala Leu Lys Cys Lys Gly Phe His Val Pro Ala Lys Glu Gln Val Glu Gly Met Gly Ala Ala Leu (434)

AUG UCC AUC AAG CUC CAG UUU UGG GCC CCA AUG ACC AGA UCU GGG GGG AAC GAA GUA GGU GGA GAC GGA GGG UCU GGU CAA AUA AGU UGC AGC CCC GUG UUU GUA GAA GAG AGA (1474)
 Met Ser Ile Lys Leu Gln Phe Thr Ala Pro Met Thr Arg Ser Gly Gly Asn Gln Val Gly Gly Asp Gly Gly Ser Gly Gln Ile Ser Cys Ser Pro Val Phe Ala Val Glu Arg (472)

CCU AUU GCU CUA AGC AAG CAA GCU GUA AGA AGA AUG CUG UCA AUG AAU AUU GAG GGA CGU GAU GCA GAU GUC AAA GGA AAU CUA CUC AAG AUG AUG AAU GAU UCA AUG GCU AAG (1588)
 Pro Ile Ala Leu Ser Lys Gln Ala Val Arg Arg Met Leu Ser Met Asn Ile Glu Gly Arg Asp Ala Asp Val Lys Lys Gly Asn Leu Leu Lys Met Met Asn Asp Ser Met Ala Lys (510)

AAA ACC AAU GGA AAU GCU UUC AUU GGG AAG AAA AUG UUU CAA AUA UCA GAC AAA AAC AAA AUC AAU CCC GUU GAU AUU CCA AUU AAG CAG ACC AUC CCC AAU UUC UUC UUU GGG (1702)
 Lys Thr Asn Gly Asn Ala Phe Ile Gly Lys Lys Lys Met Gln Ile Ser Asp Lys Asn Lys Ile Asn Pro Val Asp Ile Pro Ile Lys Gln Thr Ile Pro Asn Phe Phe Gly Gly (548)

AGG GAC ACA GCA GAG GAU UAU GAU GAC CUC GAU UAU UAAAGCAACAAAUAAGACACUAUGGGCUGACUGUUCAGUAGUUGAAUGUGGGUUAUCUUAUUGAAUAAAUAUAAAUAUUGUGUUGUUCUACU-3' (1842)
 Arg Asp Thr Ala Glu Asp Tyr Asp Asp Leu Asp Tyr C-terminus (560)

Fig. 5. NP gene. The ca and wt sequences are presented as in Fig. 2, with identical conventions for numbering and primer location, with one exception. The primer sequence starting at position 221 is actually ACC ACA AGC AGT GAA (see text).

references). Both methods were applied directly to vRNA isolated from purified virus. Thus, the determined sequences represent consensus sequences based on the population of vRNA segments. The chemical sequencing method identified the first 50–90 nucleotides of each gene and overlapped the first primer-determined sequence by at least 15 nucleotides. Therefore the sequences on the 5' side of the first primer positions were determined exclusively by the direct chemical sequencing method. Oligo-

deoxynucleotide primers were usually 15 nucleotides in length and the sequence from each preceding primer overlapped the next determined sequence by at least 15–20 nucleotides. Their positions in the sequences are shown by underlining of the wt cRNA sequence of each gene. Only the differences are presented for the ca RNA and predicted ca polypeptide sequences, and these are shown above and below the corresponding wt sequences, respectively. The results for each individual gene are presented below.

```

B/AA/1/66 ca M cRNA (+) 5'-----
B/AA/1/66 wt M cRNA (+) 5'-AGCAGAAGCACGCACUUUCUAAA AUG UCG CUG UUU GGA GAC ACA AUU GCC UAC CUG CUU UCA CUA ACA GAA GAU GGA GAA GGC AAA GCA GAA CUA (96)
B/AA/1/66 wt M1 Protein N-terminus Met Ser Leu Phe Gly Asp Thr Ile Ala Tyr Leu Leu Ser Leu Thr Glu Asp Gly Glu Gly Lys Ala Glu Leu (24)
B/AA/1/66 ca M1 protein N-terminus .....

GCA GAA AAA UUA CAC UGU UGG UUC GGU GGG AAA GAA UUU GAC CUA GAC UCU GCU UUG GAA UGG AUA AAA AAC AAA AGA UGC CUA ACU GAU AUA CAA AAA GCA CUA AUU GGU (207)
A1a Glu Lys Leu His Cys Trp Phe Gly Gly Lys Glu Phe Asp Leu Asp Ser Ala Leu Glu Trp Ile Lys Asn Lys Arg Cys Leu Thr Asp Ile Gln Lys Ala Leu Ile Gly (61)

... ..U (213)..
GCC UCC AUC UGC UUU UUA AAA CCC AAA GAC CAA GAA AGA AAA AGA AGA UUC AUC ACA GAG CCC CUG UCA GGA AUG GGA ACA ACA GCA ACA AAG AAG AAA GGC CUG AUU CUA (318)
A1a Ser Ile Cys Phe Leu Lys Pro Lys Asp Gln Glu Arg Lys Arg Arg Phe Ile Thr Glu Pro Leu Ser Gly Met Gly Thr Thr Ala Thr Lys Lys Lys Gly Leu Ile Leu (98)

... ..G (345)..
GGU GAG AGA AAA AUG AGA AGA UGU GUA AGU UUU CAU GAA GCA UUU GAA AUA GCA GAA GGC CAU GAA AGC UCA GCA CUA CUA UAU UGU CUU AUG GUC AUG UAC CUG AAC CCU (429)
A1a Glu Arg Lys Met Arg Arg Cys Val Ser Phe His Glu Ala Phe Glu Ile Ala Glu Gly His Glu Ser Ser Ala Leu Leu Tyr Cys Leu Met Val Met Tyr Leu Asn Pro (135)

... ..U (435)..
GGA AAC UAU UCA AUG CAA GUA AAA CUA GGA ACG CUU UGU GCU UUA UGC GAG AAA CAA GCA UCA CAU UCA CAC AGA GCU CAU AGC AGA GCA GCA AGA UCU UCA GUG CCU GGA (540)
Gly Asn Tyr Ser Met Gln Val Lys Leu Gly Thr Leu Cys Ala Leu Cys Glu Lys Gln Ala Ser His Ser His Arg Ala His Ser Arg Ala Ala Arg Ser Ser Val Pro Gly (172)

... ..A (501)..
Gln (159)
B/AA/1/66 wt 2ORF- Gln Ser Ser Lys Ile Phe Ser Ala Trp Ser (10)
B/AA/1/66 ca 2ORF- .....

... ..G (571)
GUG AGG CGA GAA AUG CAG AUG GUU UCA GCU AUG AAC ACA GCA AAA ACA AUG AAU GGA AUG GGG AAG GGA GAA GAC GUC CAA AAA CUG GCA GAA GAG CUG CAA AGC AAC AUU (651)
Val Arg Arg Glu Met Gln Met Val Ser Ala Met Asn Thr Ala Lys Thr Met Asn Gly Met Gly Lys Glu Asp Val Gln Lys Leu Ala Glu Glu Leu Gln Ser Asn Ile (209)

... ..Val (183)..
Glu Ala Arg Asn Ala Asp Gly Phe Ser Tyr Glu His Ser Lys Asn Asn Glu Trp Asn Gly Glu Gly Arg Arg Arg Pro Lys Thr Gly Arg Arg Ala Ala Lys Gln His Trp (47)
... ..Cys (20)..

... ..A (702)..
GGA GUA UUG AGA UCU CUG GGG GCA AGU CAA AAG AAU GGA GAA GGA AUU GCC AAG GAU GUA AUG GAA GUG CUA AAG CAG AGC UCU AUG GGA AAU UCA GCU CUU GUG AAG AAA (762)
Gly Val Leu Arg Ser Leu Gly Ala Ser Gln Lys Asn Gly Glu Gly Ile Ala Lys Asp Val Met Glu Val Leu Lys Gln Ser Ser Met Gly Asn Ser Ala Leu Val Lys Lys (246)

... ..Ser (64)..
Ser Ile Glu Ile Ser Gly Gly Lys Ser Lys Glu Trp Arg Arg Asn Cys Gln Gly Cys Asn Gly Ser Ala Lys Ala Glu Leu Tyr Gly Lys Phe Ser Ser Cys Glu Glu Ile (84)

... ..UAC CUA UAAUG CUC GAA CCA UUU CAG AUU CUU UCA AUU UGU UCU UUC AUU UUA UCA GCU CUC CAU UUC AUG GCU UGG ACA AUA GGG CAU UUG AAU CAA AUA AAA AGA GGA GUA (875)
Tyr Leu M1 C-terminus (248)
... ..M1 C-terminus (248)
Pro Ile Met Leu Glu Pro Phe Gln Ile Leu Ser Ile Cys Ser Phe Ile Leu Ser Ala Leu His Phe Met Ala Trp Thr Ile Gly His Leu Asn Gln Ile Lys Arg Gly Val (121)

... ..A (977)..
AAC CUG AAA AUA CGA AUA AGA AAU CCA AAU AAA GAG ACA AUA AAC AGA GAG GUA UCA AUU UUG AGA CAC AGU UAC CAA AAA GAA AUC CAA GCC AAA GAA ACG AUG AAG GAA (986)
Asn Leu Lys Ile Arg Ile Arg Asn Pro Asn Lys Glu Thr Ile Asn Arg Glu Val Ser Ile Leu Arg His Ser Tyr Gln Lys Glu Ile Gln Ala Lys Glu Thr Met Lys Glu (158)

... ..G (1082)..
GUA CUC UCU GAC AAC AUG GAG AUA UUG AGU GAC CAC AUA GUA AUU GAG GGG CUU UCU GCU GAA GAG AUA AUA AAA AUG GGU GAA ACA GUU UUG GAA GUA GAA GAA UUG CAG (1097)
Val Leu Ser Asp Asn Met Glu Ile Leu Ser Asp His Ile Val Ile Glu Gly Leu Ser Ala Glu Glu Ile Ile Lys Met Gly Glu Thr Val Leu Glu Val Glu Glu Leu Gln (195)

... ..A (1099)..
UGAACCAAUU--CACCGUAUUUCUUGCUAUGCAUUUAAGCAAUUGUAUCAAUGUCAGCAAUAAACUGGAAAAGUGCGUUGUUCUACU-3' (1188)
2ORF C-Terminus (1190)
2ORF C-Terminus (1188)

```

Fig. 6. M gene. The ca and wt sequences are presented as in Fig. 2, with the addition the second open reading frame (2 ORF) presented as two additional lines as indicated in the figure.

PB2 gene

Both wt and ca influenza B/AA/1/66 PB2 genes were 2396 nucleotides long. They encoded polypeptides of 770 amino acids. The first methionine codon in the open reading frame began at position 24 (in from the 5' end of the cRNA) and the first termination codon began at position 2334. A consecutive run of five adenosine residues began at position 2375 and was the probable polyadenylation site. The predicted wt and ca PB2 proteins had molecular weights of 88,035 and 88,072 Da, and both had a net charge of +24.5 at pH 7.0, with no large hydrophilic domains. The ca PB2 gene varied from the wt gene by 26 base substitutions. Of these changes, 24 occurred in the coding region. Only 4 of these changes were reflected in amino acid changes. Of these 4 changes, 1 was unique to the ca PB2 gene,

and 3 were identical to sequences noted in other wt influenza B virus PB2 genes (B/Melbourne/43 and B/AA/1/86), see Fig. 8. A reassortant containing the PB2 gene of ca B/AA/1/66 virus, and all other genes from a wt virus, B/Houston/76, is non-ts, non-ca, and as virulent as the wt virus (Donabedian *et al.*, 1987). Hence, the PB2 gene of the ca donor virus apparently contributed to none of the three phenotypes of interest in an independent manner. It is possible that it may be involved in any of these phenotypes in combination with one or more other ca B/AA/1/66 genes, especially those involved in the replication complexes.

PB1 gene

The general characteristics of the influenza B PB1 gene have been previously described by Kemdirim and

PB2 Gene					PB1 Gene				PA Gene					NP Gene					M Gene												
#	Pos.	Me1	wt	ca	AA	#	Pos.	Lee	wt	ca	#	Pos.	wt	ca	HN	Sing	#	Pos.	Lee	wt	ca	Sing	#	Pos.	Lee	wt	ca	Sing			
1	1913	AGC	AGC	AGA	AGC	1	1972	AUA	AUA	GUA	1	1320	GUG	AUG	GUG	GUG	1	221	ACC	ACC	GCC	ACC	1	501	CAU	CAC	CAA	CAC			
		Ser	Ser	Arg	Ser			Ile	Ile	Val			Val	Met	Val	Val			Thr	Thr	Ala	Thr			M1	His	His	Gln	His		
2	2362	A	A	G	A	2	1319	AAA	AGA	AAA	2	1518	UUA	CAU	UUA	UUA	2	399	GUG	GUG	GCG	GUG	2	571	AUG	AUG	GUG	AUG			
		3'-NONCODING						Lys	Arg	Lys			Tyr	His	Tyr	Tyr			Val	Val	Ala	Val			M1	Met	Met	Val	Met		
3	1334	CAA	CAA	CAG	CAA	3	2272	UAC	CAC	UAC	3	507	CAC	UCC	UCC	UCC	3	1287	CCU	CCU	CAU	CCU	2	571	UAU	UAU	UGU	UAU			
		Gln	Gln	Gln	Gln			Tyr	His	Tyr			His	Ser	Ser	Ser			Pro	Pro	His	Pro			20RF	Tyr	Tyr	Cys	Tyr		
4	256	CAA	CGA	CAA	CAA	4	345	CAG	CAA	CAG	5	841	AGC	AAC	AAC	AAC	4	1583	GCA	GCU	ACU	GCU	3	213	UCU	UCC	UCU	UCU			
		Gln	Arg	Gln	Gln			Gln	Gln	Gln			Ser	Asn	Asn	Asn			Ala	Ala	Thr	Ala			M1	Ser	Ser	Ser	Ser		
5	828	AUC	GUC	AUC	AUC	5	399	AGA	AGG	AGA	6	1514	AUA	AUG	AUG	AUG	5	1078	GCU	GCU	GCC	GCU	4	465	CUC	CUU	CUC	CUC			
		Arg	Val	Arg	Ile			Arg	Arg	Arg			Ile	Met	Met	Met			ALA	ALA	ALA	ALA			M1	Leu	Leu	Leu	Leu		
6	50	UUA	UUG	UUA	UUA	6	537	GAU	GAC	GAU	7	1796	GAU	GAA	GAA	GAA	6	1804	C	C	U	C	5	977	ACA	ACG	ACA	ACA			
		Leu	Leu	Leu	Leu			Asp	Asp	Asp			Asp	Glu	Glu	Glu			3'-Non-coding							20RF	Thr	Thr	Thr	Thr	
7	626	ACU	ACC	ACU	ACU	7	801	AAA	AAG	AAA	8	53	AAU	AAC	AAC	AAC	7	239	GAU	AUU	GAU	GAU	6	1082	GAG	GAA	GAG	GAG			
		Thr	Thr	Thr	Thr			Lys	Lys	Lys			Asn	Asn	Asn	Asn			Asp	Ile	Asp	Asp			20RF	Glu	Glu	Glu	Glu		
8	1008	AGA	AGG	CGG	CGG	8	930	ACA	ACG	ACA	9	668	GAC	GAU	GAU	GAU	9	1663	GAG	GAU	GAG	GAG	7	1107?	Insertion						
		Arg	Arg	Arg	Arg			Thr	Thr	Thr			Asp	Asp	Asp	Asp			Glu	Asp	Glu	Glu									
9	1052	GAU	GAC	GAU	GAU	9	1146	AAU	AAC	AAU	10	791	CCC	CCU	CCU	CCU	10	1658	AUU	GUU	AUU	GUU	8	300	AAG	AAG	AAA	AAA			
		Asp	Asp	Asp	Asp			Asn	Asn	Asn			Pro	Pro	Pro	Pro			Ile	Val	Ile	Val			M1	Lys	Lys	Lys	Lys		
10	1088	AAG	AAG	AAA	AAG	10	1326	UAC	UAU	UAC	11	800	UUA	UUG	UUG	UUG	11	187	AAC	AAU	AAC	AAC	9	345	GUA	GUA	GUG	GUG			
		Lys	Lys	Lys	Lys			Tyr	Tyr	Tyr			Leu	Leu	Leu	Leu			Asn	Asn	Asn	Asn			M1	VAL	VAL	VAL	VAL		
11	2003	UCC	UCU	UCC	UCC	11	1557	GGG	GGG	GGG	12	935	CCA	CCG	CCG	CCG	12	232	AGU	AGC	AGU	AGU	10	408	CUU	CUU	CUC	CUC			
		Ser	Ser	Ser	Ser			Gly	Gly	Gly			Pro	Pro	Pro	Pro			Ser	Ser	Ser	Ser			M1	Leu	Leu	Leu	Leu		
12	572	AUG	AUG	AUA	AUA	12	1695	UAC	UAU	UAC	13	1044	UUG	CUG	CUG	CUG	13	1228	UUC	UUU	UUC	UUC	11	435	AAC	AAC	AAU	AAU			
		Met	Met	Ile	Ile			Tyr	Tyr	Tyr			Leu	Leu	Leu	Leu			Phe	Phe	Phe	Phe			M1	Asn	Asn	Asn	Asn		
13	1160	GGC	AGU	AGC	AGC	13	1782	GGU	GGC	GGU	14	1214	AAU	AAC	AAC	AAC	14	1411	AAU	AAC	AAU	AAU	12	702	GCC	GCC	GCA	GCA			
		Gly	Ser	Ser	Ser			Gly	Gly	Gly			Asn	Asn	Asn	Asn			Asn	Asn	Asn	Asn			M1	Ala	Ala	Ala	Ala		
14	341	CCC	CCC	CCA	CCA	14	570	ACA	UCA	UCG	15	1655	GGU	GGC	GGC	GGC	15	1471	GAA	GAG	GAA	GAA	13	702	CAA	CAA	AAA	AAA			
		Pro	Pro	Pro	Pro			Thr	Ser	Ser			Gly	Gly	Gly	Gly			Glu	Glu	Glu	Glu			20RF	Gln	Gln	Lys	Lys		
15	359	GGG	GGG	GGU	GGU	15	822	GAG	GAG	GAA	16	1718	AAC	AAU	AAU	AAU	16	444	UAC	UAC	UUC	UUC	14	1099	UGA	UGA	UAA	UAA			
		GLY	GLY	GLY	GLY			Glu	Glu	Glu			Asn	Asn	Asn	Asn			Tyr	Tyr	Phe	Phe			20RF	End	End	End	End		
16	476	CCC	CCC	CCU	CCU	16	846	GGA	GGA	GGG	17	2235	A	G	G	G	17	1650	GUC	AUC	ACC	ACC									
		Pro	Pro	Pro	Pro			Gly	Gly	Gly			3'-NONCODING						Val	Ile	Thr	Thr									
17	728	CAU	CAU	CAC	CAC	17	1098	AGU	AGU	AGC	18	160	AAG	AAG	AAA	GAA	18	160	AAG	AAG	AAA	GAA									
		His	His	His	His			Ser	Ser	Ser			Lys	Lys	Lys	Arg			Lys	Lys	Lys	Arg									
18	899	GAC	GAC	GAU	GAU	18	1647	ACG	ACG	ACA	19	238	ACC	GCG	GCU	GCU	19	238	ACC	GCG	GCU	GCU									
		Asp	Asp	Asp	Asp			Thr	Thr	Thr			Thr	Ala	Ala	Ala			Thr	Ala	Ala	Ala									
19	983	AAA	AAA	AAG	AAG	19	1836	AUU	AUU	AUC	20	211	CCA	CCA	CCG	CCG	20	211	CCA	CCA	CCG	CCG									
		Lys	Lys	Lys	Lys			Ile	Ile	Ile			Pro	Pro	Pro	Pro			Pro	Pro	Pro	Pro									
20	1058	GAG	GAG	GAA	GAA	20	2112	GCU	GCU	GCC	21	1188	CUG	UUG	UUA		21	523	UAU	UAU	UAC	UAC									
		Glu	Glu	Glu	Glu			Ala	Ala	Ala			Leu	Leu	Leu				Tyr	Tyr	Tyr	Tyr									
21	1184	UUG	CUG	CUA	CUA	21	1188	CUG	UUG	UUA	22	607	AAG	AAG	AAA	AAA	22	607	AAG	AAG	AAA	AAA									
		Leu	Leu	Leu	Leu			Leu	Leu	Leu			Leu	Leu	Leu	Leu			Lys	Lys	Lys	Lys									
22	2081	AGA	AGA	AGG	AGG	22	2081	AGA	AGA	AGG	23	973	AAC	AAC	AAU	AAU	23	973	AAC	AAC	AAU	AAU									
		Arg	Arg	Arg	Arg			Arg	Arg	Arg			Asn	Asn	Asn	Asn			Asn	Asn	Asn	Asn									
23	2138	GAU	GAU	GAC	GAC	23	2138	GAU	GAU	GAC	24	1270	ACG	ACC	ACA	ACA	24	1270	ACG	ACC	ACA	ACA									
		Asp	Asp	Asp	Asp			Asp	Asp	Asp			Thr	Thr	Thr	Thr			Thr	Thr	Thr	Thr									
24	2141	CCG	CCG	CCA	CCA	24	2141	CCG	CCG	CCA	25	1387	GCC	GCC	GCU	GCU	25	1387	GCC	GCC	GCU	GCU									
		Pro	Pro	Pro	Pro			Pro	Pro	Pro			Ala	Ala	Ala	Ala			Ala	Ala	Ala	Ala									
25	2273	GAU	GAU	GAC	GAC	25	2273	GAU	GAU	GAC																					
		Asp	Asp	Asp	Asp			Asp	Asp	Asp																					
26	2364	U	U	C	C	26	2364	U	U	C																					
		3'-NONCODING						3'-NONCODING																							

FIG. 8. Unique and nonunique wt to ca changes. The sequence differences between the wt and ca B/AA/1/66 nonsurface genes are compared to sequences of other known influenza type B viruses. They are presented as part of the codon of which they are a member, rather than just a single nucleotide difference. The encoded amino acid is listed below the codon. Below each gene the sites are arranged as follows: changes unique to ca B/AA/1/66 followed by changes that match the sequence in one or more other influenza type B virus. These nonunique changes have been further grouped by matching to one or another known viral sequence. The columns labeled Pos. refer to the nucleotide position of the change as found in Figs. 2-7. Those nucleotide differences that result in amino acid differences have been enclosed in boxes. The viruses listed are as follows: Me1, B/Melbourne/43; Lee, B/Lee/40; HN, B/Houston/76; Sing, B/Singapore/222/79; AA, B/Ann Arbor/1/86; wt, wt B/AA/1/66; and ca, ca B/AA/1/66.

B/Lee/40 (Briedis and Tobin, 1984) and B/Singapore/72 (Londo *et al.*, 1983) left only four of these changes as unique to the ca NP gene (Fig. 8).

The primer sequence starting at position 221 is actually ACC ACA AGC AGT GAA. Thus its sequence varies from that of the wt sequence at position 232, and from that of the ca sequence at position 221. This primer was made based on B/Lee/40 sequences, but despite the one nucleotide difference for each B/AA/1/66 virus, it worked well. (All other primers which were made prior to actual B/AA/1/66 viral sequence knowledge and which were later shown to vary by even one nucleotide, would not function.)

M gene

There was a total of 12 nucleotide substitutions between wt and ca B/AA/1/66 M RNA. In addition, the ca B/AA/1/66 M gene had an insertion of two uridine nucleotides within positions 1107–1110. The insert position was ambiguous due to the presence of two uridine nucleotides at positions 1107–1108 in the wt B/AA/1/66 M gene. This insertion was in the 3'-noncoding region for M₁ protein and was also beyond the conserved second open reading frame (2 ORF) previously described for both B/Lee/40 (Briedis *et al.*, 1982) and B/Singapore/222/79 (Hiebert *et al.*, 1986) M genes. In fact, both of these viruses have a run of four uridine nucleotides at this position like the ca M gene. There were two amino acid changes in the M₁ protein region, and two predicted amino acid changes in the possible product of the 2 ORF. None of the M₁ changes are found in the other known influenza B M gene sequences, while one of the changes in the 2 ORF is also seen in B/Singapore/222/79 M gene (Hiebert *et al.*, 1986). Thus, at this time only the two changes in the M₁ protein and possibly one in the 2 ORF appeared to be unique to the cold-adapted virus (Fig. 8).

NS gene

The NS gene had the lowest number of changes between wt and ca B/AA/1/66 viruses. Only two base substitutions were present, both exclusively in the NS₁ coding region, and neither of these resulted in amino acid changes in the NS₁ protein. One change, at position 80, occurred in the 5'-donor-splice region, but it was an acceptable change based on the consensus signal from Mount (1982). Unless the wt NS gene itself was inherently attenuating, cold-adapted, or temperature-sensitive, we would predict that this gene contributes little to the overall phenotypes of the ca B/AA/1/66 virus.

Unique changes of ca B/AA/1/66 virus

The total nucleotide and amino acid changes between the six nonsurface genes of wt and ca B/AA/1/66 viruses are given in Fig. 8 showing a comparison to the corresponding site in other known influenza B sequences. Except for those changes which occur in the noncoding regions of the genes, the sites are presented in their codon, with their encoded amino acid. Where possible the changes have been divided into groups: (1) those nucleotide changes that were unique to ca B/AA/1/66 virus (the top group for each gene), and (2) those changes that were also seen in one or more other wt viruses (the bottom two groups for each gene). Sequences for PB2, NP, M, and NS genes have been determined from viral isolates that occurred both earlier and later in time than the B/AA/1/66 viruses. Only one other sequence for the PB1 gene has been reported to which B/AA/1/66 viral PB1 sequences can be compared, and it was determined from an earlier virus isolate (B/Lee/40). Therefore, the group of changes, PB1 positions 14–21, may contain some unique ca nucleotide changes even though no amino acid change results. The PA gene has two other PA sequences to which it can be compared, but both genes are from later viral isolates. The nucleotide changes that result in amino acid changes are enclosed in boxes. Out of 15 unique nucleotide changes, 10 of these resulted in 11 unique amino acid changes. (One change in the M gene caused a unique amino acid change both in M₁ and in the 2 ORF.) The 89 remaining nonunique nucleotide changes resulted in only 15 additional amino acid changes.

Total changes

Table 1 presents a summary of the total numbers of nucleotide and amino acid changes that were selected in the cold-adaptation of wt B/AA/1/66 virus. The lack of changes occurring in the 5'-noncoding regions was probably not indicative of an active conservation of sequence, since overall, only one or two changes would have been predicted by the observed, average rate of change: i.e., 199 nucleotides/5'-noncoding region \times 0.0094 changes/nucleotides = 1.88 changes/5'-noncoding region. More disturbing is the lack of change in the NS gene. While all other genes range around 0.1 change/nucleotide, that of the NS gene is fivefold less. The reasons for this difference are not known. Only 26 amino acid changes are predicted to occur from a total of 105 nucleotide changes. This indicates a high degree of functional constraint, since on a random basis, 105 changes should result in approximately two-thirds as many amino acid changes,

TABLE 1
SUMMARY OF B/AA/1/66 WT TO CA SEQUENCE CHANGES

	cRNA				%
	5'-	Coding region	-3'	Total	
PB2	0/23	24/2310	2/63	26/2396	(1.11)
PB1	0/21	21/2256	0/92	21/2369	(0.89)
PA	0/29	16/2178	1/101	17/2308	(0.74)
NP	0/58	24/1680	1/104	25/1842	(1.36)
M	0/24	12/1073 ^a	2/93	14/1190	(1.18)
NS	0/44	2/1021 ^b	0/33	2/1098	(0.18)
Total	0/199	99/10,518	6/486	105/11,203	(0.94)

	Protein	
	Total changes/length	Unique changes/length
PB2	4/770 (0.52%)	1/770 (0.13%)
PB1	3/752 (0.40%)	1/752 (0.13%)
PA	6/726 (0.83%)	2/726 (0.28%)
NP	9/560 (1.61%)	4/560 (0.71%)
M1	2/248 (0.81%)	2/248 (0.81%)
2 ORF ^c	2/195 (1.03%)	1/195 (0.51%)
NS1	0/281 (0%)	0/281 (0%)
NS2	0/122 (0%)	0/122 (0%)
Total	26/3654 (0.71%)	11/3654 (0.30%)

^{a,b} Coding region length includes both open reading frames.

^c 2 ORF, Second open reading frame, and length is given for second open reading frame.

i.e., 71. Of the 26 total amino acid differences that exist between the six pairs of wt and ca genes, only 11 are unique to ca B/AA/1/66 virus.

DISCUSSION

As seen in Fig. 1, the initial CK₁ isolate of wt B/AA/1/66 virus had not been cloned by plaquing prior to the cold-adaptation process. Therefore, it was most likely a pool of viruses whose individual genomic sequences were highly related, but to some extent different for each virus. This assumption is based on the data of Parvin and colleagues (1986) which give a mutation rate of 1.5×10^{-5} mutations/nucleotide/generation for influenza A virus. At this rate of mutation, they estimate that every virion in a virus pool will vary by an average of one mutation per genome after five generations. The further removed in time a virus line is from its last plaquing step, the more diverse its individual viruses. Thus, when a virus pool is sequenced directly using either dideoxynucleotide chain-termination or chemical methods and purified vRNA, it is a consensus sequence that is determined and not the absolute sequence of a particular virus.

It should be noted that mutations which occur during every generation of the virus become fixed in the dominant line (and hence appear in the consensus sequence) only if they are in the infecting virus selected by a plaquing step, or if they confer a very great growth advantage compared to the standard population. Mutations which confer growth advantages at suboptimal temperatures have a better chance of predominating in the population in the cold adaptation process, and hence a better chance of becoming selected during plaquing. In addition to these selected mutations, nonlethal mutations may also be selected at random by each plaquing step. Thus, in comparing the final two populations of ca and wt B/AA/1/66 viruses, the sequence changes that exist will primarily represent (1) sequence differences existing between the particular wt virus, that was the progenitor of the ca virus, and the wt virus consensus sequence, (2) cold-selected mutations (these differences may have already existed in the wt pool or have occurred spontaneously during the cold-adaptation process and then been selected, and (3) other spontaneous nonlethal mutations that occur randomly with each generation, and become fixed by the mechanism of plaquing as part of the final ca virus consensus sequence.

Thus the cold-adaptation procedure and the plaquing procedures not only selected changes that resulted in the ca, ts, and att phenotypes, but also at the time of the first plaquing, a particular individual virus was selected from the mixed wt pool that would inherently contain a number of sequence differences from the consensus sequence of the wt virus pool, which probably have no bearing on these three phenotypes. It is conceivable that a change in the nucleic acid sequence might enable the expression of the ca phenotype without a corresponding amino acid change, but it is more likely that the cold-adaptation procedure operates by selecting viruses whose proteins function better at the low temperature. Thus, those nucleotide changes that were cold-selected should most likely be reflected in an amino acid change. This condition certainly fits the set of 15 unique nucleotide changes, since they were responsible for 11 of the 26 total amino acid changes. The other 15 amino acid changes were encoded by a subset of the other 90 nonunique nucleotide changes. While it is impossible to go back now and determine which changes occurred at which stage in the cold adaptation process, it is very likely that a large part of the unique changes represented cold-selected changes, while most non-unique changes represented nonlethal random site differences that existed between the wt consensus

sequence and an individual wt virus that was the progenitor of the ca virus.

The nonunique changes can be ordered into groups as mentioned earlier, in which the ca sequence remains the same as that of another wt virus. Sequences for the PB2, NP, M, and NS genes have been determined from isolates which bracket the B/AA/1/66 sequences in time. Thus in these genes it can be determined if the ca virus sequence change resembles that of an earlier or later virus. It should be mentioned that in those sites where the ca sequence is identical to an earlier virus, except in one instance (NP gene, position 1658) it is identical to a later virus as well, with the wt B/AA/1/66 sequence being unique. There are 22 sites in which the ca sequence is identical to an earlier virus, and 33 sites in which the ca sequence resembles that of a later virus for these four genes. Because B/AA/1/66 ca virus is not an evolutionary progenitor to the later viruses, these changes either existed between the sequence of the individual wt virus that was to be the progenitor of the ca virus and the wt consensus sequence prior to the cold-adaptation process, or there is some unknown mechanism which causes convergent changes at specific sites. This latter possibility is unlikely for two reasons: (1) Convergent changes might appear to happen if some sites have a higher than normal rate of mutation and simply switch back and forth between purine to purine, or pyrimidine to pyrimidine, manifesting either as one or the other depending upon the isolate sequenced. However, if this were the case, then out of the 22 sites in which the ca virus sequence is identical to the earlier virus sequence, a good number should show the wt B/AA/1/66 site identical to the later virus isolate's sequence. As mentioned above, this sequence configuration actually occurred only once in 22 times. (2) True convergent change implies that all the changes that appeared in the ca virus sequence occurred as mutations to the wt sequence. Based on the mutation rate determined previously for influenza A virus (Parvin *et al.*, 1986), the 55 differences for these four internal protein genes could not have arisen during the limited number of passages that occurred during the cold-adaptation and plaquing process. From the original wt isolate there were seven chick kidney passages, nine egg passages, and seven plaquing steps to arrive at the virus pool responsible for the ca consensus sequence. Estimating five generations for each of these steps as in Parvin and colleagues (1986), we would calculate 23 steps \times 5 generations/steps = 115 generations. Using the mutation rate of 1.5×10^{-5} mutations/nucleotide/generation (Parvin *et al.*, 1986), and a total of 6526 nucleotides for the PB2, NP, M, and NS

genes, we can determine the total number of mutations expected to have arisen during the cold-adaptation procedure: 1.5×10^{-5} mutations/nucleotide/generation \times 6526 nucleotides \times 115 generations = 11.26 mutations expected to have occurred during the ca selection and plaquing procedure, not 55 changes. This mutation rate was determined for influenza A virus. Comparison of the rate of change for influenza A and B viruses had previously indicated that influenza A has a higher rate of change than influenza B virus, by as much as fivefold for known HA gene sequences (Palese and Young, 1983). Thus, it is even more unlikely that all 55 mutations occurred during the cold-adaptation process. Thus, while some of the nonunique nucleotide changes that exist between the wt and ca B/AA/1/66 viruses may be accounted for by mutations that occurred during the ca process, they would be predicted to account for only a small percentage of the number actually observed. The mechanism of convergent mutation cannot be entirely ruled out, however, since the number of viral generations in man prior to isolation are not known, and the convergent changes could have occurred then.

Of the six genes examined, the NS gene was quite restricted in the rate of change/nucleotide from the other five genes. If this restriction were due to a functional constraint inherent to the NS gene, then a similar restriction might be observed between wt NS genes as well. The NS gene of B/Lee/40 is 1096 nucleotides long. Both ca and wt B/AA/1/66 NS genes had an additional two nucleotides due to the insertion of two additional adenosine residues within a run of three existing adenosine residues at positions 39 to 41 of the B/Lee/40 sequence. This two-nucleotide insertion occurred in the 5'-noncoding region of mRNA and does not affect the NS1 and NS2 protein sequences. In addition to the two unmatched adenosine residues, there are 61 base mismatches between the B/AA/1/66 wt virus and B/Lee/40 virus. This amount of change is roughly the same as that found between the other known genes of B/Lee/40 and B/AA/1/66 viruses (DeBorde *et al.*, 1987), indicating that the minimal nucleotide change occurring between wt and ca B/AA/1/66 NS gene pair is probably not due to an inherent functional restriction in the NS gene, but to some other reason. Whether it is specific to cold-adaptation or random chance is not known.

Genetic studies utilizing reassortants that were made between ca B/AA/1/66 and wt B/HN/76 have indicated that the PA gene is the primary determinant of the ts and att phenotypes (Donabedian *et al.*, 1987). A combination of other ca B/AA/1/66 genes (without the PA gene) could also contribute to attenuation. A

reassortant that contained the NP gene from ca B/AA/1/66 virus and all other genes from B/Texas/1/84 was shown to be ca and ts in PCK cells, and partially attenuated in ferrets (unpublished data, this laboratory). In the study of Donabedian and colleagues (1987) no single-gene NP reassortant was isolated, but reassortants containing the NP and other genes from B/AA/1/66 ca virus, without the PA gene, were shown to have a 2-log₁₀ reduction in growth at 39° compared to growth at 33°. One of these reassortants is also restricted for viral growth in ferret lungs, a probable marker for attenuation (Donabedian *et al.*, 1987). Thus the NP gene may also carry mutations involved in the final level of ts and att phenotypes of ca B/AA/1/66 virus and its ca vaccine reassortants. From the studies of the reassortants made with the wt B/HN/76 virus, the ca marker appears to arise from a combination of the genes of ca B/AA/1/66, since no individual gene conferred this phenotype. However, with the wt B/Texas/1/84 parent, as mentioned above, a single-gene NP reassortant did become cold-adapted in PCK cells. In addition to being tested in PCK instead of MDCK cells, one other possible reason for this inconsistency is that because the ca mutations *enable* growth, any wt gene which cannot function efficiently at 25° will abrogate the appearance of the ca phenotype, obscuring any number of ca enabling genes. In our studies, all genes except the NS gene, showed at least one unique amino acid change in its predicted protein sequence. Thus, every ca viral gene which has been implicated in the direct expression of one or more of the three phenotypes, by the above genetic studies, was shown to contain nucleotide and amino acid changes from those of the wt gene. Recently, a non-ts revertant of the PA gene was isolated and sequenced, and the sequence reversion has been shown to involve the codon position, 1320, PA gene (Fig. 8). The experiments detailing this revertant gene are presented in the companion paper (Donabedian *et al.*, 1988).

We have presented a catalog of changes that existed between the wt and ca B/AA/1/66 viruses that should provide a valuable basis for understanding the mechanisms by which the ca donor virus confers the ca, ts, and att phenotypes to new ca vaccine reassortants. The changes that have occurred in the sequences of genes from phenotypic revertant viruses can be compared to those of the wt and ca B/AA/1/66 viruses in order to link specific mutations to specific phenotypes. This sequence information also provides a basis for the design of site mutagenesis experiments aimed at establishing directly the importance of some sequence sites to phenotypes not amenable to sin-

gle-gene or revertant analysis (like the ca phenotype). Finally, understanding which of the six nonsurface genes of ca B/AA/1/66 are essential for the stability and attenuation of the ca reassortant live virus vaccine, and which genes are not, can lead to the laboratory creation of more effective ca reassortant donor viruses.

ACKNOWLEDGMENTS

We thank Susan Donabedian for the expert technical help in the preparation and sequencing of viral RNA. The work in this paper was supported by Contract 1-AI-52564, National Institute of Allergy and Infectious Diseases, Development and Application Branch, Bethesda, Maryland 20892, and by a Biomedical Research Support Grant RR 05584-21.

REFERENCES

- AKOTO-AMANFU, E., SIVASUBRAMANIAN, N., and NAYAK, D. P. (1987). Primary structure of the polymerase acidic (PA) gene of an influenza B virus (B/Sing/222/79). *Virology* **159**, 147-153.
- 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 NS₁ and NS₂ proteins. *J. Virol.* **42**, 186-193.
- BRIEDIS, D. J., LAMB, R. A., and CHOPPIN, P. W. (1982). Sequence of RNA segment 7 of the influenza B virus genome: Partial amino acid homology between the membrane proteins (M₁) of influenza A and B viruses and conservation of a second open reading frame. *Virology* **116**, 581-588.
- BRIEDIS, D. J., and TOBIN, M. (1984). Influenza B virus genome: Complete 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.
- COX, N. J., KITAME, F., KLIMOV, A., KOENNECKE, I., and KENDAL, A. P. (1986). Comparative studies of wild-type and cold-mutant (temperature-sensitive) influenza virus: Detection of mutations in all genes of the A/Ann Arbor/6/60(H2N2) mutant vaccine donor strain. *Microbiol. Pathol.* **1**, 387-397.
- 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**, 275-282.
- DEBORDE, D. C., NAEVE, C. W., HERLOCHER, M. L., and MAASSAB, H. F. (1987). 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. *Virus Res.* **8**, 33-41.
- DONABEDIAN, A. M., DEBORDE, D. C., COOK, S., SMITKA, C. W., and MAASSAB, H. F. (1988). A mutation in the PA protein of cold-adapted (ca) B/Ann Arbor/1/66 influenza virus associated with reversion of the ts phenotype and attenuated virulence. *Virology* **163**, 444-451.
- DONABEDIAN, A. M., DEBORDE, D. C., COOK, S., SMITKA, C. W., and MAASSAB, H. F. (1988). A mutation in the PA protein of cold-adapted (ca) B/Ann Arbor/1/66 influenza virus associated with reversion of the ts phenotype and attenuated virulence. *Virology* **163**, 444-451.
- HIEBERT, S. W., WILLIAMS, M. A., and LAMB, R. A. (1986). Nucleotide sequence of RNA segment 7 of influenza B/Singapore/222/79:

- Maintenance of a second large open reading frame. *Virology* **155**, 747–751.
- KEITEL, W. A., CATE, T. R., and COUCH, R. B. (1986). Evaluation of a cold-recombinant influenza B vaccine. In "Options for the Control of Influenza" (A. P. Kendal and P. A. Patriarca, Eds.), pp. 287–291. A. R. Liss, New York.
- KEMDIRIM, S., PALEFSKY, J., and BRIEDIS, D. J. (1986). Influenza B virus PB1 protein: Nucleotide sequence of the genome RNA segment predicts a high degree of structural homology with the corresponding influenza A virus polymerase protein. *Virology* **152**, 126–135.
- LONDO, D. R., DAVIS, A. R., and NAYAK, D. P. (1983). Complete nucleotide sequence of the nucleoprotein gene of influenza B virus. *J. Virol.* **47**, 642–648.
- MAASSAB, H. F., DEBORDE, D. C., DONABEDIAN, A. M., and SMITKA, C. W. (1985). Development of cold-adapted "master" strains for type B influenza virus vaccines. In "Vaccines 85" (R. A. Lerner, R. M. Chanock, and F. Brown, Eds.), pp. 327–332. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- MAASSAB, H. F., DEBORDE, D. C., DONABEDIAN, A. M., and SMITKA, C. W. (1986). Prospects for influenza type B live attenuated vaccines. In "Options for the Control of Influenza" (A. P. Kendal and P. A. Patriarca, Eds.), pp. 271–286. A. R. Liss, New York.
- MARTIN, R. (1987). Overcoming DNA sequencing artifacts: Stops and compressions. *Focus* **9**, 8–10.
- MOUNT, S. M. (1982). A catalog of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472.
- PALESE P., and YOUNG, J. F. (1983). Molecular epidemiology of influenza virus. In "Genetics of Influenza Viruses" (P. Palese and D. W. Kingsbury, Eds.), pp. 321–336. Springer-Verlag, New York.
- PARVIN J. D., MOSCONA, A., PAN, W. T., LEIDER, J. M. and PALESE, P. (1986). Measurement of the mutation rates of animal viruses: Influenza A virus and poliovirus type 1. *J. Virol.* **59**, 377–383.
- PEACOCK, A. C., and DINGMAN, C. W. (1968). Molecular weight estimation and separation of ribonucleic acid by electrophoresis in agarose-acrylamide composite gels. *Biochemistry* **7**, 668–674.
- PEATTIE, D. A. (1979). Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. USA* **76**, 1760–1764.
- QUEEN, C., and KORN, L. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* **12**, 581–599.
- REEVE, P., PIBERMANN, M., BACHMAYER, H., LIEHL, E., MORITZ, A., GANZINGER, U., HOFMANN, H., and KUNZ, C. (1982). Studies in man with a cold-recombinant live influenza B virus vaccine. *J. Med. Virol.* **9**, 1–9.