VIROLOGY 163, 429–443 (1988)

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. MAASSAB[2]

*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.

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.

0042-6822/88 $3.00

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 suboptimal 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 nonsurface 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 Scientific 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

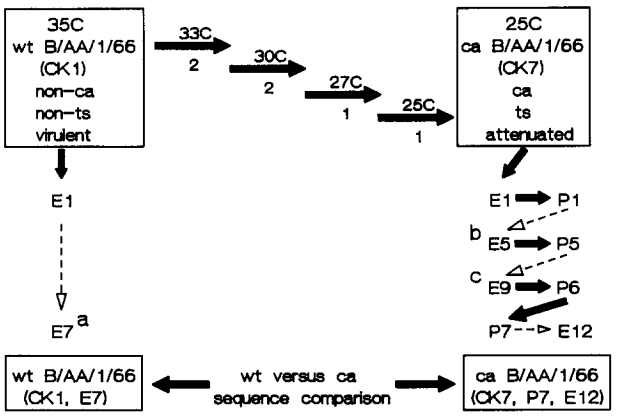

FIG. 1. Cold-adaptation and passage history of wt and ca B/AA/1/66 viruses. The cold-adaptation process is represented by the staggered arrows connecting the top two boxes. The passages were done in PCK cells at the temperatures indicated above each arrow. The number below the arrow represents number of passages at each temperature. The dashed arrows represent a number of replicative steps as follows: [a]Consecutive embryonated egg passages of the wt virus pool at high dilution. [b]Consecutive plaquing and egg amplification of the ca virus pool. [c]Consecutive 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 μl, not 50 ml as erroneously stated in that paper.

End-labeling of vRNA

Purified vRNA (30 to 50 μg) and 100 μCi of [^{32}P]cytidine bis-phosphate (pCp) were dried down together in a vacuum concentrator and resuspended in 5 μl of 2× RNA ligase buffer, (100 m*M* HEPES, pH 7.5, 20 μg/ml BSA, 6.6 μ*M* ATP, 30 m*M* $MgCl_2$), 1 μl dimethylsulfoxide, and 1 μl 0.5 m*M* ATP. Thirty to thirty-five pCp ligation units of T4 RNA ligase was added, and the total volume was adjusted to 10 μl with H_2O 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 μl 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 H_2O and mixed with an equal volume of loading buffer (0.2× TBE, 0.2% SDS, 20% sucrose, and 0.01% bromphenol 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× TBE, (0.09 *M* Trizma base, 0.09 *M* boric acid, and 2 m*M* Na_2EDTA) (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× TBE buffer containing 0.0l% SDS and 1 μg/ml ethidium bromide for 25 min at room temperature. The gel was then removed from the ethidium bromide solution, rinsed with 1× 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× TBE buffer plus 0.01% sarcosine. The high salt retention buffer was 3.0 *M* sodium acetate with 0.03% sarcosine and 0.01% bromphenol blue. Elution proceeded at 100 V, constant voltage, for l-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

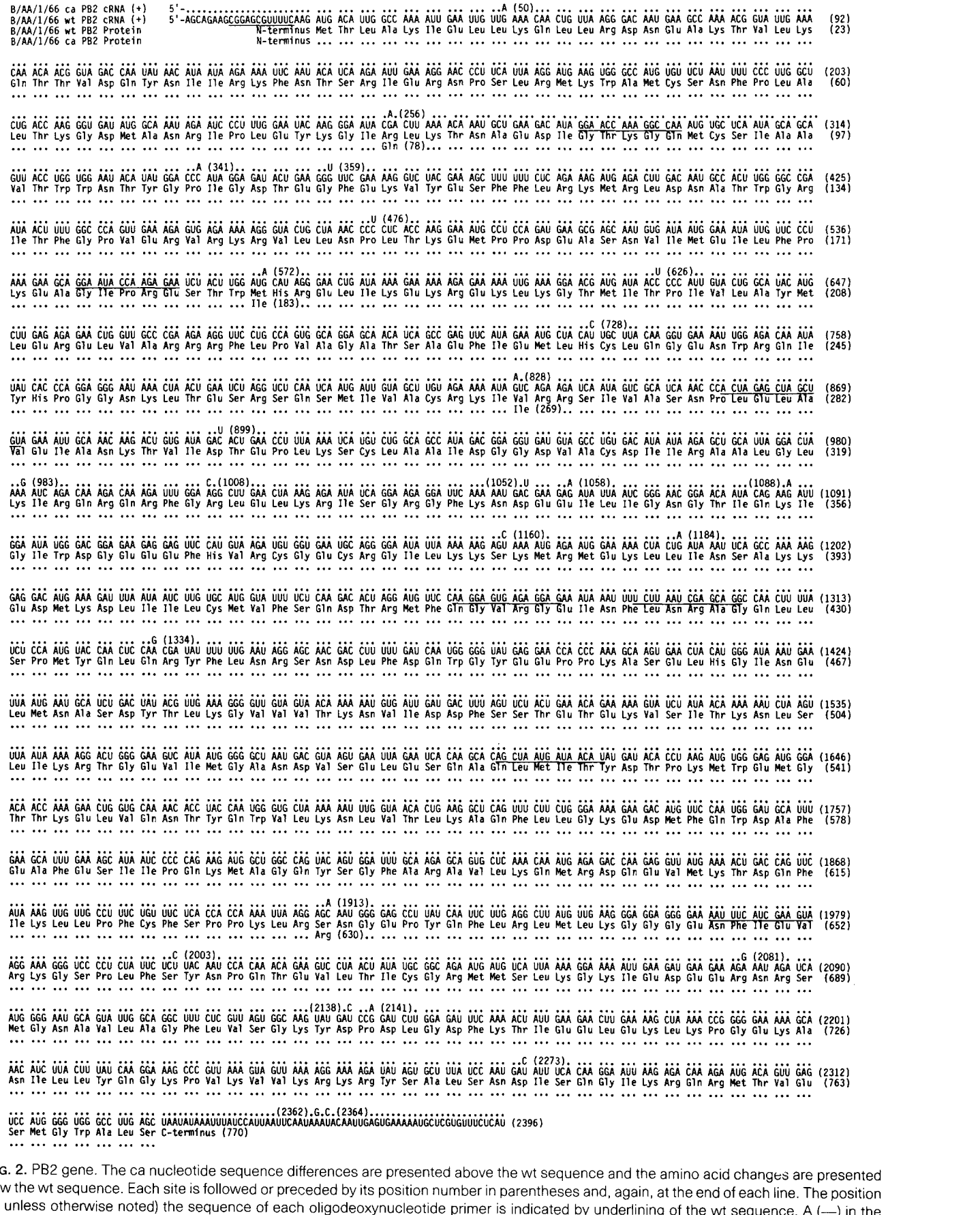

FIG. 2. PB2 gene. The ca nucleotide sequence differences are presented above the wt sequence and the amino acid changes are presented below the wt sequence. Each site is followed or preceded by its position number in parentheses and, again, at the end of each line. The position (and unless otherwise 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.

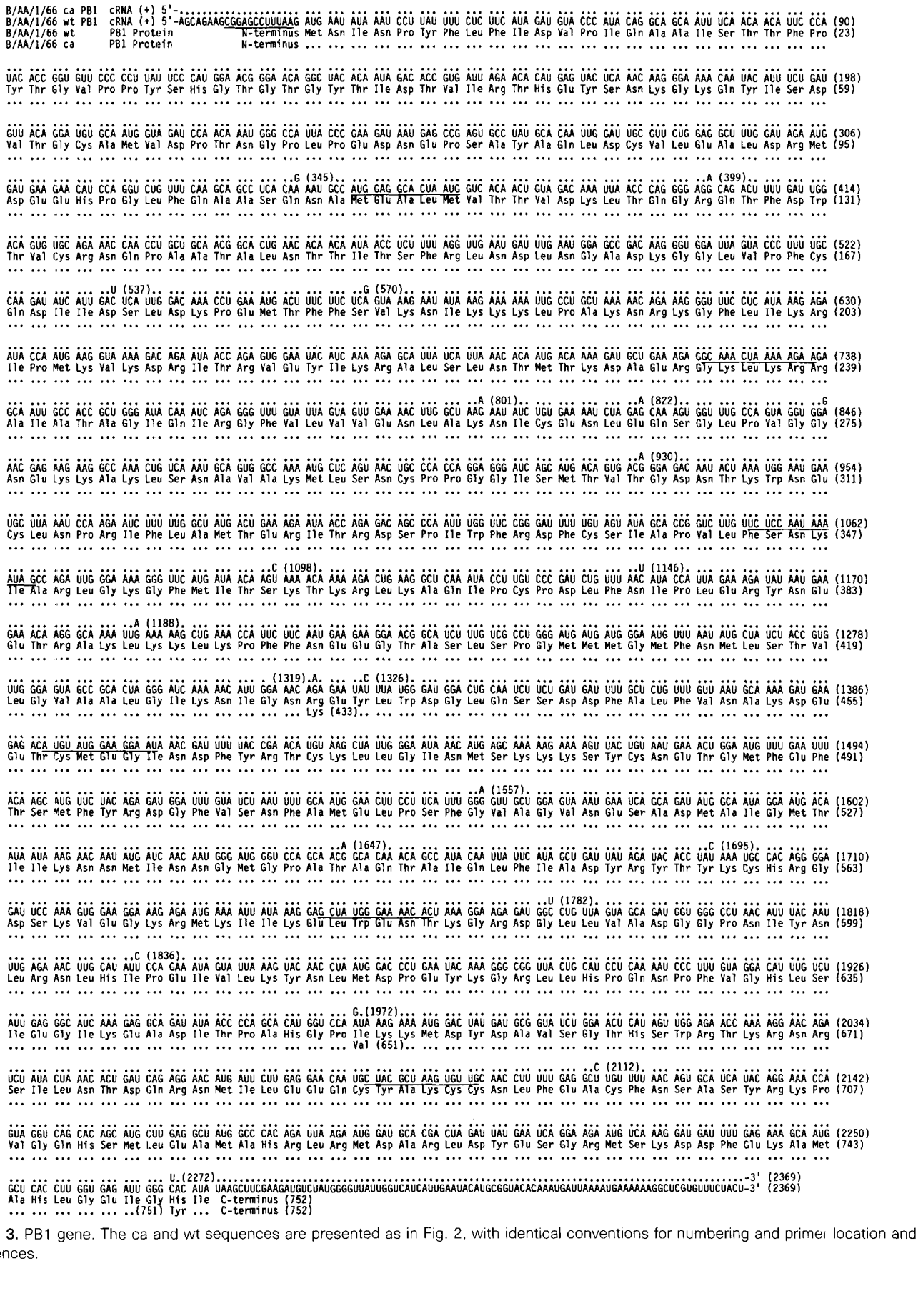

FIG. 3. PB1 gene. The ca and wt sequences are presented as in Fig. 2, with identical conventions for numbering and primer location and sequences.

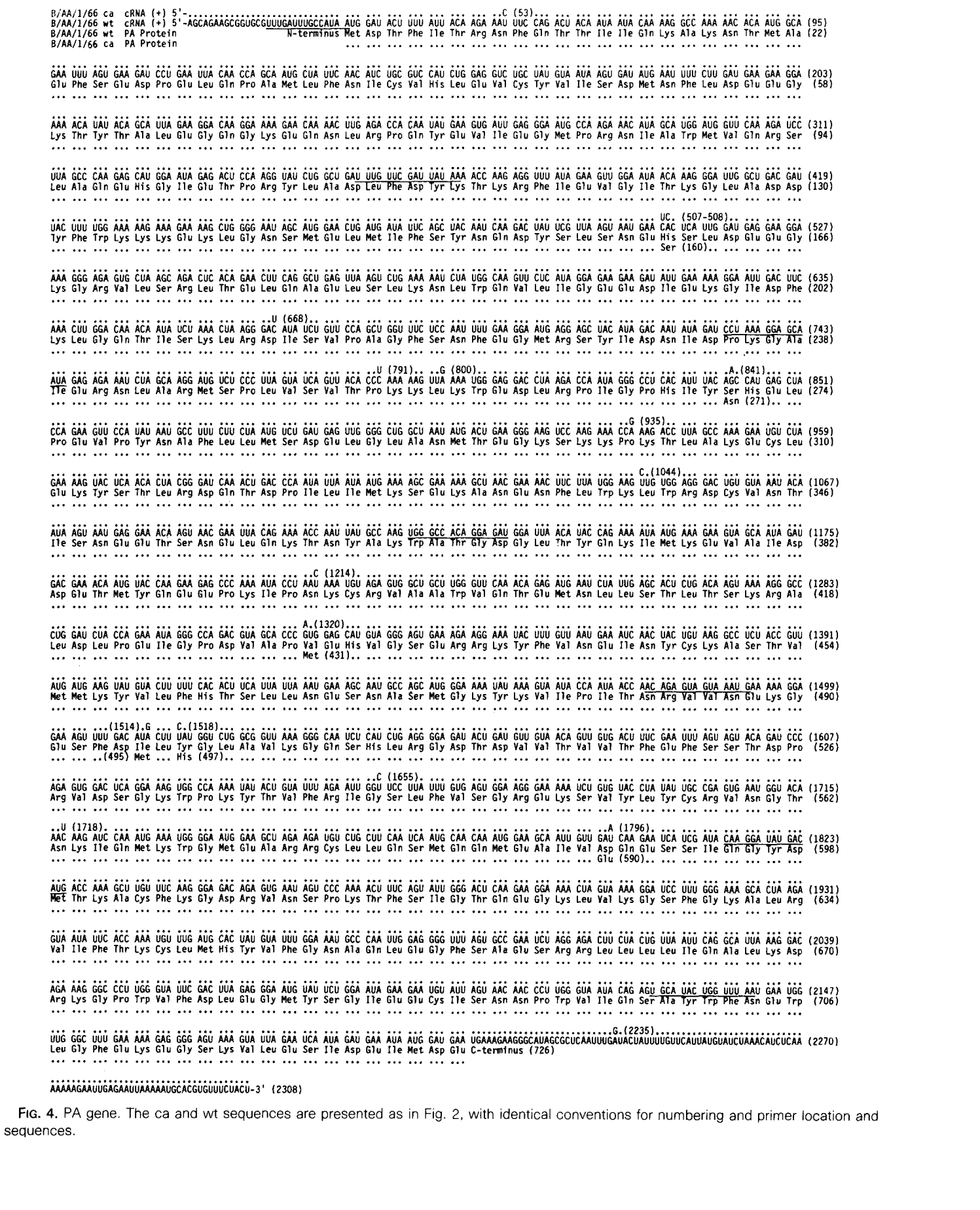

FIG. 4. PA gene. The ca and wt sequences are presented as in Fig. 2, with identical conventions for numbering and primer location and sequences.

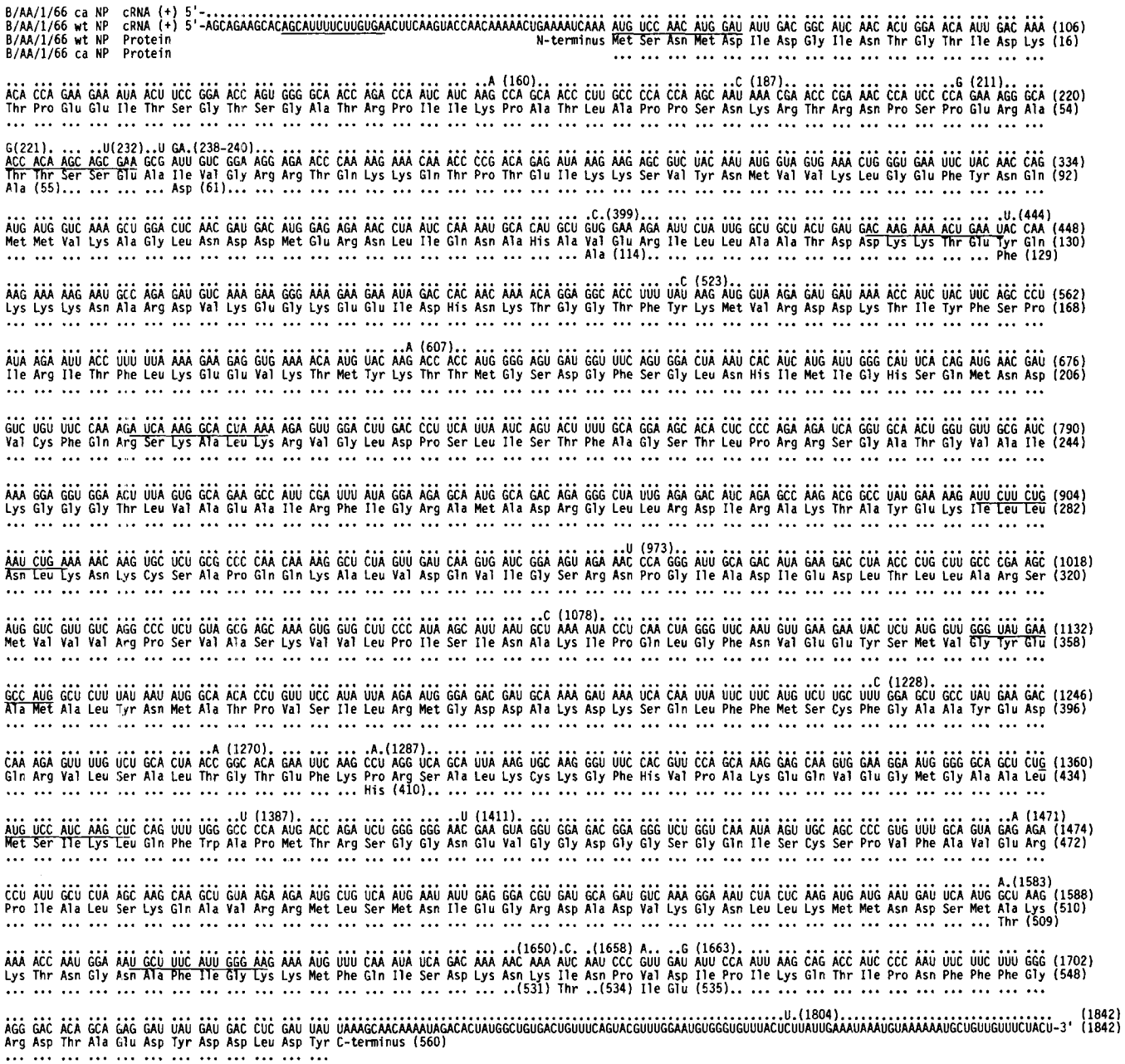

FIG. 5. NP gene. The ca and wt sequences are presented as in Fig. 2, with identical conventions for numbering and primer location, with one exception. The primer sequence starting at position 221 is actually ACC ACA AGC AGT GAA (see text).

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. Oligodeoxynucleotide 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.

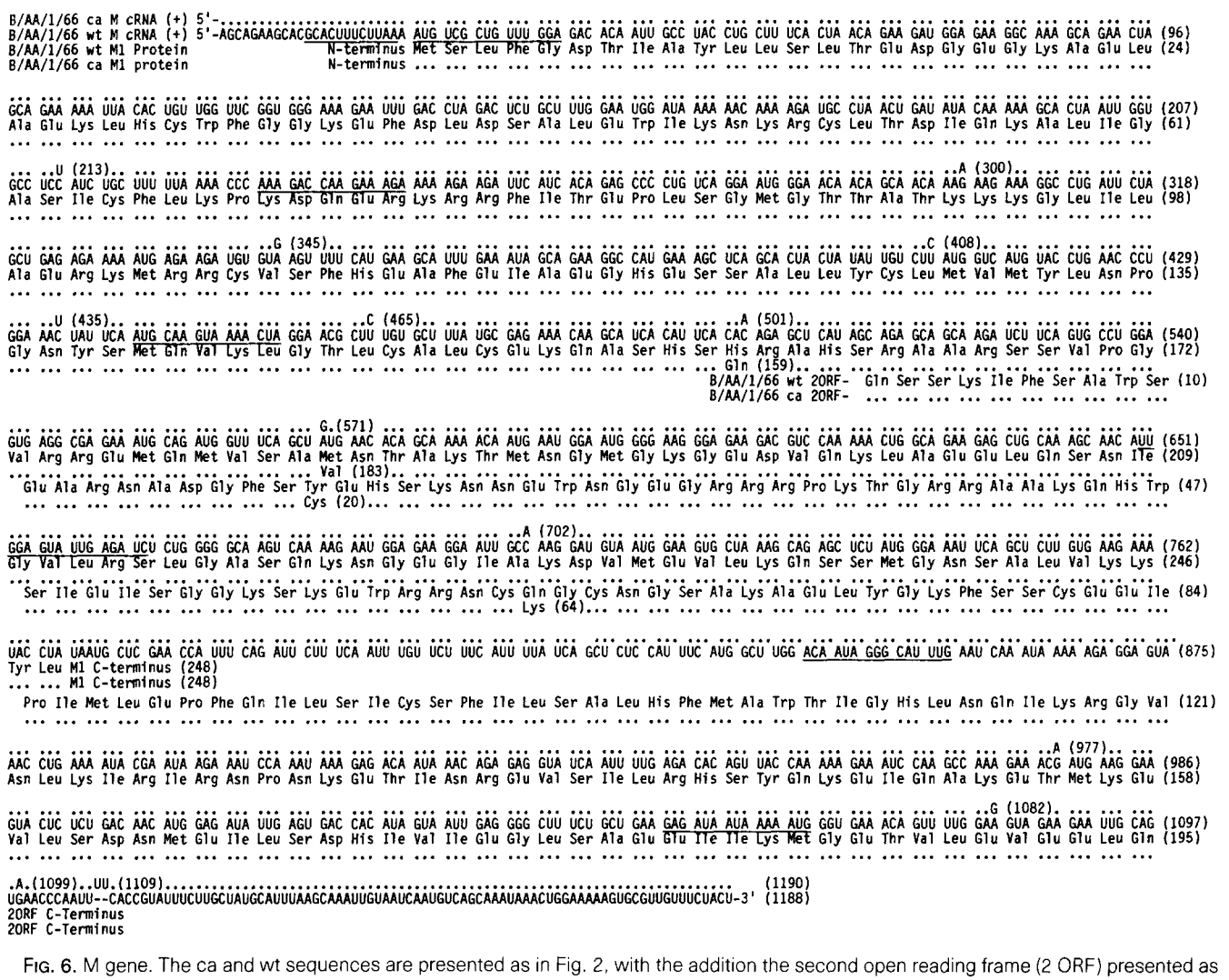

FIG. 6. M gene. The ca and wt sequences are presented as in Fig. 2, with the addition the second open reading frame (2 ORF) presented as two additional lines as indicated in the figure.

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 indentical 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

colleagues (1986) for B/Lee/40 virus and by DeBorde and colleagues (1987) for B/AA/1/66 wt virus. The ca B/AA/1/66 PB1 gene varied from the wt PB1 gene by 21 base substitutions all within the coding region of the gene, which resulted in a total of three predicted amino acid changes. By comparison of the ca B/AA/1/66 gene sequence to that of B/Lee/40 (Kemdirim *et al.*, 1986) only one of these amino acid changes remained unique to the cold-adapted virus (Fig. 8). From the reassortment analysis of Donabedian and colleagues (1987), PB1, like PB2, apparently played no significant independent role in the establishment of the three phenotypes.

PA gene

The PA gene sequence has been reported for wt B/AA/1/66 virus (DeBorde *et al.*, 1987) and is also known for B/Singapore/79 (Akoto-Amanfu *et al.*, 1987). Our laboratory has also sequenced the PA gene of B/Houston/76 virus. The ca PA gene varied from the wt gene by 17 base substitutions, one of which was in the cRNA 3′-noncoding region. The 16 changes in the coding region resulted in six amino acid changes between the wt and ca PA polypeptides. Comparison with the PA sequences of B/Houston/76 and B/Singapore/79, showed that only two of the six amino acid changes were unique to the ca B/AA/1/66 virus. As shown in the following article (Donabedian *et al.*, 1988), the codon of one of these unique positions (position 1320) underwent a change resulting in the reversion of both ts and att phenotypes.

NP gene

The NP gene has the largest number of base substitutions per nucleotide of any of the six gene pairs examined. Of the 25 changes in the RNA, 24 of them were in the coding region, accounting for nine amino acid changes from wt B/AA/1/66 NP protein to ca B/AA/1/66 NP protein. Comparison of these nine changes with the known NP gene sequences of

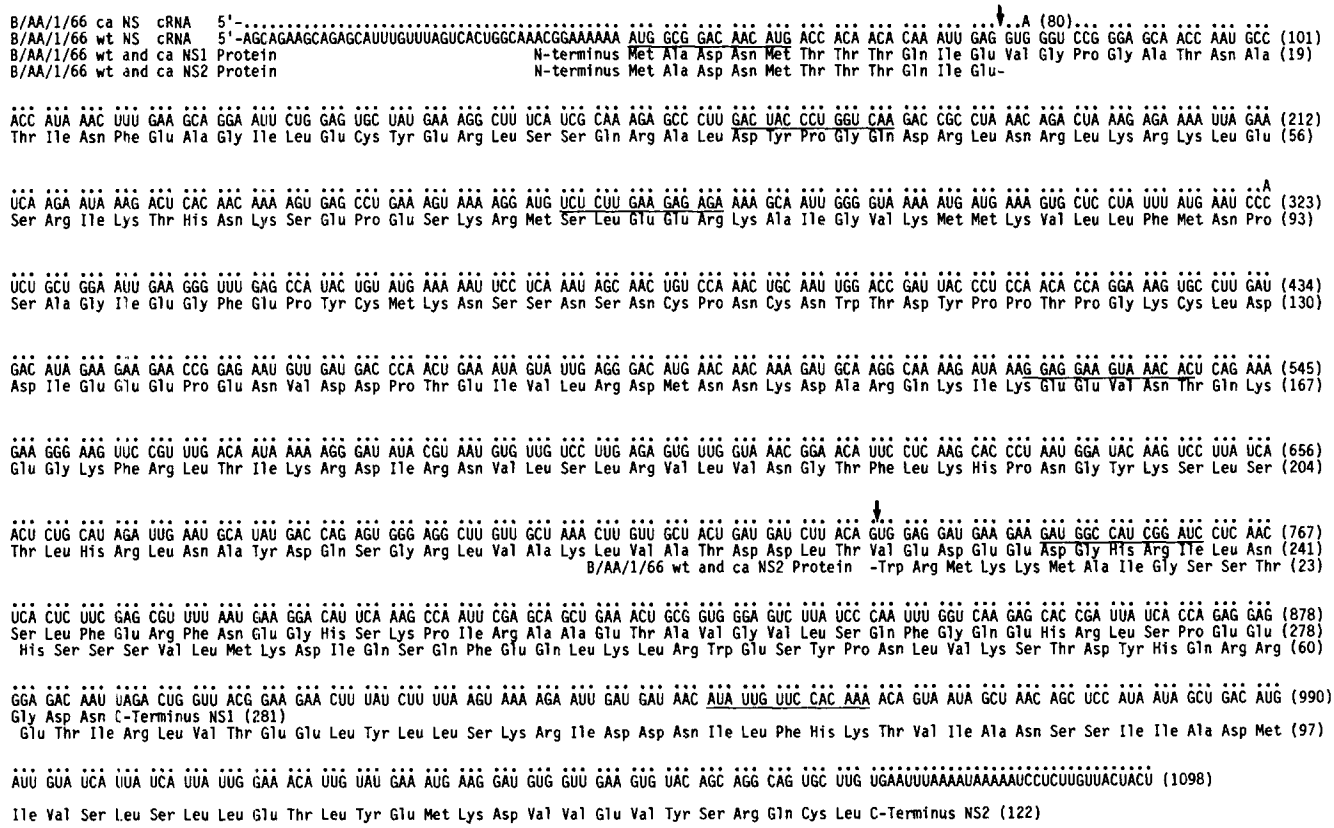

FIG. 7. NS gene. The two nucleotide substitutions in the ca virus NS gene sequence are shown above the wt gene sequence. Because neither base change resulted in an amino acid change, the ca and wt NS1 and NS2 proteins are shown together rather than separately as in the other sequence figures. The arrows represent the presumed splicing sites based on the sequence data given for B/Lee/40 virus (Briedis and Lamb, 1982).

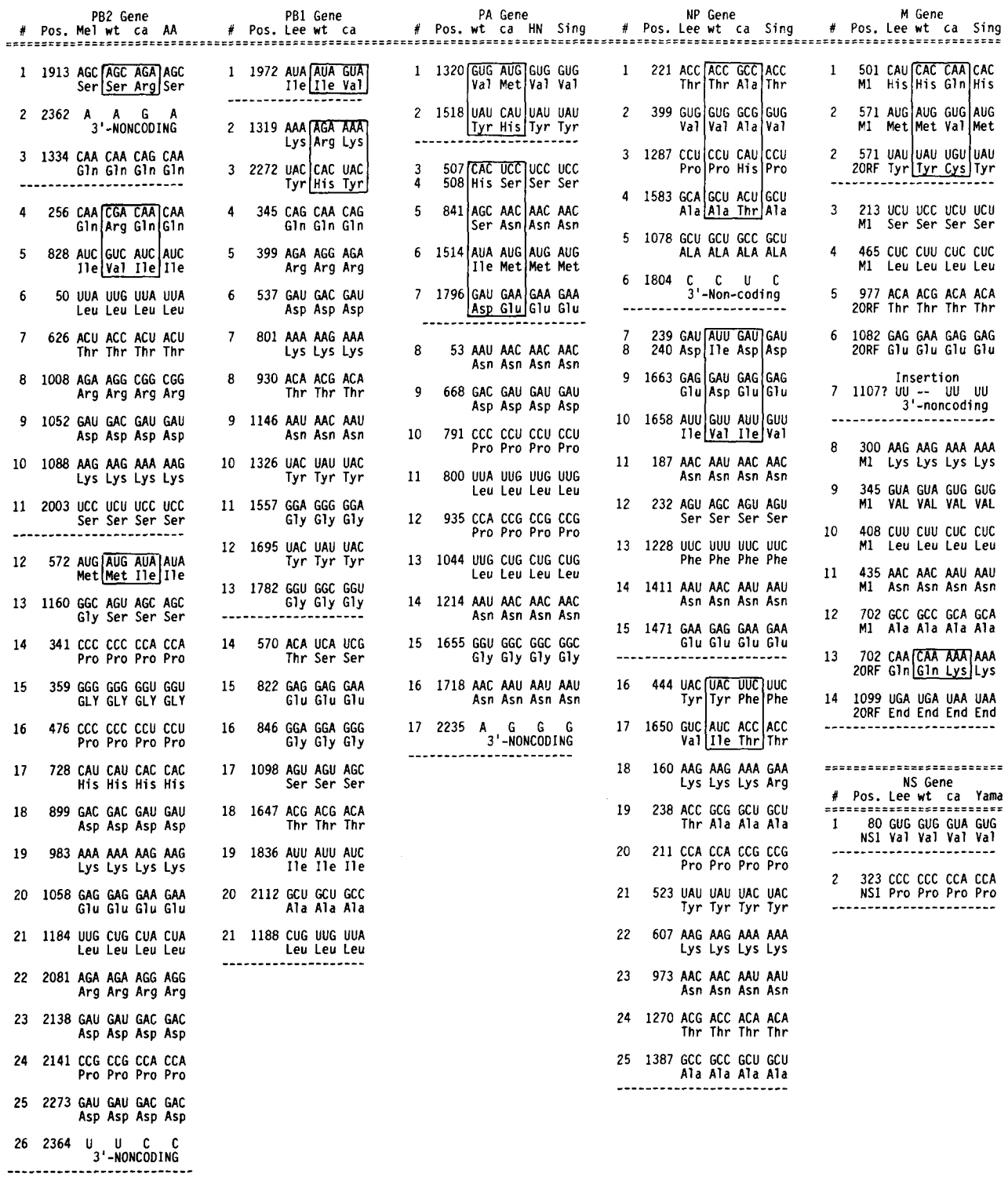

FIG. 8. Unique and nonunique wt to ca changes. The sequence differences between the wt and ca B/AA/1/66 nonsurface genes are compared to sequences of other known influenza type B viruses. They are presented as part of the codon of which they are a member, rather than just a single nucleotide difference. 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: Mel, 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 M_1 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 M_1 protein region, and two predicted amino acid changes in the possible product of the 2 ORF. None of the M_1 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 M_1 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 NS_1 coding region, and neither of these resulted in amino acid changes in the NS_1 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 M_1 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 × 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 CK_1 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 nonunique 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-$\log_{10}$ 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 single-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 NS_1 and NS_2 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 (M_1) 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 seqeunce 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.