INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

U·M·I Dissertation Information Service

University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106

Buonagurio, Deborah A.

EVOLUTIONARY PATTERNS OF INFLUENZA A AND C VIRUSES IN MAN

City University of New York PH.D. 1986

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages __✓__
2. Colored illustrations, paper or print _____
3. Photographs with dark background __✓__
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __✓__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

_____

_____

University
Microfilms
International

EVOLUTIONARY PATTERNS OF INFLUENZA A AND C VIRUSES IN MAN

by

Deborah A. Buonagurio

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1986

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

June 20 1986
Date

Peter Palese, Ph.D.
Chairman of Examining Committee

July 23, 1986
Date

Terry A. Krulwich, Ph.D.
Executive Officer

Jerome L. Schulman, M.D.

Edwin D. Kilbourne, M.D.

James G. Wetmur, Ph.D.

Edward J. Bottone, Ph.D.

Vincent R. Racaniello, Ph.D.

Supervisory Committee

The City University of New York

ABSTRACT

EVOLUTIONARY PATTERNS OF
INFLUENZA A AND C VIRUSES IN MAN

by

Deborah A. Buonagurio

Advisor: Dr. Peter Palese

The evolution of influenza A viruses was examined by comparative nucleotide sequencing of the NS gene of fifteen human viruses isolated over a 53 year period. Nucleotide substitutions are distributed throughout the NS gene and many base changes found in the earliest isolates are retained in subsequent strains. These data suggest a common evolutionary history for the NS genes of the viruses. Using the maximum parsimony method, a phylogenetic tree was constructed which allowed for the calculation of an evolutionary rate for the NS gene of influenza A viruses. The rapid rate of nucleotide change is approximately 2×10^{-3} substitutions per site per year, a million-fold higher than DNA genes of the eukaryotic host. Additionally, the uniform rate of substitution in the NS gene represents a good molecular clock in nature. The observed pattern of sequence changes in the NS gene is consistent with an evolutionary model for influenza A viruses whereby variants emerge successively with time through rapid accumulation of genetic changes. We suggest

that positive selection plays a significant role in the unusual and rapid evolution of influenza A viruses in nature.

The genetic variability of influenza C viruses was analyzed by nucleotide sequencing of the HA and NS genes of eight human viruses isolated between 1947 and 1983. The pattern of sequence changes observed in the influenza C virus genes is different from the influenza A viruses in that many of the base substitutions are not cumulative with time. The extent of nucleotide differences among the HA and NS genes of the influenza C viruses does not correlate with the isolation dates of the strains. An evolutionary model for influenza C viruses was proposed which states that influenza C variants belonging to multiple phylogenetic lineages appear to cocirculate in nature. The evolutionary pattern of influenza C viruses is therefore quite different from that of influenza A viruses, but similar to that of most other RNA and DNA viruses.

In summary, the influenza A viruses evolve sequentially over time by rapidly accumulating changes in their genomes. Influenza C viruses, in contrast, appear to cocirculate and to evolve more slowly in nature.

ACKNOWLEDGEMENTS

I would like to thank Peter Palese for his support and encouragement throughout my graduate study. I have learned from him what it means to be a good scientist and my training in his laboratory has given me a solid foundation for future successes.

Many thanks to Drs. Edwin Kilbourne, James Wetmur, and Jerome Schulman for support and good advice throughout my stay at Mount Sinai.

Very special thanks to Ed Bottone for always being there to listen and to guide through all the ups and downs.

Many thanks to all my good friends and colleagues on the 16th floor!

--To Gerry Norton--whose everlasting supply of salami and provolone got me through many days.

--To Mark Krystal--for getting me started on the right foot (He should be the one thanking me for being a good target for ice bullets!)

--To Susumu Nakada--for always being there to help solve any problem--If this thesis could have a coauthor, it would be Dr. Nakada.

--To Peter Graves--for many interesting conversations, anecdotes, numerous pictures, and for allowing me to look through his binoculars at some strange sights outside the office window.

--To Tony Bonilla--for rescuing my thesis from the jaws of the IBM PC and for remedying many, many, many other computer maladies.
--To Jo Ellen Barnett--for all her help in many matters and for great instruction in the fine arts which should put me a notch above the level of ignorant!

I would like to extend special thanks and love to my parents, Rose and Joe, and family for standing by me.

....and very special thanks and love to Billy (Pos!) for helping me in sc many ways to achieve this degree and for always being there.

Special thoughts and prayers in memory of Marlene Lin--who was always eager to make life a little easier and more pleasurable for all of us students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

Most of the data presented in the Results section of this thesis have been published or are in press:

1. Krystal, M., D. A. Buonagurio, J. F. Young, and P. Palese. 1983. Sequential muatations in the NS genes of influenza virus field strains. J. Virol. 45:547-554.

2. Buonagurio, D. A., M. Krystal, P. Palese, D. C. DeBorde, and H. F. Maassab. 1984. Analysis of an influenza A virus mutant with a deletion in the NS segment. J. Virol. 49:418-425.

3. Krystal, M., S. Nakada, D. A. Buonagurio, D. C. DeBorde, H. F. Maassab, and P. Palese. 1984. The nonstructural gene segment of influenza A virus: Expression of NS1 protein in mammalian cells; Analysis of a deletion mutant. In: Proceedings of the 5th International Symposium on Negative Strand Viruses, D. H. L. Bishop and R. W. Compans, eds., Elsevier, New York, pp. 147-157.

4. Buonagurio, D. A., S. Nakada, U. Desselberger, M. Krystal, and P. Palese. 1985. Noncumulative sequence changes in the hemagglutinin genes of influenza C virus isolates. Virology 146:221-232.

5. Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch. 1986. Evolution of human influenza A viruses over 50 years: Rapid, uniform rate of change in NS gene. Science 232:980-982.

6. Buonagurio, D. A., S. Nakada, W. M. Fitch, and P. Palese. Epidemiology of influenza C virus in man: Multiple evolutionary lineages and low rate of change. Virology (in press).

7. Norton, G. P., T. Tanaka, K. Tobita, S. Nakada, D. A. Buonagurio, D. Greenspan, M. Krystal, and P. Palese. Infectious influenza A and B virus variants with long carboxyl terminal deletions in the NS1 polypeptides. Virology (submitted).

INTRODUCTION

The influenza viruses, or orthomyxoviridae, are enveloped animal viruses containing a segmented RNA genome. The influenza viruses are classified into three types, A, B, and C, based on their immunologically distinct nucleoprotein (NP) and matrix (M) protein antigens. Influenza A viruses are further grouped into antigenic subtypes according to their specific hemagglutinin (HA) and neuraminidase (NA) surface proteins. The first human influenza A virus was isolated in 1933. In addition to man, influenza A viruses have been isolated from avian species and a wide variety of mammals including horses, pigs, and seals. The only known reservoir for influenza B viruses is man and up until recent times, influenza C viruses were only found in the human population.

Influenza A viruses are intensely studied due to their association with periodic pandemic and epidemic outbreaks of disease in man. The problem of prevention of widespread disease is a reflection of the unique property of influenza A viruses to undergo rapid changes in antigenic character. Variability is not restricted to the major surface antigens, HA and NA, of the virus, but also occurs in genes coding for nonstructural proteins as well as proteins comprising the core of the virus particle. As soon as an immune response can be generated to an antigenic variant, the virus can mutate into a "new"

variant, rendering the antibodies raised to the previous challenge ineffective. In this way, the influenza virus can evade the host immune response, thus making it quite difficult to develop an effective vaccine against the virus. Today's attenuated strains used to vaccinate individuals, may not provide protection against tomorrow's newly emerging epidemic variants.

Influenza B viruses are associated with epidemic disease and do undergo antigenic variability, which appears not to be as extreme or frequent as that found in influenza A viruses. Influenza C viruses only cause mild respiratory infections in man. Antigenic differences have been demonstrated among influenza C virus variants, but variation in these viruses and their epidemiology in nature has not been well studied.

The major focus of this thesis is to further characterize the evolution of natural variants of human influenza A viruses and to provide information on the epidemiology and evolution of influenza C viruses in nature.

1. Structure of the Influenza A Virus Particle

Intact influenza virus particles contain approximately 1% RNA, 70% protein, 6% carbohydrate, and 23% lipid (Frommhagen et al., 1959). The virions have a spherical shape when visualized in the electron microscope

using negative staining techniques. The virus particles are 80-120 nm in diameter and possess a lipid envelope which is covered by spikes (Wrigley, 1979). The spikes, of two morphologically distinct types, represent the more abundant HA and the NA proteins (Laver and Valentine, 1969; Murti and Webster, 1986). Both the HA and NA are glycoproteins which are the major antigens of the virus. Recurrent influenza epidemics are associated with changes in the antigenic structure of the HA and the NA. The HA is involved in the binding of the virus to cellular sialic acid containing receptors (Hirst, 1941) and mediating virus entry into cells by a low pH-induced membrane fusion event in endosomal vesicles (White et al., 1982; Daniels et al, 1985b; Gething et al., 1986; Wharton et al., 1986). The NA hydrolyzes sialic acid from the HA receptors on the cell surface, thereby aiding in the release of the progeny virions from the cell membrane (Varghese et al., 1983; Webster et al., 1984).

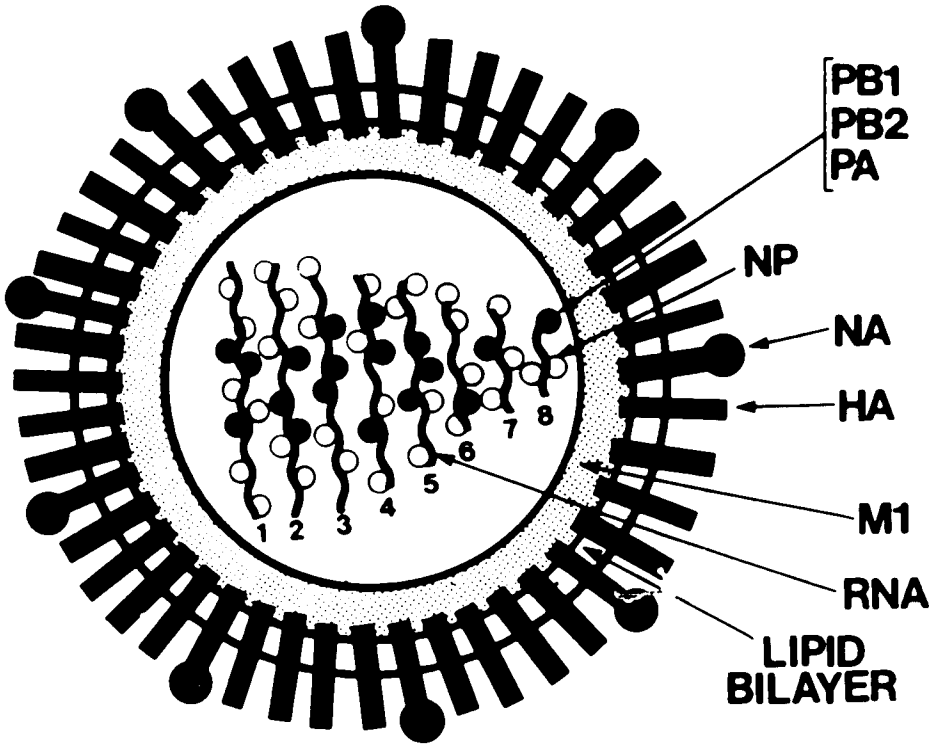
Influenza viruses derive their lipid envelope by budding at the plasma membrane of the infected cell. Directly beneath the viral lipid bilayer is an electron-dense layer composed of the most abundant virion protein, M1. It is believed that the M1 contributes to the structural integrity of the virion and also plays a role in the maturation of the virus at the plasma membrane prior to budding (Gregoriades et al., 1984).

Disruption of the outer coat of the virus reveals the internal ribonucleoprotein (RNP) stucture which contains the eight single-stranded RNA segments of the viral genome in close association with four proteins. The NP is the predominant polypeptide of the nucleocapsid. The other minor constituents of the RNP complex are the polymerase (P) proteins, PB1, PB2, and PA. The three P proteins form a complex which has RNA-dependent transcriptase activity required for initiating viral RNA transcription in the infected cell (Braam et al., 1983; Kato et al., 1985). A schematic diagram of the structure of the virion is shown in Fig. 1.

2. Structure of the Genome

Influenza A viruses possess a segmented genome of eight single-stranded RNAs of negative polarity (Palese, 1977). Partial RNA sequencing of the 5' and 3' ends of the eight RNAs from influenza A viruses reveals a conserved sequence of 13 nucleotides at the 5' terminus of each segment and a conserved 12 base sequence at the 3' terminus of each segment (Skehel and Hay, 1978a; Desselberger et al., 1980). The genetic information of the virus codes for at least ten polypeptides: seven proteins that have both structural and enzymatic functions are found in the virion (PB1, PB2, PA, HA, NA, NP, and M1) and three nonstructural proteins (NS1, NS2, and M2) are synthesized only in infected cells (reviewed by Lamb, 1983). The

Fig. 1. Schematic diagram of the structure of an influenza virus particle. Proteins PB1, PB2, PA, and NP comprise the virion core and are associated with the genome RNAs. M1 is in close contact with the lipid bilayer. The HA and NA spikes are surface glycoproteins.


PB1
PB2
PA
NP
NA
HA
M1
RNA
LIPID
BILAYER
1
2
3
4
5
6
7
8

complete genome of the influenza A virus has been mapped and recombinant DNA technology has made possible the determination of the complete nucleotide sequence of the entire genome of the influenza virus A/PR/8/34 (Fields and Winter, 1982; Winter and Fields, 1982; Winter et al., 1981a, 1981b; Winter and Fields, 1981; Fields et al., 1981; Winter and Fields, 1980; Baez et al., 1980). Over recent years, many additional nucleotide sequences of influenza virus RNA segments have been elucidated which have extended our knowledge on the molecular biology of the virus. Furthermore, considerable information on the structure of influenza virus proteins has been generated, highlighted by the determination of the 3-dimensional structure at approximately 3A resolution of the HA (Wilson et al., 1981) and NA (Varghese et al., 1983) proteins. The structure of each gene segment and the polypeptide(s) it encodes are described in the sections below and summarized in Table 1. The segments are numbered 1 to 8 in decreasing order of their electrophoretic mobility in polyacrylamide gels.

2A. RNA Segments 1, 2, and 3: Polymerase Proteins PB1, PB2, and PA

The proteins of greatest molecular weight of the virus are the three polymerase proteins, originally designated P1, P2, and P3 in decreasing order of their electrophoretic

TABLE 1. INFLUENZA A VIRUS RNAS AND PROTEINS[a]

Segment	Length[b] (Nucleotides)	Encoded Protein	Amino Acid[c] Length (Nascent Protein)	Approx. No. Molecules per Virion
1	2,341	PB2	759	30-60
2	2,341	PB1	757	30-60
3	2,233	PA	716	30-60
4	1,778	HA	566	500
5	1,565	NP	498	1,000
6	1,413	NA	454	100
7	1,027	M1	252	3,000
		M2	97	
8	890	NS1	230	
		NS2	121	
	13,588			

[a]Adapted from Lamb (1983)

[b]Values for A/PR/8/34 strain

[c]Deduced from nucleotide sequence

mobility in polyacrylamide gels (apparent MW 94-82,000)(Lamb and Choppin, 1976; Inglis et al., 1976; Plotch et al., 1981). Analysis of recombinants of strains A/PR/8/34 and A/HK/68 revealed that RNA segment 1 coded for P3, RNA segment 2 for P1, and RNA segment 3 for P2 (Palese et al., 1977b; Ritchey et al., 1977). A problem in P protein nomenclature based on gel mobility arose when it was discovered that another strain, A/FPV/Rostock/34, had a different coding assignment of P proteins to RNA segments. A less confusing biochemical identification of the P proteins was found by isoelectric focusing of the proteins using 2-dimensional gel electrophoresis (Horisberger, 1980). This technique showed that two of the P polypeptides were relatively basic and one was relatively acidic. The faster migrating species of the two basic proteins coded by RNA segment 1 is called PB2, the basic P protein with the slower mobility coded by RNA segment 2 is called PB1, and the acidic P protein coded by RNA segment 3 is designated PA.

The precise functions of the polymerase proteins in transcription and replication of viral RNAs have not been completely elucidated. Studies with temperature-sensitive (ts) mutants defective in RNA synthesis have suggested that PB1 and PB2 are involved in messenger RNA (mRNA) synthesis and PA in virion RNA (vRNA) production (Krug et al., 1975; Palese et al., 1977a; Scholtissek and Bowles, 1975).

Detailed experiments using an *in vitro* transcription system have indicated that the role of PB2 is in binding to the cap structure of the host cell mRNA primer (Ulmanen et al., 1981, 1983; Blaas et al., 1982a, 1982b; Penn et al., 1982) and that of PB1 in initiating transcription (Ulmanen et al., 1981; Plotch et al., 1981; Braam et al., 1983).

2B. RNA Segment 4: Hemagglutinin Protein (HA)

The HA protein is encoded by RNA segment 4 of the influenza virus. The HA protein was originally named because of the ability of the virus to agglutinate erythrocytes (Hirst, 1941) by attachment to glycoprotein receptors containing sialic acid (Hirst, 1942). The HA protein is an integral membrane glycoprotein responsible for the attachment of the virus to cells. Antibodies directed against the HA protein can neutralize virus infectivity.

In addition, the HA molecule plays a role in the penetration of the virus into the host cell. The HA protein undergoes a low pH-induced conformational change that initiates the fusion of the virus envelope with the endosome membrane, liberating the genome into the cytoplasm (White et al., 1982; White et al., 1983; Doms et al., 1985; Daniels et al., 1985b; Wharton et al., 1986). The fusion activity requires post-translational processing of the HA precursor into the HA1 and HA2 disulfide-linked subunits. This cleavage event exposes a very hydrophobic amino acid

sequence at the N-terminus of the HA2 subunit (fusion peptide) which has been implicated in participating in the fusion activity (Wilson et al., 1981; Skehel et al., 1982). Recently, several stages of the mechanism of HA-mediated membrane fusion have been delineated through analysis of naturally occurring fusion variants (Doms et al., 1986) and mutant HAs expressed in SV40 virus vectors, engineered by oligonucleotide-directed mutagenesis within the fusion peptide coding region of the HA cDNA (Gething et al., 1986).

On the surface of the virion envelope, the HA protein exists as a trimer of non-covalently linked monomers (Wilson et al., 1981). A monomer of HA (MW 77,000) is translated as a single polypeptide chain which is processed proteolytically into two polypeptide subunits, HA1 (MW 50,000) and HA2 (MW 27,000)(Lazarowitz et al., 1973; Wilson et al., 1981) held together by disulfide bonds (Wilson et al., 1981; Waterfield et al., 1981; Selimova et al., 1982). The cleavage of the HA is required for infectivity of the virus (Lazarowitz and Choppin, 1975), fusion activity (White et al., 1983), and spread of infection in the host (Rott et al., 1980).

Complete nucleotide sequences of HA genes of several variants of the H1, H2, and H3 subtypes have been determined (Table 2). The overall basic structure of the different HA types are very similar. The nascent HA

TABLE 2. COMPARISON OF THE HA GENE STRUCTURE AMONG HUMAN VIRUS SUBTYPES[a]

	H1[b]	H2[c]	H3[d]
Nucleotide Length	1778	1773	1765
Coding Nucleotides	1698	1686	1698
5' Untranslated Region (NT)	32	43	29
3' Untranslated Region (NT)	48	44	38
Coding Region (AA)	566	562	566
Signal Sequence (AA)	17	15	16
HA1 (AA)	326	324	328
HA2 (AA)	222	222	221
Amino Acids Lost Between HA1 and HA2	1	1	1

[a]Modified from Lamb (1983)

[b]A/PR/8/34 strain (Winter et al., 1981a)

[c]A/Jap/305/57 strain (Gething et al., 1980; Waterfield et al., 1980)

[d]A/Aichi/2/68 strain (Verhoeyen et al., 1980; Ward and Dopheide, 1981a)

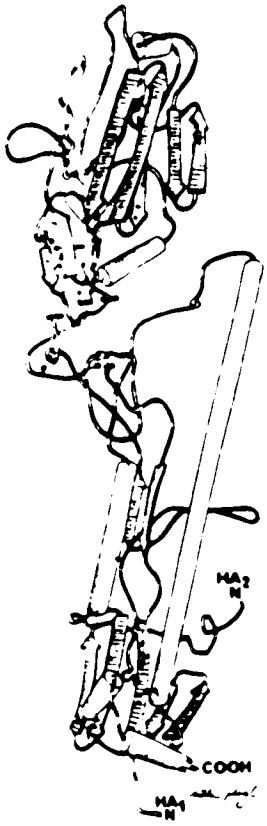
polypeptide contains an N-terminal hydrophobic signal sequence (15-17 amino acids) that is proteolytically removed after translocation across the endoplasmic reticulum membrane. The signal sequence is required for glycosylation of the HA and transport to the cell surface (Gething and Sambrook, 1982; Sekikawa and Lai, 1983). The HA molecule is further processed by a trypsin or trypsin-like protease to generate the HA1 and HA2 subunits with the loss of a single arginine residue between the chains (Garten et al., 1981). The removal of the arginine residue suggests that an additional exopeptidase cleavage is involved in HA activation. The N-terminus of the HA2 chain consists of a hydrophobic sequence that is conserved between strains within the same or different subtype (Min Jou et al., 1980; Wilson et al., 1981). The N-terminus of HA2 has been implicated in the process of virus penetration of the host cell membrane to initiate infection (Daniels et al., 1985b). A second hydrophobic sequence (approximately 24 residues) is located near the C-terminus of HA2. This sequence spans the lipid bilayer and anchors the HA in the lipid envelope (Gething and Sambrook, 1982; Sveda et al., 1982). The primary amino acid sequence of the HA contains several possible glycosylation signals (Asn-X-Ser/Thr) where carbohydrate can attach to asparagine residues via N-glycosidic linkages. There are also several sites where

disulfide bonds may form between cysteine residues within HA1, HA2, or between the two chains.

2B.1. The 3-Dimensional Structure of the HA

Cleavage of the A/HK/68 (H3) HA with bromelain releases an antigenically and structurally intact molecule which has lost only the C-terminal hydrophobic region of HA2. This molecule can be crystallized and the atomic structure has been determined to 3A resolution by X-ray crystallography (Wilson et al., 1981). The molecule, shown in Fig. 2, is an elongated cylinder 135 A long consisting of a) a long fibrous stem extending 76 A from the membrane containing two antiparallel alpha-helices which terminate near the membrane in a compact five-stranded antiparallel beta-sheet globular fold and b) a distal globular region of antiparallel beta-sheet structure. The distal globular head is comprised of HA1 residues and is connected to the HA2 fibrous stem by only two antiparallel chains from HA1. The C-terminus of HA1 is 21 A from the N-terminus of HA2, indicating a great conformational change arises when the HA is proteolytically cleaved. The N-terminus of HA1 is located very close to the membrane. Four antigenic sites (A, B, C, and D) have been proposed which are found at the top of the globular region (Wiley et al., 1981; Webster and Laver, 1980). These sites were derived from an analysis of amino acid changes in naturally occurring antigenic variants and laboratory variants selected with monoclonal

Fig. 2. Drawing of the 3-dimensional structure of the Hong Kong (H3) HA monomer showing folding of the HA1 and HA2 polypeptides. Flat, twisted arrows represent beta-sheets; cylinders represent alpha-helices; filled circles represent disulfide bonds (from Wilson et al., 1981).


COOH

antibodies (Wilson et al., 1981; Laver et al., 1980, 1981; Both and Sleigh, 1981; Yewdell et al., 1979; Webster and Laver, 1980). Although the 3-dimensional structure has only been elucidated for the H3 subtype HA, sequence comparisons with the H1 and H2 HAs suggest similar 3-dimensional configurations based on conserved amino acids (22% of HA1; 45% of HA2) including six disulfide bridges and many structurally important residues such as proline and glycine. In addition, when the positions of amino acid variation of an H1 HA (A/PR/8/34) are plotted on the 3-dimensional structure of the H3 HA monomer, clusters of variant amino acids are seen which in general, correspond to the defined antigenic groupings (Gerhard et al., 1981; Caton et al., 1982; Concannon et al., 1984; Raymond et al., 1986).

2C. RNA Segment 5: Nucleoprotein (NP)

RNA segment 5 codes for the NP protein which interacts with the viral RNA and the three P proteins to form RNP particles (Compans et al., 1972). The NP comprises up to 25% of the virion protein (Skehel and Schild, 1971). It is believed that the NP constitutes the backbone of the RNP complex. Based on the total length of the viral genome and the number of NP molecules per virion, it can be estimated that a single NP molecule interacts with approximately 20 nucleotides (Winter and Fields, 1981). Although the precise

function of NP is unclear, the analysis of ts NP mutants has suggested that NP is involved in viral RNA synthesis (Krug et al., 1975; Ritchey and Palese, 1977; Thierry and Danos, 1982). In addition, Van Wyke et al. (1981) have demonstrated that monoclonal antibodies directed against two antigenic domains of the NP can inhibit in vitro transcription of viral RNA. Recently, Kato et al. (1985) have purified an RNA polymerase-RNA complex from influenza virus that lacks NP protein. They have shown that this complex is active in synthesizing viral messenger RNAs (mRNA) in vitro. However, the synthesis of full-length complementary RNA (cRNA) transcripts may require the complex containing NP protein.

The NP protein has been shown to be modified post-translationally by phosphorylation and proteolytic cleavage. NP is phosphorylated at a single serine residue per molecule by cellular phosphokinases, but it is not clear what role the phosphate group plays (Petri and Dimmock, 1981; Kistner et al., 1985). Zhirnov and Bukrinskaya (1981, 1984) have reported that two forms of phosphorylated NP protein are detected in cells infected with human influenza viruses: one uncleaved 56,000 MW form (NP56) and a cleaved form of 53,000 MW (NP53) derived from the uncleaved product. The NP56 form appears to be preferentially incorporated into budding virions,

suggesting that RNPs containing NP56 are selected for virus assembly.

The NP protein accumulates in the nucleus of virus-infected cells (Briedis et al., 1981a; Hamaguchi et al., 1985). Davey et al. (1985) have identified an NP nuclear accumulation signal sequence (amino acids 327-345) by in vitro mutagenesis of NP cDNAs which directed the synthesis of mutant NP proteins after injection into Xenopus oocytes.

The NP is the type specific antigen of influenza viruses used to classify viruses into types A, B, and C. Minor antigenic differences among NPs of influenza A viruses have been detected (Schild et al., 1979) and studies with monoclonal antibodies to NP have elucidated at least three non-overlapping antigenic sites on the molecule (Van Wyke et al., 1980).

2D. RNA Segment 6: Neuraminidase (NA)

Neuraminidase is an integral membrane glycoprotein of the virus encoded by RNA 6. NA constitutes about 5% of the total virion protein (Skehel and Schild, 1971) and is present on the virion surface in less abundance than the HA. The NA protein has an enzymatic activity which can hydrolytically cleave the ketosidic linkage between N-acetylneuraminic acid (sialic acid) and an adjacent sugar residue (Gottschalk, 1957). The NA exists on the viral membrane as a tetrameric complex (MW 220,000) which forms a mushroom-shaped structure with a stalk and head (Wrigley, 1979; Varghese et al., 1983; Colman et al., 1983). The individual polypeptide chains of MW 56,000 are held together by disulfide bonds (Lazdins et al., 1972; Varghese et al., 1983).

There are a number of possible functions associated with the enzymatic activity of the NA in virus replication and maturation: 1) release of budding virus particles from the host cell membrane (Palese and Schulman, 1974; Webster et al., 1984); 2) prevention of self-aggregation of the virus (Palese and Schulman, 1974); 3) exposing the HA to proteolytic cleavage, thereby altering viral virulence (Schulman and Palese, 1977); and 4) penetration of virus into the host cell via low pH-induced membrane fusion (Huang et al., 1985). Antibodies to NA do not neutralize or prevent virus infection (Kilbourne et al., 1968), but do

act to reduce the extent of pathology in the host by restricting multiple cycles of replication (Schulman et al., 1968).

The complete nucleotide sequence of the NA gene has been determined for both N1 and N2 subtypes of influenza A viruses (Fields et al., 1981; Hiti and Nayak, 1982; Markoff and Lai, 1982; Bentley and Brownlee, 1982). Even though the N1 and N2 NAs have very low amino acid sequence homology, there are structural similarities between the two proteins. The deduced amino acid sequences of the NA protein reveal an N-terminal hydrophobic extended signal sequence which serves to anchor the protein in the virus membrane and is long enough to span the lipid bilayer (Blok et al., 1982). In addition, this sequence is required for translocation of the nascent NA polypeptide into the rough endoplasmic reticulum, which leads to glycosylation and cell surface expression (Bos et al., 1984; Markoff et al., 1984; Jones et al., 1985). Unlike the HA protein, post-translational cleavage of the NA polypeptide does not occur. The signal sequence of the NA is not proteolytically removed following translocation across the membrane and even the initiating methionine is retained (Blok et al., 1982). Processing at the C-terminus of the molecule has also not been observed.

2D.1. The 3-Dimensional Structure of the NA

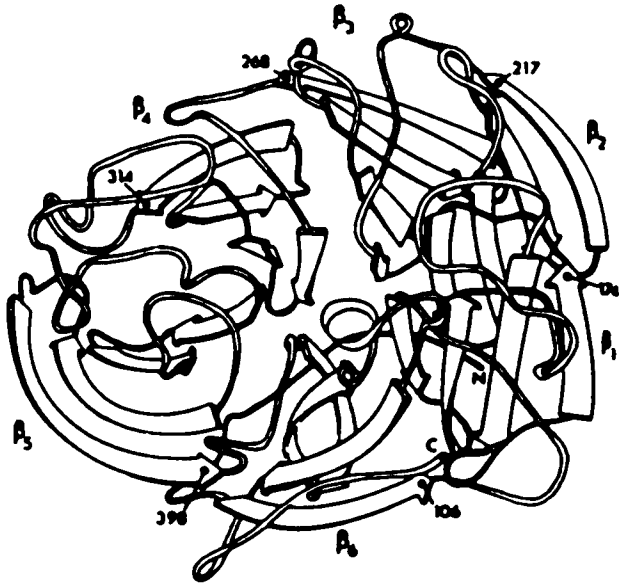
The 3-dimensional structure of the pronase-released N2

NA head of the A/Tokyo/3/67 strain was solved at 2.9 A resolution using X-ray crystallography (Varghese et al., 1983). The molecule has a box-shaped globular head with four-fold symmetry, connected to the virus membrane by a long slender stalk. The polypeptide chain is considered to be unique. Viewed from above the head, each monomer contains six topologically identical beta-sheets arranged in a propeller formation (Fig. 3). Similar to the work described for the HA, the locations of amino acid substitutions in the NA of a series of drifted field variants and in laboratory variants selected by monoclonal antibodies, have been mapped on the 3-dimensional structure to define antigenic determinants on the NA molecule (Laver et al., 1982; Webster et al., 1982a; Colman et al., 1983; Webster et al., 1984; Air et al., 1985). The antigenic sites cluster preferentially into the distal surface loops which connect the various strands of the beta-sheets. The catalytic and sialic acid binding sites of the NA have also been defined on the 3-dimensional structure.

2E. RNA Segment 7: Matrix Protein (M1); Nonstructural Protein (M2)

RNA segment 7 codes for two proteins, M1 and M2, and possibly a small peptide M3. Three separate mRNAs derived from RNA 7 are found in virus-infected cells, one of which is a colinear transcript of the genome segment and two are

Fig. 3. Drawing of the 3-dimensional structure of the N2 NA. The folding of a single polypeptide subunit is shown as viewed from above. The six beta-sheet regions are indicated (from Varghese et al., 1983).


β3
268
217
β4
β2
314
104
β1
N
β5
C
318
106
β6

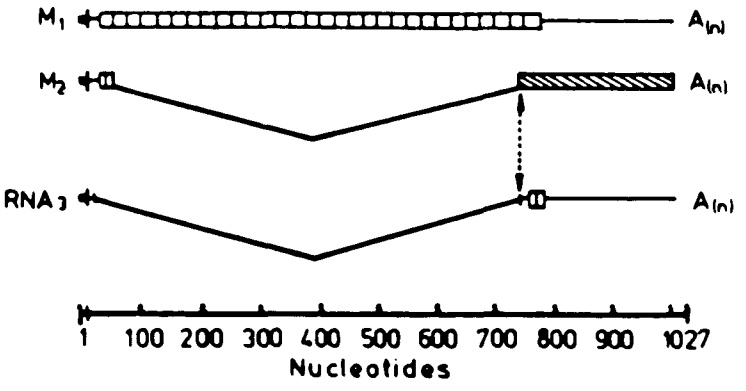
spliced mRNAs probably derived from the colinear transcript (Fig. 4).

The M1 is the most abundant protein in the virion, comprising approximately 35% of the total protein (Skehel and Schild, 1971). M1 is a type-specific antigen which allows serological discrimination of type A, B, and C influenza viruses. Electron microscopy of virions reveals that the M1 forms an electron-dense layer beneath the lipid bilayer. In addition to providing structural stability to the virion envelope, M1 may recognize the viral glycoproteins and form a domain on the inner surface of the plasma membrane which subsequently provides a binding site for the RNP segments during virus assembly (Choppin et al., 1972; Lohmeyer et al., 1979). The mechanism of virus particle formation has not been elucidated.

Gregoriades et al. (1984) have detected two forms of M1, containing similar tryptic peptides, in virus particles and infected cells. Both proteins are phosphorylated at predominantly serine residues. The function of phosphorylation of M1 is unknown.

The complete sequence of RNA segment 7 has been determined for three human strains (Winter and Fields, 1980; Lamb and Lai, 1981; Ortin et al., 1983). In addition to the long open reading frame encoding the M1 protein, there is a second open reading frame (+1 frame) at the 5' end of the vRNA which could code for 97 amino acids and

Fig. 4. Structure of the mRNAs derived from RNA 7 (M gene) of influenza A virus. The filled-in boxes at the 5' ends of the mRNAs represent nucleotides derived from host cell mRNAs. The thin lines at the 5' and 3' termini of the mRNAs represent noncoding regions. The hatched boxes designate coding nucleotides. The M2 mRNA is translated in a different reading frame from that used for the M1 in the region 740-1004. The V-shaped lines represent nucleotides removed by splicing in the M2 and $mRNA_3$ transcripts (from Lamb et al., 1981).


M1
A(n)
M2
A(n)
RNA 3
A(n)
1 100 200 300 400 500 600 700 800 900 1027
Nucleotides

which overlaps the M1 reading frame by 68 nucleotides (Winter and Fields, 1980; Allen et al., 1980; Lamb and Lai, 1981). A protein product derived from this reading frame, M2 (MW 15,000), was identified in influenza virus-infected cells and shown to be encoded by RNA segment 7 by hybrid-arrest translation experiments and by analysis of different recombinants (Lamb and Choppin, 1981; Palese et al, 1981). S1 nuclease mapping and nucleotide sequencing of mRNAs of RNA segment 7 showed that the M2 mRNA contained an interrupted region of 688 nucleotides (Lamb et al., 1981) as illustrated in Fig. 4. The 5' end of the M2 mRNA contains 51 bases which are the same as those found at the 5' end of the M1 mRNA. Following this shared sequence, the M2 mRNA body is 268 nucleotides that is 3'-coterminal with the M1 mRNA. The M1 and M2 proteins share nine amino acids at their N-termini, but the remainder of the M2 is translated in a different reading frame (+1) from that of the M1. The M1 and M2 polypeptides overlap by 14 codons.

The M2 protein has only been identified in infected cells and is therefore, a nonstructural protein (Lamb and Choppin, 1981). Detailed studies have revealed that the M2 is an integral membrane protein, not N-glycosylated, that is expressed abundantly at the cell surface (Lamb et al., 1985; Zebedee et al., 1985). A minimum of 18 N-terminal amino acids of M2 are exposed at the cell surface. Ten of these residues are conserved in all strains of influenza

virus (human and avain) whose M sequences are available (Winter and Fields, 1980; Allen et al., 1980; Lamb and Lai, 1981; McCauley et al., 1982; Ortin et al., 1983). The M2 protein possesses a single hydrophobic domain (amino acids 25-43) which is thought to anchor the protein in the membrane. The remainder of the M2 protein, approximately 54 C-terminal amino acids, comprises a long tail on the cytoplasmic side of the membrane. The function of the M2 in the virus life cycle has not been defined, but it is likely that the protein plays a role in organizing the assembly of the virion.

A second spliced mRNA derived from RNA segment 7 has been isolated from virus-infected cells (Lamb et al., 1981; Inglis and Brown, 1981). M $mRNA_3$ has a leader sequence of 11 bases that are the same as the 5' end of the M1 and M2 mRNAs, followed by a body of 268 nucleotides that is the same as that of the M2 mRNA. The M $mRNA_3$ can code for a peptide of only nine amino acids (identical to the C-terminus of the M1) whose existence has yet to be proved.

2F. RNA segment 8: Nonstructural Proteins (NS1 and NS2)

The smallest RNA segment of the influenza virus genome codes for two overlapping polypeptides, NS1 and NS2 (Inglis et al., 1979; Lamb and Choppin, 1979), that are only found in infected cells (Dimmock, 1969; Krug and Soeiro, 1975, Lazarowitz et al., 1971). NS1 is a phosphoprotein with

phosphate attached to one or two threonine residues per molecule (Privalsky and Penhoet, 1981). The NS1 polypeptide (MW 26,000) is synthesized in large amounts early in infection and accumulates in the cell nucleus (Briedis et al., 1981a). Within the nucleus, the NS1 is localized predominantly in the nucleoli or the nucleoplasm depending on the time of infection and also the strain of influenza virus (Young et al.,1983). The smaller of the two nonstructural proteins, NS2 (MW 14,000), is made late in infection and also has been shown by immunofluorescence analysis to be localized in the cell nucleus (Greenspan et al., 1985).

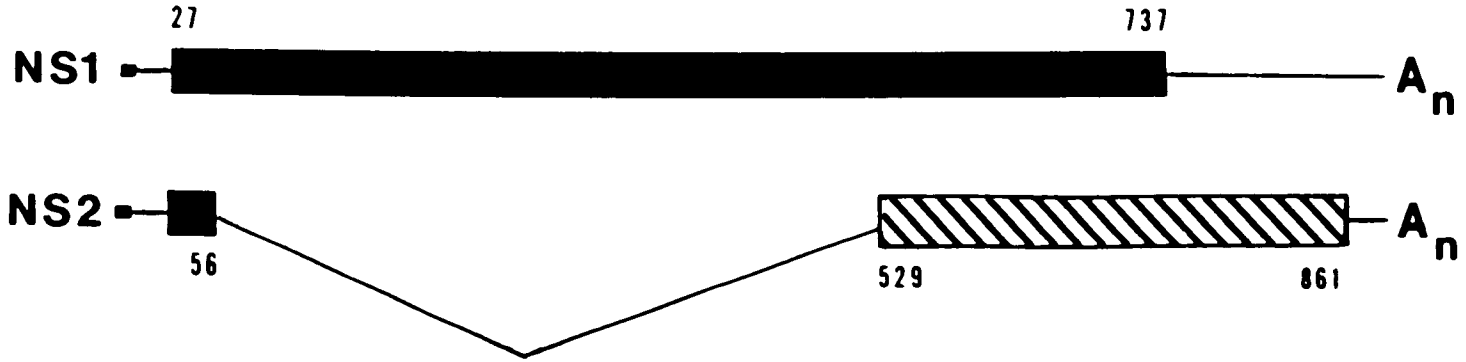
The functions of NS1 and NS2 in virus replication have not been elucidated. Examination of ts mutants of A/FPV/Rostock/34 defective in RNA segment 8 has revealed a variety of phenotypic characteristics including reduced M1 synthesis, defects in vRNA synthesis, and decreased HA production at the nonpermissive temperature (Robertson et al., 1983; Wolstenholme et al., 1980; Koennecke et al., 1981).

Previous to the present work, RNA segment 8 had been sequenced for only two human influenza A virus strains (Baez et al., 1980; Lamb and Lai, 1980; Winter et al., 1981b). The NS segment is 890 nucleotides in length. There is a 5' noncoding region of 26 bases followed by the NS1 open reading frame coding for 230 or 237 amino acids

depending on the virus strain (Fig. 5). The predicted amino acid sequence agrees well with the amino acid composition determined for the NS1 polypeptide of the A/WSN/33 strain (Shaw and Compans, 1978). A second open reading frame having the potential to code for 132 amino acids is found at the 5' end of the vRNA. This coding frame had been predicted from S1 nuclease mapping of NS mRNAs, hybridization blotting, and hybrid-arrest translation experiments using cloned DNA of RNA segment 8 (Lamb et al., 1980; Inglis et al., 1980).

There are two mRNAs corresponding to RNA segment 8 that are found in influenza virus-infected cells (Fig. 5). Nucleotide sequencing of the longer transcript, that codes for NS1, revealed a mRNA of about 860 bases which is colinear with the vRNA segment. Sequence analysis of the NS2 mRNA showed that it contains an interrupted region of 473 nucleotides. The first 56 virus-specific nucleotides at the 5' end of the NS2 mRNA are the same as those found at the 5' end of the NS1 mRNA. This NS2 leader sequence contains the initiation codon for protein synthesis plus information to code for nine amino acids which are common to NS1 and NS2. The remainder of the NS2 mRNA (340 bases) is translated in the +1 reading frame to yield an NS2 of 121 residues. The NS1 and NS2 proteins overlap by 70 codons translated from different reading frames.

Fig. 5. Schematic diagram of the NS1 and NS2 mRNAs derived from RNA 8 (NS gene) of the A/Udorn/72 virus (adapted from Lamb and Lai, 1980). The filled-in boxes at the 5' ends of the mRNAs represent nucleotides derived from cellular mRNA primers. The thin lines at the 5' and 3' termini of the mRNAs represent noncoding regions. The thick boxes designate protein coding regions. The nucleotide positions defining the coding regions are indicated. The body of the NS2 protein is translated in a different reading frame (cross-hatched box) from that used for the NS1 protein. The V-shaped line represents the region from nucleotide positions 56-529 which are spliced out of the NS2 mRNA.


27
737
NS1
An
NS2
An
56
529
861

The nucleotides at the 5' and 3' junctions of the interrupted NS2 mRNA are similar to the consensus splice signals present at the exon/intron borders of spliced eukaryotic mRNAs (Mount, 1982) indicating that the NS2 mRNA is probably generated from the NS1 transcript by a splicing mechanism.

3. Influenza B Virus

Influenza B viruses are similar in morphology and genetic composition to the influenza A viruses. The genome of the type B virus is composed of eight RNA segments (Ritchey et al., 1976a) and codes for polypeptides analogous to those of the type A viruses (Ueda et al., 1978; Racaniello and Palese, 1979). The gene segments of influenza B virus are all larger than the corresponding genes of influenza A viruses (Desselberger and Palese, 1978). Complete nucleotide sequences of segment 4 (HA), segment 5 (NP), segment 6 (NA), segment 7 (M), and segment 8 (NS) of influenza B viruses have been determined (Krystal et al., 1982; Hovanec and Air, 1984; Berton et al., 1984; Londo et al., 1983; Briedis and Tobin, 1984; Shaw et al., 1982a; Briedis et al., 1982; Briedis and Lamb, 1982). Recently, Kemdirim et al. (1986) have reported on the cloning and sequencing of RNAs 1, 2, and 3 of influenza B/Lee/40 virus. Cloned DNAs were identified as coding for PB1, PB2, and PA by two-dimensional analysis of *in vitro* translation products of mRNA selected by hybridization with

cDNA. Information on the genes and proteins of influenza B virus is summarized in Table 3.

A difference in coding capacity between the influenza A and B virus genomes is found in the RNA segment encoding the NA protein. In influenza B virus-infected cells, a glycoprotein product (MW 11,000) has been identified which is translated from the same mRNA as the NA protein, but in a different reading frame (Shaw et al., 1983; Shaw and Choppin, 1984). The B virus NA is translated from an open reading frame initiating at the second AUG codon from the 5' end of the mRNA. The first AUG is followed by an open reading frame of 100 amino acids that codes for the NB protein. An analogous open reading frame has not been found in influenza A virus NA genes.

The gene organization of RNA 7 of influenza B virus has the capacity to code for two overlapping proteins in different reading frames, but a protein product analogous to the M2 of influenza A virus has not been identified in B virus infected cells (Briedis et al., 1982).

The smallest RNA segment of influenza B virus, like that of the A virus, codes for two overlapping nonstructural polypeptides, NS1 and NS2 (Briedis et al., 1981b). Sequence analysis of the NS1 mRNA of the B/Lee/40 virus showed it to be colinear with the vRNA segment with a coding capacity of 281 amino acids (Briedis and Lamb, 1982). The NS2 mRNA shares 75 nucleotides with the NS1

TABLE 3. INFLUENZA B VIRUS RNAS AND PROTEINS[a]

Segment	Nucleotide Length		Encoded	MW	Amino Acid
	from Sequence[b]	(from PAGE)[c]	Protein	SDS-PAGE	Length[d]
1		2,800	PB2 or PA	93,000 or 80,000	
2	2,368	2,800	PB1	102,000	752
3		2,700	PB2 or PA	80,000 or 93,000	
4	1,882	2,100	HA	84,000	584
			HA1		346
			HA2		223
5	1,841	2,000	NP	66,000	560
6	1,557	1,700	NA	66,000	466
			NB	11,000[e]	100
7	1,191	1,150	M1	25,000	248
			M2	?	
8	1,096	1,000	NS1	40,000	281
			NS2	11,500[f]	122

[a]Modified from Air and Compans (1983)

[b]B/Lee/40 virus (Krystal et al., 1982; Shaw et al., 1982a; Briedis et al., 1982; Briedis and Lamb, 1982; Kendirim et al., 1986)

[c]Racaniello and Palese (1979)

[d]Deduced from nucleotide sequence

[e]Unglycosylated

[f]Briedis and Lamb (1982)

mRNA at the 5' end, resulting in a common N-terminal sequence of 11 amino acids for the NS1 and NS2 proteins. After this leader sequence, the NS2 mRNA contains an interrupted sequence of 655 nucleotides, probably generated by splicing. The remainder of the NS2 mRNA (350 bases) is translated in the +1 reading frame, resulting in an NS2 of 122 amino acids. The NS1 and NS2 coding regions overlap by 52 codons translated from different reading frames.

4. Influenza C Virus

Influenza C viruses share many features of structure and replication with other orthomyxoviruses, however, there are important differences in virion morphology, the number of RNA segments, and protein products.

Electron microscopic examination of C type virions reveals that the surface glycoproteins are packed in regular hexagonal arrays (Herrler et al., 1981; Hewat et al., 1984), a feature not found with influenza A and B particles. The spikes of influenza C virions are composed of a single glycoprotein species (gp88) in a trimer configuration, and this arrangement may be essential for formation of a regular surface lattice of glycoprotein spikes.

4A. Viral RNAs and Proteins

Influenza C viruses contain a segmented genome of seven RNA species (Palese et al., 1980). The 3' and 5'

terminal sequences of the C virus RNA segments show partial homology with the corresponding sequences in influenza A and B viruses (Desselberger et al., 1980; Clerx-van Haaster and Meier-Ewert, 1984), indicating a close evolutionary relationship of these viruses.

The polypeptides of the influenza C virion are similar to those reported for influenza A and B viruses. The major structural protein species are the nucleoprotein (NP) and the matrix (M) protein (Compans et al., 1977; Yokota et al., 1983). In addition, three minor high molecular weight proteins designated P1, P2, and P3 have been identified which probably constitute the viral RNA transcriptase (Petri et al., 1980; Meier-Ewert et al., 1981a). In contrast to type A and B viruses, C type viruses possess a single surface glycoprotein species, gp88 (Herrler et al., 1981; Nakada et al., 1984a; Pfeifer and Compans, 1984), and lack a neuraminidase activity (Kendal, 1975). The gp88 designation is based on the molecular weight of the glycoprotein in SDS-polyacrylamide gels of 88,000. The gp88 is synthesized as a single polypeptide chain and can be converted into two disulfide-linked subunits of molecular weights 65,000 (gp65) and 30,000 (gp30) (Herrler et al., 1981). Cleavage of the gp88 precursor is essential for viral infectivity. The gp88 appears to possess receptor-binding activity (Herrler et al., 1981), a low pH-induced fusion function (Ohuchi et al., 1982), and a

receptor-destroying enzyme (RDE) activity (Kendal, 1975; Herrler et al, 1985b).

Analysis of the polypeptides synthesized in influenza C virus-infected cells has revealed the existence of nonstructural proteins that may be analogous to the NS1 and NS2 products encoded by RNA 8 of influenza A and B viruses (Petri et al., 1980; Yokota et al., 1983).

Complete nucleotide sequences of cloned cDNAs have been determined for RNA segment 4 (Nakada et al., 1984a; Pfeifer and Compans, 1984), RNA segment 5 (Nakada et al., 1984b), and RNA segment 7 (Nakada et al., 1985) of influenza C virus. Analysis of protein structure from deduced amino acid sequences and hybrid-arrest translation experiments with cloned cDNAs have led to the construction of a partial genetic map for influenza C viruses: RNA 4 codes for the gp88 or HA, RNA 5 codes for the NP, and RNA 7 codes for the NS1 and NS2 (Nakada et al., 1986). The proteins encoded by the remainder of the genome segments have not been proven experimentally, however, assignments of specific RNAs to viral polypeptides have been made based on analogies to influenza A and B viruses. The generation of cDNA clones and sequences corresponding to RNAs 1, 2, 3, and 6 of influenza C virus may aid in completing the genetic map for the C type virus. Detailed information on the C virus RNAs and proteins is summarized in Table 4.

TABLE 4. INFLUENZA C VIRUS RNAS AND PROTEINS[a]

Segment	Nucleotide Length from Sequence[b]	Nucleotide Length (from PAGE)	Encoded Protein[c]	MW SDS-PAGE	Amino Acid Length[d]
1		2,350	P1	89,000	
2		2,350	P2	85,000	
3		2,150	P3	82,000	
4	2,071	2,000	HA	64,000[e]	654
			HA1		431
			HA2		209
5	1,809	1,750	NP	60,000	565
6		1,150	M	30,000	
7	934	975	NS1	25,000	286
			NS2	14,000	121

[a]Adapted from Air and Compans (1983)

[b]C/Cal/78 strain (Nakada et al., 1984a; Nakada et al., 1984b; Nakada et al., 1985; Nakada et al., 1986)

[c]Definitive coding assignments of RNAs 1, 2, 3, and 6 have not been elucidated

[d]Deduced from nucleotide sequence

[e]Unglycosylated

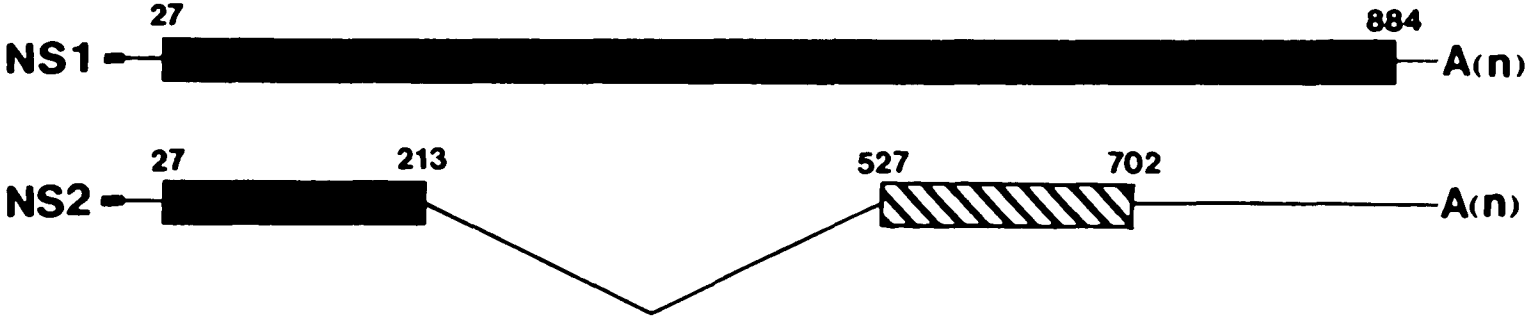
4A.1. RNA Segment 4: Hemagglutinin (HA)

The nucleotide sequences of RNA 4 of the C/CAL/78 (Nakada et al., 1984a) and the C/JHG/66 (Pfeifer and Compans, 1984) strains have been determined. RNA 4 of the C/CAL/78 virus contains 2,071 nucleotides. After a 5' noncoding region of 21 nucleotides, there is a large open reading frame that could code for a protein of 654 amino acids. There are eight potential glycosylation sites (Asn-X-Ser/Thr) in the RNA 4 protein suggesting that the protein is glycosylated *in vivo*. Comparison of the nucleotide sequence of RNA 4 with the HA genes of influenza A and B viruses revealed no convincing sequence homologies. However, conserved structural features between the RNA 4 protein and the HAs of type A and B viruses indicated that RNA segment 4 of influenza C virus codes for the HA protein. The conserved features include: 1) a hydrophobic signal peptide (14 amino acids), 2) an arginine cleavage site between the HA1 and HA2 subunits, 3) hydrophobic regions at the amino and carboxyl termini of the HA2 subunit, and 4) several conserved cysteine residues. It is suggested that the HA precursor of 654 amino acids is processed by removal of the signal peptide and then post-translational cleavage leads to two subunits, HA1 (431 amino acids) and HA2 (209 amino acids). As of this time, mapping of the antigenic sites or receptor binding site of the C virus HA molecule has not been accomplished.

4A.2. RNA Segment 7: Nonstructural Proteins (NS1 and NS2)

The nucleotide sequence of a cloned cDNA corresponding to the smallest RNA of the C/CAL/78 virus has been obtained (Nakada et al., 1985). The gene is 934 nucleotides in length. The RNA has a 5' untranslated leader of 26 bases which is followed by a long open reading frame initiating with an AUG that can code for a 286 amino acid protein. A definitive coding assignment of this gene to an NS protein (NS1-MW 28,500) was achieved by hybrid-arrest cell-free translation with influenza C virus mRNA preparations and cloned cDNA of RNA 7 (Nakada et al., 1985). Nakada et al. (1986) have recently identified a second NS product (NS2) encoded by RNA segment 7 of influenza C virus. Primer extension of a segment 7-specific oligonucleotide annealed to mRNA from influenza C virus-infected cells generated two cDNA products which differed in length by approximately 300 nucleotides. Partial sequencing of the extension products revealed that the shorter transcript contained the same 5' and 3' ends as the longer transcript, but contained an internal deletion of 313 nucleotides (Fig. 6). The shorter transcript is most likely generated from the longer mRNA by a splicing mechanism. The 5' ends of the two transcripts share a sequence of 213 bases. This region contains an initiation codon for protein synthesis that could code for 62 amino acids (shared N-terminus of NS1 and NS2). Following the splice junction, the 3' end of the

Fig. 6. Schematic representation of the NS1 and NS2 mRNAs derived from RNA 7 of influenza C virus. The filled-in boxes at the 5' ends of the mRNAs represent nucleotides derived from cellular mRNAs. The thin lines at the 5' and 3' termini of the mRNAs represent noncoding regions. The coding regions are represented by the thick boxes and are defined by the indicated nucleotides. The carboxy-terminal half of the NS2 protein is translated in a different reading frame (cross-hatched box) from that used for the amino-terminal half of the NS2 and the entire NS1 protein. The V-shaped line indicates the region (nucleotides 214-526) removed from the NS2 mRNA by splicing. (Adapted from Nakada et al., 1986.)


27
884
NS1
A(n)
27
213
527
702
NS2
A(n)

short mRNA also contains an open reading frame coding for 59 additional amino acids translated in a different reading frame (+1). The putative protein product (NS2) translated from this mRNA consists of 121 amino acids. The overall arrangement of the C virus NS gene is quite different from that of A and B virus NS genes in that the second (+1) open reading frame of the C virus NS gene is completely overlapped by that of the NS1 protein.

To demonstrate that a second polypeptide encoded by RNA 7 was synthesized in C virus-infected cells, an oligopeptide whose amino acid sequence was based on the predicted sequence of the second (+1) open reading frame, was used to elicit antibodies in rabbits. This antiserum specifically immunoprecipitated a nonstructural protein of MW 15,000 from infected cells. Additionally, hybrid-arrest translation experiments showed that the 15,000 MW protein was the NS2 encoded by RNA segment 7 of influenza C virus.

4A.3. Receptor Destroying Enzyme (RDE)

A difference between the RDE of influenza C virus and other orthomyxoviruses was first suggested by Hirst (1950), who showed that influenza A and B viruses destroyed their own receptors on erythrocytes without affecting the receptor for influenza C virus. Conversely, influenza C virions elute from erythrocytes with destruction of their receptors, whereas the receptors of influenza A and B

viruses are not affected. Evidence that the RDE of influenza C virions is not an alpha-neuraminidase was provided by Kendal (1975), who demonstrated that it did not release neuraminic acid from any known series of neuraminidase substrates. The conclusion that C type viruses lack a neuraminidase is also supported by the finding of sialic acid in the carbohydrate components of influenza C glycoproteins (Nakamura et al., 1979). In contrast, influenza A viruses lack sialic acid (Klenk et al., 1970). Recently, Herrler et al. (1985a, 1985b) showed that the RDE of influenza C virus is a neuraminate-O-acetylesterase that releases O-acetyl groups from sialic acids. Rogers et al. (1986) have obtained direct evidence that 9-O-acetyl-N-acetylneuraminic acid is the primary cellular receptor for influenza C virus.

5. Transcription and Replication of Influenza A Virus RNA

Analysis of influenza virus-infected cell extracts revealed two distinct classes of RNA transcripts which are complementary to the virion RNA (vRNA) segments (Skehel and Hay, 1978b; Hay et al., 1977, 1982; McCauley and Mahy, 1983). The first class of complementary RNA (cRNA) contains a 3' poly (A) tail and a 5'-terminal 7-methyl guanosine (m^7GpppNm) cap structure. There are two groups of cRNA transcripts which comprise this class. One group of transcripts are colinear copies of the eight vRNA segments which are incomplete because they lack a sequence of 17-22

bases present in a conserved region at the 5' termini of the vRNAs. The second set of cRNA transcripts are also incomplete, but they are spliced transcripts derived from the colinear transcripts complementary to RNAs 7 and 8. This first class of RNA transcripts is associated exclusively with polysomes and therefore, represent the viral mRNAs. The second class of cRNAs found in the infected cell are complete transcripts of the vRNAs which are unpolyadenylated. These complete cRNAs serve as templates for the synthesis of vRNAs which are incorporated into progeny virus particles. The virion-associated transcriptase complex is responsible for the synthesis of both classes of cRNAs in the nucleus of infected cells (Herz et al., 1981; Jackson et al., 1982; Beaton and Krug, 1984). In addition to the structural differences between the two types of cRNAs, differences in kinetics of synthesis of individual transcripts and regulation have been identified.

The synthesis of influenza viral mRNAs is controlled during infection with respect to the relative amount of each mRNA synthesized and to the time at which each mRNA is synthesized in greatest amount (Hay et al., 1977; Barrett et al., 1979; McCauley and Mahy, 1983; Enami et al., 1985). Immediately after infection, similar amounts of all eight viral mRNAs are detected and their production is independent of viral protein synthesis (primary

transcription). This is followed by an early phase of secondary transcription, dependent on viral protein synthesis, during which the production of the NP and NS1 mRNA predominates. Subsequently, during the late phase of secondary transcription, the rate of synthesis of NS1 mRNA relative to NP mRNA decreases while that of the M, HA, and NA mRNAs is greatly amplified. At all times except immediately after infection, the rate of synthesis of the three P protein mRNAs remains low relative to that of the other viral mRNAs. The relative rates of synthesis of the various viral mRNAs at different times of infection correlates closely with the relative rates of synthesis of the proteins encoded by these mRNAs, indicating that viral gene expression is regulated mainly at the level of transcription (Lamb and Choppin, 1976; Hay et al., 1977; Inglis and Mahy, 1979; McCauley and Mahy, 1983; Enami et al., 1985).

The mRNAs encoding the M2 and NS2 polypeptides of influenza virus are produced in infected cells through splicing of colinear transcripts from vRNA segments 7 (Lamb et al., 1981) and 8 (Lamb and Lai, 1980), respectively. It appears that the spliced transcripts are formed through the action of cellular RNA processing enzymes since unspliced and correctly spliced (as in vivo) RNAs are expressed from SV40 virus vectors containing a cDNA copy of either the M or NS gene in the absence of other influenza virus products

(Lamb and Lai, 1982; Lamb and Lai, 1984). During the course of influenza virus infection, there is an increase in the level of virus-specific spliced mRNAs, most evident for the NS2 mRNA, relative to that of their unspliced precursors (Inglis and Brown, 1984; Smith and Inglis, 1985). It appears that virus-specific products may somehow regulate the production of spliced mRNAs in influenza virus-infected cells.

Unlike the viral mRNAs, the production of full-length transcripts is not regulated. Approximately equimolar amounts of each of the eight cRNA transcripts are synthesized throughout infection (Hay et al., 1977). Influenza virus infection in the presence of cycloheximide results in the formation of only mRNA (primary transcription), indicating that newly synthesized viral or possibly host proteins are required for the synthesis of the full-length cRNAs (Barrett et al., 1979; Hay et al., 1977). The nascent virus protein may act to modify the transcriptase complex in some way so that transcription initiates without a primer and continues past the termination site utilized during viral mRNA synthesis (Beaton and Krug, 1984).

A unique feature of influenza virus replication, not found in other nononcogenic RNA viruses, is the requirement of a functional host nuclear RNA polymerase II (Lamb and Choppin, 1977; Spooner and Barry, 1977). This enzyme

synthesizes the precursors to cellular mRNAs. It was found that a specific inhibitor of RNA polymerase II, alpha-amanitin, inhibits influenza virus replication at the level of viral RNA transcription (Lamb and Choppin, 1977; Mark et al., 1979).

5A. Viral mRNA Synthesis

The mechanism of how influenza viruses synthesize mRNAs has been dissected using *in vitro* transcription systems which, in most respects, reflect the process occurring *in vivo*. The initiation of influenza viral mRNA transcription requires 5' capped primers derived from nascent host cell mRNAs synthesized by RNA polymerase II (Beaton and Krug, 1981; Krug, 1981; Plotch et al., 1981). The host mRNAs are cleaved by a virus specific endonuclease at a purine residue 10 to 15 nucleotides from the capped 5' end. It is this fragment containing the 5'-terminal m^7GpppNm (cap 1 structure) that serves as a primer and is transferred to the 5' ends of the influenza virus mRNAs during transcription --"cap snatching" (Plotch et al., 1981; Kawakami et al., 1983). Shaw and Lamb (1984) have reported that host RNA primers containing a 3'-terminal Py-G-C-A sequence before the presumed endonuclease cleavage site are preferred as primers in influenza virus mRNA synthesis. The viral PB2 protein functions in cap recognition during the endonuclease reaction (Ulmanen et

al., 1981; Blaas et al., 1982a, 1982b; Penn et al., 1982). The capped fragments do not have to be hydrogen-bonded to the 3' end of the vRNA template to prime viral transcription (Krug et al., 1980).

The second step in the reaction is the initiation of transcription via incorporation of a G residue onto the 3' end of the capped primer fragment, directed by the penultimate C residue at the 3' end of each vRNA template (3' U-C-G...). It is the PB1 protein that catalyzes the initiation step of transcription (Ulmanen et al., 1981; Horisberger, 1982; Romanos and Hay, 1984). After initiation with a G residue, chain elongation proceeds. It is not known which P protein(s) functions in this process, but there is evidence to suggest that PB1 may be involved (Braam et al., 1983; Romanos and Hay, 1984). Elongation is terminated at a tract of U residues (polyadenylation signal) about 17-22 nucleotides from the 5' end of the template and poly (A) addition occurs at the 3' end of the mRNA (Robertson et al., 1981).

Recently, an RNA polymerase-viral RNA complex has been purified from influenza virions by cesium trifluoroacetate centrifugation and phosphocellulose column chromatography (Kawakami and Ishihama, 1983; Kato et al., 1985). The complex, composed of PB1, PB2, PA, and vRNA, has been shown to be as catalytically active in viral mRNA synthesis <u>in</u>

vitro as native RNP cores. These data indicate that NP protein is not required for mRNA synthesizing activity.

5B. Synthesis of Full-length cRNA Transcripts

The unpolyadenylated cRNAs are complete copies of the vRNA segments and consequently, the termination of transcription during mRNA synthesis which occurs 17-22 bases from the 5' ends of the vRNA templates, is inoperative during the synthesis of these transcripts. In contrast to mRNA synthesis, initiation of full-length transcripts *in vivo* occurs without a primer at the 3'-terminal U of the vRNA templates (Hay et al., 1982; Beaton and Krug, 1984). The synthesis of full-length cRNA requires ongoing viral protein synthesis. Since these transcripts are complete copies of the vRNA segments and lack host-derived primer sequences, they most likely function as the templates for vRNA synthesis. To aid in deciphering the mechanism of full-length cRNA and vRNA production, *in vitro* systems have been established in which full-length cRNA, in addition to incomplete mRNA, can be synthesized (Beaton and Krug, 1984; del Rio et al., 1985). Beaton and Krug (1984) have found an antitermination factor in the cytoplasmic extract of virus-infected cells that enables the transcriptase complex to continue transcription past the site at which mRNA synthesis terminates. In addition, this cytoplasmic extract allows the transcriptase complex to initiate transcription without the capped primers used

in viral mRNA synthesis. The identification and purification of the active cytoplasmic factor(s) may help to characterize the switch mechanism from mRNA to full-length cRNA synthesis.

5C. Synthesis of vRNA (Replication)

There is little information available concerning the synthesis of vRNA and these data include conflicting reports. The synthesis of minus strand vRNA most likely occurs in the nucleus of infected cells (Jackson et al., 1982). Smith and Hay (1982) have shown that the eight vRNA segments are produced in dissimilar amounts and in different relative proportions at various times of infection. Virion RNAs 5 and 8 are preferentially synthesized early in infection, similar to mRNA synthesis of these templates. In contrast, Enami et al. (1985) have reported that all eight segments of minus strand RNA are produced coordinately at nearly equimolar ratios. Analysis of ts mutants defective in vRNA production has implicated at least two of the P proteins in vRNA synthesis (Mahy et al., 1981), but the precise role of these P proteins and possibly of other viral proteins (NP) in catalyzing vRNA synthesis is not known. Furthermore, it has not been discovered how approximately equimolar amounts of the different vRNA segments are packaged into virions.

5D. Effect of Interferon on Influenza Virus Replication

Interferon induces an antiviral state against influenza virus, but the mechanism of inhibition of viral replication in not well understood (Horisberger and de Staritzky, 1985). The mechanism of inhibition is of particular interest because influenza virus is the sole nononcogenic RNA virus which requires host cell nuclear function for replication. Ransohoff et al. (1985) have investigated the molecular mechanism whereby interferon induces resistance to influenza virus by monitoring the accumulation of both plus and minus strand viral RNAs in infected MDBK cells treated with human alpha A interferon. This group found that influenza viral primary transcripts failed to accumulate in interferon-treated cells. This result is consistent with observations made by Krug et al. (1985), who have reported that the interferon-induced Mx gene product inhibits influenza viral mRNA synthesis in mouse embryo cells bearing the Mx gene.

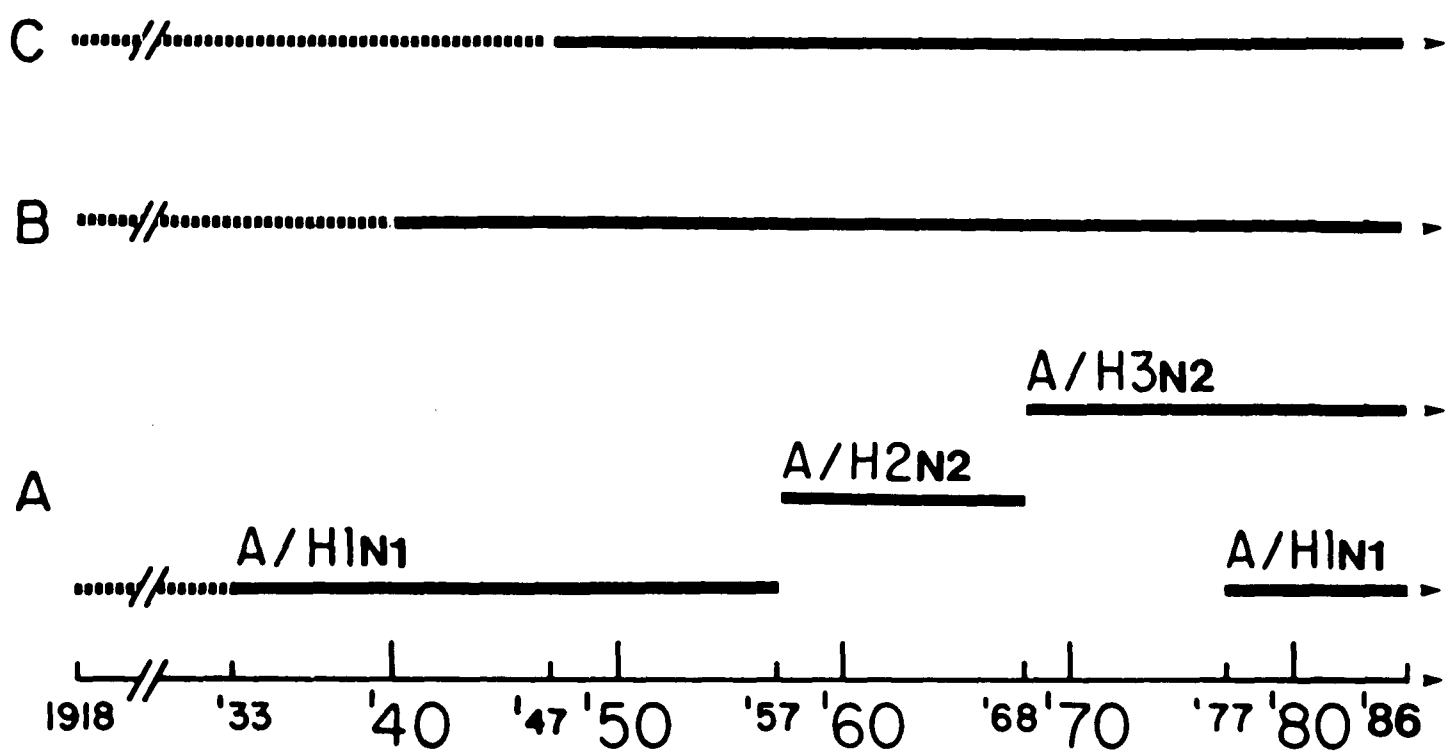
In mouse cells carrying the dominant influenza resistance allele, Mx^+, interferon alpha/beta induces a 75,000 MW protein (Mx) which accumulates in the cell nucleus and is responsible for an enhanced antiviral state specifically against influenza virus (Horisberger et al., 1983; Horisberger and Hochkeppel, 1985; Dreiding et al., 1985; Staeheli et al., 1986). There are conflicting reports concerning the stage of viral replication at which the Mx

protein exerts its inhibitory effect. In contrast to the results of Krug et al. (1985), Meyer and Horisberger (1984) claim that interferon-treated Mx-bearing macrophages exhibit inhibition of influenza virus replication at the level of protein translation. Staeheli and Haller (1985) have detected an interferon-induced 80,000 MW protein in human cells which appears to be the human homolog of Mx protein in mouse cells. The significance of the human protein for defense against influenza viruses in man has yet to be investigated.

6. Epidemiology of Influenza Viruses

The epidemiology of influenza A viruses has been well characterized due to the association of these viruses with pandemic outbreaks of disease in man. The pandemics usually result from the appearance of a new subtype strain (antigenic shift) containing a novel HA and/or NA that is immunologically different from that of previous circulating isolates. During this century, there appears to have been at least three occasions when new subtypes have been introduced in the human population (Fig. 7). In 1918, a severe pandemic occurred which resulted in the death of an estimated 20 million people worldwide. The "new" viruses responsible for this widespread mortality are serologically related to viruses first isolated from pigs in 1931. Consequently, these strains are referred to as the swine viruses or H1N1 subtype strains. Viruses of the H1N1 subtype continued to circulate for approximately 40 years until a second subtype shift occurred in 1957. It should be noted that the H1N1 strains did accumulate more subtle changes in antigenic character (antigenic drift) over the 40 year period of their circulation, so that strains isolated at the beginning of the H1N1 period were slightly different from those H1N1 viruses isolated at the end of the subtype period. In 1957, viruses of the H2N2 subtype (Asian strains) appeared whose HA and NA proteins were not cross-reactive with their counterparts of the H1N1 strains.

Fig. 7. Epidemiology of influenza A, B, and C viruses in nature. Broken lines indicate that virus isolates are not available from these periods. The periods when influenza A viruses of a certain subtype (H1N1, H2N2, H3N2) were prevalent are designated by the thick lines. (Modified from Palese and Young, 1982.)


C
B
A/H3N2
A/H2N2
A
A/H1N1
A/H1N1
1918
'33
'40
'47
'50
'57
'60
'68
'70
'77
'80
'86

For the next 11 years, H2N2 strains spread and gradually changed until the introduction of the new H3N2 subtype (Hong Kong strains) in 1968. Viruses of the H3N2 subtype retained the N2 NA of the Asian strains, but possessed a novel HA antigen. Up until 1968, the three subtypes circulated sequentially in man. Prior to 1977, viruses of the H3N2 subtype were the sole circulating strains.

In 1977, a unique event occurred with respect to influenza virus epidemiology. Viruses containing surface antigens of a previous subtype, H1N1, reappeared and these viruses (Russian strains) cocirculated with the H3N2 strains prevalent at the time. The "new" H1N1 strains are serologically and genetically similar to H1N1 strains which had circulated around 1950 (Nakajima et al., 1978; Kendal et al., 1978). It is unusual that the new H1N1 viruses did not replace the H3N2 strains introduced in 1968, but coexist with them.

The epidemiology of influenza B and C viruses appears to be less complex than that of the A type viruses (Fig. 7). Influenza B viruses have been isolated since 1940 and there is no evidence to suggest that subtypes exist. However, there is evidence in support of antigenic drift of B viruses over time (Krystal et al., 1983b; Bao-lan et al., 1983; Oxford et al., 1984). There is even less data on the epidemiology of influenza C viruses. It appears that C virus infections occur in an endemic pattern and not in the

form of epidemic outbreaks. Antigenic subtypes of influenza C viruses have not been identified, however, minor antigenic differences among isolates have been observed (Chakraverty, 1978; Guo et al., 1983).

6A. Mechanisms Leading to New Subtypes

6A.1. Reassortment of Surface Protein Genes

Extensive changes in the HA and/or NA which lead to new subtypes of influenza viruses are known as antigenic shifts. Sequence comparisons between the H1, H2, and H3 HA genes (Krystal et al., 1982) and between the N1 and N2 NA genes (Markoff and Lai, 1982) reveal such low levels of nucleotide homology that it is highly unlikely that these different HAs and NAs arose from one another by gradual accumulation of point mutations over time. Rather, it is believed that the appearance of novel surface antigens of influenza viruses are the result of reassortment (recombination) events between previous circulating human viruses and influenza viruses of avian or animal origin. There is ample evidence supporting reassortment between human viruses and avian strains *in vivo* (Webster et al., 1971), between avain strains in nature (Desselberger et al., 1978), and among human viruses (Young and Palese, 1979; Bean et al., 1980; Cox et al., 1983). The nonhuman virus is the donor of the HA and/or NA genes which encode surface proteins which are foreign to man's immune system

and the human virus provides the rest of the genetic complement which allows for replication and virulence in the human host. In addition to the three human HA subtypes (H1, H2, H3) and two NA subtypes (N1, N2), there are ten antigenically distinct HAs and seven NAs which have been identified in birds, horses, and pigs. It is conceivable that any of these foreign HA and NA genes could be incorporated into a human virus via reassortment, thereby resulting in a new human antigenic subtype.

The best evidence for reassortment generating new human subtypes involves the origin of the H3 HA. The Hong Kong H3N2 strain of influenza has been shown to be a reassortant (Laver and Webster, 1973). This virus contains the NA and all other genes from an Asian (H2N2) strain and an HA which is antigenically related to that of the A/duck/Ukraine/63 (H3N8) and A/equine/2/Miami/63 (H3N8) viruses (Fang et al., 1981; Ward and Dopheide, 1981a, 1981b). The amino acid sequence homology between the HAs of the A/duck/Ukraine/63 and A/Aichi/2/68 (H3N2) viruses is 96%.

In addition, it is likely that the H2N2 strains emerged from H1N1 viruses through reassortment. RNA:RNA hybridization data has revealed that the H2N2 strains derive their six nonsurface protein coding genes from the H1N1 viruses (Scholtissek et al., 1978a).

6A.2. Reemergence of Previously Circulating Strains

A second mechanism leading to the introduction of a new subtype of influenza A virus involves the reappearance of a subtype which had circulated at an earlier time. A well documented example of this phenomenon is the reemergent H1N1 strains that reappeared in Anshan in northern China in May of 1977 and subsequently spread to the rest of the world. Serologically, these new H1N1 viruses possess HA and NA antigens related to those of H1N1 influenza A viruses that infected man around 1950 (Kendal et al., 1978). Even more remarkable, genetic analysis of the new H1N1 strains by RNase T1 oligonucleotide mapping, RNA:RNA hybridization studies, and sequencing (Nakajima et al., 1978; Young et al., 1979; Scholtissek et al., 1978b; Concannon et al., 1984; this thesis) revealed that the entire genomes of the reemergent H1N1 strains are very similar to those of H1N1 viruses isolated in 1950.

The question is, how did the new H1N1 strains remain relatively unchanged during a 27 year period? Several mechanisms have been postulated: 1) influenza viruses are capable of latent or persistent infection in man in conditions in which the genetic information of the virus is highly conserved; 2) the genetic information of the virus could have been preserved by sequential passage in an animal reservoir in which influenza viruses replicate without rapid genetic change; and 3) a 1950 virus was

frozen in nature or elsewhere and recently introduced into man.

Additional evidence for antigenic recycling comes from serologic data. Serum samples collected from aged persons before the H3N2 pandemic in 1968 contained antibodies to H3-like agents and stored serum samples from people born before 1887 cross-reacted with H2N2 strains isolated after 1957 (Davenport et al., 1969; Masurel, 1969). These data in conjunction with that of the reemergent H1N1 strains may indicate that there are a restricted number of HA and NA genes that can contribute to variation of human influenza viruses.

6A.3. Animal Virus becomes Virulent for Man

A third way that new viruses could appear in man would be that an animal or bird virus acquires mutations which confer infectivity in the human population. The intense outbreak of influenza in 1976 among soldiers at Fort Dix, New Jersey caused by swine viruses (H1N1), may have resulted from such a mechanism (Kendal et al., 1977). Furthermore, unequivocal evidence for the infection of humans by swine viruses (H1N1) from pigs was obtained by the isolation of genetically identical swine viruses from pigs and man on a farm in Wisconsin (Hinshaw et al., 1978).

6B. Variation within Subtype Strains

6B.1. Genetic Drift via Point Mutations

Following the emergence of a new subtype, viruses of a single subtype show minor changes in antigenic character probably brought about by immune selection pressure. These minor antigenic changes (antigenic drift) have been shown to be the result of an accumulation of amino acid changing nucleotide substitutions in the genes encoding the HA and NA proteins. In addition to variation in the surface protein coding genes, it has been documented that variation among strains also occurs in the genes encoding nonsurface proteins of the influenza virus (Young et al., 1979; Ortin et al., 1980; Palese et al., 1981). This fact is reflected in the detection of migrational differences on SDS-PAGE in the NP, M, and NS1 proteins among virus isolates (Oxford et al., 1981; Petri et al., 1982; Parvin et al., 1983; Donatelli et al., 1985). A detailed description of what is known about genetic drift in influenza A virus RNAs is summarized in the following sections.

6B.1.1. P Genes

Information on genetic drift of the three polymerase genes is extremely limited. Nucleotide sequences of the PB1, PB2, and PA genes of the A/PR/8/34 (Winter and Fields, 1982; Fields and Winter, 1982) and the A/NT/60/68 (Bishop et al., 1982a, 1982b; Jones et al., 1983) isolates have

been examined. Sequence variation per year for PB1, PB2, and PA is 0.47%, 0.26%, and 0.21%, respectively. The amino acid sequences of the deduced P proteins change at a rate of 0.09% (PB1), 0.12% (PB2), and 0.11% (PA) amino acid change per year.

6B.12. HA Gene

Precise information on the evolution of the HA gene within a subtype has been generated using nucleotide sequencing analysis of HA genes of viruses isolated over time. Partial sequencing of the 3' end of HA genes of viruses of one subtype revealed that the genes were accumulating nucleotide changes (Air, 1981). A more detailed examination of HA evolution was conducted by Both et al. (1983) in which nucleotide sequences of the HA1 coding region and deduced amino acid sequences of 14 H3 subtype field isolates were compared with previously available H3 HA sequences. This extensive study showed that from 1968 to 1980, new H3 antigenic variants emerged by accumulating sequential amino acid changes within antigenic regions of the HA molecule. The amino acid substitutions alter the antigenic sites in such a way that the antibodies generated are not protective against infection.

Similarly, Raymond and et al. (1983) reported sequential amino acid substitutions in the HA1 of a series

of drifted H1N1 viruses isolated between 1977 and 1980. More recently, Raymond et al. (1986) have expanded their study on the evolution of the H1N1 strains by examining viruses from both the 1950 to 1957 era and the 1977 to 1983 period. The results of this study support the previous finding of cumulative nucleotide and amino acid changes in the HA.

Additional evidence for the mechanism of antigenic drift in the HA via successive substitutions in the gene has been provided by sequence analysis of variants selected in the laboratory in the presence of monoclonal antibodies to the HA (Webster et al., 1979; Webster and Laver, 1980; Gerhard et al, 1981).

6B.13. NA Gene

Blok and Air (1980, 1982a, 1982b) have reported on the molecular basis of antigenic drift in the NA gene through partial sequencing of the 3' end of NA genes of viruses of the human subtypes N1 and N2, in addition to NA subtypes associated with nonhuman influenza viruses. The results of these studies showed that point mutations (some of which resulted in amino acid changes) existed in the NA genes of viruses of the same NA subtype and some of the changes which appeared in earlier isolates were also found in later strains.

To further characterize variation in the NA gene, complete sequences have been determined and analyzed. Only

two complete nucleotide sequences of the N1 NA of human viruses are available and therefore, the evolutionary pattern of the N1 NA can not be accurately examined at this time. However, six complete NA sequences of natural variants of the N2 subtype isolated between 1957 and 1979 have been determined. Comparative analysis of these NA genes (Laver et al., 1982; Martinez et al., 1983) along with sequence information obtained from variants of the A/Tokyo/3/67 (H2N2) strain selected by monclonal antibodies to the NA (Laver et al., 1982; Webster et al., 1982a; Colman et al., 1983), revealed that antigenic drift of the N2 NA occurs via the same mechanism as that proposed for the HA polypeptide. Sequential point mutations are observed in the N2 NA genes which alter amino acids in regions of the NA protein associated with antigenic sites.

6B.14. NP Gene

Immunological analysis of NP proteins has shown that antigenic differences can be detected in the molecule within and between virus subtypes (Schild et al., 1979; Van Wyke et al., 1980). In addition, migrational differences among NP proteins of virus isolates have been observed, suggesting protein variability (Oxford et al., 1981). However, genetic drift in NP genes has not been well defined. Only two complete human NP nucleotide sequences are available, one of a virus of the H1N1 subtype (Winter

and Fields, 1981) and the other of an H3N2 isolate (Huddleston and Brownlee, 1982). Based on the high level of nucleotide (91.8%) and amino acid (94.2%) homology among these NP genes and the deduced proteins, it appears that the NP gene has been conserved during the major antigenic shifts.

6B.15. M Gene

Three complete nucleotide sequences of M genes (Winter and Fields, 1980; Lamb and Lai, 1981; Ortin et al., 1983) and four partial M sequences (Hall and Air, 1981) of human influenza viruses isolated over 45 years (1934-1979) have been obtained. The viruses analyzed represent all three human antigenic subtypes. The nucleotide data illustrate that substitutions occur in the M gene and many of the changes appear to accumulate sequentially. The M gene is highly conserved through evolution (0.1-0.2% sequence divergence per year). On the amino acid level, divergence among M1 polypeptides is less that 0.12% per amino acid per year. The M2 proteins are less conserved and amino acid changes of 0.25-0.30% per year have been reported. The amino acid conservation among M1 proteins is reflected in studies using monoclonal antibodies to detect antigenic variation in M1 proteins of both human and animal viruses of different subtypes (Van Wyke et al., 1984). In this study involving 26 isolates, antigenic variation in the M1 protein could only be found in three strains.

The high degree of homology among M sequences of viruses of the H1N1, H2N2, and H3N2 subtypes provides evidence that RNA segment 7 was retained through the two antigenic shifts from H1 to H2 and H2 to H3 (Hall and Air, 1981).

6B.16. NS Gene

Previous to this work, there was little known information on genetic drift of the NS gene of influenza A viruses. Complete nucleotide sequences of RNA 8 of the A/PR/8/34 (Baez et al., 1980; Winter et al., 1981) and A/Udorn/72 (Lamb and Lai, 1980) strains have been determined. In addition, partial sequences of NS genes are available (Hall and Air, 1981). On the nucleotide level, the NS genes show a sequence divergence of 0.23-0.26% per year. The deduced NS1 and NS2 protein sequences are changing at a rate of approximately 0.3% and 0.19% amino acid substitutions per year, respectively. From these data, it appears that the NS2 amino acid sequence is more conserved than that of the NS1. However, the NS1 protein itself is highly conserved as demonstrated by extensive cross-reactivity among NS1 polypeptides of viruses belonging to different human subtypes with monospecific polyclonal NS1 antiserum (Shaw et al., 1982b). Subsequently, possible antigenic sites on the NS1 molecule were mapped using monoclonal anti-NS1 antibodies and these reagents were used to probe antigenic variation among NS1 proteins of both human and animal viruses (Brown et al., 1983). Antigenically variant NS1 proteins were detected only from avian sources, once again illustrating the antigenic conservation of the protein among human isolates.

The high level of genetic homology among NS genes of viruses belonging to different subtypes revealed by both sequencing (Hall and Air, 1981) and RNA:RNA hybridization studies (Scholtissek et al., 1978a), has indicated that the NS gene, like the NP and M genes, was conserved at the two major antigenic shifts of this century. Hall and Air (1981) have suggested that sequence changes in the NS gene accumulate over time, however, this statement is based on partial NS sequence data from only three human strains. More extensive information at the nucleotide level has been generated in this work to assess the mechanism of genetic drift in the NS gene.

In addition to NS nucleotide sequences of human viruses, two complete sequences of the NS gene of the avian strains, A/FPV/Rostock/34 (Porter et al., 1980) and A/duck/Alberta/60/76 (Baez et al., 1981) are known. Surprisingly, the FPV/34 NS gene is 92% homologous in sequence to the NS gene of the human virus A/PR/8/34. This avian NS segment and its human counterpart may possibly be derived from a common ancestral gene. In contrast, the duck/Alberta/76 virus NS gene is only 72.7% homologous to the PR/34 sequence. Clearly, the duck virus NS gene has a quite different evolutionary history than the human virus homolog.

6B.2. Reassortment of Genes Coding for Nonsurface Proteins

As described in an earlier section, exchange of surface protein genes among strains via reassortment leads to antigenic shifts resulting in new subtypes. Similarly, reassortment of nonsurface protein genes among cocirculating strains can lead to genetic diversity among strains of a single subtype (Young and Palese, 1979; Bean et al., 1980; Palese and Young, 1982; Cox et al., 1983).

A well documented case involves strains of the reemerged H1N1 subtype that circulated in the winter of 1978 to 1979 in California (Young and Palese, 1979). Genetic analysis of these strains revealed that they were quite different from H1N1 viruses isolated in the previous year. A more detailed examination of the prototype A/Cal/10/78 H1N1 isolate using oligonucleotide mapping of isolated genes along with analysis of infected-cell polypeptides by SDS-PAGE, suggested that the virus arose by reassortment between an earlier H1N1 virus and a cocirculating H3N2 parent. The A/Cal/10/78 strain derives its HA, NA, M, and NS genes from the H1N1 parent and its NP and three P genes are from an H3N2 strain.

In addition, a second type of reassortant H1N1 virus has been described (Palese and Young, 1982). The prototype strain of this type is the A/Aberdeen/v1340/78 virus. This virus contains five H3N2-derived genes and an HA, NA, and NS gene from an A/USSR/90/77-like H1N1 strain. These data

illustrate that reassortment among coexisting strains is another mechanism that can explain diversity within a subtype. Furthermore, biochemical evidence has been provided in favor of recombination among avian strains leading to new virus variants in nature (Desselberger et al., 1978; Hinshaw et al., 1980).

6B.3. Deletions/Insertions in HA and NA Genes

Comparison of different HA gene sequences of the H3 subtype, demonstrated that some strains had inserted or deleted one or more codons in their HA molecules. The A/Vic/3/75 HA, for example, contains an additional amino acid (Asn) at position 8 not present in any other H3 HAs which have been sequenced to date (Verhoeyen et al., 1980). More extensive deletions/insertions have also been observed in the HAs of viruses of different A subtypes (Hiti et al., 1981) and in HAs of strains of influenza B virus (Krystal et al., 1982).

Additionally, deletions/insertions have been detected in the NA genes. Partial sequencing of the 3' end of several N1 NAs has demonstrated length variation in this region (Blok and Air, 1982b). Comparison of complete sequences of N1 and N2 NA genes also revealed deletions/insertions during the evolution of these genes (Fields et al., 1981; Hiti and Nayak, 1982; Markoff and Lai, 1982).

Deletions/insertions occurring in both the HA and NA genes is yet another mechanism contributing to variation of influenza viruses.

7. Epidemiology of Human Influenza C Viruses

Influenza C viruses, first isolated from man in 1947, generally cause only infrequent outbreaks of mild respiratory illness (Katagiri et al., 1983). A very high percentage of the young adult population possesses antibodies to these viruses (O'Callaghan et al., 1980), indicating that infection occurs at a rate considerably greater than is generally recognized.

In contrast to influenza A and B viruses, there have been few reports, up until the present work, on the epidemiology and genetic variation of influenza C viruses in nature. Serological studies using hemagglutination inhibition (HI) assays of C viruses isolated over time in different geographical locations have shown that antigenic variation does occur among isolates, even though hemagglutinin subtypes of C viruses have not been observed (Chakraverty, 1978; Guo et al., 1983). Analysis of influenza C virus proteins by SDS-PAGE and one-dimensional peptide mapping, has revealed minor migrational differences in the HA, NP, and M polypeptides among isolates (Sugawara et al., 1983; Elliott et al., 1984; Goto et al., 1984). In addition, oligonucleotide mapping of genome RNAs of influenza C strains isolated over 34 years showed that all

strains were distinguishable, but were highly conserved over time (Meier-Ewert et al., 1981b; Guo and Desselberger, 1984). These data taken collectively, suggest that the genomes of influenza C viruses are remarkably stable and undergo less variation than influenza A viruses isolated over comparable time periods.

On the nucleotide level, few data are available regarding variation in influenza C virus gene segments. Two complete nucleotide sequences of the HA gene of viruses isolated 12 years apart have been determined (Nakada et al., 1984a; Pfeifer and Compans, 1984). Comparison of the deduced amino acid sequences of these HA proteins revealed an overall difference of 3.9%. This lies in the range observed between the A virus HAs of A/NT/60/68 and A/Bangkok/1/79 (Both and Sleigh, 1981) and is typical of the differences found among H3 subtype strains isolated a decade apart. However, the comparison of the C virus HAs involves only two strains, whereas many more isolates have been analyzed among the A type viruses. Additional sequence information on the genes of influenza C viruses is needed to define the extent and mechanism of genetic variation in these viruses.

7A. Isolation of Influenza C Viruses from Pigs

It was generally thought that, in contrast to influenza A viruses, there was no natural animal reservoir for

influenza C viruses. However, Guo et al. (1983) reported the isolation of C viruses from pigs in Beijing, China in 1981. This group was able to show that pigs could be experimentally infected by C virus and that pig-to-pig transmission of influenza C virus strains occurred. It was therefore, of interest to determine whether there were biochemical differences between the swine isolates of 1981-1982 and human influenza C strains. Comparative analysis of RNAs by oligonucleotide mapping and gel electrophoresis (Guo and Desselberger, 1984) and of polypeptides, by SDS-PAGE, synthesized in infected cells (Elliott et al., 1984), demonstrated that the swine isolates were closely related to each other, but also highly homologous to the human viruses. The question of whether pigs serve as a natural reservoir for human influenza viruses cannot be answered from the available data.

SPECIFIC AIMS

A. Genetic drift in influenza A virus genes has been best studied for the HA and NA genes which code for the surface proteins of the virus. Specifically, Both et al. (1983) characterized changes in the HA genes of a series of drifted natural variants of the H3 Hong Kong subtype. This group found that antigenic variation in the HAs of these strains was the result of accumulated point mutations in the HA gene that led to sequential amino acid substitutions in key antigenic regions of the HA molecule.

We asked the question of whether sequential nucleotide changes were also occurring in a virus gene coding for nonsurface proteins. The NS gene was chosen for study because there was evidence suggesting that this segment was conserved through the major antigenic shifts of this century. Therefore, variation in the NS gene could be investigated over many years. The complete nucleotide sequences of 15 NS genes of viruses isolated over 53 years (1933-1985) and representing the H1, H2, and H3 HA serotypes were examined. Included in the analysis were H1N1 subtype strains of the 1950 period and the reemerged viruses of 1977 up until 1985. We wanted to determine how closely related on the nucleotide level was the NS gene of the prototype A/USSR/90/77 virus to that of the A/FW/1/50 strain. Additionally, we asked the question of whether

evolutionary changes in the NS genes of the 1950-1957 H1N1 strains were similar to changes found in the NS genes of the 1977-1985 H1N1 viruses. The reintroduced H1N1 viruses provided a unique opportunity to study the evolution of the H1N1 subtype in two different eras. An evolutionary tree of the NS genes of the viruses was constructed from the sequence data and an evolutionary rate for the NS gene of influenza A virus was calculated. The NS gene evolutionary rate was compared to that of other influenza virus genes.

B. A second major objective of this work was to examine the extent and nature of variation in the genes of human influenza C viruses. Prior to this report, little was known about the epidemiology of C viruses and the mechanisms leading to variation among strains. The HA and NS genes of eight C viruses isolated in 1947 to 1983 were examined. Comparative sequence analysis yielded information on the evolution of both a surface and nonsurface protein coding gene in the identical virus strains. Using the HA and NS nucleotide data, evolutionary trees depicting genetic relationships among strains were constructed. The extent of nucleotide changes in the influenza C virus genes was also compared to that in the corresponding segments of influenza A viruses. The results of the C virus analysis led to the proposal of an evolutionary model for the propagation of influenza C variants in nature that is quite different from that of

other orthomyxoviruses. Influenza C variants derived from multiple evolutionary lineages appear to cocirculate, whereas, influenza A and possibly B variants emerge successively with time along a single lineage.

C. The influenza virus host range mutant CR43-3, derived by recombination from the A/ALA/6/77 and the cold-adapted and temperature-sensitive A/AA/6/60 viruses, has previously been shown to possess a defect in the NS gene (Maassab and DeBorde, 1983). The precise nature of the NS defect was determined by nucleotide sequencing of cloned cDNAs and phenotypic properties of the virus were investigated.

MATERIALS AND METHODS

1. Virus and Virus Purification

The influenza A, B, and C viruses used in this work are listed in Table 5. Influenza viruses were propagated in embryonated chicken eggs according to the protocols outlined below.

1A. Growth and Purification of Influenza A and B Viruses

An appropriate dilution (10^{-2}-10^{-4}) of virus seed was made in virus diluent which was phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4) containing 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2.6H_2O$, 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, and 0.2% bovine albumin (Miles Scientific Laboratories, Inc.; Elkhart, IN). Virus (0.1 ml) was inoculated into the allantoic cavity of 10-11 day old embryonated eggs. Eggs were incubated at either 35°C (all B viruses) or 37°C for 40-48 h and then chilled overnight at 4°C. Allantoic fluid was harvested, pooled, and clarified by centrifugation at 3,000 rpm for 10'. The supernatants were removed and the virus titer was determined by hemagglutination assay (see below). Virus was concentrated by centrifugation at 4°C in a Ti45 rotor at 35,000 rpm for 1 h. The supernatants were discarded and the virus pellets were resuspended in several ml of NTE buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM EDTA). Virus solution was layered on 30 ml continuous

TABLE 5. INFLUENZA VIRUSES USED IN THIS STUDY

Virus	Subtype	Source	Abbreviation
A/WSN/33	H1N1	Mt. Sinai	--
A/Puerto Rico/8/34	H1N1	Mt. Sinai	A/PR/34
A/Bellamy/42	H1N1	Mt. Sinai	A/BEL/42
A/Fort Monmouth/1/47	H1N1	Mt. Sinai	A/FM/47
A/Fort Warren/1/50	H1N1	Mt. Sinai	A/FW/50
A/Denver/1/57	H1N1	Mt. Sinai	A/DEN/57
A/Ann Arbor/6/60	H2N2	Mt. Sinai	A/AA/60
A/Berkeley/1/68	H2N2	Mt. Sinai	A/BERK/68
A/Udorn/72	H3N2	Mt. Sinai	--
A/Alaska/6/77	H3N2	H.F. Maassab	A/ALA/77
A/Houston/24269/85	H3N2	H. Six	(H3) A/HT/85
A/USSR/90/77	H1N1	Mt. Sinai	A/USSR/77
A/Maryland/2/80	H1N1	Mt. Sinai	A/MD/80
A/Houston/18515/84	H1N1	H. Six	A/HT/84
A/Houston/23284/85	H1N1	H. Six	(H1) A/HT/85
CR43-3 reassortant of A/AA/6/60 (ca)(ts) x A/ALA/6/77 (wt)	H3N2	H.F. Maassab	--
A/AA/6/60 (ca)(ts)	H2N2	H.F. Maassab	--
B/Yamagata/1/73		K. Tobita	B/YA/73
B/Clone 201		K. Tobita	clone 201
C/Taylor/1233/47		Institute of Virology China National Center for Preventive Medicine	C/TAY/47
C/Ann Arbor/1/50		H.F. Maassab	C/AA/50
C/Great Lakes/1167/54		W. Dowdle	C/GL/54
C/Johannesburg/1/66		Mt. Sinai	C/JHG/66
C/California/1/78		Mt. Sinai	C/CAL/78
C/Mississippi/80		A.P. Kendal	C/MS/80
C/Yamagata/10/81		K. Nakamura	C/YA/81
C/England/892/83		P. Chakraverty	C/ENG/83
C/pig/Beijing/10/81		Guo et al. (1983)	C/P/10/81
C/pig/Beijing/115/81		Guo et al. (1983)	C/P/115/81
C/pig/Beijing/439/82		Guo et al. (1983)	C/P/439/82

gradients of 30-60% (w/v) sucrose in NTE buffer. Gradients were formed in 25 x 89 mm ultra-clear SW27 tubes (Beckman Instruments, Inc; Palo Alto, CA). Each gradient contained material harvested from approximately 25 eggs. Sucrose gradients were centrifuged in an SW27 rotor at 4°C for 3 h at 25,000 rpm. After centrifugation, the virus band was visualized in the dark using a high intensity lamp and aspirated with a 10 ml syringe by side puncture. Virus was concentrated by diluting the band at least 1:5 in NTE and pelleting in the Ti45 rotor at 35,000 rpm for 1 h (SW27 rotor, 25,000 rpm, 90'). Virus pellets were resuspended in 3-10 ml of NTE buffer for subsequent RNA extraction.

1B. Growth and Purification of Influenza C Viruses

Influenza C virus seed was diluted as described for influenza A and B viruses and 0.1 ml was inoculated into the amniotic sac of 9-10 day old embryonated eggs. Eggs were incubated at 35°C for 48 h then chilled overnight at 4°C. Amniotic fluids were harvested and virus was purified according to the protocol described for A and B viruses with the following modifications. Due to the small volume of amniotic fluid obtained from infected eggs, the pooled fluid from a 100 egg preparation, for example, was directly applied onto a 15 ml 30-60% continuous sucrose gradient (one-half the volume of the A, B virus gradients). Centrifugation was at 4°C for 3 h at 25,000 rpm in the SW27

rotor. The virus band was isolated from the gradient and the RNA was extracted.

1C. Preparation of Seed Virus

The virus sample was diluted serially ten-fold and inoculated into eggs. After appropriate incubation, the eggs were chilled and infected fluids were harvested using sterile technique. The fluids were titered for virus by the HA microtiter assay. The samples showing the highest HA titers at preferably the highest dilutions were kept and aliquoted into 1.8 ml cryotubes. The sterility of the samples were tested by inoculation of 20 ul onto a TSA blood agar plate which was incubated overnight at 37°C. Aliquots of sterile seeds were quick-frozen in an ETOH/dry ice bath and stored at -70°C.

1D. Quantitation of Virus by Hemagglutination Assay

Virus was quantitated by hemagglutination (HA) assay adapted to a microtiter scale (Sever, 1962). Virus diluent (25 ul) was added to a row of wells in a 96 well Linbro microtitration plate (Flow laboratories, Inc.; McLean, VA). Virus sample (25 ul) was added to the first well only. Two-fold serial dilutions of the sample were made using 25 ul stainless steel diluters. An additional 25 ul of diluent was added to all wells followed by 50 ul of 0.5% human type O red blood cells to a final concentration of 0.25%. The contents of the wells were

mixed on a Cooke microtiter mixer for a few seconds and the plate was incubated on ice for 1 h to allow development of the assay. The hemagglutination titer, expressed as hemagglutinating units (HAU) per ml is the reciprocal of the highest dilution exhibiting hemagglutination. Under these conditions, the first well of the microtitration plate contains a 1:8 dilution of virus. It should be noted that when titering influenza C virus samples, adult chicken red blood cells (Flow Laboratories, Inc.; McLean, VA) were sometimes used to increase the sensitivity of the assay.

1E. Quantitation of Influenza A Virus by Plaque Assay

Influenza A viruses were routinely plaqued in MDCK cells as described by Tobita et al. (1975). Confluent monolayers of MDCK cells in 35 mm dishes were infected with 0.1 ml of serial ten-fold dilutions of virus sample made in virus diluent. Virus was allowed to adsorb to the cells for 1 h at 37°C. After adsorption, 3 ml of agar overlay medium was added per dish consisting of REM (M.A. Bioproducts; Walkersville, MD) supplemented with 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, 0.2% bovine albumin, 2 mM $NaH_2PO_4.H_2O$, 10 mM HEPES, Earle's balanced salt solution, 0.1% (w/v) DEAE dextran, 0.6% (w/v) Oxoid agar (Oxoid Ltd; Hampshire, England), 0.1% (w/v) $NaHCO_3$, and 1 ug/ml 1:250 trypsin (Difco Laboratories; Detroit MI). The agar was allowed to harden on the bench top then the

dishes were inverted and incubated at 37°C for 2-3 days. When plaques were visible, the agar overlay was removed with a spatula and the monolayers were stained with a solution of 0.1% crystal violet (Fisher Scientific; Springfield, NJ), 20% MEOH for 10'. The stain was removed and the monolayers were washed with tap water. The titer of virus expressed in plaque forming units (PFU) per ml is equal to the number of plaques on a plate x the virus dilution factor x 10 (volume adjustment).

2. Tissue Culture

2A. Maintenance of Madin-Darby Canine Kidney (MDCK) Cells

Madin-Darby canine kidney (MDCK) cells were propagated in 75 cm^2 plastic flasks. Cell growth medium was minimal essential medium (MEM) with Earle's salts (Gibco Laboratories; Grand Island, NY) supplemented with 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, 10% heat inactivated fetal calf serum (56°C for 1 h), 2 mM L-glutamine (M.A. Bioproducts; Walkersville, MD), and 0.15% (w/v) $NaHCO_3$. Confluent cell monolayers were split by removing the growth medium and washing with 5 ml PBS. Five ml of trypsin-EDTA solution (0.1% 1:250 trypsin (Difco Laboratories), 0.1% (w/v) EDTA, 0.025% (w/v) $NaHCO_3$ in PBS) was added to release the cells from the solid support. Following a 5' incubation at 37°C and gently tapping of the flask, the cells were dissociated from the flask wall.

Five ml of growth medium was added and the cells were mixed by aspiration using a 10 ml pipet. The cell suspension was transferred to a 50 ml plastic Falcon conical tube (Becton Dickinson; Oxnard, CA) and centrifuged for 10' at 1,500 rpm. The cell pellet was resuspended in 4 ml of growth medium. A 1:10 dilution of the cell suspension was made and the cells were counted with a hemocytometer under an inverted microscope. After an appropriate dilution of cells with growth medium, the cells were seeded into 75 cm^2 flasks (4 x 10^6 cells/flask), 35 mm dishes (2.5 x 10^5 cells/dish), or 13 mm glass cover slips (10 cover slips in a 60 mm dish/5 x 10^5 cells). Confluent monolayers were obtained following incubation for 3 days at 37°C in a humidified CO_2 incubator.

2B. Preparation of Primary Chicken Kidney (PCK) Cells

Embryonated chicken eggs were incubated at 37°C and allowed to hatch (day 21). The newly emerged chicks were used for kidney culture preparation at 1-2 days of age as described by Odigiri et al. (1982). It is important that the chicks were not given food or water before use. The chicks were sacrificed and the kidneys were carefully removed with forceps and transferred to a 60 mm dish containing 5 ml of PBS prewarmed to 37°C. Extraneous tissue was removed from the kidneys and they were washed three times with PBS. The kidneys were finely minced in the dish with scissors and transferred to a 50 ml plastic

Falcon conical tube. The tissue was washed three times with 10 ml aliquots of PBS. At this point, a prewarmed solution of 0.25% trypsin in PBS was added to the kidneys (5 ml added per chick). The tube was shaken and incubated on its side at 34°C for 7'. The tube was shaken again and incubated at 34°C for an additional 7'. The tube was removed from the incubator and mixed vigorously by hand for 4' at room temperature. Five ml of 10% fetal calf serum was added to inactivate the trypsin. The cells were centrifuged at 1,000 rpm for 10'. The cell pellet was resuspended in 10 ml 199 medium (Gibco Laboratories; Grand Island, NY) with Earle's salts containing 10% fetal calf serum, 200 U/ml penicillin G, 200 ug/ml streptomycin sulfate, 2 mM L-glutamine, MEM nonessential amino acids, 0.1% $NaHCO_3$, and fungizone (Flow Laboratories Inc.; McLean, VA). The cell suspension was filtered through sterile gauze and additional medium was added to a total of 12.5 ml per chick. The cells were thoroughly mixed and seeded at a high density onto 35 mm dishes (5 dishes/chick) or 13 mm glass cover slips. Cells were incubated at 34°C in an atmosphere of 5% CO_2 and after two days the medium was removed and replaced with fresh medium. By the third or fourth day after seeding, the monolayers (80-90% confluent) were ready to use for viral protein labeling experiments and immunofluorescence studies.

3. Extraction of Viral RNA

Virion RNA was extracted according to Ritchey et al. (1976a) in a 30 ml glass Corex tube fitted with a rubber stopper. All glassware was baked at 250°C to inactivate ribonucleases and sterile disposable pipets were used. To the virus solution, 10% SDS (Pierce Chemical Co.; Rockford, IL) was added to a concentration of 0.33%. After vigorous mixing, one-tenth volume of 10X RSB buffer (0.1 M Tris-HCl, pH 7.4, 0.1 M KCl, 0.015 M $MgCl_2.6\ H_2O$) and 5 mg/ml proteinase K (Beckman Instruments, Inc.; Palo Alto, CA) in RSB were added. The tube was vortexed and incubated in a 56°C water bath for 15'. One-tenth volume of LiCl buffer (5% SDS, 0.1 M acetate buffer, pH 4.9, 1.4 M LiCl) and one volume of phenol (redistilled and saturated with RSB) were added. The tube was mixed and incubated at 56°C for 5'. One volume of chloroform was added (1:1 ratio of phenol to chloroform) and the tube was mixed for 20' on a wrist-action shaker (Burrell Corporation; Pittsburgh, PA). The tube was centrifuged for 10' at 3,000 rpm to separate the phases. The top aqueous layer was removed and transferred to a new Corex tube. The phenol/chloroform extraction was repeated until the interface formed after centrifugation was clear. The RNA was precipitated by the addition of 2.5 volumes of 100% ETOH. After incubation at -70°C for 30', the RNA was pelleted by centrifugation at 4°C for 20' at 11,000 rpm (Sorvall SS-34 rotor). The supernatant was

removed and the RNA pellet was dried by lyophilization in a Speed Vac Concentrator (Savant Instruments, Inc.; Hicksville, NY). The RNA was resuspended in 0.4 ml of 0.3 M NaOAc with a sterile transfer pipet and placed in a 1.5 ml Eppendorf polypropylene tube. The RNA solution was extracted twice with two volumes of ethyl ether and residual ether was removed by evaporation with nitrogen gas. The RNA, found in the lower phase after ether extraction, was precipitated with 2.5 volumes of 100% ETOH and collected by centrifugation at 12,000 x g for 15' in an Eppendorf model 5412 microfuge. The RNA pellet was rinsed with 70% ETOH (centrifuge 5'). The RNA was lyophilized and resuspended in 100 ul of sterile double distilled water (DDW). A dilution of the sample was made and the RNA concentration and purity was determined by spectrophotometric analysis by scanning over the wavelengths from 230-290 nm. One $O.D._{260}$ unit for single-stranded RNA is equal to a 40 ug/ml concentration.

4A. Gel Electrophoresis of Viral RNAs

Influenza Virus RNAs were routinely electrophoresed in polyacrylamide vertical slab gels using a Tris-borate-EDTA (TBE) buffer system modified from Peacock and Dingman (1967). The 10X TBE buffer of pH 8.3 consisted of 0.89 M Trizma base, 0.89 M boric acid, and 0.025 M disodium-EDTA. A 40% stock of polyacrylamide was prepared in sterile DDW with acrylamide (>99.9%, Bio-Rad Laboratories; Rockville

Centre, NY) and N,N'-methylenebisacrylamide (ultra pure, Bethesda Research Laboratories; Gaithersburg, MD) in a ratio of 38:2. The stock solution was filtered using a Nalgene filtration device (Nalge Co.; Rochester, NY). A 10 M urea (ultra pure, Schwarz/Mann Biotech; Cleveland, OH) stock was prepared in sterile DDW which was gravity filtered through Whatman qualitative paper. Seventy-five ml of gel solution was made up which was 1X TBE, 3% polyacrylamide, and 6 M urea. Gel polymerization was initiated by the addition of 10% (w/v) ammonium persulfate (AP) to a concentration of 0.15% and N,N,N'N'-tetramethylethylenediamine (TEMED)(Bio-Rad Laboratories) at 1 ul per ml of gel solution. The gel was cast between autoclaved glass plates, 26 x 16 x 0.5 cm. One plate was notched on both edges of one end and the other was sandblasted to a rough finish on one side. The sandblasted side of the plate faced the inside of the gel mold. The plates were separated by plexiglass spacers (Aquebogue, NY) of 1.5 mm thickness. A 1.5 mm thick plexiglass comb of 15 teeth (6 mm wide separated by 1.6 mm) was used to form the sample wells.

4B. Preparation of RNA Samples for Electrophoresis

For analysis of isolated RNA, approximately 1 ug of RNA was added to a 1.5 ml Eppendorf tube. An equal volume of 2X running mix (40% (w/v) glycerol, 2X TBE, 0.05% (w/v) XC,

0.05% (w/v) BPB) was added. Total volume of the sample was about 10-15 ul. The sample was heated at 56°C for 5' and immediately loaded onto the gel with a Hamilton syringe. The gel was run at room temperature for 12 h at 175 V (constant voltage).

4C. Silver Staining of RNA Separated on Polyacrylamide Gels

At the end of the electrophoresis run, the gel plates were pried apart with a spatula and the gel was silver stained according to the method of Whitton et al. (1983). The silver stain technique, adapted from histochemical procedures, is used to detect proteins and nucleic acids separated on polyacrylamide gels (Merril et al., 1981). Silver staining is more sensitive than the conventional Coomassie brilliant blue R-250 organic stain for proteins and more sensitive than ethidium bromide which is routinely used to detect nucleic acids. The basis of the staining reaction is that silver cations can interact with negatively charged nucleic acid. After the addition of a reducing reagent, the silver cations are converted to metallic silver which is visualized as a black deposit in the gel at the position of the nucleic acid bands.

The gel was fixed for 30' in a solution of 25% ETOH, 10% glacial acetic acid followed by 10% ETOH, 0.5% glacial acetic acid for 30'. The fixative was removed and the gel was submerged in a freshly prepared solution of 500 ml DDW containing 0.9 g of silver nitrate crystal for 1.5 h.

After the staining step, the gel was washed 3-4 times with DDW until a milky white precipitate no longer eluted from the gel. A reduction step followed by the addition of 500 ml of a freshly prepared mixture in DDW consisting of 15 g NaOH pellets, 45 mg $NaBH_4$, and 3.75 ml 37% (w/v) formaldehyde, added after the other ingredients had dissolved. The gel was rocked from side to side and as soon as the RNA bands appeared, it was transferred to a stop solution of 5% glacial acetic acid for 30'.

5A. Isotopic Labeling of Viral Proteins

Influenza virus proteins were pulse-labeled in MDCK or PCK cells as described by Ritchey et al. (1976b). Confluent monolayers of MDCK or PCK cells in 35 mm dishes were infected with 0.1 ml of undiluted virus seed. Virus was adsorbed to the cells for 1 h at the appropriate temperature. After adsorption, the inoculum was left on and 2 ml per dish of maintenance medium was added. Maintenance medium for MDCK cells was MEM with Earle's salts supplemented with 0.21% bovine albumin, 2 mM L-glutamine, 10 mM HEPES, 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, 0.12% $NaHCO_3$, and 1 ug/ml trypsin if multi-cycle virus replication was required. Maintenance medium for PCK cells was 199 with Earle's balanced salt solution containing 0.21% bovine albumin, 10 mM HEPES, 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, and

0.12% $NaHCO_3$. The cell monolayers were incubated at the appropriate temperature for various times until the desired time post-infection (p.i.) for labeling was reached. At this time, the maintenance medium was removed and the monolayers were washed with PBS. Protein labeling medium was added (1.5 ml per dish) which consisted of Hanks' balanced salt solution, 0.5% $NaHCO_3$, 0.004% (w/v) phenol red, 100 U/ml penicillin G, 100 ug/ml streptomycin sulfate, 0.2% (w/v) glucose supplemented with 200 uCi/ml L-[^{35}S]-methionine (specific activity 1,100 Ci/mmole; New England Nuclear; Boston, MA). The cells were pulse-labeled for 1 h. The labeling medium was removed and the monolayers were washed with PBS. Two hundred ul of protein loading buffer (30% glycerol, 3% 2-mercaptoethanol, 2% SDS, 0.025% BPB, 0.01X buffer 11B --see next section) was added per dish to lyse the cells. The cells were scraped off the dish with a rubber policeman and transferred to a 1.5 ml Eppendorf tube.

5B. Gel Electrophoresis of Viral Proteins

Virus-infected pulse-labeled cell lysates were electrophoresed on 7-14% linear gradient SDS-polyacrylamide slab gels (Young and Palese, 1979). Gradient gels and electrophoresis conditions were according to the method described by Maizel (1971).

Stock solutions of polyacrylamide (ratio of acrylamide to bisacrylamide of 30:0.8), 10% SDS, 60% glycerol, and

10% AP were made up in DDW. Buffer 11A (20X resolving gel buffer) was 36.3 g of Trizma base dissolved in <100 ml of DDW. The pH was made 8.9 by addition of 12 N HCl and then the volume was adjusted to 100 ml with DDW. Buffer 11B (8X stacking gel buffer) was 5.7 g Trizma base dissolved in <100 ml of DDW. The pH was adjusted to 7.1 with 14.6 M phosphoric acid and the volume was brought up to 100 ml. The 7.0% gel solution (24 ml) was prepared containing 7.0% polyacrylamide, 5.2% buffer 11A, and 0.1% SDS. The 14.0% gel solution (24 ml) consisted of 14.0% polyacrylamide, 5.2% buffer 11A, 0.1% SDS, and 20% glycerol. Additionally, SDS-polyacrylamide gels containing urea were formed. In this case, ultra pure urea was dissolved in both gel solutions at a concentration of 7 M. To start polymerization, 10% AP and TEMED were added to a final concentration of 0.03% and 0.075%, respectively. The gradient gel was cast using a gradient former between glass plates, 26 x 16 x 0.5 cm, separated by 1.5 mm spacers. The gel mold was filled to within 4 cm of the top of the plate and immediately 1 ml of isobutanol was layered on top of the gel solution to insure a smooth edge. When polymerization was complete, the isobutanol was washed out with several rinses of DDW and the inside of the gel mold was dried with 3 MM chromatography paper. The stacking gel solution was prepared which consisted of 4.6% polyacrylamide, 12.5% buffer 11B, 0.2% SDS, 0.15% AP, and

0.15% TEMED. The stacking gel was poured directly on top of the resolving gel and a 15 slot (6 mm wide separated by 1.6 mm) comb was inserted to form the sample wells. The electrophoresis running buffer was 0.025 M Tris - 0.19 M glycine (Sigma Chemical Co.; St. Louis, MO), pH 8.25, and 0.1% SDS.

The protein samples in loading buffer were boiled for 5' to denature the proteins and immediately loaded into the sample wells. The proteins were electrophoresed for a total of 1600 voltage hours (i.e., 130 V for 12 h) until the BPB tracking dye ran off the gel. Following electrophoresis, the stacking gel was removed and the resolving gel was fixed in 25% methanol, 8% glacial acetic acid for 30' with gentle rocking at room temperature. The fixed gel was then dried on a Bio-Rad slab gel dryer and exposed to Cronex X-ray film (Dupont; Wilmington, DE).

6. Indirect Immunofluorescence of Virus-Infected Cells

Localization of the NS1 protein in influenza virus-infected cells was determined by immunofluorescence analysis as described by Young et al. (1983). MDCK and PCK cells were seeded onto 13 mm glass cover slips and grown to 60-100% confluency. Monolayers were infected with 0.1 ml of approximately 10^8 PFU/ml of virus for 1 h at the appropriate temperature. At various times p.i., the coverslips were harvested, washed with PBS, and allowed to

air dry. The cells were then fixed in cold acetone for 10' at room temperature, air dried, rinsed with PBS, and air dried again. The monolayers were incubated for 30' at 37°C in a moist environment with a prewarmed monospecific rabbit antiserum (diluted 1:50) which was raised to E. coli-produced NS1 protein (Young et al., 1983). The primary antibody was aspirated and the coverslips were rinsed with three changes of PBS. The cells were treated for 30' at 37°C with a diluted fluorescein-conjugated goat antibody to rabbit 7S immunoglobulin G (Hyland Diagnostics; Dearfield, Il.). The secondary antibody was aspirated off and the coverslips were washed with three changes of PBS and blotted dry on filter paper. The coverslips were inverted and mounted in phosphate buffered glycerol onto microscope slides. NS1-specific immunofluorescence was visualized using a Leitz fluorescence microscope.

7. Preparation of Influenza Virus-Specific ds cDNA

The protocol for the synthesis of virus-specific cDNA involved the use of reverse transcriptase and synthetic oligonucleotide primers as described by Baez et al. (1980). The procedure takes advantage of the fact that the 3' and 5' ends of all the gene segments of influenza A viruses have conserved sequences (Desselberger et al., 1980). Therefore, a dodecamer primer complementary to the 3' terminal nucleotides of the vRNA and one the same sense as the 5' end of the vRNA, effectively prime first and second

strand cDNA synthesis, respectively, of all eight viral genes.

7A. Synthesis of First Strand cDNA (ss cDNA)

The reaction to synthesize single-stranded cDNA complementary to total virion RNA was in a 300 ul volume and contained the following components: 20 ug of purified vRNA, reverse transcriptase (RT) buffer (50 mM Tris-HCl, pH 8.3, 60 mM NaCl, 20 mM dithiothreitol (DTT), 12 mM $MgCl_2$), 0.5 U/ul RNasin (Promega Biotec; Madison, WI.), 100 ug/ml actinomycin D (Sigma Chemical Co.; St. Louis, MO), 0.5 mM each of dATP, dGTP, dTTP, dCTP (Pharmacia P-L Biochemicals Inc.; Milwaukee, WI), 10 ug 3' end primer d(AGCAAAAGCAG)rG (Collaborative Research, Inc.; Waltham, MA.), 100 uCi [alpha-^{32}P]-dATP in Tricine (specific activity 3,000 Ci/mmole; New England Nuclear; Boston, MA), and 500 U/ml AMV RT (Molecular Genetic Resources, Inc.; Tampa, FL). The reaction was incubated at 42^{o}C for 1.5 h and then terminated by making it 0.2% SDS, 0.02 M EDTA, and 0.3M NaOAc. The mixture was extracted two times with an equal volume of redistilled phenol/chloroform (1:1) followed by two extractions of the aqueous phase with two volumes of ether. The RNA/DNA hybrids were ethanol precipitated by the addition of 2.5 volumes of 100% ETOH, incubated at -70^{o}C for 20'. The nucleic acid was collected by centrifugation for 15' at 12,000 x g and the pellet was washed with 70%

ETOH and lyophilized. The pellet was resuspended in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) and the RNA template of the hybrids was removed by digestion with pancreatic RNase A (25 ug/ml) at 37°C for 30'. The mixture was made 0.3 M NaOAc and extracted several times with phenol/chloroform (1:1) and then with ether. The aqueous phase was passed through a Sephadex G-75 (Pharmacia Fine Chemicals; Piscataway, NJ) gel filtration column (20 x 0.8 cm, polypropylene; Kontes; Vineland, NJ) equilibrated with NTE, to separate the ss cDNA from unincorporated dNTPs and digested RNA mononucleotides. The first radioactive peak to elute from the column contained the ss cDNA and the fractions containing this peak were counted by measuring Cerenkov radiation. Based on the incorporated radioactivity and the ug of dNTP put into the reaction, the amount of ss cDNA was quantitated. The ss cDNA was ethanol precipitated and the DNA pellets were resuspended in 20 ul of DDW and pooled into a single Eppendorf tube. At this point in the procedure, an aliquot of the ss cDNA was put aside for subsequent gel analysis.

7B. Synthesis of Second Strand cDNA (ds cDNA)

One ug of 5' end primer d(AGTAGAAACAAG), complementary to the 3' end of the first strand cDNA, was added to the ss cDNA product of the first strand reaction. These two components were incubated at 90°C for 2' and then cooled to room temperature for 5'. RT buffer, 0.5 mM each of dATP,

dGTP, dTTP, dCTP, and 500 U/ml RT were added to give a final reaction volume of 100 ul. The mixture was incubated at 37°C for 1.5 h. The reaction was stopped by addition of 10% SDS, 0.5 M EDTA, and 3 M NaOAc to 0.2%, 0.02 M, and 0.3 M, respectively. The reaction was phenol/chloroform (1:1) extracted, ether extracted, and the ds cDNA peak was collected after gel filtration in NTE buffer on Sephadex G-75 (20 x 0.8 cm, polypropylene column; Kontes; Vineland, NJ). The radioactivity in the peak was counted and the fractions were ethanol precipitated. The ds cDNA pellets were resuspended in DDW, pooled into a single tube, and ethanol precipitated a second time. The final pellet was rinsed with 70% ETOH, lyophilized, and resuspended in 20 ul DDW. An aliquot of the ds cDNA preparation was saved for further analysis.

Aliquots (5,000 cpm) of the ss and ds cDNA products were treated with the single-strand specific nuclease S1 (Sigma Chemical Co.; St. Louis, MO) to analyze the efficiency of conversion of ss to ds cDNA and also to check if full-length S1-resistant copies of the viral genes were generated. In addition, the ds cDNA was digested with a variety of restriction endonucleases having a single recognition sequence in the plasmid cloning vector, pBR322, to check for internal restriction sites within the genes. The S1- and restriction enzyme- digested samples, along with control size markers, were electrophoresed in a 4%

polyacrylamide (stock solution acrylamide to bisacrylamide, 29:1)/TBE gel (26 x 16 x 0.5 cm, 1.5 mm thick) until the XC tracking dye had migrated 17 cm down the gel.

8. Modification of ds cDNA for Insertion into pBR322

The NS gene-specific ds cDNAs of the A/FM/1/47, A/FW/1/50, CR43-3, and A/Alaska/6/77 viruses did not contain internal EcoRI sites. Therefore, the unique EcoRI site of pBR322 was used for cloning these NS genes.

The 5' ends of the ds cDNAs were phosphorylated so that EcoRI synthetic linkers, d(GGAATTCC) (Collaborative Research, Inc.; Waltham, MA) could be attached. The linkers were also phosphorylated.

The cDNA was phosphorylated in a 20 ul reaction mixture containing 1 mM ATP, 10 U T4 PNK (Bethesda Research Laboratories; Gaithersburg, MD) in Wu buffer (66 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 15 mM DTT, 0.1 mM ATP). The reaction was incubated at 37°C for 60'. Another 10 U aliquot of PNK was added and the incubation was continued for an additional 60'.

To phosphorylate the linkers, 50 uCi of [gamma-^{32}P]-ATP (specific activity 3,000 Ci/mmole; New England Nuclear; Boston, MA.) was lyophilized in an Eppendorf tube. Two ug of EcoRI linkers and 10 U of PNK in kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA) were added to give a volume of 20 ul. The

reaction was incubated at 37°C for 30'. Ten mM ATP was added to a concentration of 1 mM along with 10 U of PNK. The incubation was continued for 60'.

The cDNA and linker kinase reactions were incubated at 65°C for 10' to inactivate the kinase enzyme. The cDNA and linkers were added together. The mixture was made 0.3 M NaOAc, phenol/chloroform (1:1) extracted, ether extracted, and ethanol precipitated. The pellet was ethanol rinsed, lyophilized, and resuspended in DDW. A 100 ul ligation reaction containing the kinased cDNA and linkers in Wu buffer, 0.5 mM ATP, and 400 U T4 DNA ligase (New England Biolabs; Beverly, MA.) was prepared. The ligation was incubated for 15 h at 14°C and then for 5' at 70°C to inactivate the ligase. An aliquot of the reaction was saved for analysis and the rest was digested with an excess of 100 U of EcoRI restriction enzyme (New England Biolabs; Beverly, MA) according to the manufacturer's specifications to convert polymerized linkers to monomers. The linker ligation to the cDNA and subsequent polylinker digestion were analyzed by autoradiography following electrophoresis of aliquots of both reactions in a 4% polyacrylamide/TBE gel (26 x 16 x 0.5 cm, 1.5 mm thick) until the BPB tracking dye had migrated 12 cm.

The linker digest was loaded directly onto a Sepharose CL-6B (Pharmacia Fine Chemicals; Piscataway, NJ) column (30 x 1.8 cm) in NTE buffer to separate the cDNA with attached

linker from the digested linker. The radioactive cDNA peak (first to elute) was collected and 3 M NaOAc was added to a concentration of 0.3 M. The cDNA was ethanol precipitated, rinsed with 70% ETOH, lyophilized, and resuspended in 200 ul sterile DDW.

9. Preparation of Cloning Vector and Ligation of cDNA

The vector pBR322 was linearized by complete digestion with EcoRI. The 5' ends of the plasmid were dephosphorylated to prevent recirculariza tion of plasmid without insert in the ligation of the cDNA to the vector. The dephosphorylation protocol and reagents were obtained from a 5' DNA terminus labeling system (Bethesda Research Laboratories; Gaithersburg, MD). The 100 ul cocktail contained 25 ug of linearized pBR322 (18 pmole 5'termini) and 1 ul of 256 U/ul bacterial alkaline phosphatase (BAP) in BAP buffer (10 mM Tris-HCl, pH 8.0, 120 mM NaCl). The mixture was incubated at 65^{o}C for 60' and then cooled to room temperature. The reaction was extracted several times with an equal volume of phenol/chloroform (1:1) followed by two ether extractions. The DNA was ethanol precipitated, washed with 70% ETOH, lyophilized, and resuspended in sterile DDW.

Routinely, one-half of the cDNA preparation (100 ul) was divided into two ligation reactions. One reaction contained 80 ul of cDNA and the other, 20 ul. The 100 ul cocktail also consisted of 200 ng BAP-treated pBR322 and

400 U T4 DNA ligase (New England Biolabs; Beverly, MA) in Wu buffer supplemented with 0.5 mM ATP. The ligation was at 14°C for 15 h.

10. Preparation of Competent Bacterial Cells and Transformation

A 20 ml overnight culture of E. coli C600 cells was propagated in TY medium (10 g Difco Bacto-tryptone, 10 g Difco Bacto-yeast extract, 5 g NaCl per liter). A small aliquot of the culture was inoculated into 50 ml of fresh TY medium and the cells were grown at 37°C to logarithmic phase ($O.D._{590}$ = 0.6). The cells were made transformation competent according to the method described by Dagert and Ehrlich (1979). The bacterial cells were pelleted by centrifugation at 3,000 rpm for 10' and resuspended in 10 ml of cold, sterile 100 mM $CaCl_2$, 10 mM Tris-HCl, pH 8.0. The bacteria were incubated on ice for 60' and pelleted at 4°C. The cell pellet was resuspended in 2 ml of 100 mM $CaCl_2$, 10 mM Tris-HCl, pH 8.0.

A 100 ul aliquot of competent E. coli cells was added to the 100 ul ligation reaction along with 100 ul of TCM (10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM $CaCl_2$). This mixture was chilled on ice for 30' followed by a heat shock at 42°C for 2'. Five hundred ul of TY medium plus ampicillin (50 mg per liter) was added and the cells were incubated at 37°C for 45'. Aliquots of 20, 50, and 100 ul

of the transformation mixture were plated onto TY agar plates (20 g Difco Bacto-agar per liter) supplemented with ampicillin. The plates were inverted and incubated at 37°C for 16 h. A typical transformation using one-half of the total cDNA preparation (10 ug of starting vRNA) yielded approximately 500 bacterial colonies (10-fold greater than the background control: BAP-treated pBR322 plus ligase).

11. Identification of Recombinant Clones

Bacterial colonies containing influenza virus-specific sequences were identified using the in situ colony hybridization method of Grunstein and Hogness (1975). Bacterial transformants (isolated single colonies) were streaked onto TY plus ampicillin agar plates (50 colonies per plate) and incubated at 37°C for 15 h. The bacterial streaks were replica-plated using velvet onto several nitrocellulose membrane filters (0.45 um, 82.5 mm circles; Schleicher & Schuell; Keene, NH) which were layered on top of the agar plates. The bacteria on the filters were grown at 37°C for 15 h and the replica plates were stored at 4°C until they were ready to be used.

Two sheets of Whatman 3 MM chromatography paper per tray were placed in four large trays lined up on the bench top. The 3 MM paper was saturated with a series of solutions in the order listed: 0.5 N NaOH, 1 M Tris-HCl, pH 8.0, 1 M Tris-HCl, pH 8.0, and 10X SSC (1.5 M NaCl, 0.15 M $Na_3C_6H_5O_7.2\ H_2O$). The nitrocellulose filters

containing the bacteria were transferred through each of the solutions, 10' per tray, to lyse the cells, denature the DNA, and adjust the pH back to neutral. The filters were air dried on two sheets of 3 MM paper and then transferred to a tray of chloroform. The tray was agitated for a few minutes and in this time, the bacterial streaks first turned white and then were no longer visible. The filters were air dried on 3 MM paper and floated on a solution of 10X SSC. Once the filters became thoroughly wet from beneath, they were submerged in the 10X SSC. The nitrocellulose filters were partially air dried and baked for 2 h in a vacuum oven at 80^{o}C to permanently fix the DNA to the filters.

11A. Preparation of a Hybridization Probe by Nick Translation

To identify bacterial clones containing plasmids with NS gene-specific sequences, a cDNA copy of the NS gene of the A/PR/8/34 virus was nick translated (Rigby et al., 1977) and used as a hybridization probe. The NS gene of the A/PR/8/34 virus was previously cloned into the EcoRI site of pBR322 (Baez et al., 1980).

Nick translation was carried out using a ^{32}P nick translation system (New England Nuclear; Boston, MA). The reaction volume was 25 ul containing 100 uCi of [alpha-^{32}P]-dATP (specific activity 3,000 Ci/mmole), 0.5 ug of

insert DNA (in TE buffer), 20 uM each of cold dGTP, dCTP, dTTP, 48 U/ml *E. coli* DNA polymerase I, and 0.02 U/ml pancreatic DNase I in 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$. The reaction cocktail was incubated at 14°C for 2 h. The reaction was stopped by addition of 100 ul of nick translation stop buffer containing 0.5 M EDTA, 0.2% SDS, and 0.3 M NaCl. The labeled DNA was separated from the unincorporated radioactivity by Sephadex G-75 chromatography (20 x 0.8 cm, polypropylene column; Kontes; Vineland, NJ) in NTE buffer. The DNA peak was the first to elute from the column and the 0.4 ml peak fractions were counted and pooled.

Baked nitrocellulose filters were prehybridized at 65°C in a water bath shaker for 2 h, in hybridization solution: 1X Denhardt's (0.02% Ficoll-70, 0.02% polyvinylpyrrolidone-40, 0.02% BSA-fraction V; Sigma Chemical Co.; St. Louis, MO), 3X SSC, and 0.1% SDS. The filters were placed in a heat-sealable plastic bag (eight filters per bag). The probe was added to 15 ml of the hybridization solution in a 50 ml Falcon tube and denatured by boiling for 10'. The probe solution was poured into the plastic bag, the air bubbles were removed, and the bag was sealed. The hybridization was in a 65°C water bath with continuous slow shaking for 12 h.

The probe was removed from the bag and the filters were washed three times at 65°C with shaking for 30' per wash with two solutions. The filters were first dipped in 200 ml of prewarmed 3X SSC, 0.1% SDS then transferred to 800 ml of the same solution for the first wash. The filters were washed a second time with 3X SSC, 0.1% SDS. The final wash was 0.1X SSC, 0.1% SDS. Two ml of 1% methyl green was added during the last wash to visualize the positions of the colony streaks. The filters were air dried on 3 MM paper, mounted to a rigid support and exposed at -70°C to Cronex X-ray film with a lightning plus intensifying screen (Dupont; Wilmington, DE). Bacterial colonies containing NS-specific DNA were identified from the autoradiogram.

11B. Rapid Screen Analysis of Bacterial Colonies

The hybridization positive colonies were picked from the master plate and inoculated into 3 ml of TY plus ampicillin medium. The mini cultures were grown at 37°C for 15 h and were screened by the rapid alkaline method (Birnboim and Doly, 1979) for the presence of plasmids with full-length NS inserts.

One-half of the culture was transferred to an Eppendorf tube and the cells were pelleted by centrifugation at 12,000 x g for 2'. The bacterial cells were resuspended in 100 ul of solution I (25 mM Tris-HCl, pH 8.0, 50 mM glucose, 10 mM EDTA) with lysozyme (Sigma Chemical Co.; St.

Louis, MO) freshly added to 2 mg/ml. The incubation was on ice for 30'. Two hundred ul of solution II (0.2 N NaOH, 1.0% SDS) was added and the tube was gently vortexed. At this time, the suspension became almost clear and slightly viscous. The tube was maintained on ice for 5' and then 150 ul of solution III (3 M NaOAc, pH 4.8) was added. The contents of the tube were mixed and incubated on ice for 60' to allow protein, high molecular weight RNA, and chromosomal DNA to precipitate. After centrifugation at 12,000 x g for 10', the supernatant was transferred to a second tube. An equal volume of isopropanol was added to precipitate the plasmid DNA (-70^{o}C, 15'). The DNA was collected by centrifugation, resuspended in 100 ul of 0.1 M NaOAc, and precipitated with 2.5 volumes of ethanol. The pellets obtained after centrifugation were rinsed with 70% ETOH and lyophilized.

The DNA was resuspended in 50 ul DDW. One-half of the preparation was digested with EcoRI in a 150 ul reaction containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, and 0.1 ug/ul RNase A. After digestion at 37^{o}C for 60', the DNA was ethanol precipitated, ethanol rinsed, lyophilized, and resuspended in 10 ul of DDW. Ten ul of 2X sample loading buffer (0.1% BPB, 0.1% XC, 20% glycerol in 2X TBE) was added and the sample was applied to a 6 x 1.5 mm well formed in a 1.0% agarose (ultra pure; Bethesda Research Laboratories; Gaithersburg, MD) gel

containing 0.5 ug/ml ethidium bromide (EtBr). The agarose gel was a 3 mm thick horizontal slab gel. The electrophoresis buffer was TBE with 0.5 ug/ml EtBr. The rapid screen samples, along with molecular weight size markers, were electrophoresed at 4°C for 2 h at 75 mA (constant current). Following the run, the gel was photographed under short wave 254 nm UV illumination with a Tiffen (Hauppauge, NY) dark red filter (#29). The film was Polaroid 3000 speed, black and white land pack film, type 107C. The exposure was for 1/4 second at f 4.5. Bacteria harboring recombinant plasmids which appeared to possess complete EcoRI-released NS inserts, based on gel mobility, were grown up in large quantities so that plasmid DNA could be obtained for further analysis. The plasmids generated that possess full-length copies of the NS gene are described in Table 6.

12. Large Scale Purification of Plasmid DNA

Plasmid DNA from a one liter bacterial cell culture (grown to $O.D._{590}$ = 0.6) was isolated using the alkaline method (Birnboim and Doly, 1979) as described previously for mini cultures. Cells were harvested by centrifugation at 4°C at 8,000 rpm for 15'. The bacteria were resuspended in 40 ml of NTE and pelleted once again. The cell pellet was resuspended in 14 ml of solution I and transferred to two Beckman Type 45 Ti tubes. Seven ml of solution I with

TABLE 6. PLASMIDS CONTAINING INFLUENZA VIRUS NS GENE SEQUENCES

Plasmid	NS Gene Orientation*	Virus
pG33	V	A/FM/47
pQ8	M	A/FW/50
pZ22	M	CR43-3
pQ7	M	A/ALA/77

*In the M orientation, base 1 of the NS gene (plus sense orientation) is adjacent to base 4361 of the EcoRI site of pBR322. V defines the opposite orientation.

4 mg/ml lysozyme was added per tube. The tube contents were mixed and incubated on ice for 30'. 28 ml of solution II was added. The solution was mixed thoroughly and maintained on ice for 5'. Once the mixture cleared, 21 ml of solution III was added per tube and the tube was mixed and kept on ice for 60'. Chromosomal DNA was pelleted by centrifugation at 4°C in the Type 45 Ti rotor at 30,000 rpm for 20'. The supernatant was transferred to two Type 45 Ti tubes and the plasmid DNA was precipitated twice with an equal volume of isopropanol. The plasmid DNA was collected by centrifugation at 12,000 rpm for 20'. The lyophilized DNA pellet was brought up in 4 ml of 10 mM Tris-HCl, pH 7.5 and RNase A was added to a concentration of 100 ug/ml. The RNA was digested at 37°C for 30'. The solution was made 0.3 M NaOAc and extracted with phenol/chloroform (1:1) until the interface was devoid of protein. The DNA was ethanol precipitated and resuspended in 1 ml of NTE containing 10% glycerol. The sample was applied to a Bio-Gel A-50m agarose (Bio-Rad Laboratories; Rockville Centre, NY) column (30 x 1.8 cm) equilibrated with NTE buffer, to separate the DNA from the digested RNA. The DNA void peak was monitored by UV absorbance at 254 nm and collected. The DNA solution was made 0.3 M NaOAc and ethanol precipitated. Following centrifugation at 11,000 rpm (Sorvall SS-34 rotor) for 30', the plasmid pellet was resuspended in 400 ul of 0.3 M NaOAc and transferred to an

Eppendorf tube. The DNA solution was extracted several times with phenol/chloroform (1:1) and then with ether. Ethanol (2.5 volumes) was added to precipitate the DNA which was collected by centrifugation at 12,000 x g for 15'. The resultant pellet was ethanol rinsed, lyophilized, and resuspended in 200 ul TE buffer. The DNA was quantitated by UV absorbance at 260 nm of a diluted aliquot of the sample. The DNA was adjusted to 1 ug/ul.

The plasmid DNA isolated by this method was of sufficient purity for most enzymatic manipulations. However, for some experiments, it was necessary to further purify the DNA by CsCl density gradient centrifugation.

12A. Purification of Plasmid DNA on CsCl-EtBr Density Gradients

The DNA peak eluted from the Bio-Gel A-50m agarose column in the protocol described above could be further purified by CsCl density gradient centrifugation as described by Maniatis et al. (1982). For every ml of DNA solution, 1 g of solid CsCl (Sigma Chemical Co.; St. Louis, MO) was added and dissolved. For every 10 ml of CsCl solution, 0.8 ml of 10 mg/ml EtBr was added in the dark to a concentration of approximately 600 ug/ml. The solution was mixed well. The final density of the DNA mixture was 1.55 g/ml (refractive index = 1.3860). The CsCl solution was transferred to Beckman Type 50 Ti tubes. Centrifugation was at 20^{o}C in the Type 50 Ti rotor at

40,000 rpm for 48 h. The lower visible band in the middle of the gradient was the closed circular plasmid DNA. This band was extracted by side-puncture with a syringe. The EtBr was removed from the DNA by extraction with an equal volume of n-butanol saturated with water. The phases were mixed vigorously and then separated by centrifugation at room temperature at 3,000 rpm for 3'. The lower aqueous phase was transferred to a clean glass tube. The extraction was repeated until the pink color was no longer visible in the aqueous solution. The aqueous layer was dialyzed against several changes of TE buffer, pH 8.0, to remove the CsCl. The DNA was concentrated by ethanol precipitation, resuspended in a small volume of DDW and quantitated.

13. End-labeling DNA Fragments for Maxam and Gilbert Sequencing

Restriction enzyme maps of the NS inserts of pG33, pQ8, pZ22, and pQ7 were generated to develop sequencing strategies. Most of the experiments to generate labeled fragments of suitable size for Maxam and Gilbert reactions involved restriction enzyme cleavage of 25 ug of plasmid DNA (or 1.5 ug of insert DNA), 5' end-labeling of the cleavage products, followed by digestion with a second enzyme to obtain the desired fragments with a single 5' labeled terminus.

5' end-labeling of DNA fragments was carried out with a 5' DNA terminus labeling system (Bethesda Research Laboratories; Gaithersburg, MD). Restricted DNA was first treated with BAP to remove the 5' phosphate groups. The 100 ul reaction contained DNA (20 pmole of 5' extensions), 200 U BAP in BAP buffer (10 mM Tris-HCl, pH 8.0, 120 mM NaCl). The mixture was incubated at 65^{o}C for 60' and cooled to room temperature. The reaction was extracted several times with an equal volume of phenol/chloroform (1:1). The aqueous phase was extracted with ether and the DNA was ethanol precipitated. The dephosphorylated DNA was collected by centrifugation at 12,000 x g for 10'. The DNA pellet was rinsed with 70% ETOH and resuspended in 15 ul of sterile DDW. Two hundred uCi of [gamma-^{32}P]-ATP (specific activity 3,000 Ci/mmole; New England Nuclear; Boston, MA.) was lyophilized in an Eppendorf tube. The dephosphorylated DNA was added and a 50 ul reaction consisting of 8 U T4 PNK in forward reaction buffer (60 mM Tris-HCl, pH 7.8, 15 mM 2-mercaptoethanol, 10 mM $MgCl_2$, 0.33 uM ATP) was set up. The incubation was at 37^{o}C for 30'. A second 8 U aliquot of PNK was added and the incubation was continued for 30'. The reaction was extracted with phenol/chloroform (1:1), ether, and ethanol precipitated. The 5' end-labeled DNA was digested with an appropriate second restriction enzyme to obtain fragments with a uniquely labeled end. Following digestion, the

mixture was phenol/chloroform (1:1) extracted, ether extracted, and the DNA was collected by ethanol precipitation. The end labeled fragments were isolated by polyacrylamide gel electrophoresis in TBE buffer and visualized by autoradiography.

A second strategy for generating DNA fragments uniquely labeled at one end was to cleave the DNA with a restriction enzyme and fill the 3' ends with reverse transcriptase in the presence of the appropriate [alpha-^{32}P]-dNTP. Two hundred uCi of [alpha-^{32}P]-dNTP (specific activity 3,000 Ci/mmole) was dried down and a 20 ul reaction was prepared, consisting of the restricted DNA, 0.5 mM each of the other three cold dNTPs, and 20 U of RT (Molecular Genetic Resources; Tampa, FL) in RT buffer. The reaction cocktail was incubated at 37^{o}C for 60'. The reaction was made 0.3 M NaOAc and phenol/chloroform (1:1) extracted. The aqueous layer was ether extracted and the DNA was ethanol precipitated. A second enzyme digest was required to obtain DNAs labeled at a single end.

A third protocol to generate DNAs labeled at a single terminus was to digest the NS insert and label the 5' ends of the fragments as described above. The lyophilized pellet resulting from ethanol precipitation was resuspended in 40 ul of strand separation running mix (30% glycerol, 1 mM EDTA, 0.05% XC, 0.05% BPB). The sample was heated to 90^{o}C for 10', quick-chilled in an ice water bath, and

loaded immediately into a 3 cm slot of a 5% polyacrylamide/TBE gel of 1.5 mm thickness. The gel was cast between 41 x 20 x 0.5 cm glass plates. The denatured double strands were electrophoresed at 4°C at 300 V for 16 h. The complementary DNA strands of a fragment have distinct secondary structures dictated by base composition and therefore, the strands migrate to different positions in the gel.

Subsequent to gel electrophoresis of end labeled DNAs and autoradiography, the desired bands were excised from the gel and the radioactivity (cpm) of each fragment was counted. The DNA molecules were electroeluted in TAE buffer (40 mM Tris-acetate, pH 8.0, 1 mM EDTA). The eluted DNA was made 0.3M NaOAc, phenol/chloroform (1:1) extracted, ether extracted, and ethanol precipitated twice. The final DNA pellet was ethanol rinsed and lyophilized. The cpm per labeled fragment were measured to determine the efficiency of electroelution.

14. Maxam and Gilbert Sequencing of End-Labeled DNA

The basis of the Maxam and Gilbert (1980) protocol is that each nucleotide of a ^{32}P-end labeled DNA molecule can be specifically cleaved with a series of chemical reagents. The base-specific cleavage is limited so that every DNA molecule contains only one or two modified bases. Most chemical schemes which cleave at one or two of the four bases involves three consecutive steps: modification of a

base, removal of the modified base from its sugar, and DNA strand scission at that sugar. It is important to note that each step is contingent on the reaction preceding it. The DNA will break only at a sugar without a base and only an appropriately modified base can be removed from its sugar.

In most of the base-specific cleavage protocols, one reaction disrupts the electronic structure of a DNA base, making it possible for a second reaction to break the bond between that base and its sugar. Specifically, dimethyl sulfate methylates the 7-nitrogen of guanine resulting in an opening of the guanine ring between carbon 8 and nitrogen 9. Piperidine then displaces the ring-opened 7-methylguanine from its sugar. The nucleophile hydrazine splits a thymine or cytosine ring, leaving ring fragments which can be displaced from the sugar by piperidine. The presence of molar sodium chloride during the limited hydrazine reaction causes only cytosine residues to react appreciably. In other reactions specific for two bases, acid weakens the glycosidic bonds of adenine and guanine by protonating purine ring nitrogens and alkali treatment opens adenine and cytosine rings. Piperidine is then able to displace the modified base from its sugar.

In the final step of all reaction schemes, piperidine catalyzes the beta-elimination of phosphates from the empty sugar to finally cleave the DNA strand. The reaction

leaves a free sugar, one phosphate on an unlabeled fragment, and the other phosphate on a fragment containing the original ^{32}P end label. Partial chemical cleavage produces nested end-labeled products. For example, in the guanine modification reaction, fragments originating from a common labeled end which terminate at the preceding base of every guanine residue in the DNA fragment will be represented. The same is true for each of the three other specific base modification reactions. If the four different nested sets of products are electrophoresed in parallel on a high resolution gel and autoradiographed, the sequence of the DNA fragment can be read off the four meshing band patterns on the film.

14A. Modification at Guanines (G)

Five ul of end labeled DNA in DDW, 1 ul of herring sperm carrier DNA (1 mg/ml), and 200 ul of G buffer (50 mM sodium cacodylate, pH 8.0, 1 mM EDTA) were combined in a 1.5 ml siliconized Eppendorf tube. The tube was chilled on ice and 1 ul of dimethyl sulfate, 99% (Aldrich Chemical Co.; Milwaukee, WI) was added. The reaction was vortexed and incubated at 20°C for 7'. Fifty ul of chilled G stop solution (1.5 M sodium acetate, pH 7.0, 1 M 2-mercaptoethanol, 100 ug/ml herring sperm carrier DNA) was added and the DNA was precipitated by the addition of 750 ul of 95% ETOH. After a 5' incubation in a dry

ice/ETOH bath, the DNA was collected by centrifugation at 12,000 x g for 10'. The supernatant was removed and the DNA pellet was resuspended in 250 ul of 0.3 M NaOAc. The DNA was ethanol precipitated a second time and the resultant pellet was rinsed with 1 ml of 95% ETOH and lyophilized.

14B. Modification at Guanines and Adenines (G + A)

Ten ul of end labeled DNA in DDW and 1 ul of herring sperm carrier DNA (1 mg/ml) were combined with 10 ul of DDW in a 1.5 ml siliconized Eppendorf tube. Two ul of 1.0 M piperidine formate, pH 2.0 (4% (v/v) formic acid, adjusted to pH 2.0 with piperidine) was added and the reaction was incubated at 30°C for 75'. The sample was freezed and lyophilized. The DNA was redissolved in 20 ul of DDW and lyophilized again.

14C. Modification at Pyrimidines

For base modification at cytosines and thymines (C + T), 10 ul of end labeled DNA, 1 ul of herring sperm carrier DNA (1 mg/ml), and 10 ul of DDW were combined in a 1.5 ml siliconized Eppendorf tube. For the cytosine (C) specific reaction, 5 ul of end labeled DNA and 15 ul of 5 M NaCl were added to an Eppendorf tube. The reaction was initiated by the addition of 30 ul of hydrazine, 95% (Eastman Kodak; Rochester, NY). The tube was vortexed and maintained at 30°C for 7' (C + T) or 10' (C). Two

hundred ul of chilled hydrazine stop solution (0.3 M NaOAc, 0.1 mM EDTA, 25 ug/ml herring sperm carrier DNA) and 750 ul of 95% ETOH were added. The reaction was mixed and kept in a dry ice/ETOH bath for 5'. The DNA was pelleted by centrifugation at 12,000 x g for 10'. The supernatant was removed and the pellet was lyophilized.

14D. Modification at Adenines and Cytosines (A > C)

Five ul of end labeled DNA in DDW was combined with 1 ul of herring sperm carrier DNA (1 mg/ml) and 100 ul of A > C buffer (1.2 N NaOH, 1 mM EDTA) in a 1.5 ml siliconized Eppendorf tube. The tube was heated at 90°C for 6'. The reaction was stopped by addition of 150 ul of 1 N acetic acid, 5 ul herring sperm carrier DNA (1 mg/ml), and 750 ul of 95% ETOH. The tube was incubated in a dry ice/ETOH bath for 5' and centrifuged at 12,000 x g for 10'. The supernatant was removed and the pellet was rinsed with 1 ml of 95% ETOH and lyophilized.

14E. Piperidine Cleavage of Base Modification Reactions

The lyophilized DNA pellets obtained from the base modification procedures outlined above were resuspended in 100 ul of 1 M piperidine (Fisher Scientific; Springfield, NJ; 99%, freshly diluted 1:10 in DDW). The reactions were heated at 90°C for 30' to allow for strand scission to go to completion. The 100 ul mixtures were transferred to fresh tubes and 100 ul of 0.6 M NaOAc and 2.5 volumes of

ETOH were added to precipitate the DNA. After centrifugation at 12,000 x g for 10', the pellets were rinsed with 95% ETOH, lyophilized, redissolved in 40 ul DDW, and dried again. The DNA pellets were resuspended in 10 ul of loading mix (80% (v/v) deionized formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) XC, 0.1% (w/v) BPB). The reactions were incubated at 90° C for 1', quick-chilled in ice water, and loaded immediately onto sequencing gels with a 10 ul Hamilton syringe.

15. Preparation of Oligonucleotide Primers for Dideoxy Sequencing

Influenza virus-specific deoxyribonucleotide primers, 14-20 bases in length, were synthesized using the Sam One automated DNA synthesizer (Biosearch, Inc.; San Rafael, CA). The two chemistries used for solid phase oligo synthesis are the phosphoramidite (Beaucage and Caruthers, 1981) and the phosphotriester (Itakura et al., 1975; Khorana, 1979) methods. The starting material for both synthetic routes is a solid support (silica or controlled pore glass) which has derivatized to it, the 3' base of the desired sequence. The nucleoside is attached to the support via its 3'-hydroxyl group. The 5'-hydroxyl group of the nucleoside is derivatized to a dimethoxytrityl (DMTr) protecting group. The DNA chain is elongated in the 3' to 5' direction.

15A. Phosphoramidite Method

The acid-labile DMTr group is cleaved from the 5'-hydroxyl group of the nucleoside bound to the solid support by treatment with dilute trichloroacetic acid in methylene chloride. In an acidic solution, the DMTr cation is a bright orange and effluents containing the DMTr cation are collected after each detritylation step, measured spectrophotometrically at 498 nm and compared to the previous detritylations to calculate a coupling efficiency. The next nucleoside phosphoramidite to be added is activated and coupled to the chain with tetrazole. A reactive phosphite is generated after the coupling which was oxidized to a more stable phosphate with a solution of iodine in tetrahydrofuran, pyridine, and water. At this time, any 5'-hydroxyl groups that did not react with the activated phosphoramidite in the previous coupling are modified by a capping procedure so they can not react in subsequent cycles. The capping protocol converts the unreacted 5'-hydroxyls to acetate esters with acetic anhydride and dimethylaminopyridine in pyridine and tetrahydrofuran. After cleavage of the second DMTr group, the coupling/oxidation/capping/cleavage cycle is continued until the desired oligonucleotide is generated.

15B. Phosphotriester Method

The phosphotriester method differs from the phosphoramidite method in the following respects: 1) the

condensation reaction is carried out in the presence of pyridine and N-methylimidazole; 2) the monomers used to extend the DNA chain are triethylammonium salts of protected phosphodiesters; 3) the coupling reagent is mesitylene-2-sulfonyl chloride (MSCl); 4) no oxidation step is required since condensation results in the formation of a stable triester linkage, not a reactive phosphite; and 5) the capping procedure is not employed.

15C. Processing of Oligonucleotide

After the final 5' DMTr group was cleaved from the oligo, the solid support was treated with 1 ml of concentrated NH_4OH at 55°C for 5 h to remove the oligo from the support and to cleave off the protecting groups from the bases. The NH_4OH supernatant was transferred to a fresh tube and lyophilized to dryness. The pellet was resuspended in 300 ul of 0.3 M NaOAc. The solution was extracted two times with an equal volume of n-butanol and the DNA was ethanol precipitated from the aqueous phase. The pellet was rinsed with 70% ETOH, dried, and dissolved in 200 ul of DDW. The DNA concentration was determined by reading absorbance at 260 nm. One $O.D._{260}$ unit of single-stranded oligonucleotide corresponds to a 35 ug/ml solution. To analyze the product of the DNA synthesis, 1 ug of the crude sample was 5' end-labeled with T4 PNK and [gamma-^{32}P]-ATP, electrophoresed in a 12% polyacrylamide

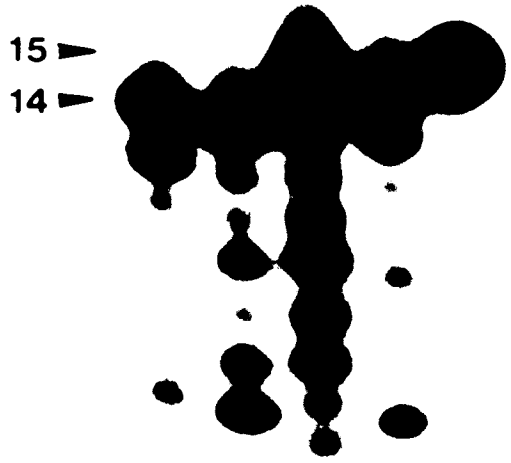
(acrylamide to bisacrylamide, 38:2)/TBE gel (26 x 16 x 0.5 cm, 1.5 mm thick) containing 7 M urea, and autoradiographed (Fig. 8).

15D. Large Scale Purification of Oligonucleotide

An equal volume of Maxam and Gilbert loading mix was added to a 150 ug aliquot of crude oligonucleotide. The sample was heated at 90°C for 2' and loaded into a 3 cm long slot of a 1.5 mm thick, 12% polyacrylamide/TBE gel (26 x 16 x 0.5 cm) containing 7 M urea. Electrophoresis was at 500 V until the BPB tracking dye was 3 cm from the gel bottom. The gel was removed from the glass plates and placed directly onto a thin layer chromatography cellulose sheet containing a fluorescent indicator (Eastman Kodak; Rochester, NY). The oligonucleotide band was visualized by UV shadowing using a shortwave 254 nm UV lamp. The band was excised and the DNA was eluted in 2 ml of 0.1 M NaCl at 37°C for 16 h with gentle rocking. The DNA was desalted over a NACS prepac ion exchange minicolumn (Bethesda Research Laboratories; Gaithersburg, MD). The oligo was eluted from the column in 400 ul of 1 M NaCl and 2.5 volumes of ethanol were added to precipitate the DNA. The DNA pellet was recovered by centrifugation at 12,000 x g for 10'. The pellet was rinsed with 70% ETOH, lyophilized, and dissolved in 50 ul of DDW. The concentration of the DNA was determined and the purified oligonucleotide

Fig. 8. Autoradiogram of synthetic oligonucleotides specific for the HA gene of influenza C virus. The oligonucleotides were 5' end-labeled with T4 PNK and [gamma-^{32}P]-ATP and electrophoresed in a 12% polyacrylamide gel containing 7 M urea. Lanes 1 and 2: crude preparation of two 14-mers; lanes 3 and 4: crude preparation of two 15-mers. Note that many smaller products of the DNA synthesis (n-1, n-2, etc.), differing by a single base in length are visible in lanes 1 through 4. Lane 5 shows a single species of 15 nucleotides obtained after gel purification and NACS column chromatography of the crude sample in lane 4.

1 2 3 4 5



(Fig. 8) was ready to be used in dideoxy sequencing reactions.

16. Dideoxy Sequencing of Influenza Virus RNAs

Many of the nucleotide sequences presented in this work were obtained using the dideoxy chain termination protocol (Sanger et al., 1977) with modifications. The Sanger method employs 2',3'-dideoxynucleoside triphosphate analogs of the normal deoxynucleoside triphosphates. The dideoxy analog is incorporated correctly into a growing DNA chain, but once incorporated, the DNA chain can not be further extended because the dideoxy nucleotide lacks the 3'-hydroxyl group necessary for chain extension.

To obtain influenza virus sequences, purified vRNA was used as template and an oligonucloetide specific for the gene of interest was used to prime cDNA synthesis using reverse transcriptase. The primer was annealed to the template and the mixture was divided into four dideoxy reactions. Each reaction contained all four deoxynucleoside triphosphates, one of which was alpha-^{32}P-labeled, and one of the four dideoxy analogs at a lower concentration. In the presence of reverse transcriptase, the primer was extended to generate a cDNA copy of the vRNA template. In each reaction, there was a base-specific partial incorporation of a terminating analog onto the 3' ends of the extending transcripts throughout the sequence. Parallel fractionation by gel electrophoresis of the size

ranges of terminated labeled transcripts from each reaction, each with a common 5' end of the primer, allowed a sequence to be deduced (Fig. 9).

Twenty uCi of [alpha-^{32}P]-dCTP (specific activity 3,000 Ci/mmole; New England Nuclear; Boston, MA) was aliquoted to the A reaction tube and 20 uCi of [alpha-^{32}P]-dATP was aliquoted to the G, T, and C reaction tubes. The radioactivity was lyophilized and 5 ul of the appropriate 2X cocktail was added to each tube. The 2X mixes were prepared as shown in Table 7.

Five ug of vRNA (from sucrose gradient purified virus) and 200 ng of NACS-purified oligonucleotide primer were combined in an Eppendorf tube. The volume was adjusted to 18 ul with sterile DDW. The mixture was heated at 90^{o}C for 3' then cooled to room temperature for 5'. Four ul of the above hybridization mix was added to each of the four reaction tubes. RT at 20 U/ul (Molecular Genetic Resources; Tampa, FL) was diluted to 4 U/ul in RT diluent (10 mM KH_2PO_4, pH 7.2, 2 mM DTT, 0.2% triton X-100, 10% glycerol) and 1 ul of the diluted enzyme was added to each tube to start the reaction. The reactions were incubated at 42^{o}C for 30' and then 1 ul of a cold dNTP chase (1 mM each of the four dNTPs in RT buffer) was added to each tube. The incubation was continued for 30'. Eleven ul of Sanger loading mix (80% (v/v) deionized formamide, 10 mM EDTA, pH 7.0, 0.1% XC, 0.1% BPB) was added per reaction.

Fig. 9. Autoradiogram of dideoxy sequencing reactions fractionated in an 8% polyacrylamide, 7 M urea, 90 x 20 x 0.04 cm sequencing gel. The dideoxy sequencing reactions are of the NS gene of three strains of influenza C virus (left to right: C/GL/54, C/JHG/66, C/CAL/78). The primer used was a synthetic 15-mer specific for the C virus NS gene. The same base-specific reactions for the three strains were electrophoresed side-by-side in order to facilitate the detection of base differences in the NS gene among the viruses. Arrowheads indicate positions 652 and 662 where nucleotide differences exist in the NS gene sequences.

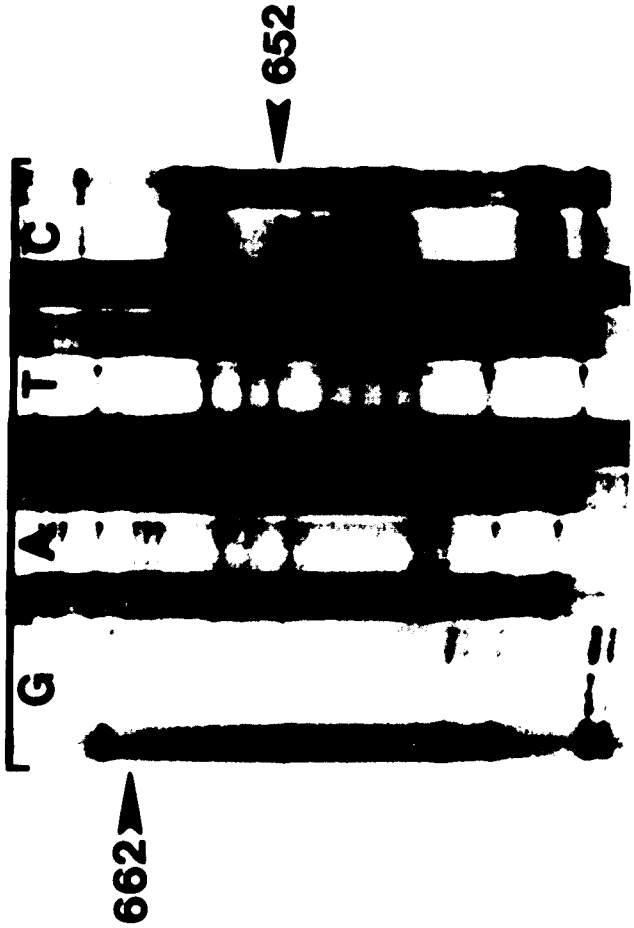

652
C
T
A
G
662

TABLE 7. 2X DIDEOXY REACTION MIXES

	A	G	C	T
RT Buffer	2X*	2X	2X	2X
dATP	200 uM	20 uM	20 uM	20 uM
dGTP	500 uM	200 uM	500 uM	500 uM
dCTP	20 uM	500 uM	200 uM	500 uM
dTTP	500 uM	500 uM	500 uM	200 uM
ddNTP	ddA,25 uM	ddG,25 uM	ddC,25 uM	ddT,25 uM
Label	20 uCi	20 uCi	20 uCi	20 uCi
alpha-^{32}P	dCTP	dATP	dATP	dATP

*2X RT Buffer (100 mM Tris-HCl, pH 8.3; 120 mM NaCl; 40 mM DTT; 24 mM $MgCl_2$)

The samples were boiled for 5', quick-chilled in ice water, and loaded onto the sequencing gel.

17. Sequencing Gels

All sequencing gel solutions were made up with a stock of 40% polyacrylamide with an acrylamide to bisacrylamide ratio of 38:2. Maxam and Gilbert cleavage reactions were electrophoresed in 20%, 7 M urea gels cast between 40 x 20 x 0.5 cm siliconized glass plates and 8%, 7 M urea gels cast between 90 x 20 x 0.5 cm siliconized glass plates. Mylar spacers were used of 0.4 mm thickness. For the long gel, 100 ml of solution was prepared containing 8% polyacrylamide, 7 M urea, and 0.06% AP in TBE. For the short gel, 50 ml of solution was prepared that was 20% polyacrylamide, 7 M urea, and 0.06% AP in TBE. Polymerization was initiated by the addition of 1 ul TEMED per ml of gel solution. The gel was cast with a 50 ml plastic syringe fitted with a 19 gauge, 1.5 inch needle. The sample slots were formed with a mylar comb of 20, 5 mm wide teeth, separated by 1 mm. Approximately 2 ul of each Maxam and Gilbert reaction for an end-labeled DNA was loaded with a 10 ul Hamilton syringe. Electrophoresis in TBE running buffer was at room temperature at 2100 V. Electrophoresis times varied, depending on the experiment, and were empirically determined. After electrophoresis, the gel was transferred to Whatman 3 MM chromatography paper,

mounted to a stiff support, and exposed to Cronex X-ray film at -70°C with a lightning plus intensifying screen (Dupont; Wilmington, DE).

The 20% short gels were used to determine sequence about 30 bases in from the labeled end of the fragment. In addition, multiple loadings on a long 8% gel of a set of reactions which were run for different lengths of time, allowed for the determination of sequence 250 bases in from the labeled terminus of the fragment.

Dideoxy sequencing reactions were electrophoresed in 0.4 mm thick, 6% polyacrylamide buffer gradient gels (Biggin et al., 1983) cast between 40 x 20 x 0.5 cm siliconized glass plates. Fifty ml of low concentration buffer gel solution was prepared consisting of 6% polyacrylamide, 7.7 M urea, 0.75X TBE, and 0.06% AP. Twenty-five ml of high concentration buffer gel solution was made up of 6% polyacrylamide, 7.7 M urea, 2.5X TBE, 0.06% AP with 2.5 g sucrose and a small amount of BPB powder. TEMED was added to start polymerization. Four ml of 0.75X gel solution was gently drawn up in a 25 ml pipet, followed by 6 ml of 2.5X gel solution. The two solutions were mixed to form a gradient by drawing up a few air bubbles through the pipet tip. The gradient was dispensed into the gel mold and the remainder of the gel was poured with 0.75X gel solution through a 50 ml syringe. Sample wells were formed with a shark's tooth comb and 1.5 ul of

sample per well was applied to the gel. The gel was run with 1X TBE buffer at 1200 V until the BPB dye was at the bottom. The gel was fixed in 10% acetic acid, 10% methanol for 10' and transferred to Whatman 3 MM paper. The gel was dried and autoradiographed at -70°C for 16 h. Usually, the first 200 bases from the oligonucleotide primer were read from the 6% buffer gradient gel. To determine nucleotide sequences beyond this point, the samples were electrophoresed for different lengths of time in 6% polyacrylamide, 7.7 M urea gels without buffer gradients.

18. Computer Analysis of Nucleotide Sequences

The complete nucleotide and predicted amino acid sequences of influenza virus genes were stored and analyzed on an IBM 370 computer at the University Computing Center of the City University of New York using published programs (Staden, 1977, 1978).

19. Maximum Parsimony Analysis of Nucleotide Sequences

Influenza virus gene sequences were analyzed by the maximum parsimony procedure (Fitch, 1971). Given a set of nucleotide sequences of homologous genes, one can determine a best fitting evolutionary tree describing their descent from a common ancestral gene. The best fitting or most parsimonious tree generated by this procedure contains the minimum number of nucleotide replacements necessary to account for the evolutionary relationship among the genes.

The computer program is available upon request from Dr. Fitch.

Upon comparison of nucleotide sequences of homologous genes (taxa), a number of different character states can be described for particular nucleotide positions. A nucleotide position which has the same base or a single character state for all the genes, does not give any information on evolutionary descent since no base substitutions are required. A singularity is defined as a nucleotide represented in only one gene in a particular nucleotide position. Singularities are unique base changes in a sequence which are represented as the side branch distance of a taxa on the constructed evolutionary tree. Singularities are uninformative in that they do not affect the construction of the most parsimonious tree. Having removed all uninformative nucleotide positions, what is left are those positions that affect the structure of the most parsimonious tree. The informative positions are called discordancies. A discordancy is the presence in at least two taxa of different, nonsingular nucleotides at a position. Theoretically, a most parsimonious solution is determined for each position in the sequences where discordancies exist among taxa.

The pair of taxa that are initially joined to construct the tree is usually the two sequences which are most similar or have the fewest number of discordancies. All

nucleotides which could possibly have existed in the immediate ancestor sequence of the joined taxa are inferred through a computer algorithm. The possible ancestral nucleotides have to be consistent with a minimum of base replacements necessary to account for the observed data. The remaining taxa are then successively linked to the first joined pair and all possible ancestral sequences at each node in the line of descent are determined. Even within the framework of the permitted nucleotides at a node, not all nucleotide replacements (linkages) are allowable in descending from node to node, in generating a most parsimonious tree. Of the valid linkages between nodes, not all substitutions are equally probable, therefore, probabilities have to be assigned to specific linkages. There can be more than one parsimonious solution for a given set of data, so that the final evolutionary tree is a weighted average of all most parsimonious solutions.

RESULTS

1. Evolution of Human Influenza A Viruses

1A. Nucleotide Sequences of NS Genes of Human Influenza A Viruses

Variation of influenza A viruses was examined by comparison of nucleotide sequences of the NS gene (890 bases) of 15 human viruses isolated over 53 years (1933-1985). The viruses represent all three human hemagglutinin serotypes (H1, H2, and H3) and except for the three Houston isolates, the strains were obtained from diverse geographical locations. Fig. 10 shows the nucleotide sequences of the NS genes of 15 human influenza A strains. The 15 sequences are easily aligned for analysis because of the size conservation of the NS gene segment of 890 bases. In addition, the 5' and 3' putative splicing signals (indicated by arrowheads in Fig. 10) required to generate spliced NS2 mRNA are conserved in all the strains examined. Nucleotide substitutions occur at 149 positions scattered throughout the NS gene and usually, once a base change is observed in a virus isolate, it is found in subsequent strains. The occurrence of sequential nucleotide changes in the NS gene suggests an evolutionary relationship among the NS genes of the 15 strains.

The cumulative nature of the NS nucleotide changes is further illustrated when pairwise comparisons among the

Fig. 10. Nucleotide sequences of 15 NS genes of human influenza A viruses. The NS sequences of the A/WSN/33 (sequenced by J. Parvin), A/Bellamy/42, A/Maryland/2/80, A/Houston/18515/84, A/Houston/23284/85, A/Denver/1/57, A/Ann Arbor/6/60, A/Berkeley/1/68, and A/Houston/24269/85 viruses were determined by the dideoxy chain terminator method. The sequencing primers correspond to the following positions of the NS gene: 10-29; 133-147; 240-255; 293-312; 344-359; 518-536; 593-612; 616-632; and 739-755. The nucleotide sequences of the A/FM/1/47, A/FW/1/50, A/USSR/90/77, and A/ALA/6/77 NS genes were determined by the Maxam and Gilbert chemical method from cloned cDNAs. The A/USSR/90/77 NS gene was cloned by M. Baez and sequenced by M. Krystal. The sequences of the A/PR/8/34 and A/Udorn/72 virus NS genes were previously published (Baez et al., 1980; Lamb and Lai, 1980).

The NS sequence of the oldest isolate, A/WSN/33, is shown in its entirety and serves as the reference sequence. Only nucleotide positions of the other strains that differ from the reference sequence are listed. The NS sequences of the A/Houston/18515/84 and A/Houston/23284/85 viruses are identical. Asterisks represent nucleotide positions that could not be determined because the first primer used for sequencing corresponds to positions 10-29 of the NS gene. The 5' terminal nucleotides of the A/WSN/33 NS gene were determined by dideoxy sequencing with an M13 clone and

Fig. 10 continued.

the universal primer. The ATG initiation codon for NS1 and NS2 protein translation is at positions 27-29. The 5' (nucleotide 56) and 3' (nucleotide 529) putative splice sites are indicated by arrowheads.

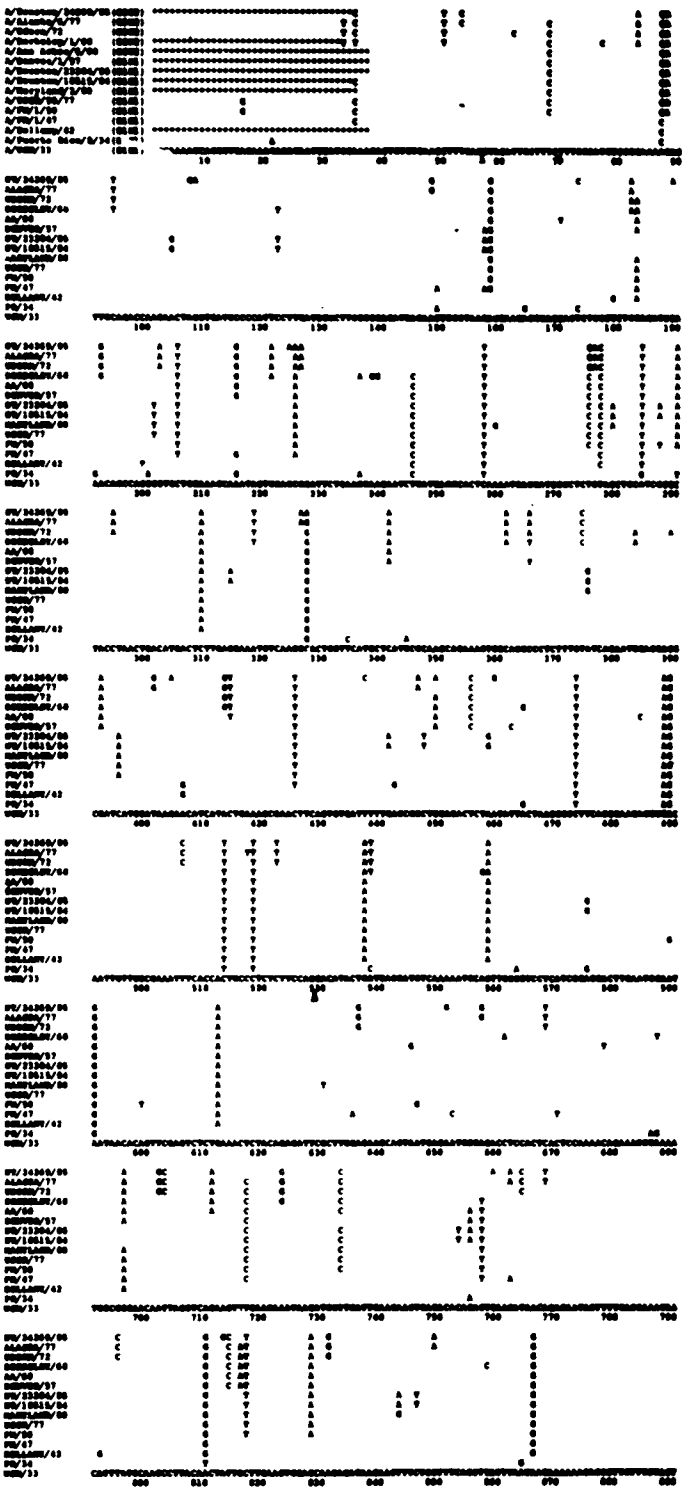


TABLE 8. NUCLEOTIDE DIFFERENCES BETWEEN INFLUENZA A VIRUS NS GENES

	PR/34	BEL/42	FM/47	FW/50	DEN/57	AA/60	BERK/68	UDORN/72	ALA/77	(H3)HT/85
WSN/33	30(3.5)*	24(2.8)	35(4.1)	38(4.5)	43(5.0)	47(5.5)	66(7.7)	69(8.1)	74(8.7)	82 (9.6)
PR/34		32(3.8)	39(4.6)	45(5.3)	46(5.4)	50(5.9)	65(7.6)	75(8.8)	80(9.4)	87(10.2)
BEL/42			17(2.0)	22(2.6)	27(3.2)	31(3.6)	50(5.9)	55(6.4)	62(7.3)	71 (8.3)
FM/47				21(2.5)	22(2.6)	28(3.3)	47(5.5)	54(6.3)	59(6.9)	70 (8.2)
FW/50					17(2.0)	19(2.2)	38(4.5)	45(5.3)	52(6.1)	65 (7.6)
DEN/57						10(1.2)	29(3.4)	37(4.3)	46(5.4)	59 (6.9)
AA/60							29(3.4)	36(4.2)	45(5.3)	58 (6.8)
BERK/68								28(3.3)	39(4.6)	52 (6.1)
UDORN/72									15(1.8)	28 (3.3)
ALA/77										17 (2.0)

NUCLEOTIDES 1 TO 37 ARE EXCLUDED FROM ANALYSIS.

*NUMBER OF NUCLEOTIDE DIFFERENCES BETWEEN TWO NS GENES (% DIFFERENCE).

viruses are made (Table 8). The per cent nucleotide difference among NS genes increases as strains isolated farther apart in time are compared. The most dissimilar NS genes (10.2 % difference) are those of the A/PR/34 and the (H3) A/HT/85 strains which were isolated 51 years apart. The most closely related NS sequences (no differences) are derived from the A/HT/84 and the (H1) A/HT/85 viruses (Table 9).

Table 9 shows the nucleotide differences among the NS genes of the H1N1 A/FW/50 virus and the reemerged H1N1 strains isolated in 1977-1984. It is quite striking from Table 9 that the NS genes of the A/FW/50 and the A/USSR/77 strains only possess five nucleotide differences, although these viruses were isolated 27 years apart. The nucleotide sequence data clearly indicate a close relationship between 1950 H1N1 viruses and H1N1 strains reintroduced in 1977 as described previously. Consistent with the data of Table 8, nucleotide changes in the NS genes of the "new" H1N1 strains accumulate over time (Table 9).

The nucleotide sequences shown in Fig. 10 reveal that the changes in the NS genes of the A/FW/50 and A/DEN/57 strains are distinct from those found in the NS genes of the A/USSR/77, A/MD/80, and A/HT/84 H1N1 viruses. These data suggest that the H1N1 subtype has evolved in different directions during its two periods of circulation in man.

TABLE 9. NUCLEOTIDE DIFFERENCES BETWEEN H1N1 INFLUENZA A VIRUS NS GENES

	USSR/77	MD/80	HT/84**
FW/50	5(0.6)*	10(1.2)	21(2.5)
USSR/77		5(0.6)	17(2.0)
MD/80			17(2.0)

Nucleotides 1-37 are excluded from analysis

*Number of nucleotide differences between two NS genes (% difference)

**HT/84 and (H1) HT/85 have identical NS sequences

1B. Predicted Amino Acid Sequences of the NS1 and NS2 Proteins

A comparison of the NS1 protein sequences deduced from the nucleotide sequences (Fig. 10) is presented in Fig. 11. Amino acid substitutions are detected throughout the NS1 molecule, and many of the changes found in earlier isolates are retained in subsequent strains. Pairwise comparison of the NS1 amino acid sequences of the strains (Table 10) reveals an accumulation of substitutions over time, similar to the observed nucleotide data in Table 8. Amino acid differences among NS1 proteins range from 1.0-12.1% over the 53 year period in which the viruses were isolated (Table 10). Amino acid differences among the NS1 proteins of the A/FW/50 H1N1 virus and the new H1N1 strains are presented in Table 11. The close relationship between the A/FW/50 virus and the A/USSR/77 isolate is further evidenced by the fact that the NS1 proteins of these strains differ at only three positions.

It is apparent from Fig. 11 that there is considerable heterogeneity in the predicted amino acid length of NS1 polypeptides of different strains, as summarized in Table 12. The size variability is confined to the carboxy terminus of the NS1 protein. The different NS1 carboxy termini observed in the strains are the result of point mutations in the NS gene (within the coding region of NS1) which lead to termination codons for protein translation in

Fig. 11. Comparison of deduced amino acid sequences of the NS1 protein of 15 human influenza A viruses. The amino acid sequence of the NS1 protein of the A/WSN/33 isolate is presented in its entirety and serves as the reference sequence. Amino acid substitutions in the other viruses, relative to the A/WSN/33 sequence are indicated. =, represents amino acids which could not by predicted because the nucleotide sequence in the region was not determined. *, represents amino acids not found in the NS1 protein of a particular strain (NS1 size heterogeneity).

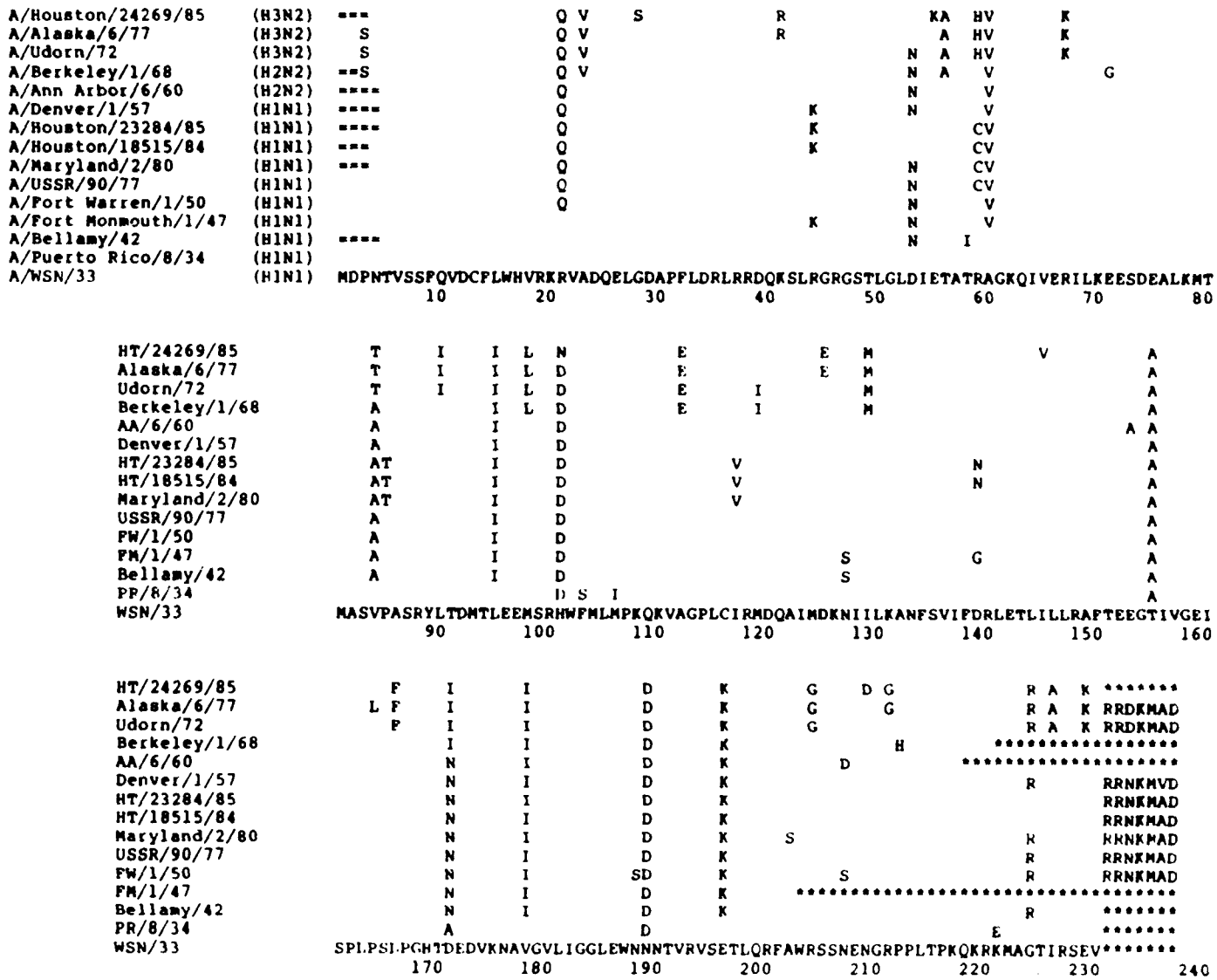

A/Houston/24269/85 (H3N2)
A/Alaska/6/77 (H3N2)
A/Udorn/72 (H3N2)
A/Berkeley/1/68 (H2N2)
A/Ann Arbor/6/60 (H2N2)
A/Denver/1/57 (H1N1)
A/Houston/23284/85 (H1N1)
A/Houston/18515/84 (H1N1)
A/Maryland/2/80 (H1N1)
A/USSR/90/77 (H1N1)
A/Port Warren/1/50 (H1N1)
A/Fort Monmouth/1/47 (H1N1)
A/Bellamy/42 (H1N1)
A/Puerto Rico/8/34 (H1N1)
A/WSN/33 (H1N1)
MDPNTVSSFQVDCFLWHVRKRVADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRAGKQIVERILKEESDEALKMT
10 20 30 40 50 60 70 80
HT/24269/85
Alaska/6/77
Udorn/72
Berkeley/1/68
AA/6/60
Denver/1/57
HT/23284/85
HT/18515/84
Maryland/2/80
USSR/90/77
FW/1/50
FM/1/47
Bellamy/42
PR/8/34
WSN/33
MASVPASRYLTDMTLEEMSRHWFMLMPKQKVAGPLCIRMDQAIMDKNIILKANFSVIFDRLETLILLRAFTEEGTIVGEI
90 100 110 120 130 140 150 160
HT/24269/85
Alaska/6/77
Udorn/72
Berkeley/1/68
AA/6/60
Denver/1/57
HT/23284/85
HT/18515/84
Maryland/2/80
USSR/90/77
FW/1/50
FM/1/47
Bellamy/42
PR/8/34
WSN/33
SPLPSLPGHTDEDVKNAVGVLIGGLEWNNNTVRVSETLQRFAWRSSNENGRPPLTPKQKRKMAGTIRSEV*******
170 180 190 200 210 220 230 240

TABLE 10. AMINO ACID DIFFERENCES BETWEEN INFLUENZA A VIRUS NS1 PROTEINS

	PR/34	BEL/42	FM/47	FW/50	DEN/57	AA/60	BERK/68	UDORN/72	ALA/77	(H3) HT/85
WSN/33	6(3.0)*	11(5.6)	13(6.6)	12(6.1)	12(6.1)	12(6.1)	18(9.1)	21(10.7)	22(11.1)	24(12.1)
PR/34		10(5.1)	12(6.1)	11(5.6)	11(5.6)	11(5.6)	17(8.6)	20(10.1)	21(10.7)	23(11.7)
BEL/42			4(2.0)	5(2.5)	5(2.5)	5(2.5)	12(6.1)	16(8.1)	19(9.6)	21(10.7)
FM/47				5(2.5)	3(1.5)	5(2.5)	12(6.1)	16(8.1)	19(9.6)	21(10.7)
FW/50					2(1.0)	2(1.0)	9(4.5)	13(6.6)	16(8.1)	18(9.1)
DEN/57						2(1.0)	9(4.5)	13(6.6)	16(8.1)	18(9.1)
AA/60							9(4.5)	13(6.6)	16(8.1)	18(9.1)
BERK/68								6(3.0)	11(5.6)	13(6.6)
UDORN/72									5(2.5)	7(3.5)
ALA/77										4(2.0)

198 Amino acids compared; residues 1-4 and those beyond amino acid 202 are excluded from analysis

*Number of amino acid differences between two NS1 proteins (% difference)

TABLE 11. AMINO ACID DIFFERENCES BETWEEN H1N1 INFLUENZA A VIRUS NS1 PROTEINS

	USSR/77	MD/80	HT/84**
FW/50	3(1.3)*	6(2.6)	9(3.9)
USSR/77		3(1.3)	6(2.6)
MD/80			5(2.1)

233 Amino acids compared; residues 1-4 are excluded from analysis

*Number of amino acid differences between two NS1 proteins (% difference)

**HT/84 and (H1) HT/85 have identical NS1 amino acid sequences

TABLE 12. LENGTHS OF NS1 PROTEINS OF DIFFERENT HUMAN INFLUENZA VIRUSES

Strain	NS1 Amino Acids
A/FM/1/47	202
A/AA/6/60	217
A/BERK/1/68	220
A/WSN/33	230
A/PR/8/34	230
A/BEL/42	230
A/HT/24269/85	230
A/FW/1/50	237
A/USSR/90/77	237
A/MD/2/80	237
A/HT/18515/84	237
A/HT/23284/85	237
A/DEN/1/57	237
A/UDORN/72	237
A/ALA/6/77	237

the NS1 mRNA. The predicted sizes of the NS1 polypeptides are in good agreement with their observed mobilities in SDS-polyacrylamide gels (Fig. 12). The smallest NS1 protein (202 amino acids) of the A/FM/47 strain has the greatest electrophoretic mobility. Similarly, if the virus-infected cell extracts are electrophoresed in SDS-polyacrylamide gels containing urea which eliminates protein secondary structure, the migrational heterogeneity of the NS1 protein is still observed (Fig. 13). It should be noted that the 237 amino acid NS1 protein of the A/USSR/77 strain has a greater mobility in this gel system (Fig. 12 and 13) than NS1 polypeptides of other strains which also consist of 237 amino acids.

The size variability of the NS1 protein in field strains suggests that the NS1 can tolerate large deletions at its carboxy terminus without apparently compromising its functional integrity in the virus replicative cycle. Recently, Norton et al. (submitted) have reported that the NS1 protein of an infectious avian influenza A isolate, A/turkey/Oregon/71, contains only 124 N-terminal amino acids.

Figure 14 shows an amino acid comparison of the NS2 proteins of the influenza A viruses. The first ten amino acids at the N-terminus of the NS2 protein are the same residues found at the N-terminus of the NS1 protein. There are 17 positions in the NS2 where amino acid substitutions

Fig. 12. Electrophoresis of influenza A virus-infected cell lysates in a 7-14% gradient SDS-polyacrylamide gel. Confluent MDCK cell monolayers were infected with virus and pulse-labeled with [^{35}S]-methionine at 33^{o}C for 1 h at 8 h p.i. as described in the Materials and Methods. Lanes: 1, A/FM/47 virus-infected cells; 2, A/AA/60 virus-infected cells; 3, A/BERK/68 virus-infected cells; 4, A/WSN/33 virus-infected cells; 5, A/PR/34 virus-infected cells; 6, A/BEL/42 virus-infected cells; 7, (H3) A/HT/85 virus-infected cells; 8, A/FW/50 virus-infected cells; 9, A/USSR/77 virus-infected cells; 10, A/MD/80 virus-infected cells; 11, A/HT/84 virus-infected cells; 12, (H1) A/HT/85 virus-infected cells; 13, A/DEN/57 virus-infected cells; 14, A/UDORN/72 virus-infected cells; 15, A/ALA/77 virus-infected cells; 16, mock-infected cells. Virus specific proteins are indicated.

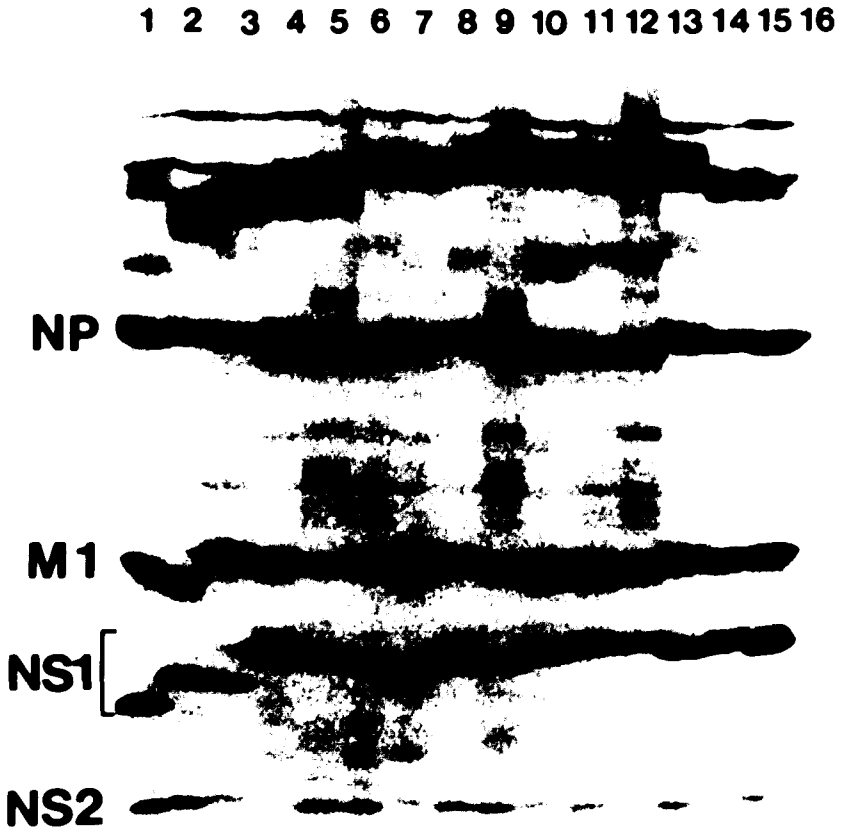

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
NP
M1
NS1
NS2

Fig. 13. Electrophoresis of influenza A virus-infected cell lysates in a 7-14% gradient SDS-polyacrylamide gel containing 7 M urea. Confluent MDCK cell monolayers were infected with virus and pulse-labeled with [^{35}S]-methionine at 33^{o}C for 1 h at 8 h p.i. as described in Materials and Methods. Lanes: 1, A/FM/47 virus-infected cells; 2, A/AA/60 virus-infected cells; 3, A/BERK/68 virus-infected cells; 4, A/WSN/33 virus-infected cells; 5, A/PR/34 virus-infected cells; 6, A/BEL/42 virus-infected cells; 7, (H3) A/HT/85 virus-infected cells; 8, A/FW/50 virus-infected cells; 9, A/USSR/77 virus-infected cells; 10, A/MD/80 virus-infected cells; 11, A/HT/84 virus-infected cells; 12, (H1) A/HT/85 virus-infected cells; 13, A/DEN/57 virus-infected cells; 14, A/UDORN/72 virus-infected cells; 15, A/ALA/77 virus-infected cells; 16, mock-infected cells. Virus-specific proteins are indicated.

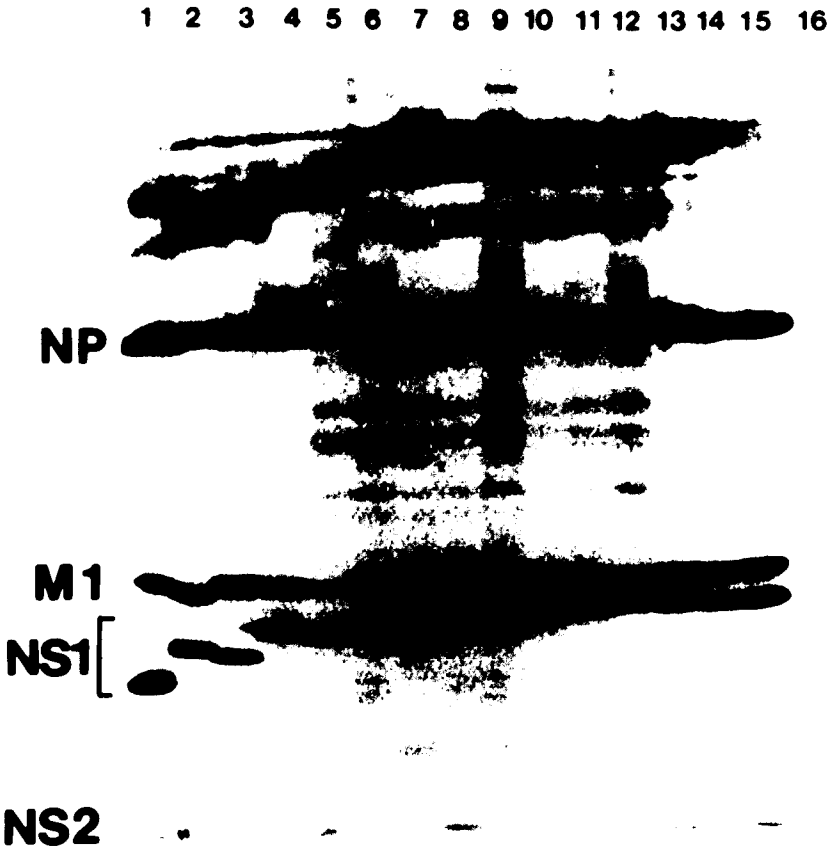
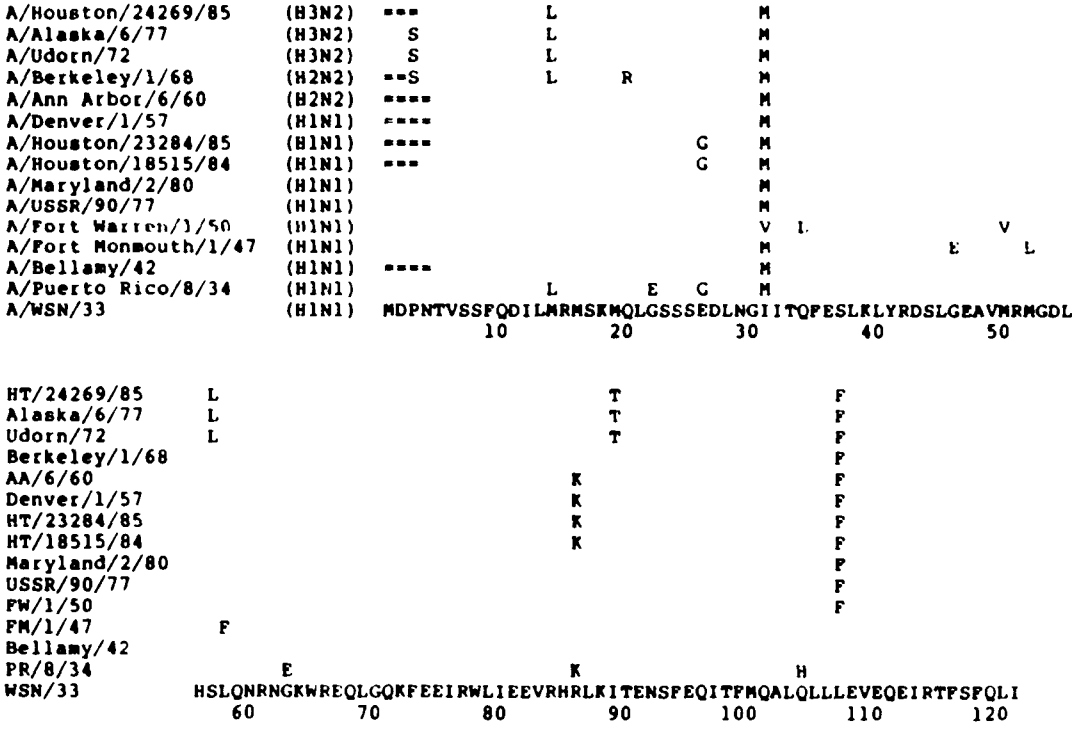

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
NP
M1
NS1
NS2

Fig. 14. Comparison of deduced amino acid sequences of the NS2 proteins of 15 human influenza A viruses. The A/WSN/33 NS2 sequence is the reference sequence and amino acid substitutions in the NS2 proteins of the other strains, relative to A/WSN/33, are indicated. =, represents amino acids which could not be predicted because the nucleotide sequence in the region was not determined. The first ten residues at the N-terminus of the NS2 protein are the same as those located at the N-terminus of the NS1 protein (Fig. 11).


A/Houston/24269/85 (H3N2) --- L M
A/Alaska/6/77 (H3N2) S L M
A/Udorn/72 (H3N2) S L M
A/Berkeley/1/68 (H2N2) --S L R M
A/Ann Arbor/6/60 (H2N2) ---- M
A/Denver/1/57 (H1N1) ---- M
A/Houston/23284/85 (H1N1) ---- G M
A/Houston/18515/84 (H1N1) --- G M
A/Maryland/2/80 (H1N1) M
A/USSR/90/77 (H1N1) M
A/Fort Warren/1/50 (H1N1) V L V
A/Fort Monmouth/1/47 (H1N1) M E L
A/Bellamy/42 (H1N1) ---- M
A/Puerto Rico/8/34 (H1N1) L E G M
A/WSN/33 (H1N1) MDPNTVSSFQDILMRMSKMQLGSSSEDLNGIITQFESLKLYRDSLGEAVMRMGDL
10 20 30 40 50
HT/24269/85 L T F
Alaska/6/77 L T F
Udorn/72 L T F
Berkeley/1/68 F
AA/6/60 K F
Denver/1/57 K F
HT/23284/85 K F
HT/18515/84 K F
Maryland/2/80 F
USSR/90/77 F
FW/1/50 F
FM/1/47 F
Bellamy/42
PR/8/34 E K H
WSN/33 HSLQNRNGKWREQLGQKFEEIRWLIEEVRHRLKITENSFEQITFMQALQLLLEVEQEIRTFSFQLI
60 70 80 90 100 110 120

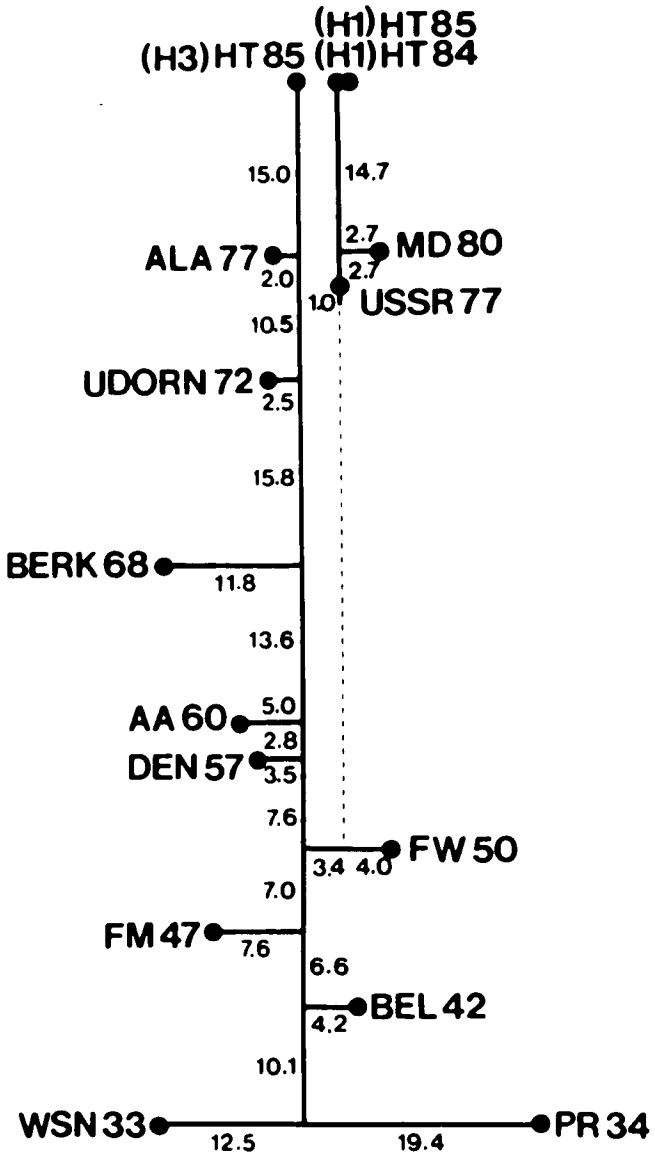
are detected. In contrast to the NS1 protein, all NS2 proteins examined have a deduced length of 121 amino acids. The conserved size of the NS2 polypeptides among strains is apparent from similar mobilities of the NS2 proteins in SDS-polyacrylamide gels without (Fig. 12) and with urea (Fig. 13).

The nucleotide sequences corresponding to the coding region of the NS1 protein (711 bases) were analyzed for silent and amino acid changing base substitutions as compared to the A/WSN/33 NS sequence. 9.8% (70/711) of the positions display silent base substitutions whereas, at 7.9% (56/711) of the base positions there are nucleotide replacements which result in amino acid changes in the sequence of the NS1 protein. The ratio of silent to nonsilent substitutions is 1.3:1.

1C. Evolutionary Analysis of NS Genes by Maximum Parsimony

The sequence information as presented in Fig. 10 was analyzed by maximum parsimony (Fitch, 1971) to determine the phylogenetic tree of minimum length. The best tree found contains a total of 186 substitutions and is illustrated in Fig. 15. The parsimony method also yielded four alternative trees containing 187 substitutions. These alternative trees contain only minor branch perturbations of the best tree (data not shown). The numbers defining genetic distances in the evolutionary tree in Fig. 15

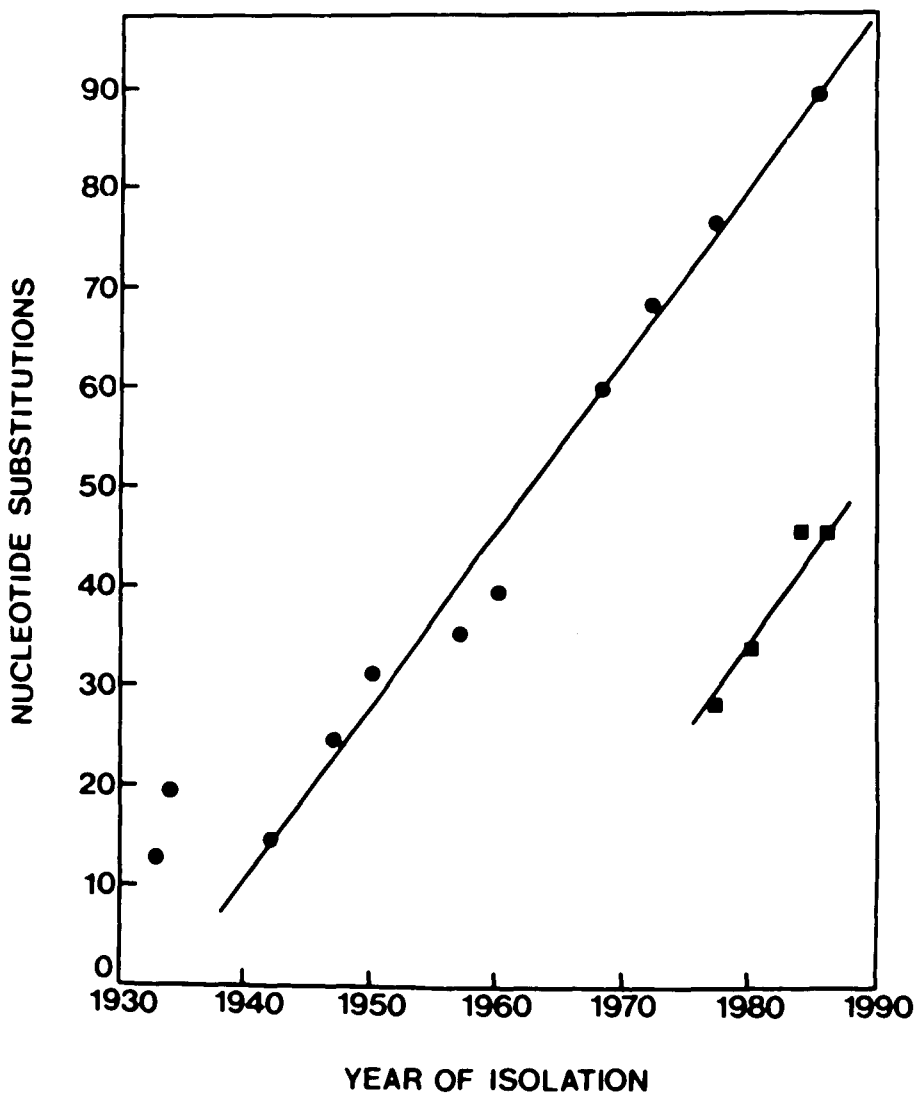
Fig. 15. Most parsimonious evolutionary tree for 15 influenza A virus NS genes containing 186 substitutions. The nucleotide sequences shown in Fig. 10 were analyzed by the method of Fitch (1971). The length of the trunk and side branches of the evolutionary tree are proportional to the minimal number of substitutions required to account for the differences in sequence. Non-integral numbers arise from averaging over all possible minimal solutions. The broken line represents the predicted number of additional substitutions between the NS genes of FW/50 and USSR/77 based on the calculated evolutionary rate (see text for discussion).


(H1)HT85
(H3)HT85
(H1)HT84
15.0
14.7
2.7
ALA77
MD80
2.0
2.7
1.0
USSR77
10.5
UDORN72
2.5
15.8
BERK68
11.8
13.6
AA60
5.0
2.8
DEN57
3.5
7.6
FW50
3.4
4.0
7.0
FM47
7.6
6.6
BEL42
4.2
10.1
WSN33
PR34
12.5
19.4

represent average values calculated by the maximum parsimony method using all the NS sequences. Therefore, some values in Fig. 15 do not exactly match the pairwise comparison data presented in Tables 8 and 9.

Figure 16 shows the number of nucleotide substitutions between the origin of the best tree (formed at the point where the A/WSN/33 and A/PR/34 branches meet) and the tip of each branch (Fig. 15) plotted against the date of isolation of the viruses whose NS gene is represented by that tip. The major line, derived by linear regression analysis, shows that these sequences (solid circles) are evolving at the steady rate of 1.73 +/- 0.08 substitutions per year or 1.94 +/- 0.09 x 10^{-3} substitutions per nucleotide site per year. The A/WSN/33 and A/PR/34 strains appear to have more substitutions per year than expected and therefore, were excluded from the evolutionary rate calculation. Since these strains were isolated before refrigeration became available in the laboratory, we believe that continuous passaging in animal hosts and in embryonated eggs (particularly in the first 10-15 years following isolation of the strains) may have introduced additional mutations not present in the original isolates. Fig. 16 also shows that the group of H1N1 subtype strains (solid squares), which reemerged in the human population in 1977 after a 27 year absence (Nakajima et al., 1978) is evolving at the same rate. These new H1N1 viruses have

Fig. 16. Linearity with time of number of substitutions in the NS genes of influenza A viruses. The abscissa represents the year of isolation of the influenza A viruses used in the analysis. The ordinate indicates the number of substitutions observed in their NS genes between the first branching point formed by the WSN/33 and PR/34 sequences in Fig. 15 and the tips of all branches of the evolutionary tree. A line, generated by linear regression analysis, is drawn through the points. The slope of the line is 1.73 +/- 0.08 substitutions per year. In addition to the sequences found on the trunk of the evolutionary tree (solid circles), the NS genes of the four new H1N1 viruses are also represented in this graph (solid squares). The line through the squares is arbitrarily drawn parallel to the other line.


NUCLEOTIDE SUBSTITUTIONS
0
10
20
30
40
50
60
70
80
90
1930
1940
1950
1960
1970
1980
1990
YEAR OF ISOLATION

been cocirculating with the H3N2 viruses since 1977 and form a separate evolutionary branch (Fig. 15). In reality, the H1N1 branch should be directly connected to the A/FW/50 branch of the main tree, since there are only five nucleotide differences between the A/FW/50 and A/USSR/77 virus NS genes. However, the viruses were isolated 27 years apart and based on the calculated evolutionary rate of 1.73 substitutions per year, we would predict approximately 46 additional substitutions in the NS gene of A/USSR/77 (represented by the broken line in Fig. 15). The observed data, thus suggest a unique epidemiology of the new H1N1 isolates.

Two points can be made following the analysis of the data. First, the uniform rate of nucleotide substitution in the NS gene suggests that the NS gene of influenza A viruses behaves as an accurate molecular clock in nature. Calibration of the molecular clock is not affected by inaccurate paleontological dates since the time of fossilization (isolation) of these strains is recorded. Thus, given only the NS gene sequence of a main line isolate, one can closely estimate the year of its isolation (Fig. 16). Although fewer points are available for measuring the rate in the new H1N1 strains (1977-1985), the data (Fig. 16, solid squares) are compatible with a molecular clock ticking at the same evolutionary rate for these NS genes. Second, the evolutionary rate of 1.94 +/-

0.09 x 10^{-3} substitutions per site per year appears to be approximately 10^{6}-fold higher than the rate for mammalian germ-line genes (Li and Gojobori, 1983). (The difference calculated on a per generation basis would be much less.)

1D. Evolutionary Analysis of H3 HA and N2 NA Genes by Maximum Parsimony

To determine an evolutionary rate for the HA gene, 16 previously published complete HA1 sequences of viruses of the H3N2 subtype (1968-1979) were examined by maximum parsimony (Verhoeyen et al., 1980; Sleigh et al., 1981; Hauptmann et al., 1983; Both et al., 1983; Newton et al., 1983). The most parsimonious tree for the 16 HA1 genes (1068 bases analyzed) contains 184 nucleotide substitutions and is illustrated in Fig. 17. Figure 18 shows the number of nucleotide substitutions between the origin of the best tree (formed at the point where NT68 and Ai68 are joined) and the tip of each branch (Fig. 17) plotted against the date of isolation of the viruses whose HA gene is represented by that tip. The slope of the line, derived by linear regression, indicates that 7.18 +/- 0.22 substitutions are occurring in the HA1 gene per year or 6.72 +/- 0.21 x 10^{-3} substitutions per nucleotide site per year. This rapid evolutionary rate of the HA1 gene is approximately three times higher than the calculated evolutionary rate for the NS gene (Fig. 16). In addition, the uniform rate of change in the HA1 gene over time

Fig. 17. Most parsimonious evolutionary tree for 16 influenza A virus H3 subtype HA1 genes. The HA nucleotide sequences have been published previously (Verhoeyen et al., 1980; Sleigh et al., 1981; Hauptmann et al., 1983; Both et al., 1983; Newton et al., 1983). The sequences were analyzed by the method of Fitch (1971). The length of the trunk and the side branches of the evolutionary tree are proportional to the minimal number of substitutions required to account for sequence differences. Fractional substitutions arise from averaging over all most parsimonious solutions. The length of the tree is 184 substitutions.

The following abbreviations (NT68, Ai68, HK71, ENG69, QU70, MEM71, ENG72, MEM72, PC73, ViC75, ViC76, AC76, TEX77, ENG77, BK179, BK279) are used for influenza viruses A/Northern Territory/60/68, A/Aichi/2/68, A/Hong Kong/107/71, A/England/878/69, A/Queensland/7/70, A/Memphis/1/71, A/England/42/72, A/Memphis/102/72, A/Port Chalmers/1/73, A/Victoria/3/75, A/Victoria/112/76, A/Allegheny County/29/76, A/Texas/1/77, A/England/321/77, A/Bangkok/1/79, A/Bangkok/2/79, respectively.

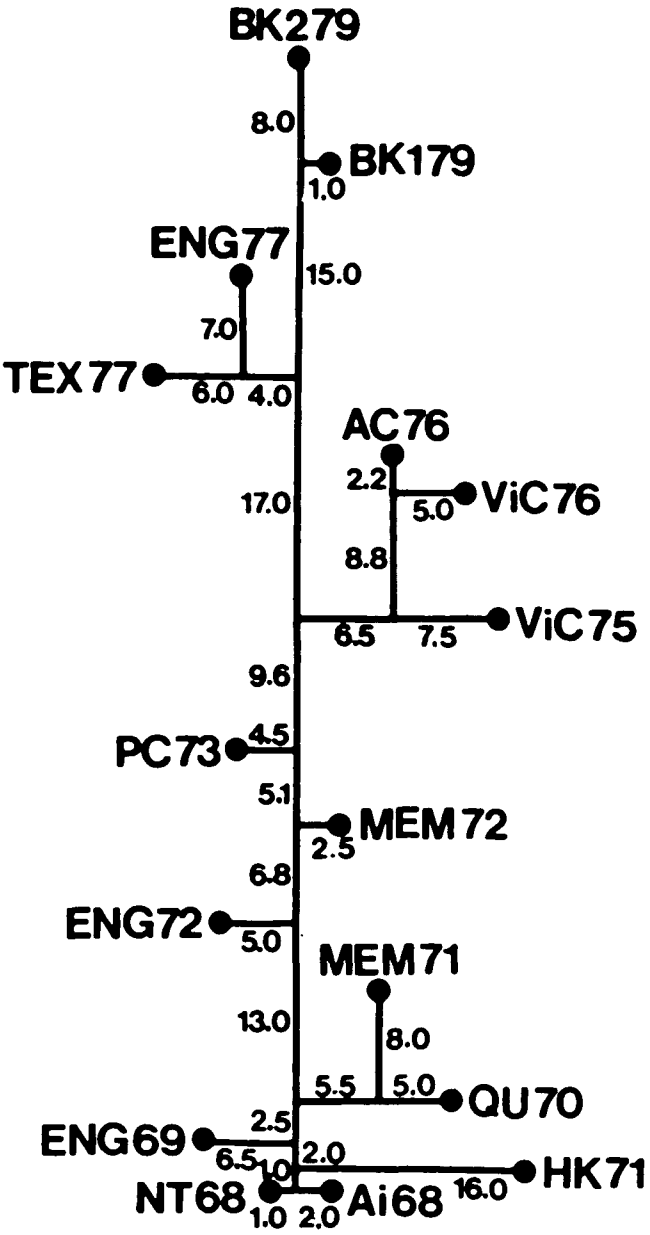
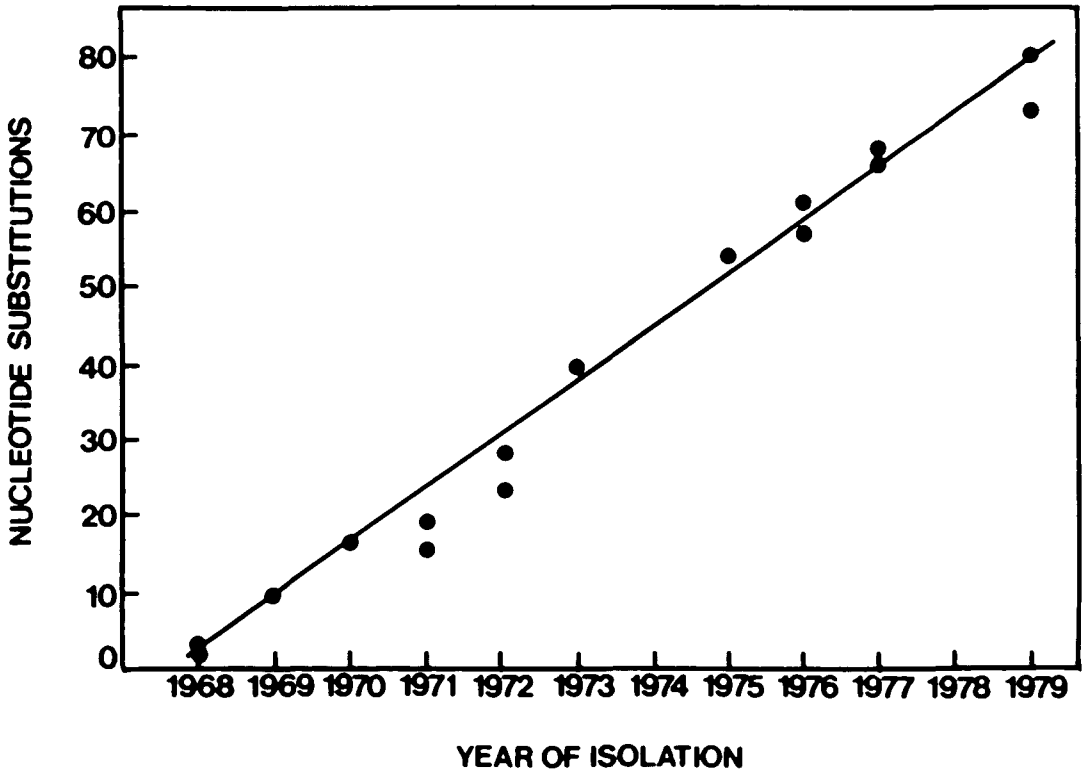

BK279
8.0
BK179
1.0
ENG77
15.0
7.0
TEX77
6.0
4.0
AC76
2.2
ViC76
17.0
5.0
8.8
ViC75
6.5
7.5
9.6
PC73
4.5
5.1
MEM72
2.5
6.8
ENG72
5.0
MEM71
13.0
8.0
5.5
5.0
QU70
2.5
ENG69
6.5
1.0
2.0
HK71
NT68
1.0
2.0
Ai68
16.0

Fig. 18. Linearity with time of number of substitutions in the H3 HA1 genes of influenza A viruses. The abscissa represents the year of isolation of the influenza A viruses used in the analysis. The ordinate indicates the number of substitutions observed in their HA1 genes between the first branching point formed by the NT68 and Ai68 sequences in Fig. 17 and the tips of all branches of the evolutionary tree. A line, generated by linear regression analysis, is drawn through the points. The slope of the line is 7.18 +/- 0.22 substitutions per year.


NUCLEOTIDE SUBSTITUTIONS
0
10
20
30
40
50
60
70
80
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
YEAR OF ISOLATION

represents an accurate molecular clock in nature, similar to what has been described for the NS gene.

Six previously published N2 subtype NA sequences (1466 bases) of viruses isolated between 1957 and 1979 were also examined by maximum parsimony (Van Rompuy et al., 1982; Markoff and Lai, 1982; Bentley and Brownlee, 1982; Elleman et al., 1982; Martinez et al., 1983; Lentz et al., 1984). The best-fit evolutionary tree contains 165 nucleotide substitutions and is presented in Fig. 19. A plot of the number of nucleotide substitutions from the origin of the tree (the point where the RI57 branch intersects the vertical axis) to the tip of each branch versus the year of isolation of the strain represented at each tip is shown in Fig. 20. An evolutionary rate for the N2 NA gene was calculated from the slope of the line, derived by linear regression. The rate of change in the NA gene is 4.65 +/- 0.30 substitutions per NA gene per year or 3.17 +/- 0.20 x 10^{-3} substitutions per nucleotide site per year. Similar to the results obtained for the NS and H3 HA genes, a linear accumulation of changes over time is observed in the N2 NA gene, although only six strains were analyzed.

2. Evolution of Influenza C Viruses

Variation in the genes of influenza C viruses was investigated by comparative nucleotide analysis of the HA and the NS genes of influenza C strains isolated over 36 years.

Fig. 19. Most parsimonious evolutionary tree for six influenza A virus N2 subtype NA genes. The NA nucleotide sequences have been published previously (Van Rompuy et al., 1982; Markoff and Lai, 1982; Bentley and Brownlee, 1982; Elleman et al., 1982; Martinez et al., 1983; Lentz et al., 1984). The nucleotide sequences (1466 bases) were analyzed by the method of Fitch (1971). The evolutionary tree contains 165 nucleotide substitutions. The length of the trunk and side branches of the tree are proportional to the minimal number of substitutions required to account for sequence differences. Fractional substitutions are generated from averaging over all possible minimal solutions. For calculation of the side branch of the RI57 virus, see legend to Fig. 20.

The following abbreviations (RI57, TOK67, NT68, UDORN72, ViC75, BK79) are used for influenza viruses A/RI/5-/57, A/Tokyo/3/67, A/Northern Territory/60/68, A/Udorn/72, A/Victoria/3/75, A/Bangkok/1/79, respectively.

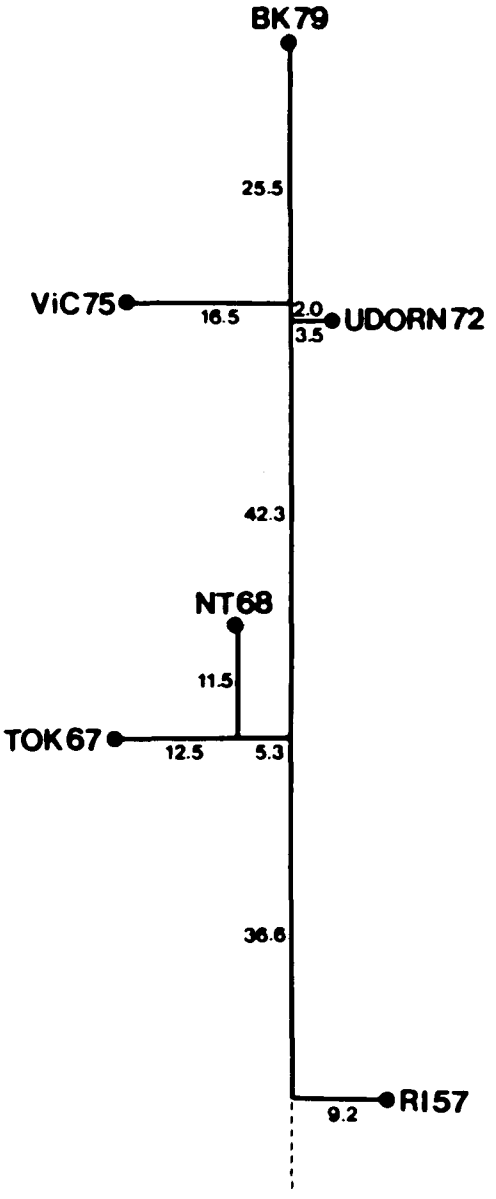
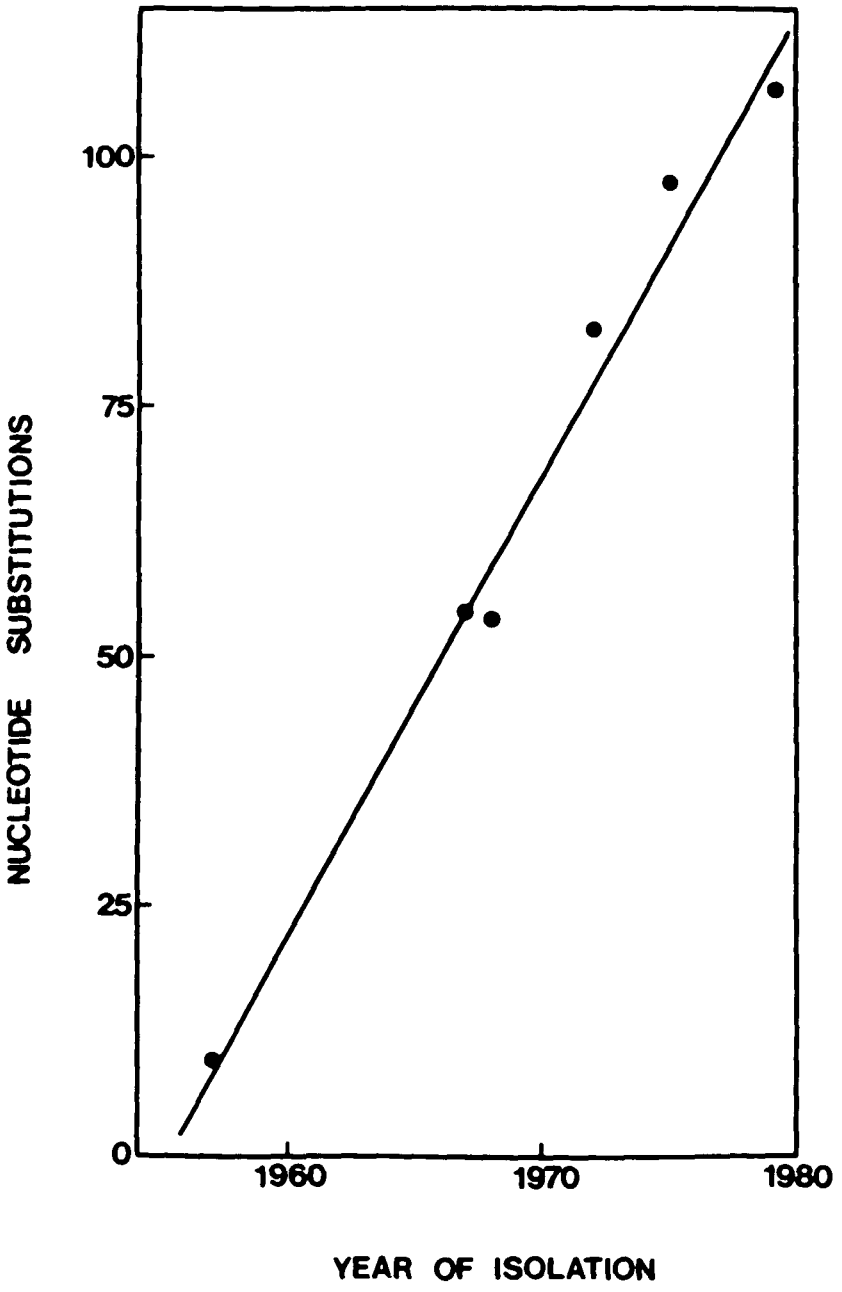

BK79
25.5
ViC75
16.5
2.0
3.5
UDORN72
42.3
NT68
11.5
TOK67
12.5
5.3
36.6
RI57
9.2

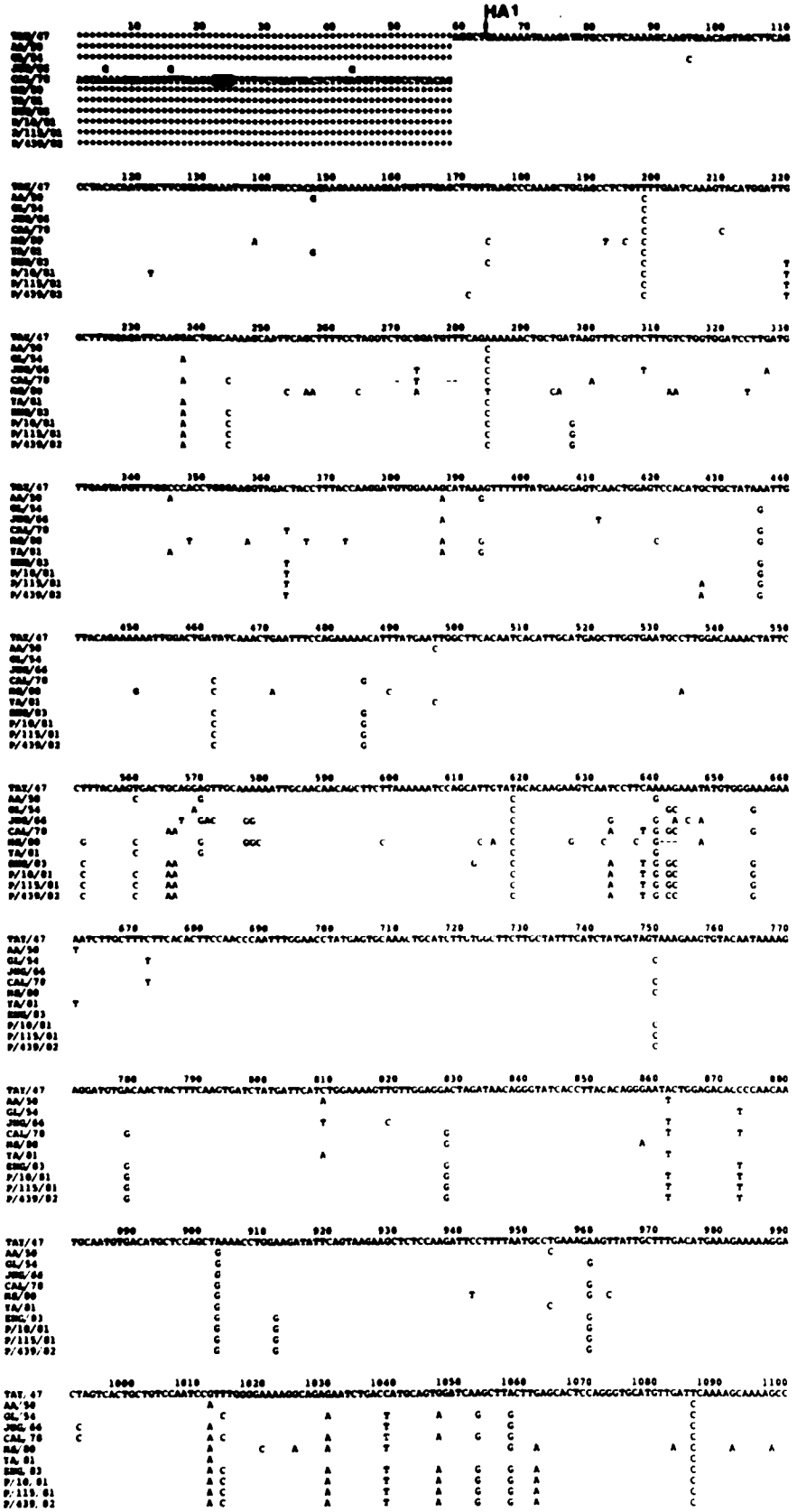
Fig. 20. Linearity with time of number of substitutions in the N2 NA genes of influenza A viruses. The abscissa represents the year of isolation of the viruses used in the analysis. The ordinate indicates the number of nucleotide substitutions observed in their NA genes between the origin of the tree in Fig. 19 and the tips of all branches of the evolutionary tree. The line drawn through the points was derived by linear regression analysis and it has a slope of 4.65 +/- 0.30 substitutions per year. The side branch of RI57 in Fig. 19 was arbitrarily assigned to be 9.2, so that RI57 would fall on the line determined by the remaining five data points.

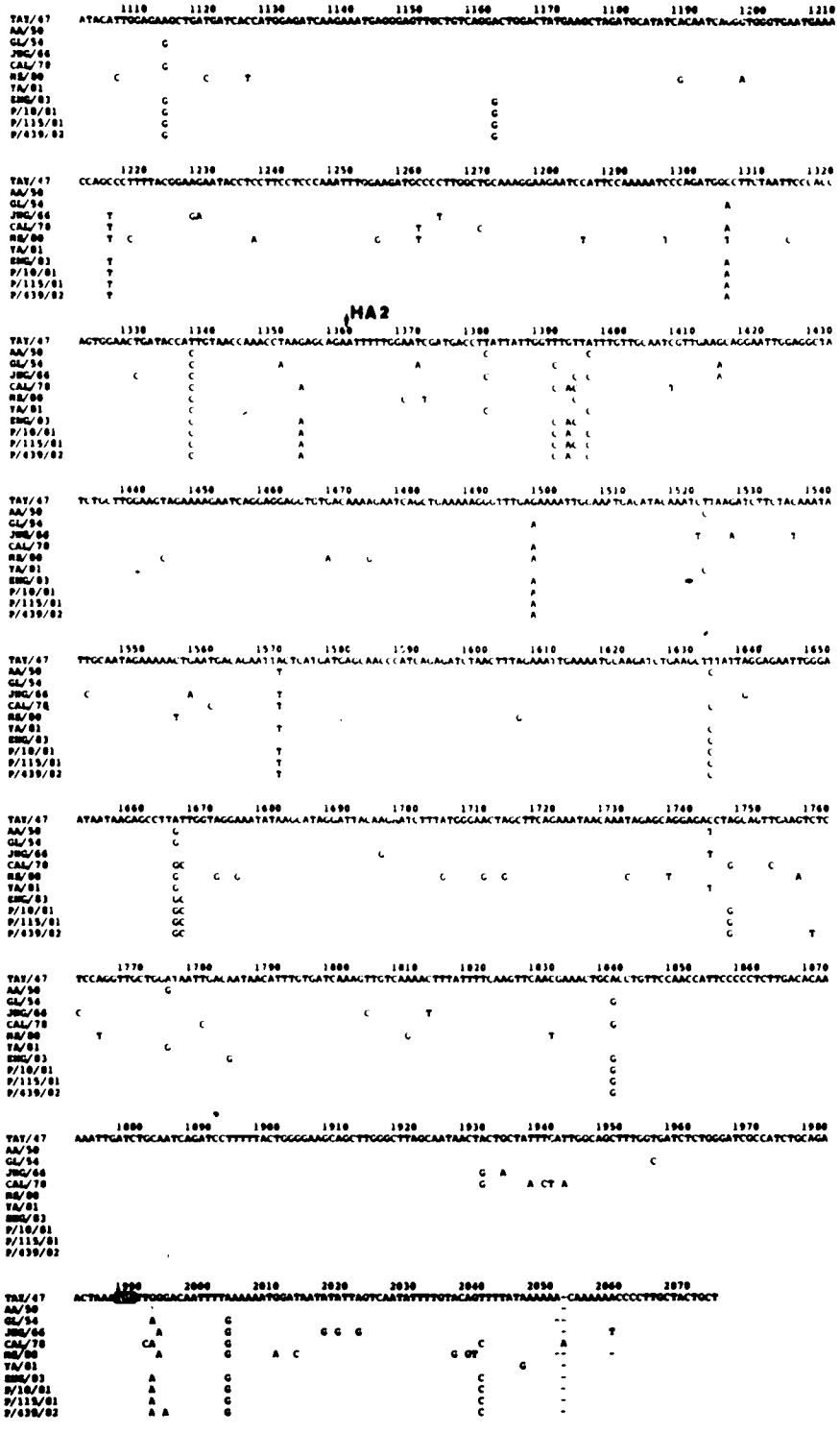

NUCLEOTIDE SUBSTITUTIONS
100
75
50
25
0
1960
1970
1980
YEAR OF ISOLATION

2A. Nucleotide Sequences of the HA Genes in Human Influenza C Viruses

The nucleotide sequences of the HA genes derived from six human influenza C strains isolated during the period of 1947-1983 were determined by the dideoxy sequencing method using synthetic oligonucleotides as primers. A comparison of these HA gene sequences including those from the previously published C/JHG/66 and C/CAL/78 viruses (Pfeifer and Compans, 1984; Nakada et al., 1984a) is shown in Fig. 21. Using the C/TAY/47 gene as the prototype sequence, the C/Cal/78 virus appears to lack three nucleotides in positions 270, 278, and 279, resulting in the effective deletion of one codon. Another deletion in a coding region has to be postulated for the C/MS/80 gene, which also leads to the loss of one amino acid in the HA1 region of the molecule (nucleotide positions 641-643). In addition, several deletions/insertions were observed in the 3' noncoding region of the genes. The C/GL/54 and C/MS/80 genes lack one (position 2051) and two nucleotides (positions 2051 and 2059), respectively. In the case of the C/CAL/78 gene, we suggest an insertion in nucleotide position 2052. It should be noted that the nucleotide comparisons do not include the 58 bases at the 5' terminal ends of the genes.

Fig. 21. Nucleotide sequences of eight human and three swine influenza C virus HA genes. The C/TAY/47 HA gene is used as the prototype sequence and only nucleotides which are changed relative to that sequence are indicated. Since most of the sequences were obtained by the dideoxy sequencing method using primers, the 5' end nucleotide sequences are not available. Thus, sequences excluding positions 1 to 58 are compared in this analysis. It should be noted that the C/JHG/66 and C/CAL/78 virus HA sequences were reported previously (Pfeifer and Compans, 1984; Nakada et al., 1984a). Nucleotide deletions and nucleotides at the 5' end which were not determined are indicated by - and *, respectively. Initiation and termination codons are enclosed within rectangles and the start of the mature HA1 chain (minus signal peptide) and of the cleavage site between HA1 and HA2 are indicated by arrows.


HA1


HA2

Nucleotide changes in the HA genes are seen throughout the molecules and analysis of these mutations shows that the nucleotide changes in these genes are noncumulative. In fact, most of the mutations (129 out of 203) are only found once in any of the genes and are thus unique. In those nucleotide positions in which some genes retain a specific nucleotide and others show differences, the changes do not appear at a specific time. For example, a nucleotide change in position 1390 from T to C is observed in HA genes from strains isolated in 1954, 1978 and in 1983, but not in strains isolated in 1947, 1950, 1966, 1980 or in 1981. Analysis of all genes derived from human influenza C viruses thus suggests that the circulating strains are not related through evolutionary changes accumulating with time.

A similar conclusion can be reached by pairwise comparison of different HA genes. Again, the total difference between two genes does not always increase when strains are compared which were isolated at increasing time intervals (Table 13). The extent of nucleotide variation between HA genes derived from human C isolates varied between 0.1 and 6.8%.

It is surprising that only two nucleotides (positions 237 and 2046) are different between the C/AA/50 and C/YA/81 HA genes, which are derived from strains isolated 31 years apart (Fig. 21 and Table 13). Furthermore, only a 1.8%

TABLE 13. NUCLEOTIDE DIFFERENCES BETWEEN INFLUENZA C VIRUS HA GENES

	AA/50	GL/54	JHG/66	CAL/78	MS/80	YA/81	ENG/83	P/10/81
TAY/47	27(1.3)*	36(1.8)	58(2.9)	71(3.5)	110(5.5)	29(1.4)	54(2.7)	58(2.9)
AA/50		49(2.4)	54(2.7)	74(3.7)	112(5.6)	2(0.1)	61(3.0)	57(2.8)
GL/54			73(3.6)	49(2.4)	112(5.6)	49(2.4)	36(1.8)	38(1.9)
JHG/66				91(4.5)	128(6.3)	56(2.8)	84(4.2)	84(4.2)
CAL/78					134(6.6)	74(3.7)	33(1.6)	31(1.5)
MS/80						114(5.7)	117(5.8)	121(6.0)
YA/81							61(3.0)	57(2.8)
ENG/83								12(0.6)

Nucleotides 1 to 58 are excluded from analysis and nucleotide deletions/insertions are counted as changes in this comparison.
*, Number of nucleotide differences between 2 HA genes (% differences).
Since the C virus isolates from pigs show a high degree of homology, only one swine virus HA gene (C/P/10/81) is included in the analysis.

difference is observed between the genes of the C/GL/54 and C/ENG/83 viruses, which were isolated 29 years apart. On the other hand, variations of HA genes between viruses isolated one (1980 and 1981) and two (1978 and 1980) years apart are 5.7% and 6.6%, respectively.

2B. Deduced Amino Acid Sequences of HA Genes in Human Influenza C Viruses

The predicted amino acid sequences of eight HA genes derived from human influenza C viruses are shown in Fig. 22. Only one deletion in the HA1 portion of the C/CAL/78 and C/MS/80 HA is required to permit alignment of all eight sequences. The HA1 polypeptide chains, except those of the C/CAL/78 and C/MS/80 strains, are 432 amino acids long and the HA2 portions all show a length of 209 amino acids. It should be noted that the amino acid comparisons do not include the signal peptide regions.

As shown in Fig. 22, like the nucleotide changes, the amino acid variation in the HA proteins of the human strains does not appear to be cumulative with time. Amino acid differences of the HA proteins are summarized in Table 14. The extent of differences among the HA proteins is independent of the year in which these viruses were isolated. For instance, the differences in the HA proteins derived from viruses isolated one (1980 and 1981) and two (1978 and 1980) years apart are 3.6% and 4.5%, respectively. In contrast, the HA proteins of a 1950 and a

Fig. 22. Comparison of deduced amino acid sequences of C virus HA proteins. The HA sequence of C/TAY/47 virus is shown as prototype and changes in the HAs of the other strains are indicated. Arrows show (possible) cleavage sites in the HA precursor molecules. Amino acid deletions which are found in the HA proteins of C/CAL/78 and C/MS/80 viruses are indicated by -, and amino acid positions which were not identified are indicated by *. The C/JHG/66 and C/CAL/78 virus HA sequences were published previously (Pfeifer and Compans, 1984; Nakada et al., 1984a).

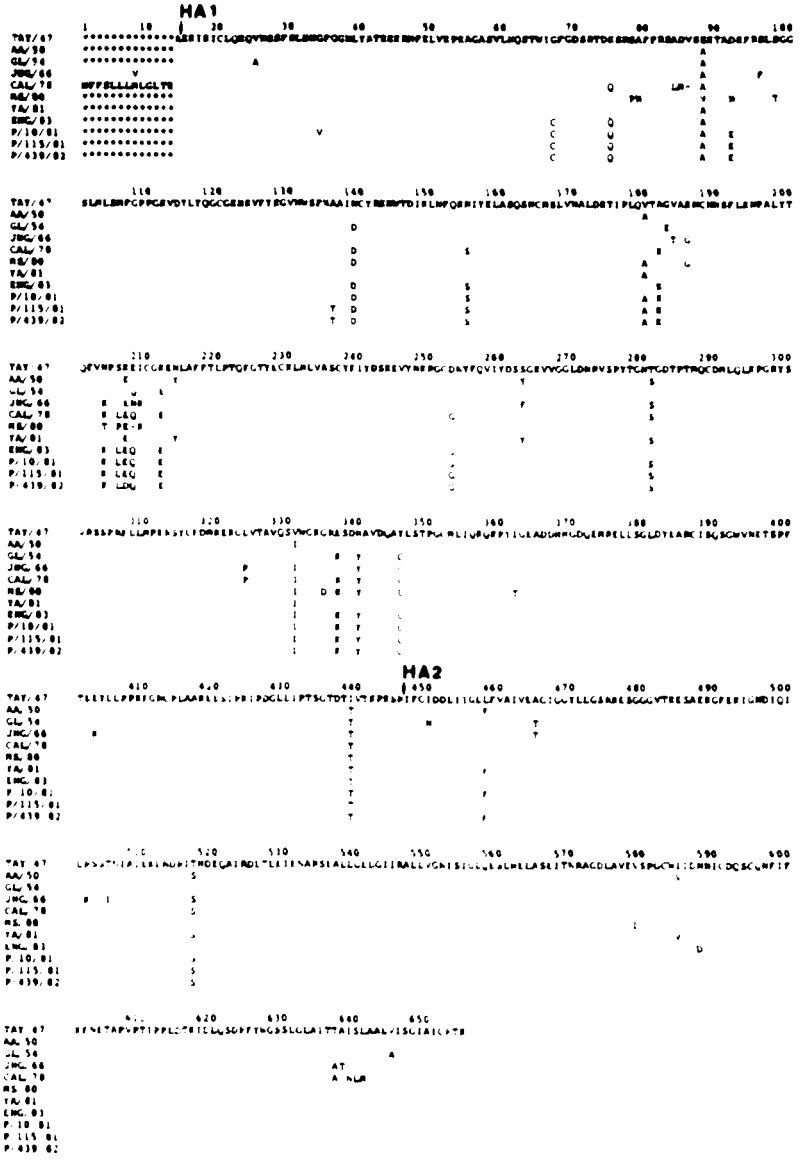

HA1
HA2

TABLE 14. AMINO ACID VARIATION OF HA PROTEINS IN INFLUENZA C VIRUSES

	AA/50	GL/54	JHG/66	CAL/78	MS/80	YA/81	ENG/83	P/10/81
TAY/47	11(1.7)*	13(2.0)	23(3.6)	26(4.1)	21(3.3)	11(1.7)	18(2.8)	23(3.6)
AA/50		20(3.1)	20(3.1)	25(3.9)	23(3.6)	0	21(3.3)	18(2.8)
GL/54			24(3.7)	23(3.6)	22(3.4)	20(3.1)	15(2.3)	20(3.1)
JHG/66				25(3.9)	26(4.1)	20(3.1)	25(3.9)	26(4.1)
CAL/78					29(4.5)	25(3.9)	12(1.9)	13(2.0)
MS/80						23(3.6)	21(3.3)	23(3.6)
YA/81							21(3.3)	18(2.8)
ENG/83								7(1.1)

Comparison of the HA proteins includes the HA1 and the HA2 regions, but not the signal peptide regions. Amino acid deletions are counted as changes in this analysis.
*, Number of amino acid differences between two HA proteins (% differences).
Since the swine C virus HAs are similar to each other, only one swine C virus HA protein (C/P/10/81) is included in this analysis.

1981 strain do not show any amino acid differences and the C/TAY/47 and C/ENG/83 HA proteins differ by only 2.8% (Fig. 22 and Table 14).

Although the total number of amino acid changes is small when all HA proteins are compared, three clusters of amino acid substitutions (positions 67-98, 180-214, and 331-346) can be observed. Specifically, the region between amino acids 204 and 214 appears to be a site which is characterized by multiple amino acid changes.

2C. Variation of the HA Genes and Proteins in Swine Influenza C Viruses

Influenza C/P/10/81, C/P/115/81, and C/P/439/82 strains were obtained from pigs in an abattoir in Beijing, China. Influenza C/P/115/81 and C/P/439/82 strains were obtained one and 18 months, respectively, following the isolation of the C/P/10/81 virus (Guo et al., 1983; Guo and Desselberger, 1984). As the sequence analysis shows, the HA genes of all three strains are closely related. For example, only three nucleotides of the C/P/115/81 (positions 122, 427, and 1393) and six nucleotides of the C/P/439/82 (positions 122, 171, 427, 642, 1758, and 1994) HA genes are different from those found in C/P/10/81 (Fig. 21).

This nucleotide difference translates into only a three amino acid difference between the C/P/10/81 and C/P/439/82

genes (Fig. 22). Interestingly, only one of these amino acid changes is located in a region of high variation which was identified in the human C virus HA genes. Comparison of the swine virus HA genes with those of the human viruses reveals that the C/ENG/83 HA gene is most closely related to the swine virus HA genes. On the other hand, the HA gene of a 1980 strain (C/MS/80) shows the highest nucleotide difference when compared to the three porcine HA genes (Table 13).

2D. Frequency of Variation in the First, Second, and Third Codon Position

The HA gene of the C/TAY/47 virus was compared pairwise to those of the other human C viruses with respect to changes in the first, second, and third codon position. This analysis revealed that first and second position changes occurred at a frequency of 17-33% and of 11-21%, respectively (Table 15). In contrast, a much higher value of 52-72% was found for changes occurring in the third position of the coding frame.

Additionally, the HA sequence (HA1 and HA2 domains) of the C/TAY/47 virus was compared to those of the other human C viruses to examine silent and amino acid changing nucleotide substitutions. We observed that 6.4% (124/1923) of the base positions exhibit silent changes, whereas only 3.2% (62/1923) of the sites display nucleotide changes

TABLE 15. FREQUENCY OF CHANGES IN CODON POSITION: COMPARISON OF THE HA GENE OF C/TAY/47 VIRUS WITH THOSE OF OTHER STRAINS

	Positions		
	First	Second	Third
AA/50	9 (33)*	4 (15)	14 (52)
GL/54	7 (21)	7 (21)	19 (58)
JHG/66	13 (25)	11 (21)	28 (54)
CAL/78	18 (27)	11 (17)	37 (56)
MS/80	17 (17)	11 (11)	73 (72)
YA/81	9 (32)	4 (14)	15 (54)
ENG/83	15 (29)	7 (14)	29 (57)
P/10/81	15 (27)	9 (17)	31 (56)

Signal sequences were excluded from analysis.
*, Numbers of nucleotide changes between the HA gene of C/TAY/47 virus and that of other strains (% differences).

which cause amino acid changes in the HA protein sequences. The ratio of silent to nonsilent changes is exactly 2:1.

2E. Nucleotide Sequences of NS Genes of Human Influenza C Viruses

The nucleotide sequences of the NS genes of seven influenza C viruses isolated between 1947 and 1983 were determined and compared with the sequence of the previously reported CAL/78 virus NS gene (Nakada et al., 1985). The sequence data are shown in Fig. 23. Base differences in the NS sequences relative to the earliest isolate, TAY/47, are indicated. In the sequenced region, there were no deletions/insertions detected in any of the NS sequences. Thus, all NS sequences examined appear to have the same length. Nucleotide substitutions in the C virus NS genes are observed throughout the segment. Of the 45 substituted positions, 25 sites are uniquely substituted in a single virus strain. Although the remaining substituted positions show the same base change in several strains, most of the changes do not correlate with the time of isolation of the virus. That is, once a nucleotide change is observed in a virus isolate, it is not consistently present in all later isolates. This is further illustrated when pairwise comparisons among the viruses are made (Table 16). For example, the TAY/47 and ENG/83 viruses isolated 36 years apart differ by only 2.1% in their NS sequences, whereas the 1978 (CAL/78) and the 1981 (YA/81) isolates possess NS genes with an overall difference of 3.0%. In addition, the NS genes of TAY/47 and JHG/66 viruses (isolated 19

Fig. 23. Nucleotide sequence comparison of influenza C virus NS genes. Nucleotide differences with respect to the TAY/47 strain are presented. All nucleotide sequences are presented in the plus sense strand. Asterisks represent the 5' forty bases that were not determined in the analysis. The arrowheads indicate the putative splice sites where NS1 mRNA is spliced to generate NS2 mRNA. The boxed nucleotides represent the following: ATG (27-29) - initiation codon for NS1 and NS2 proteins; TGA (703-705) - termination codon for NS2 protein; TAA (885-887) - termination codon for NS1 protein.

The sequence of the CAL/78 NS gene was generated by Nakada et al. (1985). The remaining NS sequences were determined by the dideoxy method using oligonucleotide primers. Nine primers were needed to obtain complete NS sequences for all strains. These primers correspond to the following nucleotide positions of the NS gene: 22-35; 115-130; 180-198; 356-370; 356-375; 495-509; 667-683; 667-684; 815-829.

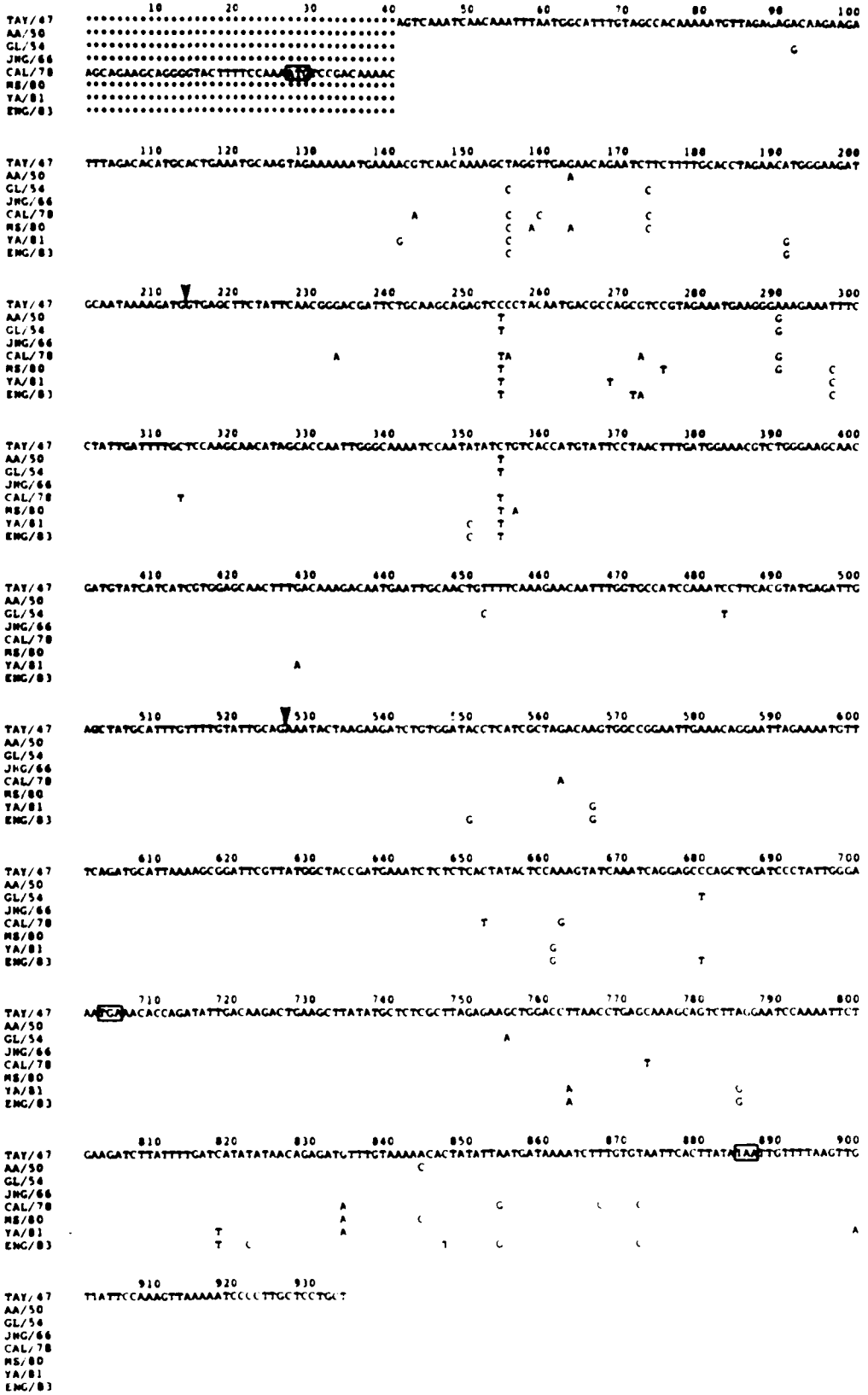

TAY/47
AA/50
GL/54
JNG/66
CAL/78
MS/80
YA/81
ENG/83

TABLE 16

NUCLEOTIDE DIFFERENCES BETWEEN INFLUENZA C VIRUS NS GENES

	AA/50	GL/54	JHG/66	CAL/78	MS/80	YA/81	ENG/83
TAY/47	5(0.6)*	10(1.1)	0	19(2.1)	12(1.3)	16(1.8)	19(2.1)
AA/50		9(1.0)	5(0.6)	18(2.0)	7(0.8)	17(1.9)	20(2.2)
GL/54			10(1.1)	19(2.1)	12(1.3)	20(2.2)	21(2.3)
JHG/66				19(2.1)	12(1.3)	16(1.8)	19(2.1)
CAL/78					19(2.1)	27(3.0)	26(2.9)
MS/80						18(2.0)	23(2.6)
YA/81							13(1.5)

NUCLEOTIDES 1 TO 40 ARE EXCLUDED FROM ANALYSIS.

*NUMBER OF NUCLEOTIDE DIFFERENCES BETWEEN 2 NS GENES (% DIFFERENCE)

years apart) show no nucleotide differences. Clearly, the extent of nucleotide differences among the NS genes does not correspond to the isolation dates of the strains.

2F. Predicted Amino Acid Sequences of the NS1 and NS2 Proteins of Human Influenza C Viruses

The NS gene of influenza C viruses has recently been shown to code for two nonstructural proteins, NS1 (Nakada et al., 1985) and NS2 (Nakada et al., 1986). The NS1 protein consists of 286 amino acids and is translated from a mRNA colinear with the virion RNA segment. The predicted amino acid sequences of the NS1 proteins of the eight human influenza C viruses are shown in Fig. 24. The first five amino acids at the N-terminus of the NS1 protein could not be deduced because the nucleotides in this region were not determined. Excluding the first five residues, all the NS1 proteins appear to have conserved amino acid length. A pairwise comparison of the NS1 proteins is shown in Table 17. As previously observed on the nucleotide level, amino acid differences among NS1 proteins are not always greater when viruses isolated farther apart in time are compared. In fact, the greatest amino acid difference of 4.3% is seen with the CAL/78 and ENG/83 strains which were isolated five years apart (Table 17). The NS1 protein of the TAY/47 virus and that of a virus isolated 36 years later, ENG/83, only show a 2.5% difference.

Fig. 24. Comparison of deduced NS1 protein sequences of eight influenza C viruses. Amino acid differences are presented with respect to the TAY/47 sequence. The asterisks represent amino acids at the N-terminus which could not be deduced because the nucleotide sequence in this region was not determined. The 62 amino acid sequence to the left of the arrowhead is the shared N-terminal region of the NS1 and NS2 proteins.

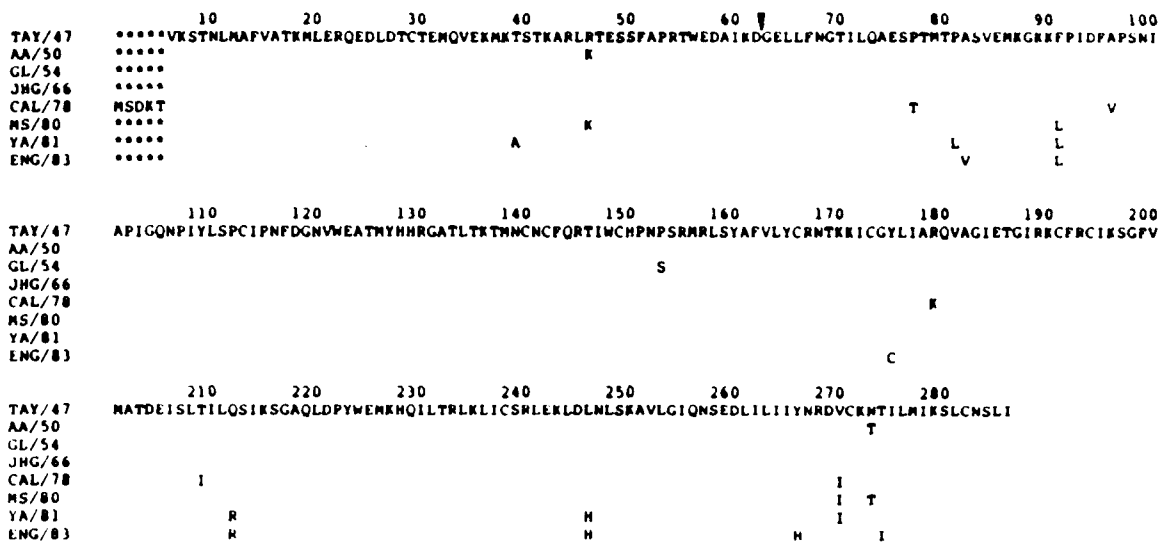

10 20 30 40 50 60 70 80 90 100
TAY/47 *****VKSTNLMAFVATKMLERQEDLDTCTEMQVEKMKTSTKARLRTESSFAPRTWEDAIKDGELLFWGTILQAESPTMTPASVEMKGKKFPIDFAPSNI
AA/50 *****
GL/54 *****
JHG/66 *****
CAL/78 MSDKT
MS/80 *****
YA/81 *****
ENG/83 *****
110 120 130 140 150 160 170 180 190 200
TAY/47 APIGQNPIYLSPCIPNFDGNVWEATMYHHRGATLTKTMNCNCFQRTIWCHPNPSRMRLSYAFVLYCRNTKKICGYLIARQVAGIETGIRKCFRCIKSGFV
AA/50
GL/54
JHG/66
CAL/78
MS/80
YA/81
ENG/83
210 220 230 240 250 260 270 280
TAY/47 MATDEISLTILQSIKSGAQLDPYWEMKHQILTRLKLICSHLEKLDLMLSKAVLGIQNSEDLILIIYNRDVCKNTILMIKSLCNSLI
AA/50
GL/54
JHG/66
CAL/78
MS/80
YA/81
ENG/83

TABLE 17

AMINO ACID DIFFERENCES BETWEEN INFLUENZA C VIRUS NS1 PROTEINS

	AA/50	GL/54	JHG/66	CAL/78	MS/80	YA/81	ENG/83
TAY/47	2(0.7)*	1(0.4)	0	5(1.8)	4(1.4)	6(2.1)	7(2.5)
AA/50		3(1.1)	2(0.7)	7(2.5)	2(0.7)	8(2.8)	9(3.2)
GL/54			1(0.4)	6(2.1)	5(1.8)	7(2.5)	8(2.8)
JHG/66				5(1.8)	4(1.4)	6(2.1)	7(2.5)
CAL/78					7(2.5)	9(3.2)	12(4.3)
MS/80						6(2.1)	9(3.2)
YA/81							7(2.5)

AMINO ACIDS 1 TO 5 ARE EXCLUDED FROM ANALYSIS.

*NUMBER OF AMINO ACID DIFFERENCES BETWEEN 2 NS1 PROTEINS (% DIFFERENCE).

The NS2 protein of influenza C viruses is translated by a spliced mRNA (Nakada et al., 1986). The N-terminal 62 amino acids of the NS2 protein are shared with those of the NS1 protein. The remaining 59 residues of NS2 are translated in a different reading frame from that of NS1 as a result of the splicing event. The putative splice sites, indicated by arrowheads in Fig. 23, and the putative branching point (positions 483-489) are conserved in the C virus NS sequences analyzed (Fig. 23). The deduced unique sequence of the NS2 protein from amino acid 63 to 121 is shown in Fig. 25. All NS2 proteins examined appear to have the same length and very few amino acid differences among strains are observed in the region unique to the NS2 protein.

The NS sequence of the C/TAY/47 virus was compared pairwise to those of the other influenza C viruses with respect to silent and nonsilent base substitutions in the NS1 coding region. 3.2% (27/847) of the nucleotide positions display silent substitutions while only 2.0% (17/847) of the sites show nucleotide changes in some strains which alter the amino acid sequence of the NS1 protein. The ratio of silent to nonsilent substitutions is 1.6:1.

2G. Evolutionary Analysis of NS and HA Genes of Eight Influenza C Viruses by Maximum Parsimony

The influenza C virus NS gene sequences presented in

Fig. 25. Comparison of the unique region of the predicted NS2 protein sequences of eight influenza C viruses. Amino acid differences with respect to the TAY/47 strain are indicated. The open triangle denotes the 62 amino acids at the N-terminus of the NS2 proteins that are identical to those at the N-terminus of the NS1 proteins (see Fig. 24).

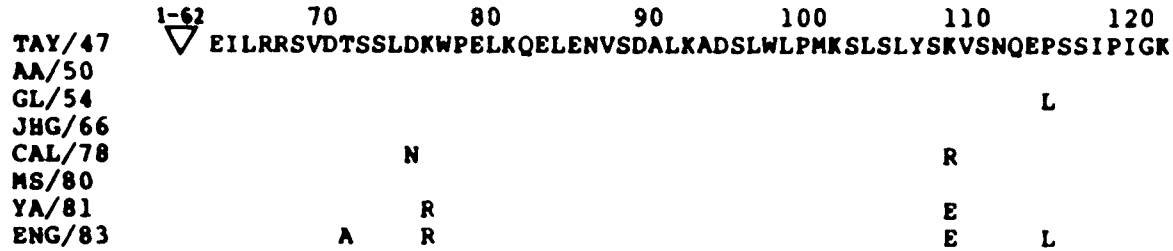

1-62
70
80
90
100
110
120
TAY/47 EILRRSVDTSSLDKWPELKQELENVSDALKADSLWLPMKSLSLYSKVSNQEPSSIPIGK
AA/50
GL/54 L
JHG/66
CAL/78 N R
MS/80
YA/81 R E
ENG/83 A R E L

Fig. 23 were examined by the maximum parsimony procedure (Fitch, 1971). For the C virus NS sequences, three equally parsimonious trees of 54 substitutions can be constructed. These different tree structures are presented in Fig. 26A. The differences among the trees arise in the placement of the GL/54 strain. In addition, one evolutionary tree containing 55 and three trees containing 56 substitutions were examined. All of these trees have different structures from those presented in Fig. 26A (data not shown). However, none of these evolutionary trees constructed for the C virus NS genes are even remotely compatible with a single gene lineage passing on its accumulated nucleotide changes to subsequent isolates. We therefore conclude that multiple NS genes coexist in nature. It should be noted that the numbers defining genetic distances in the evolutionary trees in Fig. 26 represent average values calculated by the maximum parsimony method involving all the strains used for the tree. Thus, some values in Fig. 26 do not precisely match the pairwise comparison data shown in Tables 13 and 16.

For comparison, the HA sequences of the same eight human influenza C virus strains (Fig. 21) were also analyzed by the maximum parsimony method. Fig. 26B illustrates the two equally parsimonious trees of 241 substitutions which were generated using the HA sequences. The two trees have similar structures and only differ in

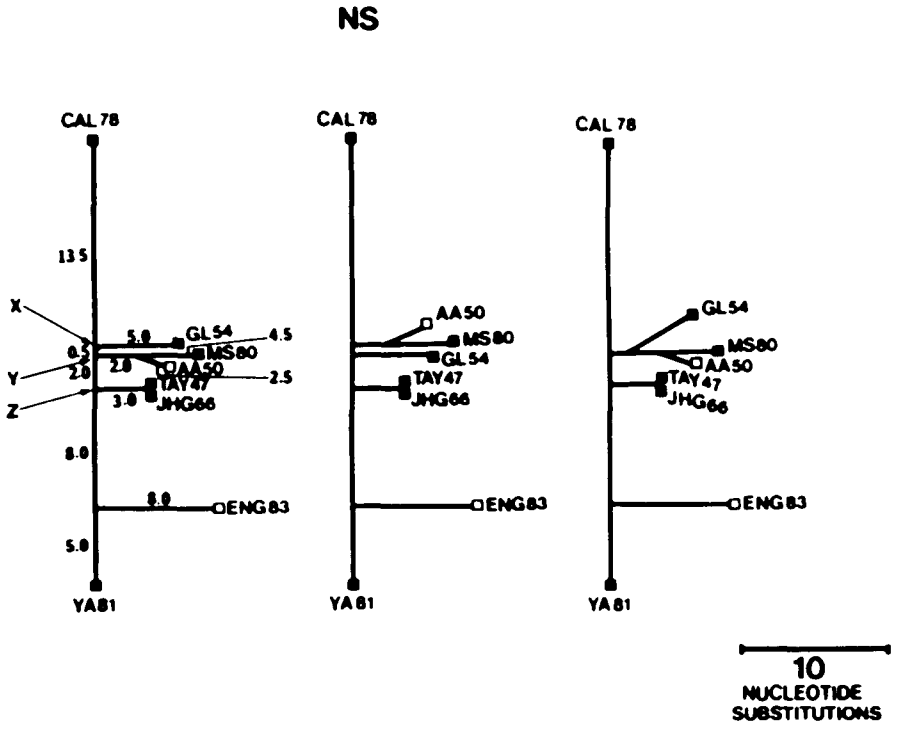
the placement of the TAY/47 strain. Four alternative HA trees containing 242, 245, 246, or 251 substitutions were also analyzed (data not shown). Once again, none of the trees is compatible with a single gene lineage passing on its accumulated nucleotide substitutions to subsequent isolates. In addition, trees obtained for the HA genes differ in a significant way from those constructed for the NS genes (Fig. 26).

Although the NS and HA evolutionary trees have distinct structures, there are similarities in gross topology. The NS and HA trees shown in Fig. 26A and 26B were drawn so as to illustrate the similarities in their topology. Comparison of the branching order of strains on both the NS and HA trees reveals the same relative placement of the CAL/78, GL/54, MS/80, TAY/47, JHG/66 and YA/81 strains (represented by solid squares in Fig. 26). The ENG/83 and AA/50 viruses (represented by open squares in Fig. 26) occupy very different positions on the NS and HA trees. These data indicate that the ENG/83 virus as well as the AA/50 strain may have arisen through reassortment among the other lineages.

A possible model illustrating the derivation of the ENG/83 strain is shown in Fig. 26C. The construction of this tree is described in detail in the legend to Fig. 26C. The evolutionary model suggests that the ENG/83 virus received its HA gene from a recent ancestor of the CAL/78

Fig. 26A and 26B. Evolutionary relationships of NS and HA genes of eight influenza C viruses. The three most parsimonious NS gene trees of 54 nucleotide substitutions each (26A) and the two most parsimonious HA gene trees of 241 nucleotide substitutions each (26B) are shown. The data were obtained using the maximum parsimony method for determining evolutionary relationships (see Materials and Methods). Branch lengths are proportional to the number of nucleotide substitutions between branch points. The C viruses represented by the solid squares at their branch tips have a similar branching order in both the NS and HA trees. The viruses represented by the open squares at their branch tips occupy very different positions in the NS and HA trees. The major nodes, X, Y and Z, defined by the taxa with the solid squares are indicated.

A


NS
CAL 78
13 5
X
Y
Z
0.5
2.0
5.0
2.8
3.0
GL 54
MS 80
AA 50
TAY 47
JHG 66
4.5
2.5
8.0
8.0
ENG 83
5.0
YA 81
CAL 78
AA 50
MS 80
GL 54
TAY 47
JHG 66
ENG 83
YA 81
CAL 78
GL 54
MS 80
AA 50
TAY 47
JHG 66
ENG 83
YA 81
10
NUCLEOTIDE
SUBSTITUTIONS

B

HA

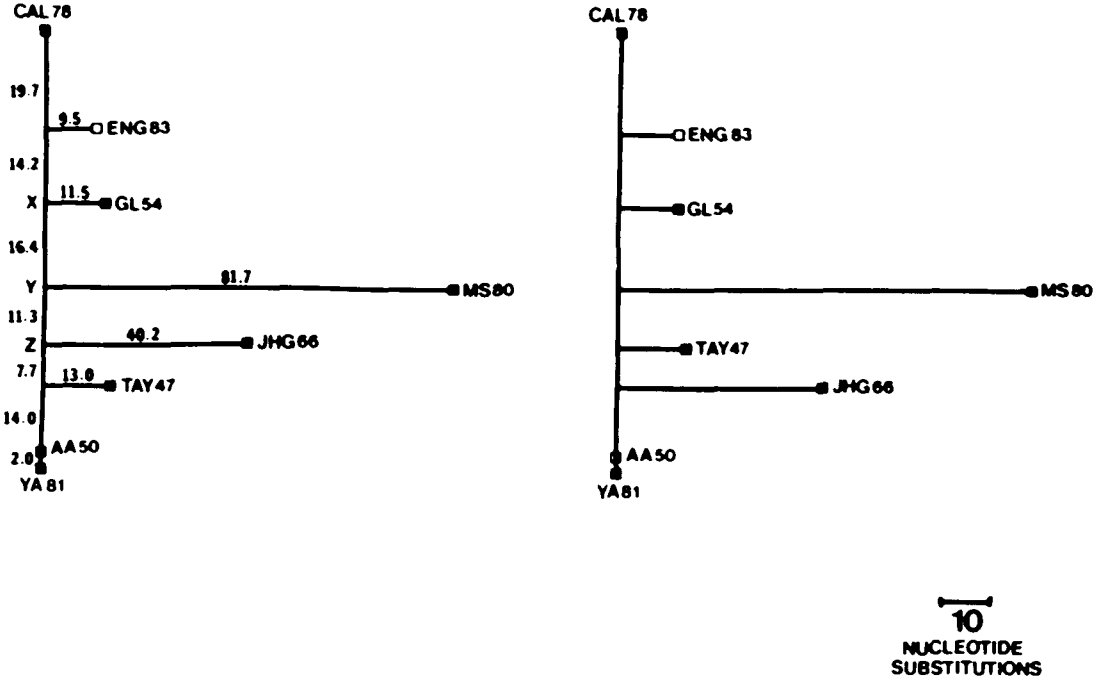
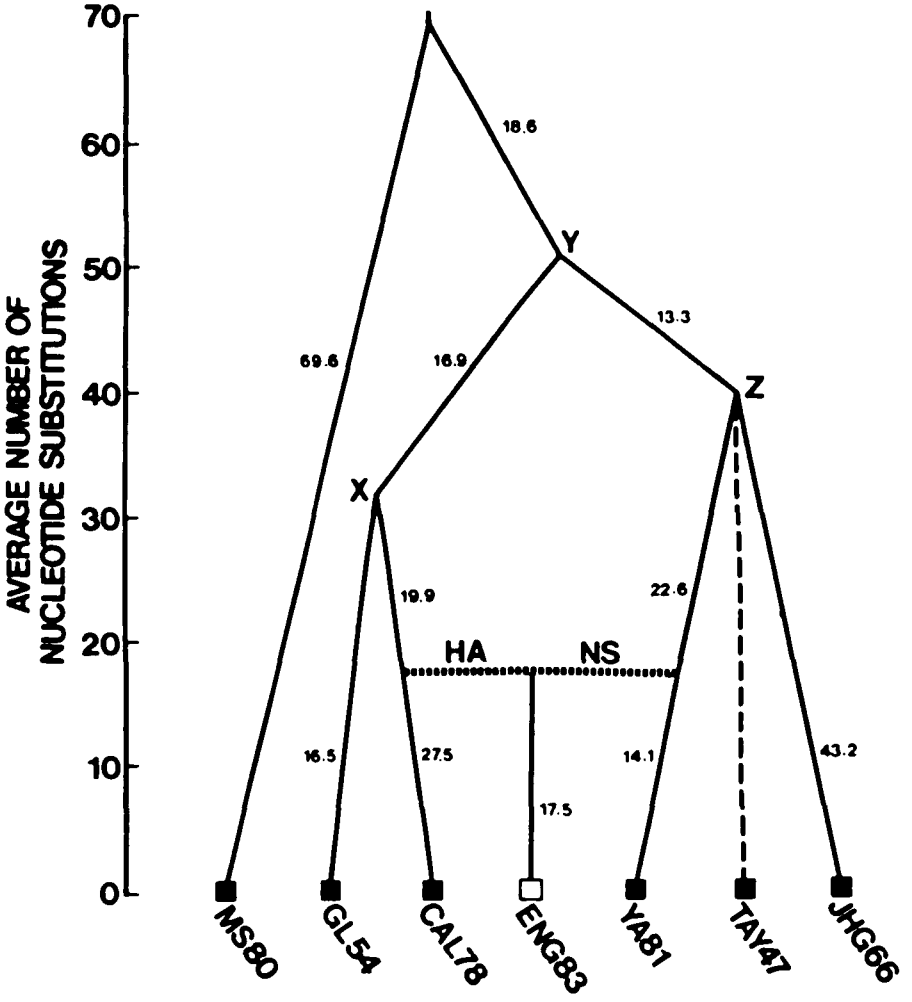

CAL78
19.7
9.5
ENG83
14.2
X
11.5
GL54
16.4
Y
81.7
MS80
11.3
Z
40.2
JHG66
7.7
13.0
TAY47
14.0
2.0
AA50
YA81
CAL78
ENG83
GL54
MS80
TAY47
JHG66
AA50
YA81
10
NUCLEOTIDE
SUBSTITUTIONS

Fig. 26C. Phylogenetic model suggesting the reassortment of genes in the influenza C/ENG/83 virus. The phylogenetic tree is based on the two topologically most congruent sets of trees for the NS and HA genes (the left-most trees of the sets in 26A and 26B). The branch lengths of this tree were obtained by summing the corresponding numbers of nucleotide substitutions on the trees for the NS (26A) and HA (26B) genes. The branch lengths are not proportional to the indicated genetic distances. The major nodes X, Y and Z in 26A and 26B are also indicated in this tree. The ENG/83 virus is shown as a reassortant that derives its HA gene from the CAL/78 line and its NS gene from the YA/81 line. On the CAL/78 side, the ENG/83 lineage is attached at a distance proportional to its distance on the corresponding HA branch in 26B. For example, the sum of the distance of CAL/78 to node X in 26A (13.5) and 26B (33.9) equals 47.4. This is also the total distance of CAL/78 to node X in 26C. In the HA tree (26B), the ENG/83 branch intersects the CAL/78 to node X branch at a distance of 19.7 (19.7/33.9 = 0.58). Therefore, in 26C, ENG/83 intersects the CAL/78 to node X branch at 0.58 x 47.4 = 27.5. On the YA/81 side, the ENG/83 lineage descends at a distance proportional to its distance on the corresponding NS branch in 26A. The divergence of the reassortant equals the sum of its branch lengths on the trees in 26A and 26B. The root of the tree is arbitrarily placed by assuming the

Fig. 26C continued.

MS/80 lineage is the oldest by virtue of its being the most distantly related to the other strains. Ancestral nodes are placed at a height corresponding to the weighted average of the number of substitutions on each of its two lines of descent. Because TAY/47 is ambiguously located in the various trees, its descent from node Z is indicated by a broken line suggesting that it could equally well be attached above that node or on either of the two branches descending from it. The genetic distance for the TAY/47 branch is not calculated.

C


AVERAGE NUMBER OF NUCLEOTIDE SUBSTITUTIONS
70
60
50
40
30
20
10
0
18.6
Y
13.3
69.6
16.9
Z
X
19.9
22.6
HA
NS
16.5
27.5
17.5
14.1
43.2
MS80
GL54
CAL78
ENG83
YA81
TAY47
JHG66

strain and its NS gene from a recent ancestor of the YA/81 virus (based on the position of ENG/83 on the HA and NS trees). One can also state that the HA gene of the ENG/83 strain is most closely related to that of the CAL/78 virus (Table 13) and that the NS gene of the ENG/83 strain is most closely related to that of the YA/81 virus (Table 16). Assuming constant evolutionary change over time, reassortment between lineages is thus a likely mechanism to explain the origin of the ENG/83 lineage.

3. Analysis of an Influenza A Virus Mutant with a Deletion in the NS Segment

The defect of an NS mutant virus that exhibits a restricted host range phenotype that is not temperature dependent was characterized by nucleotide sequencing of cloned cDNAs. The influenza virus host range mutant, CR43-3, was accidentally generated in a recombination experiment in PCK cells at the nonpermissive temperature of 38^{o} with A/ALA/6/77 and the cold-adapted (ca) and temperature-sensitive (ts) A/AA/6/60 viruses as parents (Maassab et al., 1981). CR43-3 virus displays a host range phenotype different from that of either parent. The mutant virus forms plaques in PCK cells, but not in MDCK cells, whereas the parent viruses and an isogenic recombinant virus (CR43-7) plaque equally well in both cell systems (Maassab and DeBorde, 1983). Previous analysis of the RNA gene

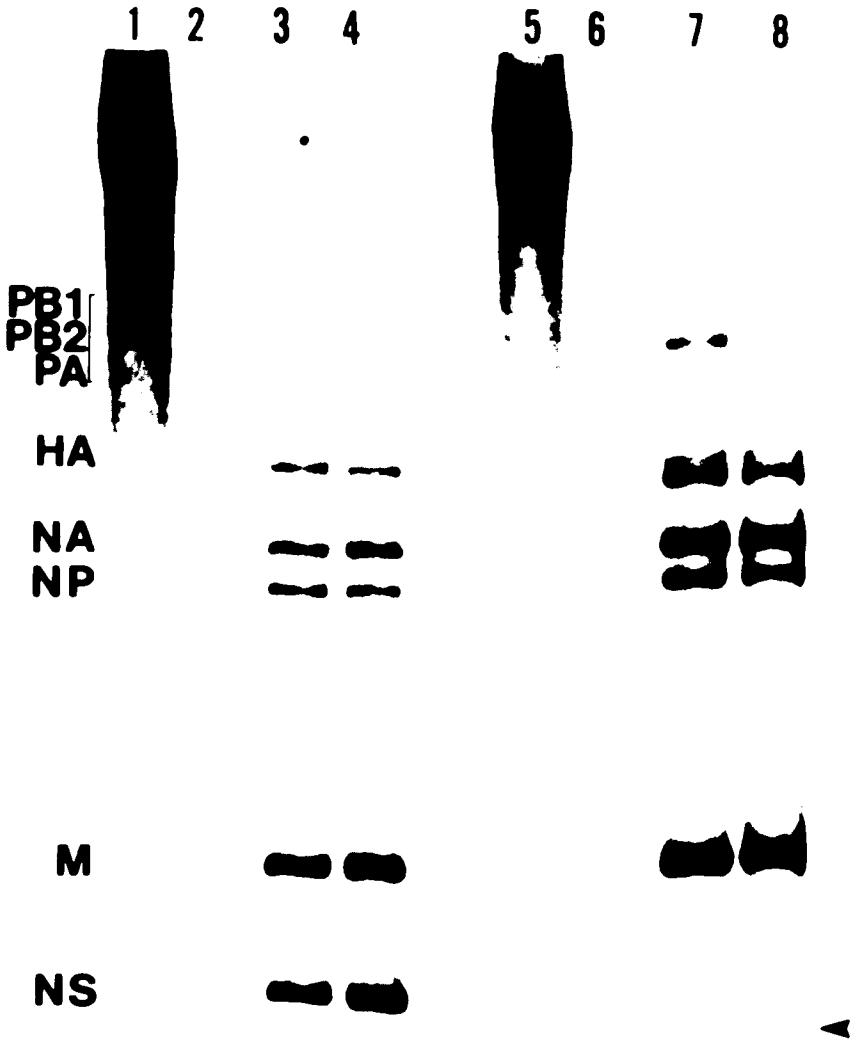
constellation of CR43-3 virus by electrophoresis in polyacrylamide gels showed that CR43-3 possesses an NS segment with an aberrant gel migration.

The precise defect in the NS gene of the CR43-3 virus was examined by nucleotide sequencing. In addition, viral protein synthesis and cellular localization of the NS1 gene product in both permissive and nonpermissive cells was investigated.

3A. cDNA Analysis of NS Segments

Polyacrylamide-urea gel electrophoresis of the glyoxalated RNAs of many influenza A virus strains has shown identical gel migration for the NS segments (Desselberger and Palese, 1978). These data indicate a size conservation among NS genes which is supported by nucleotide sequence data (Baez et al., 1980; Baez et al., 1981; Krystal et al., 1983a; Lamb and Lai, 1980; this work). In contrast, the NS segment of CR43-3 virus migrates faster than that of other influenza A viruses (D. DeBorde, personal communication). Similar results were obtained when the cDNAs of A/ALA/6/77, A/AA/6/60, and CR43-3 viruses were compared on polyacrylamide gels. Figure 27 shows the analysis of both single- and double-stranded cDNA copies of the A/AA/6/60 and CR43-3 viral genomes. The cDNA products of the first and second strand reverse transcriptase reactions were treated with S1-nuclease and electrophoresed in parallel on the gel. From the relative

Fig. 27. Electrophoresis of the reverse-transcribed cDNA copies of A/AA/60 and CR43-3 viral genes in a 5% polyacrylamide gel. Lanes: 1, A/AA/60 single-stranded cDNA; 2, A/AA/60 single-stranded cDNA after S1 nuclease treatment; 3, A/AA/60 double-stranded cDNA; 4, A/AA/60 double-stranded cDNA after S1 nuclease treatment; 5, CR43-3 single-stranded cDNA; 6, CR43-3 single-stranded cDNA after S1 nuclease treatment; 7, CR43-3 double-stranded cDNA; 8, CR43-3 double-stranded cDNA after S1 nuclease treatment. The proteins coded by the double-stranded cDNA transcripts are indicated. The faster migrating, NS-specific cDNA of CR43-3 virus is indicated by the arrowhead.


1
2
3
4
5
6
7
8
PB1
PB2
PA
HA
NA
NP
M
NS

mobilities and known sizes of the double-stranded cDNAs on the gel, we calculated that the CR43-3 NS segment is approximately 40 bases shorter than the NS segment of A/AA/6/60 virus. This suggested that the CR43-3 NS gene has a deletion. It should also be noted that the CR43-3 NS cDNA consistently showed a decreased intensity relative to the other viral gene segments. This could indicate that either the amount of RNA 8 in the viral preparation is low or that cDNA transcription of this segment is not efficient. The migration of the NS cDNA of the A/ALA/6/77 virus was identical to that of the corresponding segment of the A/AA/6/60 virus (data not shown).

3B. Nucleotide Sequences of cDNAs

To confirm that the CR43-3 virus NS gene contained a deletion mutation and to localize the exact nucleotide position of this defect, the complete nucleotide sequence of a full-length NS gene of CR43-3 virus, cloned into pBR322, was determined (Fig. 28). This NS gene was found to be 854 nucleotides long, which is 36 bases shorter than that reported for NS genes of all other influenza A viruses (Baez et al., 1980; Baez et al., 1981; Krystal et al., 1983a; Lamb and Lai, 1980; this work). Comparison of the CR43-3 NS sequence with other known NS gene sequences indicated that the deletion occurred around nucleotide position 220. To precisely define the deletion junctions

Fig. 28. Nucleotide sequence of the cloned NS genes of influenza A viruses A/ALA/77 and CR43-3. The deduced amino acid sequences of the NS1 and NS2 polypeptides, respectively, are shown below the nucleotide sequence. The region of the A/ALA/77 NS gene which is deleted in the CR43-3 NS gene is denoted by brackets. The arrowheads indicate the nucleotide positions where the NS1 mRNA is spliced to generate the NS2 mRNA. Base positions 400, 403, and 415 in the CR43-3 NS gene and position 521 in the A/ALA/77 NS gene were not unambiguously identified.

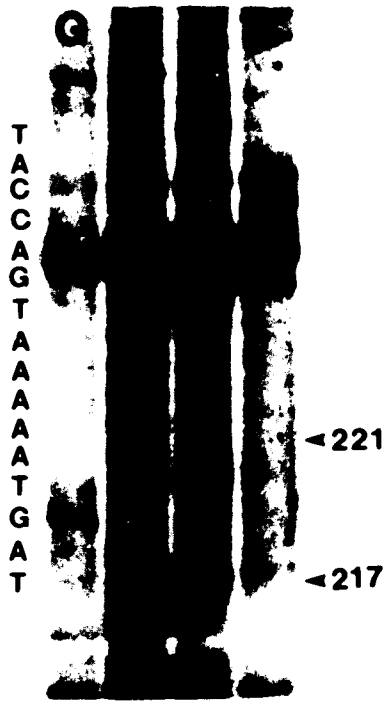
```
        10         20         30         40         50         60         70         80         90
 AGCAAAAGCAGGGTGACAAAGACATA ATG GAT TCC AAC ACT GTG TCA AGT TTC CAG GTA GAT TGC TTC CTT TGG CAT GTC CGA AAA CAA GTT
                            MET ASP SER ASN THR VAL SER SER PHE GLN VAL ASP CYS PHE LEU TRP HIS VAL ARG LYS GLN VAL

       100        110        120        130        140        150        160        170        180
GTA GAC CAA GAA CTA GGT GAT GCC CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AGG TCC CTA AGG GGA AGA GGC AGC ACT CTC GGT CTA
VAL ASP GLN GLU LEU GLY ASP ALA PRO PHE LEU ASP ARG LEU ARG ARG ASP GLN ARG SER LEU ARG GLY ARG GLY SER THR LEU GLY LEU

       190        200        210        220        230        240        250        260        270
GAC ATC GAA GCA GCC ACC CAT GTT GGA AAG CAG ATA GTA[GAG AAG ATT CTG AAG GAA GAA TCT GAT GAG GCA CTT]AAA ATG ACC ATG GCC
ASP ILE GLU ALA ALA THR HIS VAL GLY LYS GLN ILE VAL[GLU LYS ILE LEU LYS GLU GLU SER ASP GLU ALA LEU]LYS MET THR MET ALA

       280        290        300        310        320        330        340        350        360
TCC ACA CCT GCT TCG CGA TAC ATA ACT GAC ATG ACT ATT GAG GAA TTG TCA AGA GAC TGG TTC ATG CTA ATG CCC AAG CAG AAA GTG GAA
SER THR PRO ALA SER ARG TYR ILE THR ASP MET THR ILE GLU GLU LEU SER ARG ASP TRP PHE MET LEU MET PRO LYS GLN LYS VAL GLU

       370        380        390        400        410        420        430        440        450
GGA CCT CTT TGC ATC AGA ATG GAC CAG GCA ATC ATG GAG AAG AAC ATC ATG TTG AAA GCG AAT TTC AGT GTG ATT TTT GAC CGA CTG GAG
GLY PRO LEU CYS ILE ARG MET ASP GLN ALA ILE MET GLU LYS ASN ILE MET LEU LYS ALA ASN PHE SER VAL ILE PHE ASP ARG LEU GLU

       460        470        480        490        500        510        520        530        540
ACC CTA ATA TTA CTA AGG GCT TTC ACC GAA GAG GGA GCA ATT GTT GGC GAA ATC TCA CCA TTG CTT TCT TTT CCA GGA CAT ACT ATT GAG
THR LEU ILE LEU LEU ARG ALA PHE THR GLU GLU GLY ALA ILE VAL GLY GLU ILE SER PRO LEU LEU SER PHE PRO GLY HIS THR ILE GLU
                                                                                                ASP ILE LEU LEU ARG

       550        560        570        580        590        600        610        620        630
GAT GTC AAA AAT GCA ATT GGG GTC CTC ATC GGA GGA CTT GAA TGG AAT GAT AAC ACA GTT CGA GTC TCT AAA ACT CTA CAG AGA TTC GCT
ASP VAL LYS ASN ALA ILE GLY VAL LEU ILE GLY GLY LEU GLU TRP ASN ASP ASN THR VAL ARG VAL SER LYS THR LEU GLN ARG PHE ALA
 MET SER LYS MET GLN LEU GLY SER SER SER GLU ASP LEU ASN GLY MET ILE THR GLN PHE GLU SER LEU LYS LEU TYR ARG ASP SER LEU

       640        650        660        670        680        690        700        710        720
TGG GGA AGC AGT AAT GAG AAT GGG GGA CCT CCA CTT ACT CCA AAA CAG AAA CGG AAA ATG GCG AGA ACA GCT AGG TCA AAA GTT CGA AGA
TRP GLY SER SER ASN GLU ASN GLY GLY PRO PRO LEU THR PRO LYS GLN LYS ARG LYS MET ALA ARG THR ALA ARG SER LYS VAL ARG ARG
 GLY GLU ALA VAL MET ARG MET GLY ASP LEU HIS LEU LEU GLN ASN ARG ASN GLY LYS TRP ARG GLU GLN LEU GLY GLN LYS PHE GLU GLU

       730        740        750        760        770        780        790        800        810
GAT AAG ATG GCT GAT TGA AGA AGT GAG ACA CAG ACT GAA AAC AAC TGA GAA TAG TTT TGA GCA AAT AAC ATT CAT GCA AGC CTT ACA GCT
ASP LYS MET ALA ASP ***
 ILE ARG TRP LEU ILE GLU GLU VAL ARG HIS ARG LEU LYS THR THR GLU ASN SER PHE GLU GLN ILE THR PHE MET GLN ALA LEU GLN LEU

       820        830        840        850        860        870        880        890
ACT ATT TGA AGT GGA ACA GGA GAT AAG AAC TTT CTC ATT TCA GCT TAT TTA A TGATAAAAAACACCCTTGTTTCTACT

 LEU PHE GLU VAL GLU GLN GLU ILE ARG THR PHE SER PHE GLN LEU ILE ***
```

and to determine the parental origin of the CR43-3 NS gene, it was necessary to obtain sequence information of the NS genes of the parent viruses. We compared the CR43-3 NS sequence to the previously obtained partial sequence of the NS gene of the wild-type A/AA/6/60 virus (Parvin et al., 1983). In the region sequenced, 6 out of 55 nucleotides differed from those in the CR43-3 NS sequence. More recently, the complete nucleotide sequence of the wild type A/AA/6/60 virus NS gene has been determined (this work). There are 45 nucleotide differences in the NS genes of the A/AA/6/60 and A/ALA/6/77 viruses. These base differences make it very unlikely that CR43-3 derives its NS gene from the A/AA/6/60 parent. Therefore, we proceeded to clone and sequence the NS gene of A/ALA/6/77 virus. Excluding the area of the deletion (positions 222 through 257), the NS gene sequences of A/ALA/6/77 and CR43-3 viruses were identical, indicating that CR43-3 derived its NS gene from the A/ALA/6/77 parent (Fig. 28).

3C. Direct RNA Sequencing

To eliminate the possibility that the deletion in the NS-specific cDNA of CR43-3 virus was an artifact generated in the cloning procedure, RNA 8 of CR43-3 was directly sequenced by the dideoxy method in the region corresponding to nucleotides 166-290 of the A/ALA/6/77 NS gene (Fig. 29).

Fig. 29. Direct sequencing of RNA 8 of CR43-3 virus. The primer used to generate this sequence was an EcoRI-DdeI restriction fragment corresponding to nucleotide positions 1 through 153 of the cloned NS gene of A/ALA/77 virus. The bases indicated extend from nucleotide position 217 of the CR43-3 NS segment. Beyond nucleotide position 221, the CR43-3 virus NS sequence is different from that of A/ALA/77 virus. The dideoxy nucleotide-terminated DNA chains were fractionated in an 8% polyacrylamide-7 M urea sequencing gel.


G
TACCAGTAAAAATGAT
221
217

3D. Amino Acid Sequences

The deduced NS1 and NS2 amino acid sequences of the A/ALA/6/77 and CR43-3 NS genes are also presented in Fig. 28. The NS1 protein of A/ALA/6/77 consists of 237 amino acids, whereas the NS1 of CR43-3 virus consists of only 225 amino acids. The deletion in the NS1 protein of CR43-3 virus corresponds to positions 66 through 77 of the A/ALA/6/77 NS1 protein. The deduced length of the NS2 polypeptides of both A/ALA/6/77 and CR43-3 viruses was found to be 121 amino acids, which is identical to that reported for all other NS2 proteins of influenza A viruses (Baez et al., 1980; Baez et al., 1981; Krystal et al., 1983a; Lamb and Lai, 1980; this thesis).

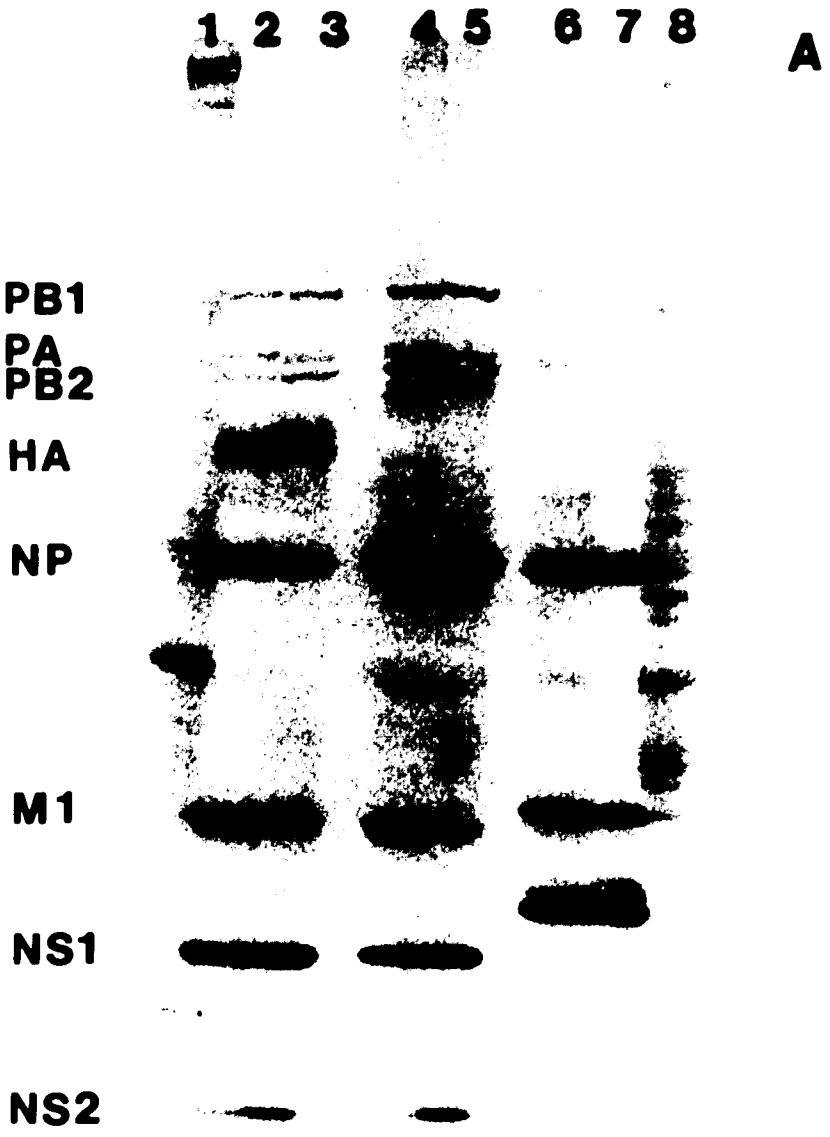
3E. Viral Protein Analysis

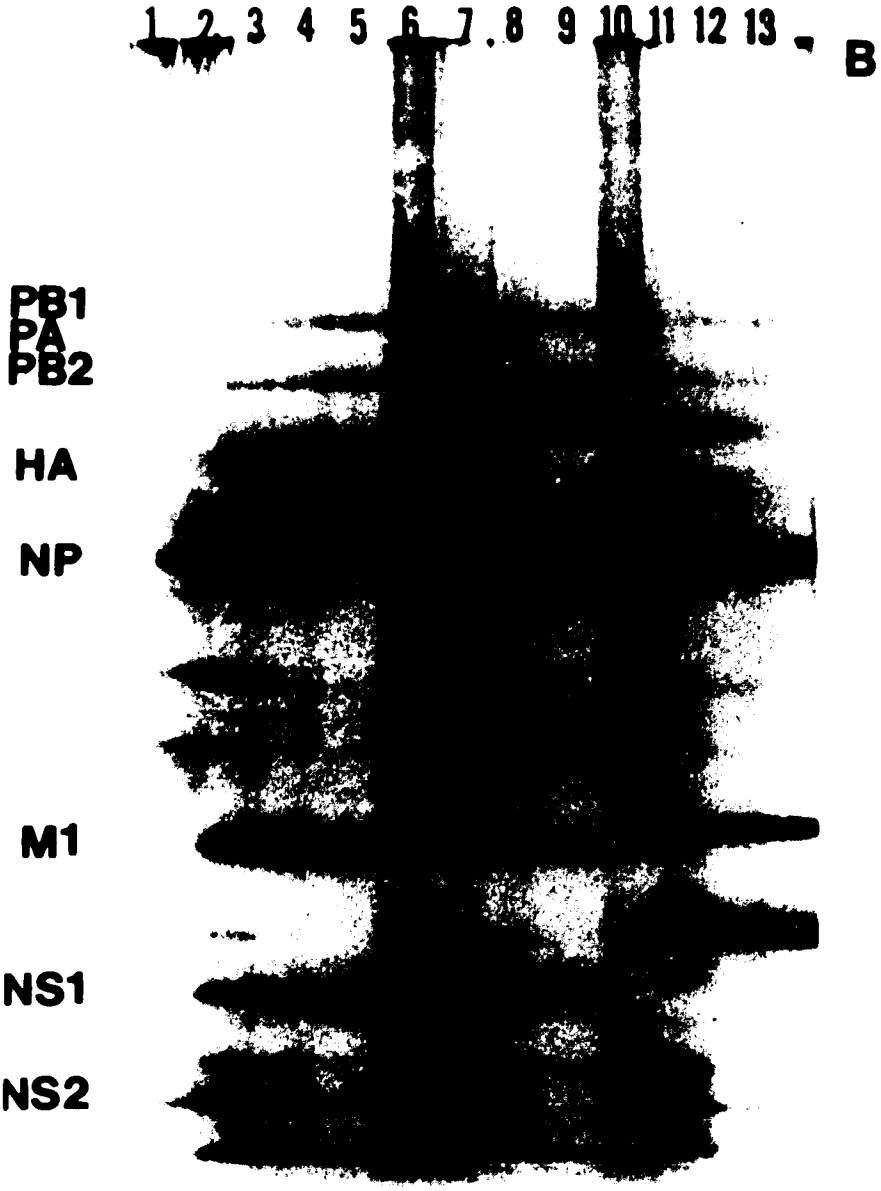
Since CR43-3 virus shows a restricted host range phenotype, [^{35}S] methionine-labeled proteins of MDCK and PCK cells infected with CR43-3 and its parent viruses were analyzed on SDS-polyacrylamide gels (Fig. 30A). No major differences in protein synthesis patterns could be detected in A/AA/6/60 virus-infected MDCK and PCK cells at 8 h p.i. (Fig. 30A, lanes 2 and 3). Similarly, cell extracts from A/ALA/6/77 virus-infected MDCK and PCK cells showed comparable protein patterns after polyacrylamide gel electrophoresis (Fig. 30A, lanes 6 and 7). This result was not unexpected, since neither virus displays a growth

restriction in the two host cells. Analysis of MDCK and PCK cells infected with CR43-3 virus (8 h p.i.) revealed that the NS1 protein is synthesized in comparable amounts in both cell systems (Fig. 30A, lanes 4 and 5). However, from this and other gels (data not shown) it was found that the M1 protein is reduced by approximately 50% relative to the NS1 polypeptide in the nonpermissive MDCK cells as compared with the permissive PCK cells infected with CR43-3 virus. This finding was obtained by quantitating densitometer tracings of different autoradiographs and was confirmed by time course experiments.

Figure 30B shows that M1 protein synthesis in CR43-3 virus-infected MDCK cells was reduced throughout the infection cycle relative to that in the parental virus-infected MDCK cells. Time points were taken at 3, 6, 9, and 12 h p.i. (Fig. 30B, lanes 2 through 13). This time course experiment further suggests that the NS1 defect in CR43-3 virus causes a reduction in M1 protein synthesis. It should be noted that the A/ALA/6/77 HA (H3 subtype) produced in MDCK or in PCK cells does not form a sharp band under the gel conditions used and thus cannot be identified on the autoradiogram. In contrast, the A/AA/6/60 HA, belonging to the H2 subtype, forms visible bands on gels obtained with extracts from infected PCK and MDCK cells (Fig. 30A and 30B).

Fig. 30. SDS-PAGE of viral proteins synthesized in PCK and MDCK cells. (A) Cells were infected with virus at 34°C and pulse-labeled for 1 h at 8 h p.i. with [^{35}S]-methionine as described in the Materials and Methods. Lanes: 1, mock-infected PCK cells; 2, A/AA/60 virus-infected PCK cells; 3, A/AA/60 virus-infected MDCK cells; 4, CR43-3 virus-infected PCK cells; 5, CR43-3 virus-infected MDCK cells; 6, A/ALA/77 virus-infected PCK cells; 7, A/ALA/77 virus-infected MDCK cells; 8, mock-infected MDCK cells. Virus-specific proteins are indicated. (B) MDCK cells were infected with virus at 34°C and pulse-labeled for 1 h at the indicated times. Lanes 2, 6, and 10 were infected for 3 h; lanes 3, 7, and 11 were infected for 6 h; lanes 4, 8, and 12 were infected for 9 h; lanes 5, 9, and 13 were infected for 12 h. Lanes: 1, mock-infected MDCK cells; 2 through 5, A/AA/60 virus-infected MDCK cells; 6 through 9, CR43-3 virus-infected MDCK cells; 10 through 13, A/ALA/77 virus-infected MDCK cells. Virus-specific proteins are indicated.

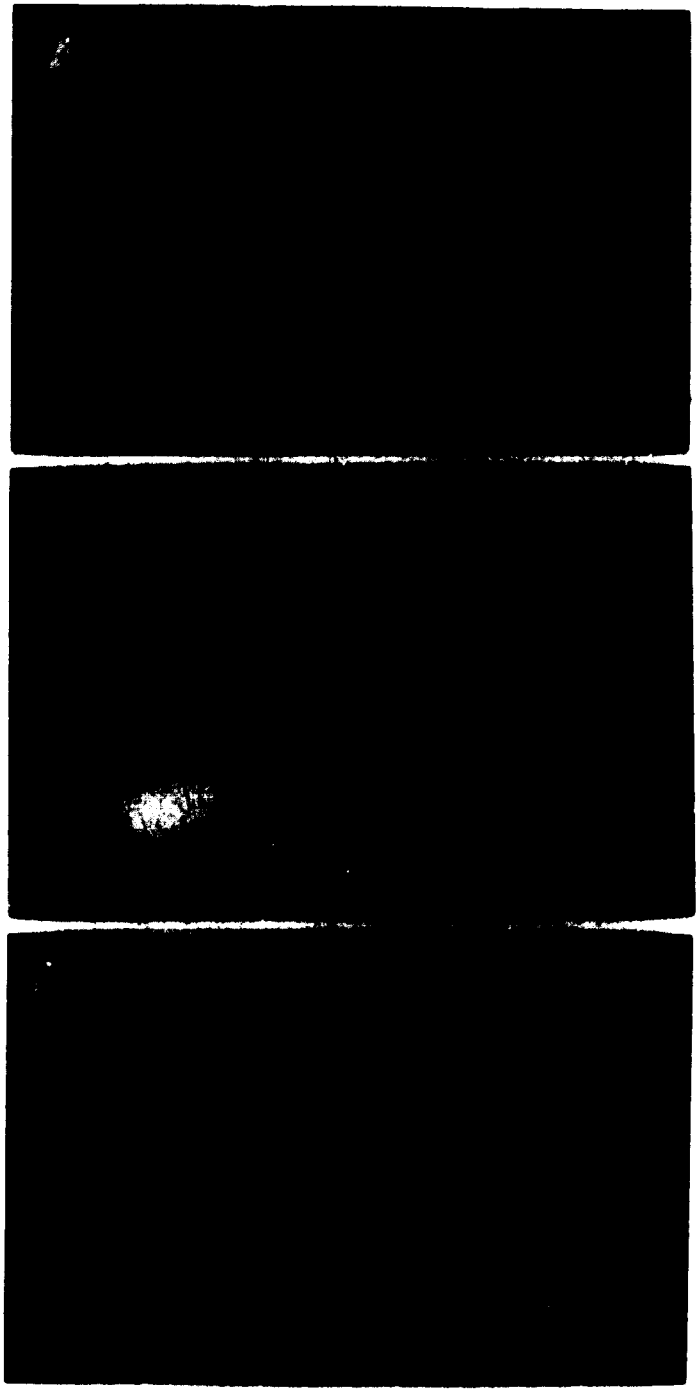

1 2 3 4 5 6 7 8 A
PB1
PA
PB2
HA
NP
M1
NS1
NS2


1 2 3 4 5 6 7 8 9 10 11 12 13
B
PB1
PA
PB2
HA
NP
M1
NS1
NS2

3F. Indirect Immunofluorescence Analysis of Virus-infected Cells

MDCK cells infected with CR43-3 and its parent viruses were acetone-fixed 3 h p.i. and stained by using monospecific rabbit antiserum against NS1 and fluorescein-conjugated goat anti-rabbit antibody. Both parents, as well as CR43-3 virus, induced the synthesis of NS1, which was found localized in the nucleoplasm of MDCK cells (Fig. 31). This indicates that the mutant protein accumulates in the same cellular compartment in which the wild-type protein is found during the course of infection. At later time points (6, 9, and 12 h p.i.) CR43-3 and A/AA/6/60 virus-infected MDCK cells continued to display intense nucleoplasmic staining. Uninfected cells treated with the anti-NS1 rabbit antiserum and fluorescein-conjugated goat anti-rabbit antibody did not exhibit any positive immunofluorescence (data not shown). It should be noted, however, that the A/ALA/6/77 virus-infected MDCK cells exhibited NS1-specific nucleolar staining at later time points (6, 9, and 12 h p.i.). Previously, it had been observed that some influenza virus strains, but not others, display nucleolar staining specific for NS1 late in infection. This observation, however, could not be correlated with any biological activity of these strains (Young et al., 1983). It thus appears that the immunofluorescence studies do not point to an unusual

Fig. 31. Indirect immunofluorescent staining of NS1 protein in influenza virus-infected MDCK cells. MDCK cell monolayers were adsorbed with virus for 1 h at 34°C. At 3 h p.i., the cells were acetone-fixed and incubated with a monospecific rabbit anti-NS1 antibody followed by an anti-rabbit fluorescein-conjugated antibody. Shown are MDCK cells infected with (A) CR43-3 virus, (B) A/ALA/77 virus, and (C) A/AA/60 virus. Cells infected with any of the three viruses show intense nucleoplasmic staining at this early time point.



behavior of the mutant NS1 polypeptide in the MDCK cells nonpermissive for CR43-3 virus. Control studies with PCK cells, which are permissive for all three viruses, showed the same staining pattern for NS1 as observed in MDCK cells (data not shown).

4. Analysis of an Influenza B Virus NS Deletion Mutant

Influenza virus clone 201 was generated in the laboratory by repeated backcrosses of influenza virus A/Aichi/2/68 (H3N2) with B/Yamagata/1/73 (B/YA/73) in MDCK cells (Tanaka et al., 1984). Clone 201 (B serotype) forms sharp plaques in MDCK cells and induces severe cell lysis early in infection. Parallel fractionation in polyacrylamide gels of the RNAs of clone 201 along with those of the parent viruses revealed that the genes of clone 201 had identical mobilities to those of the B/YA/73 parent, except for RNA segment 8 (NS). The NS gene of clone 201 migrated faster than the corresponding segment of B/YA/73. Analysis of the proteins synthesized in MDCK cells infected with B/YA/73 and clone 201 viruses showed that in clone 201-infected cells, a nonstructural protein having a similar mobility to the NS1 of B/YA/73 virus could not be detected. Instead, the predominant nonstructural protein (NS1) of clone 201 migrated much faster in the gel, near the position of the NS2 protein of B/YA/73.

To characterize the possible defect in the NS gene of clone 201, the nucleotide sequences of the NS genes of the B/YA/73 and clone 201 viruses were generated by the dideoxy method using oligonucleotide primers. A comparison of the NS sequences is shown in Fig. 32. A 13 base deletion was detected in the NS gene of clone 201, corresponding to nucleotide positions 374-386 in the B/YA/73 NS gene. The autoradiogram of the nucleotide sequence in this region is presented in Fig. 33. Except for the deletion, no other nucleotide differences were detected in the NS genes of B/YA/73 and clone 201 (Fig. 32). It appears, therefore, that clone 201 derives its NS gene from the B/YA/73 parent virus.

The NS deletion of clone 201 alters the NS1 coding region of the gene (Fig. 32). At the deletion point, the NS1 reading frame is shifted and an early amber termination codon is reached at nucleotides 437-439 (18 codons downstream of the deletion). The deduced NS1 protein of clone 201 consists of 127 amino acids, the last carboxy-terminal 17 of which are encoded in a different reading frame than that which codes for the wild type NS1 protein. In contrast, the predicted NS1 polypeptide of the B/YA/73 parent strain contains 281 amino acids, more than twice as many as the NS1 of clone 201. The deduced NS2 proteins of B/YA/73 and clone 201 are identical in amino acid sequence and consist of 122 amino acids (Fig. 32).

Fig. 32. Comparison of nucleotide and deduced NS1 and NS2 amino acid sequences of influenza B/YA/73 and clone 201 viruses. Nucleotide sequences (plus sense strand) were obtained by direct RNA sequencing using synthetic oligonucleotide primers and reverse transcriptase. The primers were initially based on the published sequence of the NS gene of B/Lee/40 virus (Breidis and Lamb, 1982). The primers correspond to the following nucleotide positions of the NS gene: 13-26; 86-101; 293-307; 645-660; 965-980. Sequences at the 5' end of the NS gene (bases 1-42) could not be determined. The ATG initiation codon for the NS1 and NS2 polypeptides is at nucleotide positions 43-45. Arrows indicate the putative splice sites generating the NS2 mRNA. The dashed line indicates a 13 nucleotide deletion of clone 201, resulting in a frame shift in the NS1 coding region as indicated by the amino acid sequence below the nucleotide sequence. Asterisks represent stop codons in the reading frames of the NS1 and NS2 proteins.

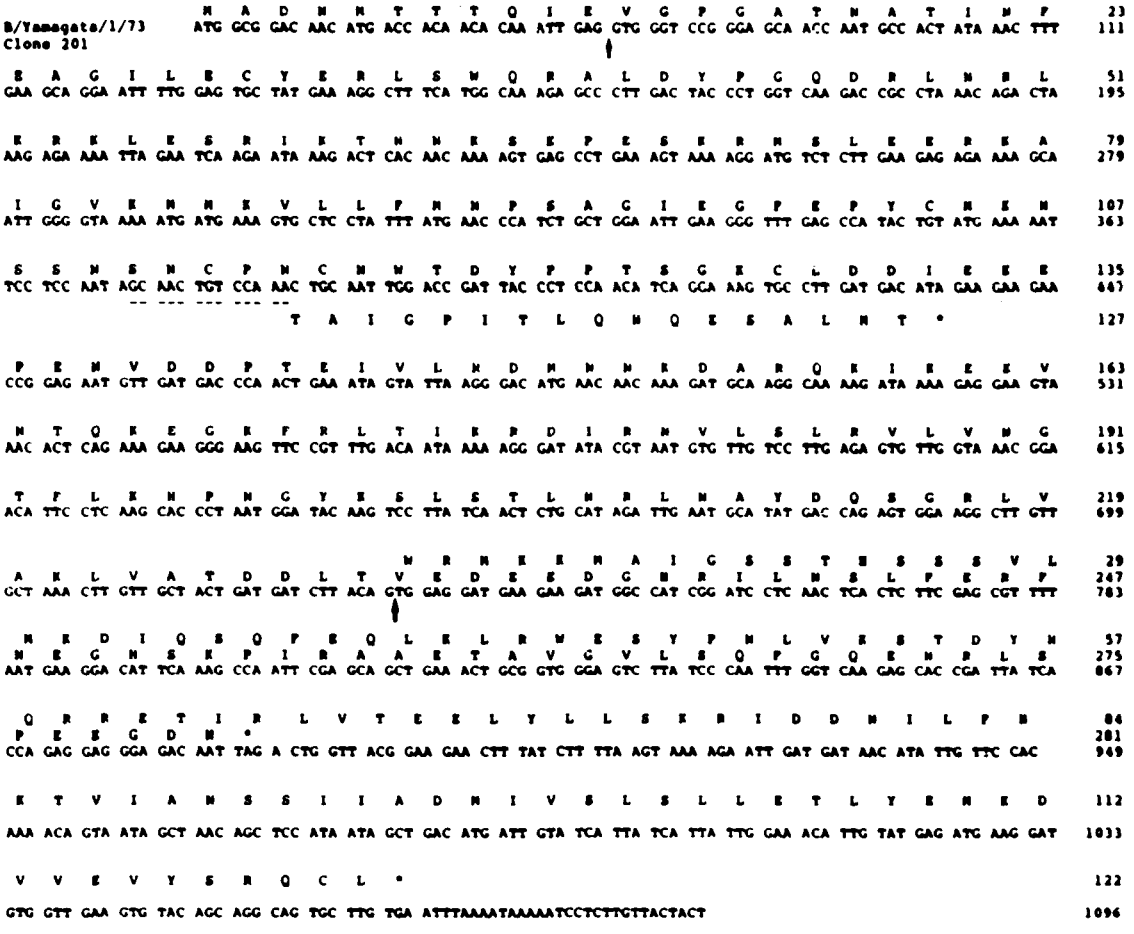
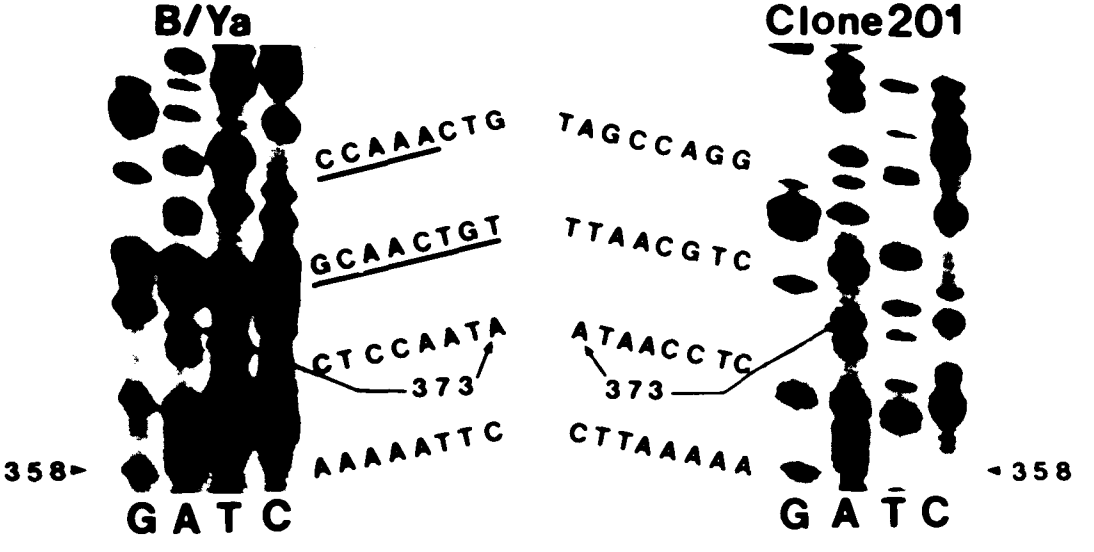

B/Yamagata/1/73
Clone 201

Fig. 33. Sequence at the deletion site of the NS gene of influenza B virus clone 201. The NS genes of clone 201 and B/YA/73 viruses were determined by the dideoxy method. The deletion site in the NS gene of clone 201 virus (right side) was compared to the B/YA/73 virus NS gene sequence (left side). Numbers correspond to nucleotide positions shown in Fig. 32. The arrows at position 373 indicate the nucleotide juxtaposed to the deletion site in the NS gene of clone 201. The nucleotides underlined in the sequence of the B/YA/73 virus NS gene are not found in the NS gene of clone 201 virus.


B/Ya
Clone201
CCAAACTG
GCAACTGT
CTCCAATA
373
AAAAATTC
358
GATC
TAGCCAGG
TTAACGTC
ATAACCTC
373
CTTAAAAA
358
GATC

Norton et al. (submitted) have demonstrated that the nonstructural protein of altered mobility synthesized in clone 201 virus-infected cells is indeed the NS1 protein through immunoprecipitation with monospecific rabbit anti-NS1 antiserum made against the purified NS1 polypeptide (281 amino acids) of B/Lee/40 virus. Additionally, Norton et al. (submitted) have shown by immunofluorescence assay that the NS1 protein of B/YA/73 and clone 201 viruses is localized in the nucleus of infected cells. Thus, the karyophilic pattern of the NS1 protein of influenza B viruses is similar to that observed for the NS1 of influenza A viruses (Young et al., 1983; Greenspan et al. 1985).

DISCUSSION

1. Evolution of Influenza A Viruses

1A. Evolutionary Rate of Influenza A Virus Genes

The evolution of human influenza A viruses was investigated by comparative nucleotide sequence analysis of the NS gene derived from viruses isolated over five decades. The NS genes of the 15 viruses belong to a common evolutionary lineage, indicated by the fact that the NS gene has accumulated sequential nucleotide changes over time. A phylogenetic tree was constructed using the maximum parsimony procedure of Fitch (1971) which subsequently allowed us to calculate an evolutionary rate for the NS gene. The rate of nucleotide substitution in the NS gene is 1.94×10^{-3} substitutions per site per year. Evolutionary rates were also calculated for the H3 HA and N2 NA genes by maximum parsimony analysis of previously published HA and NA nucleotide sequences. The nucleotide substitution rates for the H3 HA (HA1 domain) and the N2 NA genes are 6.72×10^{-3} and 3.17×10^{-3} substitutions per site per year, respectively. The higher evolutionary rates for the HA and NA genes may be due to immune selective pressure exerted on the HA and NA surface proteins of the virus. Immune pressure may not play a significant role in the evolution of the nonstructural gene products, found only inside of the virus-infected cell.

One point to be made on the evolutionary pattern observed for influenza A virus genes is that there appears to be a linear accumulation of base substitutions in the viral genes over time. This near constant rate of evolution for influenza virus genes has also been reported by Hayashida et al. (1985) and is consistent with the molecular clock model of evolution (for review, Wilson et al., 1977).

The possibility that eukaryotic macromolecular sequences evolved at constant rates was first recognized by Zuckerkandl and Pauling (1965). Fitch and Langley (1976) published a rigorous examination of the rates at which different mammalian genes evolve. They estimated phyletically the minimum number of nucleotide substitutions required to account for amino acid sequence differences found between various polypeptides of selected mammalian species. The study included cytochrome C, myoglobin, hemoglobin alpha-chain, hemoglobin beta-chain, insulin C peptide, fibrinopeptide A, and fibrinopeptide B. The results of this study showed that each protein has its own characteristic evolutionary rate, but there appeared to have been an approximately constant rate of fixation of those nucleotide substitutions which resulted in amino acid changes over the past 120 million years (the time since the most distantly related mammals diverged from a common ancestor).

More recently, Hayashida and Miyata (1983) and Miyata et al. (1982) have provided evidence that the rate of nucleotide substitution at silent positions in protein coding regions and in introns (for various mammalian genes) is approximately uniform (clock-like) for different genes. This result is in contrast to the rate of protein evolution which varies greatly for different mammalian proteins.

Reliable evolutionary rate calculations depend on reliable estimates of divergence times between species. Divergence requires the separation of one species into two non-interbreeding species (time since two species had a common ancestor). Estimates of divergence times depend on the existence of an abundant and properly interpreted fossil record. A source of error in the evolutionary rate calculations, which are the basis of the molecular clock, may thus be attributed to uncertainties involved in estimating time of divergence.

In contrast, the NS gene of influenza virus behaves as a verifiable and accurate molecular clock in nature. The clock is verifiable because its calibration does not have to rely upon inaccurate fossil dates since the time of isolation of the virus strains is known precisely. The evolutionary rate calculation for the NS gene is not plagued by errors introduced by virtue of unreliable divergence time estimates.

The evolutionary rates of the H3 HA and N2 NA genes also appear to by clock-like over time. However, evolutionary change in the HA gene is only examined over a 12-year period (1968-1980), a much shorter time span than the 53 years over which the NS gene was studied. The evolutionary data for the N2 NA gene was generated using only six strains isolated over 22 years. Clearly, more sequence information is needed to assess with greater confidence the clock-like behavior of this viral gene.

A second point concerns the extremely rapid nucleotide substitution rates for influenza A virus genes. This observation has also been reported by other laboratories (Air and Hall, 1981; Hayashida et al., 1985; Saitou and Nei, 1986). The rapid evolution of the influenza virus genome appears to be a million-fold higher than that of chromosomal genes of the eukaryotic host (Table 18). Evolutionary rates for mammalian genes are on the order of 10^{-9} substitutions per site per year (Li and Gojobori, 1983). It should be noted that these values do not take into account the high rate of somatic mutation (10^{-3} mutations per base per division) reported for immunoglobulin variable region genes (Clarke et al., 1985). Somatic mutation arises following germ-line gene rearrangements required to generate antibody secreting B cells.

TABLE 18. EVOLUTIONARY RATES (SUBSTITUTIONS/SITE/YEAR)*

Influenza A Virus:	NS	1.94×10^{-3}
	H3 HA**	6.72×10^{-3}
	N2 NA	3.17×10^{-3}
Hemoglobin:	Alpha Chain	4.50×10^{-9}
	Beta Chain	3.83×10^{-9}
Interferon:	Alpha-1	4.94×10^{-9}
	Beta-1	8.09×10^{-9}
	Gamma	11.39×10^{-9}
Histone:	H4	6.16×10^{-9}

*Mammalian gene values from Li et al., 1985

**HA1 domain

The influenza virus system provides a unique opportunity to study evolutionary processes in viral genes which are manifested over a relatively short time frame (decades). In contrast, the eukaryotic chromosomes of the host cell take millions of years to achieve the same extent of molecular variation that viral genes can attain in just a few human generations.

Influenza A viruses also appear to evolve more rapidly in nature than other RNA or DNA containing viruses. One explanation for the extreme variability of influenza viruses may be that they have a higher mutation rate than other viruses. However, many other viruses, including vesicular stomatitis virus (Holland et al., 1982; Spindler et al., 1982; O'Hara et al., 1984), parainfluenza virus (Portner et al., 1980), foot and mouth disease virus (Sobrino et al., 1983), RNA tumor viruses (Zarling and Temin, 1976; Darlix and Spahr, 1983; Gojobori and Yokoyama, 1985), and herpesviruses (Holland et al, 1983), have also been shown to possess the potential to undergo rapid genetic change. Furthermore, the mutability of these viral genomes has been assessed mainly by replicating viruses in tissue culture systems. For most viruses, passage conditions (temperature, multiplicity of infection, host cells) can be found which can lead to the generation of mutant viral genomes at high frequency.

It has been reported that viruses containing RNA genomes have higher mutation frequencies than their host chromosomes and viruses which possess their genetic information in the form of DNA (Holland et al., 1982). It is thought that RNA polymerases exhibit less fidelity of genome replication than DNA polymerases, due in part, to a lack of proofreading enzymes. The proofreading exonuclease activity of DNA polymerase is able to remove misincorporated bases from newly synthesized DNA strands so that errors in DNA replication average as low as 10^{-8} to 10^{-11} per incorporated nucleotide. In contrast, studies of the bacteriophage QB by Domingo et al. (1978) have shown that the error level per genome doubling at given base positions in the RNA genome is between 10^{-3} and 10^{-4}.

Previous attempts to calculate and compare mutation rates for RNA viruses have indicated that the mutation rate for the influenza A virus does not appear to be higher than that of other viruses. One technique to assess mutation rates of viruses is to measure the frequency of antigenic variants resistant to neutralization by a selecting monoclonal antibody. It has been reported that the frequency of resistant mutants is similar for influenza virus, Sendai virus, vesicular stomatitis virus, poliovirus, and Coxsackie virus and is in the range of 10^{-4} to 10^{-5} (Portner et al., 1980; Prabhakar et al., 1982; Lubeck et al., 1980; Emini et al., 1983; Minor et al.,

1983). This approach, however, is limited in that only a small region of the genome can be probed for changes using the monoclonal antibody selection. The target size for mutant selection is frequently unknown and may often include more than one gene. Additionally, differences in mutation frequencies generated from resistant mutant selection experiments, may by explained by differences in avidity of the discriminating monoclonal antibodies.

Steinhauer and Holland (1986) have quantitated polymerase error frequencies for vesicular stomatitis virus at one highly conserved nucleotide site using both in vivo and in vitro assays. The extremely high frequency of base misincorporation is approximately 10^{-4} substitutions per base incorporated at the site in both assays. One shortcoming of this assay is that the substitution frequency at a single nucleotide site may not be representative of all the nucleotide positions of the genome RNA.

Recently, Parvin et al. (in press) have assayed the mutation rate in tissue culture for the NS gene of influenza A virus and for the VP1 gene of type 1 poliovirus. Each gene was directly sequenced in over one hundred randomly selected viral clones which had arisen from a single virion in one plaque generation. The VP1 gene encodes one of the structural proteins comprising the capsid of the poliovirus. The results of this experiment

are quite surprising in light of the earlier data on mutation rates of RNA viruses. Seven mutants of the influenza A virus NS gene were detected, whereas, no VP1 gene poliovirus mutant was observed. The calculated mutation rates are 1.5×10^{-5} and less than 2.1×10^{-6} mutations per nucleotide per infectious cycle for the influenza A virus and the poliovirus-1, respectively. These values represent "neutral" mutation rates since the viral mutants were not compromised in their replication (compared to wild type virus) as judged by the ability to form a plaque in tissue culture.

The mutation rate of influenza A virus was found to be significantly higher than that of poliovirus-1. It is suggested that the absence of poliovirus VP1 mutants may be the result of not only greater functional constraint on the VP1 gene, but also may be a reflection of a viral RNA polymerase with enhanced fidelity of replication. These data suggest that the higher mutation rate of the influenza A virus genome is a contributing factor in the extreme variability these viruses display in nature.

However, a high mutation rate does not provide an adequate explanation for the unusual shape of the NS, HA, and NA genealogical trees. The evolutionary tree of the NS gene, for example, has a long, slender structure. The side branches of the tree do not greatly diverge from the main tree vertical axis (the average age of the side branches is

only three years). This appears to be a consequence of the short life span of any lineage other than the one that gives rise to the future generations. This is very reminiscent of periodic selection in chemostats (Atwood et al., 1951) where the number of accumulated mutants in E. coli suddenly drops to zero when a new favorable mutant replaces the older strains, so that at any point in time all coexisting strains are recently diverged from the main lineage. Animal genes and genes from viruses, except possibly those of enterovirus 70 (Tanimura et al., 1985), do not follow this pattern. Rather, they appear to have multiple surviving lineages undergoing slower change, for example, the influenza C viruses in man (see section on influenza C virus in discussion). In addition, the length of time a vaccine is effective against a viral pathogen may correlate with the evolutionary rate of the virus. Vaccines for all three poliovirus types are made with isolates obtained approximately five decades ago. Similarly, the yellow fever virus vaccine was developed more than 50 years ago and current isolates have not sufficiently changed to warrant a new vaccine formula. Also, the currently used rabies virus vaccine strain goes back to Pasteur's time, whereas the influenza A viruses used for vaccine manufacture are changed every two to three years (even in the absence of a subtype shift) to accomodate changes in the prevalent strains.

An explanation for the unusual pattern and rapid rate of NS gene evolution may be found in positive selection of influenza A virus variants. There is no evidence for immune surveillance of the NS gene products nor have other selectional forces on the NS gene been identified. It may be that only one influenza virus gene (most likely the hemagglutinin) needs to be subject to selection. In the brief time before immunity develops to a new (antigenic) variant, that strain may sweep through the population, carrying with it whatever variant of the NS gene happens by chance to be present. In this way, the NS gene's phylogeny may be linked (hitch-hiking) to that of another gene undergoing extensive positive selection. The fixation of substitutions in the NS genes is not, therefore, simply the result of random genetic drift and one probably cannot calculate a "neutral mutation rate" for the A virus NS gene from these data.

The neutralist theory (Kimura, 1983) asserts that most of the evolutionary changes occurring at the molecular level are neither beneficial nor deleterious, but neutral in terms of fitness of the organism. Evolution, therefore proceeds by the random fixation of selectively neutral mutations. The neutralists do not deny the occurrence of advantageous alleles that are subject to Darwinian natural selection, but believe that these adaptive changes are extremely rare. The neutral theory also does not deny the

occurrence of deleterious mutations which are subject to negative selection (functional constraint), but neutralists assert that it occurs more frequently than positive selection.

If neutral changes are randomly fixed in the NS gene, we would expect to see many variants cocirculating. This is not the case for influenza A viruses, where a single lineage predominates at one time. Therefore, it appears that the evolution of influenza A virus genes is shaped by selectional forces. The high mutation rate of influenza A viruses may provide the raw material upon which selection is operative, that is, the generation of multiple genetic variants. In such a diverse pool of variants, maybe only one will be successful in replication and transmission in the face of host immunity, and thus be positively selected. We suggest that positive selection represents a significant factor in the unusual pattern and high rate of influenza A virus evolution in man.

1B. NS1 and NS2 Protein Sequences

The NS1 and NS2 proteins of the 15 influenza A virus strains display a great deal of amino acid sequence heterogeneity, as would be predicted based on the corresponding nucleotide sequences. Amino acid substitutions are distributed throughout the NS1 and NS2 molecules and many of the changes appearing in the earliest virus isolates examined are retained in the protein

sequences of viruses isolated at later dates (Fig. 11 and Fig. 14).

In the NS1 polypeptide, 21.1% of the amino acid positions show amino acid changes among the viruses, whereas only 14.0% of the NS2 sites are substituted. It then appears that the NS2 protein sequence is subject to greater constraint than that of the NS1. Additionally, there is heterogeneity in the size of the NS1 proteins of the viruses (202-237 amino acids) as summarized in Table 12. The size differences have been mapped to the carboxy terminus of the NS1 protein, where the NS1 molecules of some strains have deleted some amino acids. These data suggest that the carboxy terminus of the NS1 may not be a critical region with respect to the proper function of the protein in viral replication.

In contrast, the NS2 proteins of the viruses all possess 121 amino acids. There have been no reports of an A virus field strain which appears to code for an NS2 polypeptide whose amino acid length deviates from 121 residues. It may be that any deletion in the NS2 protein cannot be tolerated, since such an alteration may disrupt proper folding of the protein, and thus render the polypeptide nonfunctional. The NS2 protein may exert its effect by an intimate interaction with another viral or host protein or RNA and any slight perturbation of the secondary structure (through amino acid substitution or

deletion) might preclude this necessary interaction. It is difficult at this time to assess the functionally important sites of the NS1 and NS2 proteins because the precise roles of these proteins in the influenza virus life cycle have not been elucidated.

1C. Reemerged H1N1 Subtype Viruses

The nucleotide sequence comparison of the NS genes of the H1N1 A/FW/50 virus and the reemerged H1N1 A/USSR/77 prototype strain revealed that only five base changes accumulated in 27 years (Table 9). These data confirm previous results suggesting that the H1N1 epidemic strains reintroduced in 1977 are closely related to H1N1 viruses which circulated around 1950 (Nakajima et al., 1978; Kendal et al., 1978; Scholtissek et al., 1978b). Similarly, the H1 HA proteins of the A/FW/50 virus and the A/USSR/77 strain only possess four amino acid differences in the 239 residues of the HA1 domain compared by Raymond et al. (1986).

The evolutionary changes in the NS genes of the 1950-1957 H1N1 viruses were compared to those in the NS genes of the new H1N1 viruses of 1977-1985. The data revealed that nucleotide substitutions in the NS genes of viruses of the 1950-1957 period (A/FW/50, A/DEN/57) were quite distinct from those detected in the NS sequences of the reemerged H1N1 strains of 1977-1985 (A/USSR/77, A/MD/80, A/HT/84,

A/HT/85). These results suggest that the NS gene of the H1N1 subtype has evolved along two divergent lineages in the two periods of its circulation in man. Raymond et al. (1986) have published similar findings which involved comparing both nucleotide and amino acid changes in the HA1 domain of the HA molecules of five H1N1 viruses of the 1950-1957 epidemic period and 14 isolates of the 1977-1983 H1N1 epidemic period. They concluded that the HAs of the earlier and later H1N1 period have followed two different evolutionary pathways. It is not surprising that divergent evolution is observed for the H1 subtype HA. In the two periods of H1N1 circulation, it is not difficult to imagine that environmental conditions were different, especially the immune status of the population. During the 1950s, the H1N1 viruses circulated in the presence of high levels of population immunity (most people have anti-H1 antibodies). In the subsequent 1977 to the present era, H1N1 viruses circulated in young people born after 1957, many of whom had no preexisting antibody. Immune pressure in the two periods probably exerted different influences on the virus, causing the H1N1 virus to evolve in dissimilar ways.

The observations on the evolution of the H1N1 subtype demonstrate the extreme flexibility of the virus in its capacity to change and what may be an unlimited potential of the influenza virus to evolve successfully in many directions and escape the host immunity. These results

suggest the improbability that one can predict the molecular makeup of future epidemiologically important influenza A viruses based on information on the nature of epidemic variants of the past.

2. Evolution of Influenza C Viruses

One specific goal of this thesis was to better understand the molecular basis of the epidemiology of influenza C viruses and to compare it with that of influenza A viruses. Earlier evidence suggested that influenza C viruses are antigenically more stable over time than A type viruses (Chakraverty, 1978; Meier-Ewert et al., 1981b) and that in this respect C viruses resemble other RNA viruses such as polio or parainfluenza viruses. Furthermore, oligonucleotide mapping analysis of the genomes of different influenza C viruses (Meier-Ewert et al. 1981b; Guo and Desselberger, 1984) also suggested less variation than was observed for influenza A viruses. Genetic variation of influenza C viruses was investigated by comparative nucleotide analysis of the HA and NS genes of eight human viruses isolated over approximately four decades (1947-1983).

A quantitative analysis of the HA genes of eight human influenza C virus strains revealed that both on the nucleotide and amino acid levels changes can be observed but they do not appear to accumulate with time. For example, as shown in Table 13 and Fig. 21, strains isolated

31 years apart (C/AA/50 and C/YA/81) may possess almost identical HA genes whereas strains isolated only one or two years apart may differ by many changes (e.g. C/MS/80 and C/YA/81 or C/Cal/78 and C/MS/80). This suggests that the strains of influenza C viruses examined do not directly share the same evolutionary pathway with respect to the HA genes. Rather, it appears that multiple evolutionary pathways exist and C virus variants of different lineages may cocirculate in nature for extended periods of time.

Variation on the nucleotide level is observed throughout the genes of the C virus HAs, including the portions coding for the HA1 and HA2 chains. It is interesting, however, that changes in the third position of codons are much more frequent (52-72%) than the first and second position substitutions taken together (Table 15). This finding suggests that nonsynonymous changes in the HA gene are subjected to negative selection, as has been observed for many other evolutionary systems (Nei, 1983; Kimura, 1983).

Variation among the different C virus HA proteins is limited, with a maximum of 4.5% overall amino acid difference observed for the HA genes of C/Cal/78 and C/MS/80 viruses (Table 14). In all of the C type HAs analyzed, the highly hydrophobic regions at the amino and carboxy termini of the HA2 portion and the arginine at the potential cleavage site between the HA1 and HA2 subunits

are preserved. Since the cysteine residues also remain conserved among the C virus HAs examined, it is likely that the different C virus HA molecules retain the same overall protein structure. In addition, all eight potential glycosylation sites identified for the C/Cal/78 virus HA (Nakada et al., 1984a) remain conserved in the other HA variants which were examined.

Amino acid comparison of all HA variants reveals three clusters of amino acid substitutions in the HA1 portions of the molecules centering on regions around amino acid positions 80, 205 and 340 (Fig. 22). Since there are no comparable clusters of amino acid changes in the HA2 portion of the proteins, it is tempting to speculate that these three clusters in the HA1 region are associated with antigenic characteristics of the influenza C virus HAs. However, in the absence of further studies, which may include the selection of antigenic variants through the use of monoclonal antibodies, the precise antigenic structure of the C virus HA remains unknown.

Recently, influenza C viruses were isolated from pigs in China (Guo et al., 1983). These swine C viruses were shown to be closely related to recent human C virus strains by protein and peptide mapping analysis (Elliott et al., 1984). A more detailed analysis based on oligonucleotide mapping data revealed differences between the human and swine C virus isolates, with the genomes of swine C virus

isolates of 1981 being the ones most closely related to each other (Guo and Desselberger, 1984). A comparison of the nucleotide and predicted amino acid sequences of the HA genes derived from the human and swine influenza C virus isolates confirms these results. Based on our sequence data it is impossible to draw conclusions as to the origin of the swine C viruses. Since at any one time different influenza C virus variants cocirculate in the human population, it is as probable, based on our results, that these swine viruses have recently passed from the human into the animal population as it is that C viruses have an animal reservoir from which the virus gets transmitted to the human population. Epidemiologic data and serologic analyses, however, suggest that the latter possibility is less likely (for review: Palese and Young, 1983).

Variation in the NS genes of eight human influenza C viruses was also examined. As shown in Fig. 23, nucleotide changes appearing in earlier viruses are not found in all later isolates. In addition, viruses isolated only three years apart can show 0.6% difference in their NS sequences (TAY/47 and AA/50) or 3.0% difference (CAL/78 and YA/81). Similarly, viruses isolated thirty years apart can also exhibit few NS differences (0.8% - AA/50 and MS/80) or many changes (2.3% - GL/54 and ENG/83) as illustrated in Table 16. These results suggest that the NS gene of influenza C viruses has multiple different representatives

occurring in the population at any one time. We believe that these genetic variants belong to different evolutionary lineages coexisting in nature. Thus, the C virus NS sequence data further support our conclusion of multiple lineages for influenza C viruses based upon HA sequences.

Analysis of the C virus NS gene sequences by the maximum parsimony method yielded three equally plausible evolutionary trees (Fig. 26A) which had very different structures from those constructed using the HA genes (Fig. 26B) from the same eight influenza C viruses. This would by expected if there is reassortment of influenza C virus genes in man. As shown in Fig. 26C, we speculate that the ENG/83 strain is the result of a reassortment event occurring across lineages of influenza C viruses. Reassortment of influenza A virus genes has been postulated to take place in man (Laver and Webster, 1973; Young and Palese, 1979; Bean et al., 1980; Cox et al., 1983) and animals (Desselberger et al., 1978) and recently it has been suggested that this process may also occur with influenza C viruses in pigs (Guo and Desselberger, 1984). Additional sequences, either of the NS and HA genes from other strains or of other genes from these isolates, can test this hypothesis. Specifically, a comparison of the NS genes with other genes which are also not subject to direct immune selection should be warranted.

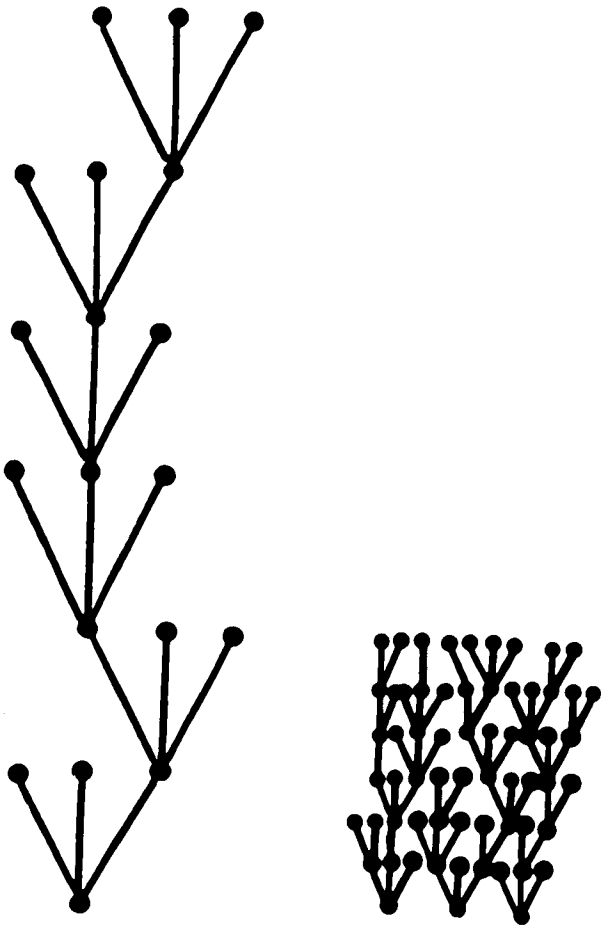
The examination of genetic variation in the NS gene of influenza C viruses allows us to further compare the genetic stability over time of influenza A and C viruses. An extensive study on the evolution of influenza A viruses isolated over long periods of time through comparative sequence analysis of NS genes has been described in this thesis. These data reveal that influenza A viruses undergo rapid evolutionary change with time. Excluding the reemergent H1N1 strains from 1977 (Nakajima et al., 1978; Kendal et al., 1978), NS genes of influenza A viruses isolated 35 years apart show sequence differences of approximately 7.5% (Table 8). In contrast, the C viruses, TAY/47 and JHG/66, which were isolated 19 years apart possess identical NS genes and NS genes of viruses isolated in 1950 (AA/50) and in 1980 (MS/80) show only 0.8% difference (Table 16). These data suggest that C virus NS genes possibly evolve more slowly and that genes belonging to the same lineage may not exhibit mutations over a period of only one or two decades.

The nucleotide data obtained for the C virus HA and NS genes then allows us to propose the following evolutionary model for the propagation of influenza C viruses in nature. We suggest that influenza C virus variants derived from multiple evolutionary pathways may cocirculate at any one time. This is schematically represented by showing three

separate evolutionary trees for influenza C viruses (Fig. 34). Sampling of several influenza C viruses isolated at one or two year intervals may thus yield viruses which have many changes. The second part of the model suggests that evolutionary changes during the passage of influenza C viruses may be less frequent than observed in influenza A viruses. Figure 34 illustrates this point by the short branches of the evolutionary trees of the influenza C viruses. One would postulate that the C/AA/50 and the C/YA/81 viruses (at least with respect to the HA genes) belong to the same evolutionary tree since only two nucleotide (and no amino acid) changes had been detected between their HA genes. Similarly, the C/TAY/47 and C/JHG/66 viruses may by part of the same lineage since no nucleotide differences were detected in their NS genes.

The evolutionary model postulated for influenza C viruses is quite different from that suggested for influenza A viruses. For the latter viruses it has been demonstrated that antigenically dominant variants emerge rapidly with time (Palese and Young, 1982; Webster et al, 1982b; Both et al., 1983). The hemagglutinins of subtype H3 and H1 viruses change so that successive variants emerge which accumulate mutations found in the HA genes of variants circulating in earlier years (Both et al., 1983; Daniels et al., 1985a; Saitou and Nei, 1986; Raymond et al., 1986). Evolutionary trees for influenza A viruses

Fig. 34. Evolutionary model for the propagation of influenza A and influenza C viruses. The length of the branches indicates relative genetic distances. Dots lying on a horizontal line represent influenza virus isolates obtained in the same season (year). The left part of the cartoon shows the emergence of influenza A virus variants lying on the same evolutionary tree. Dominant variants appear to emerge which show an accumulation of changes. The right part of the diagram depicts the cocirculation of influenza C virus variants derived from multiple evolutionary pathways. Variation appears to be slower than in the case of influenza A viruses, as demonstrated by the shorter length of the branches on the different evolutionary trees. For both influenza A and influenza C viruses an arbitrary number of seven seasonal cycles is shown on the diagram.



TYPE A VIRUSES

TYPE C VIRUSES

have also been proposed utilizing the NS gene (Krystal et al., 1983a; Saitou and Nei, 1986; this thesis) and accumulating changes have been observed in the NA (Martinez et al, 1983; Saitou and Nei, 1986) and M genes (Ortin et al., 1983; Saitou and Nei, 1986) of influenza A virus variants isolated in successive years.

A similar picture is also found with respect to nucleotide changes in influenza B virus HA genes (Krystal et al., 1982, 1983b; Verhoeyen et al., 1983; Berton et al., 1984). However, in the case of influenza B virus genes, fewer data are available and fewer changes appear to occur with time so that a definitive answer has not yet been obtained regarding the emergence of variants with cumulative changes. It is thus suggested that influenza A viruses--and possibly influenza B viruses--emerge through a series of successive variants whereby only dominant viruses survive and less successful variants disappear. (It should be noted that the model in Figure 34 represents an idealized evolutionary tree of strains, which does not take into account the occasional reassortment of genes among different influenza A viruses or the reemergence of old strains.) In contrast, influenza C viruses do not rapidly accumulate changes with time and strains of different lineages cocirculate.

What could be the nature of these differences? First, we may speculate that influenza A viruses induce an immune

response which favors selection of antigenic variants, whereas influenza C viruses do not. This would provide an explanation for the presence of different influenza C virus variants which continue to circulate in man. In addition, the polymerase of influenza A viruses may be more error-prone than that of influenza C viruses. These two factors could explain why successive variants (with many nucleotide substitutions) emerge during influenza A virus propagation and why influenza C viruses may not change much with time. Clearly, these speculations do not allow elimination of other factors which may contribute to this situation. For example, influenza A and C viruses may differ in their mode of transmission in that one or many particles are actually involved during the infection process. If transmission of the virus is the result of infection with a single infectious particle, the virus would in effect be cloned whenever it is transmitted from one person to the other. On the other hand, infection with many particles during each transmission cycle would allow for a slower selection of viral variants (Brand and Palese, 1980). Alternatively, influenza C viruses may have a long replication cycle <u>in vivo</u> which results in a decreased number of infection cycles per unit time for C viruses. Also the number of infected individuals worldwide may influence the genetic variation of these viruses. Influenza A virus infections may be more prevalent throughout the population than are

influenza virus infections caused by C viruses. All or some of these factors may lead to an altered epidemiological pattern and to an increased number of sequence changes in the genomes of influenza A viruses as compared to those of influenza C viruses.

3. Analysis of NS Deletion Mutants of Influenza A and B Viruses

3A. CR43-3

The defect in the NS segment of the influenza A virus host range mutant, CR43-3, was examined. Sequence data from cloned cDNAs revealed that the CR43-3 NS gene derived from the NS gene of the A/ALA/6/77 parent by a deletion of 36 nucleotides. Dideoxy sequencing of RNA 8 confirmed the sequence data obtained from the cloned cDNA, indicating that the deletion of the NS clone was not an artifact. The CR43-3 virus NS deletion is in the NS1-coding region (Fig. 28). Consequently, the CR43-3 NS1 protein is 12 amino acids shorter than that of the A/ALA/6/77 parent. The stretch of 12 amino acids (66-77) not found in the CR43-3 NS1 protein is conserved among NS1 proteins of 15 human influenza A virus isolates (Fig. 11) with the following exceptions. In position 67 of the NS1 of A/Udorn/72, A/ALA/6/77, and (H3) A/HT/85 viruses, an amino acid change was detected which resulted in a conservative arginine to lysine substitution. In the NS1 polypeptide of the A/BERK/68 strain, a glutamic acid to glycine change was

noted at position 71. However, in the NS genes of the A/FPV/Rostock/34 (Baez et al., 1980), A/duck/Alberta/60/76 (Baez et al., 1981), and A/swine/Iowa/15/30 (Nakajima et al., 1984) viruses, multiple substitutions are observed, indicating that functional NS proteins may not require a high degree of amino acid conservation in this region.

The mutant NS1 protein of CR43-3 virus appeared to be synthesized in approximately equal amounts in the nonpermissive (MDCK) and in the permissive (PCK) cell system (Fig. 30). In the restrictive MDCK cells, the NS1 protein of CR43-3 virus accumulates in the nuclei as shown by specific NS1 immunofluorescence of acetone-fixed MDCK cells (Fig. 31). The NS1 nucleoplasmic staining pattern displayed by CR43-3 virus-infected MDCK cells was similar to that observed for A/AA/6/60 virus-infected MDCK cells. An identical NS1 staining pattern has previously been noted for several wt influenza virus stains (Young et al., 1983). It should also be noted that a similar pattern of nucleoplasmic staining was exhibited in permissive PCK cells infected with CR43-3 virus.

Although the NS mutant CR43-3 did not show aberrant NS1 synthesis in the nonpermissive MDCK cells, a reduction in M1 protein of approximately 50% was observed in MDCK cells as compared with that in PCK cells infected with CR43-3 virus. Reduced synthesis of M1 protein has been reported for three NS mutants of fowl plague virus: ts47 (Almond et

al., 1977; Wolstenholme et al., 1980; Robertson et al., 1983), tsmN3 (Wolstenholme et al., 1980), and ts412 (Koennecke et al., 1981). Since M1 protein is synthesized late in infection, NS1 may be required for efficient expression of late viral functions. Although, ts47, tsmN3, and ts412 are all defective in the NS segment, they do not exhibit identical phenotypes at the nonpermissive temperature. In addition to a common defect in M1 synthesis, ts47 and tsmN3 show defects in viral RNA synthesis, whereas ts412 displays reduced hemagglutinin production at the nonpermissive temperature. The variable phenotypic characteristics exhibited by these NS mutants could be attributed to the fact that it has not yet been determined whether the temperature-sensitive lesion in RNA 8 of ts47, tsmN3, and ts412 affects NS1, NS2, or both polypeptides. Recently, Smith and Inglis (1985) have identified a point mutation in RNA 8 of tsmN3 which results in an amino acid change in the NS1 protein. It is not known, however, if this nucleotide substitution is the ts lesion of this virus. Thus it will be interesting to study the RNA phenotype of CR43-3 virus, which appears to have a defect in the NS1 only. At this time, however, no data are available regarding the synthesis of virus-specific vRNA and cRNA in CR43-3-infected permissive or nonpermissive cells.

The study of temperature-sensitive mutants of influenza virus has greatly aided in the assignment of functional roles to specific viral proteins. Through nucleotide sequence analysis, we have now determined that the defect of the CR43-3 virus mutant is a deletion mutation in the NS gene. It appears likely that this deletion in the NS1 polypeptide is associated with the restricted host range of the virus. With the possible exception of variants with deletions or insertions in the HA or NA genes (Blok and Air, 1982b; Caton et al., 1982; Krystal et al., 1983b), there is no precedent for a conditional lethal mutant of influenza virus with a defined deletion. Further analysis of the CR43-3 mutant may be valuable for determining the function of the NS gene products in influenza virus replication.

3B. Clone 201

The NS gene of the influenza B virus clone 201 has a deletion of 13 nucleotides in the NS1 coding region. The deletion causes a shift in reading frame of the NS1 protein so that a premature termination codon is reached 18 triplets downstream of the deletion site. The NS1 protein of clone 201 is comprised of 127 amino acids, and only the first N-terminal 117 residues are found in the NS1 protein of the parent virus B/YA/73. The NS1 polypeptide of B/YA/73 virus consists of 281 amino acids, more than twice the size of that of clone 201. Although the NS1 protein of

clone 201 is truncated, virus replication and protein synthesis do not appear to be hampered as compared with the wild type B/YA/73 virus. Additionally, the karyophilic character of the NS1 protein of clone 201 is retained.

It thus appears that only the N-terminal portion of the NS1 protein is required for functional activity. In this respect, the NS1 of influenza B virus may be similar to that of influenza A virus. As reported in this thesis, the NS1 proteins of influenza A virus field isolates are very heterogeneous in size (202-237 residues) and can tolerate large deletions in the carboxy terminus. An extreme example is the A/turkey/Oregon/71 virus which codes for an NS1 protein of only 124 amino acids (Norton et al., submitted).

An alternative speculation is that the NS1 protein may not be critical for virus replication. It may be that the other gene product encoded by RNA 8, NS2, plays a more significant role in the virus life cycle and therefore, there is greater constraint to preserve the sequence of NS2. Clone 201 has an NS2 protein of 122 amino acids, wild type length. The influenza A virus CR43-3 mutant also encodes an NS2 protein of normal length (121 residues). Even though influenza A viruses show a wide range of NS1 protein lengths, there have not been any reports of an influenza A or B virus (field or laboratory isolate) which

codes for an NS2 polypeptide of variable size. Deletions in the NS2 protein may be lethal for virus growth.

SIGNIFICANCE

The evolutionary patterns of influenza A and C viruses in nature were examined and compared. To assess evolutionary rates, nucleotide sequences of the NS, H3 HA, and N2 NA genes of influenza A viruses isolated over time were analyzed by the maximum parsimony procedure. Evolutionary trees were constructed which allowed us to calculate evolutionary rates for these genes which are approximately 10^{-3} substitutions per site per year. The rate of nucleotide substitution in influenza A virus genes is extremely rapid, a million times faster than that of mammalian genomes. In addition, the sequential nature of nucleotide changes in the influenza A virus NS gene is consistent with an evolutionary model for influenza A viruses where a single variant predominates at a given time and subsequent variants emerge successively along a common lineage. We believe that positive selection of virus variants contributes significantly to the rapid and unusual pattern of evolution that influenza A viruses display in nature.

The extreme variability of influenza A viruses is clearly a stumbling block for the development of an effective vaccine to control virus infection. The data presented in this work significantly expand our understanding of genetic drift in influenza A viruses. The genetic changes which have occurred in previously

circulating epidemic strains may provide clues on evolutionary pathways which may give rise to epidemiologically important strains of the future. The examination of molecular changes in the influenza virus genome may thus aid in the design of better vaccines or alternative strategies to protect man from this respiratory pathogen.

The influenza C viruses are generally associated with mild respiratory infections and the epidemiology of these viruses in man has not been well characterized. This thesis presents the first extensive study of genetic variation in influenza C viruses at the level of the nucleotide sequence. The HA and NS genes of human influenza C viruses isolated over approximately four decades were examined. The pattern of nucleotide changes observed in the influenza C virus genes is very different from that of influenza A viruses in that nucleotide substitutions in these genes do not accumulate sequentially over time. It appears that multiple variants of distinct lineages of influenza C viruses coexist in nature and these variants may evolve less extensively over time than the type A viruses. Although the epidemiology of influenza C viruses appears to be quite different from that of influenza A viruses, it is typical of most other RNA and DNA viruses.

REFERENCES

Air, G. M. 1981. Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza virus. Proc. Natl. Acad. Sci. USA 78:7639-7643.

Air, G. M., and R. W. Compans. 1983. Influenza B and influenza C viruses. In: Genetics of Influenza Viruses, P. Palese and D. W. Kingsbury, eds., Springer-Verlag, New York, pp. 280-304.

Air, G. M., M. C. Els, L. E. Brown, W. G. Laver, and R. G. Webster. 1985. Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. Virology 145:237-248.

Air, G. M., and R. M. Hall. 1981. Conservation and variation in influenza gene sequences. In: Genetic Variation among Influenza Viruses, D. Nayak and C. F. Fox, eds., Academic Press, New York, pp. 29-44.

Allen, H., J. McCauley, M. Waterfield, and M.-J. Gething. 1980. Influenza virus RNA segment 7 has the coding capacity for two polypeptides. Virology 107:548-551.

Almond, J. W., D. McGeoch, and R. Barry. 1977. Method for assigning temperature-sensitive mutations of influenza viruses to individual segments of the genome. Virology 81:62-73.

Atwood, K. C., L. K. Schneider, and J. Ryan. 1951. Periodic selection in Escherichia coli. Proc. Natl. Acad. Sci. USA 37:146-155.

Baez, M., R. Taussig, J. J. Zazra, J. F. Young, P. Palese, A. Reisfeld, and A. Skalka. 1980. Complete nucleotide sequence of the influenza A/PR/8/34 virus NS gene and comparison with the NS genes of A/Udorn/72 and A/FPV/Rostock/34 strains. Nucl. Acids Res. 8:5845-5858.

Baez, M., J. J. Zazra, R. M. Elliott, J. F. Young, and P. Palese. 1981. Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus NS RNA: Conservation of the NS1/NS2 overlapping gene structure in a divergent influenza virus RNA segment. Virology 113:397-402.

Bao-lan, L., R. G. Webster, L. E. Brown, and K. Nerome. 1983. Heterogeneity of influenza B viruses. Bull. WHO 61:681-687.

Barrett, T., A. Wolstenholme, and B. W. J. Mahy. 1979. Transcription and replication of influenza virus RNA. Virology 98:211-225.

Bean, W. J., N. J. Cox, and A. P. Kendal. 1980. Recombination of human influenza A viruses in nature. Nature 284:638-640.

Beaton, A. R., and R. M. Krug. 1981. Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. Nucl. Acids Res. 9:4423-4436.

Beaton, A. R., and R. M. Krug. 1984. Synthesis of the templates for influenza virion RNA replication in vitro. Proc. Natl. Acad. Sci. USA 81:4682-4686.

Beaucage, S. L., and M. H. Caruthers. 1981. Deoxynucleoside phosphoramidites--a new class of key intermediates for deoxypolyribonucleotide synthesis. Tetrahedron Lett. 22:1859-1862.

Bentley, D. R., and G. G. Brownlee. 1982. Sequence of the N2 neuraminidase from influenza virus A/NT/60/68. Nucl. Acids Res. 10:5033-5042.

Berton, M. T., C. W. Naeve, and R. G. Webster. 1984. Antigenic structure of the influenza B virus hemagglutinin: Nucleotide sequence analysis of antigenic variants selected with monoclonal antibodies. J. Virol. 52:919-927.

Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80:3963-3965.

Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7:1513-1523.

Bishop, D. H. L., J. A. Huddleston, and G. G. Brownlee. 1982a. The complete sequence of RNA segment 2 of influenza A/NT/60/68 and its encoded P1 protein. Nucl. Acids Res. 10:1335-1343.

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

Blaas, D., E. Patzelt, and E. Kuechler. 1982a. Cap-recognising protein of influenza virus. Virology 116:339-348.

Blaas, D., E. Patzelt, and E. Kuechler. 1982b. Identification of the cap binding protein of influenza virus. Nucl. Acids Res. 10:4803-4812.

Blok, J., and G. M. Air. 1980. Comparative nucleotide sequences at the 3' end of the neuraminidase gene from eleven influenza type A viruses. Virology 107:50-60.

Blok, J., and G. M. Air. 1982a. Sequence variation at the 3' end of the neuraminidase gene from 39 influenza type A viruses. Virology 121:211-229.

Blok, J., and G. M. Air. 1982b. Block deletions in the neuraminidase genes from some influenza A viruses of the N1 subtype. Virology 118:229-234.

Blok, J., G. M. Air, W. G. Laver, C. W. Ward, G. G. Lilley, E. F. Woods, C. M. Roxburgh, and A. S. Inglis. 1982. Studies on the size, chemical composition, and partial sequence of the neuraminidase (NA) from type A influenza viruses show that the N- terminal region of the NA is not processed and serves to anchor the NA in the viral membrane. Virology 119:109-121.

Bos, T. J., Davis A. R., and D. P. Nayak. 1984. NH_2-terminal hydrophobic region of influenza virus neuraminidase provides the signal function in translocation. Proc. Natl. Acad. Sci. USA 81:2327-2331.

Both, G. W., and M. J. Sleigh. 1981. Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J. Virol. 39:663-672.

Both, G. W., M. J. Sleigh, N. J. Cox, and A. P. Kendal. 1983. Antigenic drift in influenza virus H3 hemagglutinin from 1968-1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J. Virol. 48:52-60.

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

Brand, C., and P. Palese. 1980. Sequential passage of influenza virus in embryonated eggs or tissue culture: Emergence of mutants. Virology 107:424-433.

Briedis, D. J., G. Conti, E. A. Munn, and B. W. J. Mahy. 1981a. Migration of influenza virus-specific polypeptides from cytoplasm to nucleus of infected cells. Virology 111:154-164.

Briedis, D. J., and R. A. Lamb. 1982. The 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., R. A. Lamb, and P. W. Choppin. 1981b. Influenza B virus RNA segment 8 codes for two nonstructural proteins. Virology 112:417-425.

Briedis, D. J., R. A. Lamb, and P. W. Choppin. 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 M. Tobin. 1984. Influenza B virus genome: Complete nucleotide sequence of the influenza B/Lee/40 virus genome RNA segment 5 encoding the nucleoprotein and comparison with the B/Singapore/222/79 nucleoprotein. Virology 133:448-455.

Brown, L. E., V. S. Hinshaw, and R. G. Webster. 1983. Antigenic variation in the influenza A virus nonstructural protein, NS1. Virology 130:134-143.

Buonagurio, D. A., M. Krystal, P. Palese, D. C. DeBorde, and H. F. Maassab. 1984. Analysis of an influenza A virus mutant with a deletion in the NS segment. J. Virol. 49:418-425.

Buonagurio, D. A., S. Nakada, U. Desselberger, M. Krystal, and P. Palese. 1985. Noncumulative sequence changes in the hemagglutinin genes of influenza C virus isolates. Virology 146:221-232.

Buonagurio, D. A., S. Nakada, W. M. Fitch, and P. Palese. Epidemiology of influenza C virus in man: Multiple evolutionary lineages and low rate of change. Virology (in press).

Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch. 1986. Evolution of human influenza A viruses over 50 years: Rapid, uniform rate of change in NS gene. Science 232:980-982.

Caton, A. J., G. G. Brownlee, J. W. Yewdell, and W. Gerhard. 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell 31:417-427.

Chakraverty, P. 1978. Antigenic relationship between influenza C viruses. Arch. Virol. 58:341-348.

Choppin, P. W., R. W. Compans, A. Scheid, J. J. McSharry, and S. G. Lazarowitz. 1972. Structure and assembly of viral membranes. In: Membrane Research, C. F. Fox, ed., Academic Press, New York, pp. 163-179.

Clarke, S. H., K. Huppi, D. Ruezinsky, L. Staudt, W. Gerhard, and M. Weigert. 1985. Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. J. Exp. Med. 161:687-704.

Clerx-van Haaster, C. M., and H. Meier-Ewert. 1984. 3'-Terminal sequences of influenza C virion RNA. Arch. Virol. 80:239-246.

Colman, P. M., J. N. Varghese, and W. G. Laver. 1983. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature 303:41-44.

Compans, R. W., D. H. L. Bishop, and H. Meier-Ewert. 1977. Structural components of influenza C virions. J. Virol. 21:658-665.

Compans, R. W., J. Content, and P. H. Deusberg. 1972. Structure of the ribonucleoprotein of influenza virus. J. Virol. 10:795-800.

Concannon, P., I. W. Cummings, and W. A. Salser. 1984. Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene. J. Virol. 49:276-278.

Cox, N. J., Z. S. Bai, and A. P. Kendal. 1983. Laboratory-based surveillance of influenza A (H1N1) and A (H3N2) viruses in 1980-81: antigenic and genomic analyses. Bull. WHO 61:143-152.

Dagert, M., and S. D. Ehrlich. 1979. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6:23-28.

Daniels, R. S., A. R. Douglas, J. J. Skehel, and D. C. Wiley. 1985a. Antigenic and amino acid sequence analyses of influenza viruses of the H1N1 subtype isolated between 1982 and 1984. Bull. WHO 63:273-277.

Daniels, R. S., J. C. Downie, A. J. Hay, M. Knossow, J. J. Skehel, M. L. Wang, and D. C. Wiley. 1985b. Fusion mutants of the influenza virus hemagglutinin glycoprotein. Cell 40:431-439.

Darlix, J.-L., and P.-F. Spahr. 1983. High spontaneous mutation rate of Rous sarcoma virus demonstrated by direct sequencing of the RNA genome. Nucl. Acids Res. 11:5953-5967.

Davenport, F. M., E. Minuse, A. V. Hennessy, and T. Francis. 1969. Interpretations of influenza antibody patterns of man. Bull. WHO 41:453-460.

Davey, J., N. J. Dimmock, and A. Colman. 1985. Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in Xenopus oocytes. Cell 40:667-675.

Del Rio, L., C. Martinez, E. Domingo, and J. Ortin. 1985. In vitro synthesis of full-length influenza virus complementary RNA. EMBO J. 4:243-247.

Desselberger, U., K. Nakajima, P. Alfino, F. S. Pedersen, W. A. Haseltine, C. Hannoun, and P. Palese. 1978. Biochemical evidence that "new" influenza virus strains in nature may arise by recombination (reassortment). Proc. Natl. Acad. Sci. USA 75:3341-3345.

Desselberger, U., and P. Palese. 1978. Molecular weights of RNA segments of influenza A and B viruses. Virology 88:394-399.

Desselberger, U., V. R. Racaniello, J. J. Zazra, and P. Palese. 1980. The 3' and 5'-end terminal sequences of influenza A, B, and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene 8:315-328.

Dimmock, N. J. 1969. New virus-specific antigens in cells infected with influenza virus. Virology 39:224-234.

Domingo, E., D. Sabo, T. Taniguchi, and C. Weissmann. 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell 13:735-744.

Doms, R. W., M.-J. Gething, J. Henneberry, J. White, and A. Helenius. 1986. Variant influenza virus hemagglutinin that induces fusion at elevated pH. J. Virol. 57:603-613.

Doms, R. W., A. Helenius, and J. White. 1985. Membrane fusion activity of the influenza virus hemagglutinin. The low pH-induced conformational change. J. Biol. Chem. 260:2973-2981.

Donatelli, I., C. Rozera, S. Butto, G. Arangio-Ruiz, and G. Mancini. 1985. Comparative electrophoretic study of polypeptides of influenza A/H3N2 viruses isolated in circumscribed geographical areas. J. gen. Virol. 66:637-641.

Dreiding, P., P. Staeheli, and O. Haller. 1985. Interferon-induced protein Mx accumulates in nuclei of mouse cells expressing resistance to influenza viruses. Virology 140:192-196.

Elleman, T. C., A. A. Azad, and C. W. Ward. 1982. Neuraminidase gene from the early Asian strain of human influenza virus, A/RI/5-/57 (H2N2). Nucl. Acids Res. 10:7005-7015.

Elliott, R. M., Y. J. Guo, and U. Desselberger. 1984. Polypeptide synthesis in MDCK cells infected with human and pig influenza C viruses. J. gen. Virol. 65:1873-1880.

Emini, E. A., S.-Y. Kao, A. J. Lewis, R. Crainic, and E. Wimmer. 1983. Functional basis of poliovirus neutralization determined with monospecific neutralizing antibodies. J. Virol. 46:466-474.

Enami, M., R. Fukuda, and A. Ishihama. 1985. Transcription and replication of eight RNA segments of influenza virus. Virology 142:68-77.

Fang, R., W. Min Jou, D. Huylebroeck, R. Devos, and W. Fiers. 1981. Complete structure of A/Duck/Ukraine/63 influenza hemagglutinin gene: Animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin. Cell 25:315-323.

Fields, S., and G. Winter. 1982. Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. Cell 28:303-313.

Fields, S., G. Winter, and G. G. Brownlee. 1981. Structure of the neuraminidase gene in human influenza virus A/PR/8/34. Nature 290:213-217.

Fitch, W. M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. 20:406-416.

Fitch, W. M., and C. H. Langley. 1976. Protein evolution and the molecular clock. Fed. Proc. 35:2092-2097.

Frommhagen, L. H., C. A. Knight, and N. K. Freeman. 1959. The ribonucleic acid, lipid, and polysaccharide constituents of influenza virus preparations. Virology 8:176-197.

Garten, W., F. Bosch, D. Linder, R. Rott, and H.-D. Klenk. 1981. Proteolytic activation of the influenza virus hemagglutinin: The structure of the cleavage site and the enzymes involved in cleavage. Virology 115:361-374.

Gerhard, W., J. Yewdell, M. Frankel, and R. G. Webster. 1981. Antigenic structure of influenza virus hemagglutinin defined by hybridoma antibodies. Nature 290:713-717.

Gething, M.-J., J. Bye, J. J. Skehel, and M. D. Waterfield. 1980. Cloning and DNA sequence of double-stranded copies of hemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature 287:301-306.

Gething, M.-J., R. W. Doms, D. York, and J. White. 1986. Studies on the mechanism of membrane fusion: site-specific mutagenesis of the hemagglutinin of influenza virus. J. Cell Biol. 102:11-23.

Gething, M.-J., and J. Sambrook. 1982. Construction of influenza haemagglutinin genes that code for intracellular and secreted forms of the protein. Nature 300:598-603.

Gojobori, T., and S. Yokoyama. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. Proc. Natl. Acad. Sci. USA 82:4198-4201.

Goto, H., T. Tanaka, and K. Tobita. 1984. Comparison of nine strains of influenza C virus in growth characteristics and viral polypeptides. Arch. Virol. 82:111-117.

Gottschalk, A. 1957. Neuraminidase: The specific enzyme of influenza virus and Vibrio cholerae. Biochim. Biophys. Acta. 23:645-646.

Greenspan, D., M. Krystal, S. Nakada, H. Arnheiter, D. S. Lyles, and P. Palese. 1985. Expression of influenza virus NS2 nonstructural protein in bacteria and localization of NS2 in infected eucaryotic cells. J. Virol. 54:833-843.

Gregoriades, A., T. Christie, and K. Markarian. 1984. The membrane (M1) protein of influenza virus occurs in two forms and is a phosphoprotein. J. Virol. 49:229-235.

Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA 72:3961-3965.

Guo, Y. J., and U. Desselberger. 1984. Genome analysis of influenza C viruses isolated in 1981/1982 from pigs in China. J. gen. Virol. 65:1857-1872.

Guo, Y. J., F. G. Jin, P. Wang, M. Wang, and J. M. Zhu. 1983. Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. J. gen. Virol. 64:177-182.

Hall, R. M., and G. M. Air. 1981. Variation in nucleotide sequences coding for the N-terminal regions of the matrix and nonstructural proteins of influenza A viruses. Virology 38:1-7.

Hamaguchi, M., K. Maeno, T. Yoshida, Y. Nagai, M. Iinuma, and T. Matsumoto. 1985. Analysis of nuclear accumulation of influenza nucleoprotein antigen using a temperature-sensitive mutant. Microbiol. Immunol. 29:1131-1137.

Hauptmann, R., L. D. Clarke, R. C. Mountford, H. Bachmayer, and J. W. Almond. 1983. Nucleotide sequence of the haemagglutinin gene of influenza virus A/England/321/77. J. gen. Virol. 64:215-220.

Hay, A. J., B. Lomniczi, A. R. Bellamy, and J. J. Skehel. 1977. Transcription of the influenza virus genome. Virology 83:337-355.

Hay, A. J., J. J. Skehel, and J. McCauley. 1982. Characterization of influenza virus RNA complete transcripts. Virology 116:517-522.

Hayashida, H., and T. Miyata. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA 80:2671-2675.

Hayashida, H., H. Toh, R. Kikuno, and T. Miyata. 1985. Evolution of influenza virus genes. Mol. Biol. Evol. 2:289-303.

Herrler, G., A. Nagele, H. Meier-Ewert, A. S. Bhown, and R. W. Compans. 1981. Isolation and structural analysis of influenza C virion glycoproteins. Virology 113:439-451.

Herrler, G., R. Rott, and H.-D. Klenk. 1985a. Neuraminic acid is involved in the binding of influenza C virus to erythrocytes. Virology 141:144-147.

Herrler, G., R. Rott, H.-D. Klenk, H.-P. Muller, A. K. Shukla, and R. Schauer. 1985b. The receptor-destroying enzyme of influenza C virus is neuraminate- O-acetylesterase. EMBO J. 4:1503-1506.

Herz, C., E. Stavnezer, and R. M. Krug. 1981. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell 26:391-400.

Hewat, E. A., S. Cusack, R. W. H. Ruigrok, and C. Verwey. 1984. Low resolution structure of the influenza C glycoprotein determined by electron microscopy. J. Mol. Biol. 175:175-193.

Hinshaw, V. S., W. J. Bean, Jr., R. G. Webster, and B. C. Easterday. 1978. The prevalence of influenza viruses in swine and the antigenic and genetic relatedness of influenza viruses from man and swine. Virology 84:51-62.

Hinshaw, V. S., W. J. Bean, R. G. Webster, and G. Sriram. 1980. Genetic reassortment of influenza A viruses in the intestinal tract of ducks. Virology 102:412-419.

Hirst, G. K. 1941. Agglutination of red cells by allantoic fluid of chick embryos infected with influenza virus. Science 94:22-23.

Hirst, G. K. 1942. Adsorption of influenza hemagglutinins and virus by red blood cells. J. Exp. Med. 76:195-209.

Hirst, G. K. 1950. The relationship of the receptors of a new strain of virus to those of the mumps-NDV-influenza group. J. Exp. Med. 91:177-185.

Hiti, A. L., and D. P. Nayak. 1982. Complete nucleotide sequence of the neuraminidase gene of human influenza virus A/WSN/33. J. Virol. 41:730-734.

Hiti, A. R., A. R. David, and D. P. Nayak. 1981. Complete sequence analysis shows that the hemagglutinins of the H0 and H2 subtypes of human influenza virus are closely related. Virology 111:113-124.

Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol. 1982. Rapid evolution of RNA genomes. Science 215:1577-1585.

Holland, T. C., S. D. Marlin, M. Levine, and J. Glorioso. 1983. Antigenic variants of herpes simplex virus selected with glycoprotein-specific monoclonal antibodies. J. Virol. 45:672-682.

Horisberger, M. A. 1980. The large P proteins of influenza A viruses are composed of one acidic and two basic polypeptides. Virology 107:302-305.

Horisberger, M. A. 1982. Identification of a catalytic activity of the large basic P polypeptide of influenza virus. Virology 120:279-286.

Horisberger, M. A., and K. de Staritzky. 1985. Sensitivity of influenza A viruses to human interferons in human diploid cells. FEMS Microbiol. Lett. 29:207-210.

Horisberger, M. A., and H. K. Hochkeppel. 1985. An interferon-induced mouse protein involved in the mechanism of resistance to influenza viruses. J. Biol. Chem. 260:1730-1733.

Horisberger, M. A., P. Staeheli, and O. Haller. 1983. Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. Proc. Natl. Acad. Sci. USA 80:1910-1914.

Hovanec, D. L., and G. M. Air. 1984. Antigenic structure of the hemagglutinin of influenza virus B/Hong Kong/8/73 as determined from gene sequence analysis of variants selected with monoclonal antibodies. Virology 139:384-392.

Huang, R. T. C., E. Dietsch, and R. Rott. 1985. Further studies on the role of neuraminidase and the mechanism of low pH dependence in influenza virus-induced membrane fusion. J. gen. Virol. 66:295-301.

Huddleston, J. A., and G. G. Brownlee. 1982. The sequence of the nucleoprotein gene of human influenza A virus strain A/NT/60/68. Nucl. Acids Res. 10:1029-1038.

Inglis, S. C., T. Barrett, C. M. Brown, and J. W. Almond. 1979. The smallest genome RNA segment of influenza virus contains two genes that may overlap. Proc. Natl. Acad. Sci. USA 76:3790-3794.

Inglis, S. C., and C. M. Brown. 1981. Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. Nucl. Acids Res. 9:2727-2740.

Inglis, S. C., and C. M. Brown. 1984. Differences in control of virus mRNA splicing during permissive or abortive infection with influenza A (fowl plague) virus. J. gen. Virol. 65:153-164.

Inglis, S. C., A. R. Carroll, R. A. Lamb, and B. W. J. Mahy. 1976. Polypeptides specified by the influenza virus genome. 1. Evidence for eight distinct gene products specified by fowl plague virus. Virology 74:489-503.

Inglis, S. C., M.-J. Gething, and C. M. Brown. 1980. Relationship between the messenger RNAs transcribed from two overlapping genes of influenza virus. Nucl. Acids Res. 8:3575-3589.

Inglis, S. C., and B. W. J. Mahy. 1979. Polypeptides specified by the influenza virus genome: 3. Control of synthesis in infected cells. Virology 95:154-164.

Itakura, K., N. Katagiri, S. A. Narang, C. P. Bahl, K. J. Marians, and R. Wu. 1975. Chemical synthesis and sequence studies of deoxyribonucleotides which constitute the duplex sequence of the lactose operator of Escherichia coli. J. Biol. Chem. 250:4592-4600.

Jackson, D. A., A. J. Caton, S. J. McCready, and P. R. Cook. 1982. Influenza virus RNA is synthesized at fixed sites in the nucleus. Nature 295:366-368.

Jones, K. L., J. A. Huddleston, and G. G. Brownlee. 1983. The sequence of RNA segment 1 of influenza virus A/NT/60/68 and its comparison with the corresponding segment of strains A/PR/8/34 and A/WSN/33. Nucl. Acids Res. 11:1555-1566.

Jones, L. V., R. W. Compans, A. R. Davis, T. J. Bos, and D. P. Nayak. 1985. Surface expression of influenza virus neuraminidase, an amino-terminally anchored viral membrane glycoprotein in polarized epithelial cells. Mol. Cell. Biol. 5:2181-2189.

Katagiri, S., A. Ohizumi, and M. Homma. 1983. An outbreak of type C influenza in a children's home. J. Infect. Dis. 148:51-56.

Kato, A., K. Mizumoto, and A. Ishihama. 1985. Purification and enzymatic properties of an RNA polymerase-RNA complex from influenza virus. Virus Res. 3:115-127.

Kawakami, K., and A. Ishihama. 1983. RNA polymerase of influenza virus. III. Isolation of RNA polymerase-RNA complexes from influenza virus PR8. J. Biochem. 93:989-996.

Kawakami, K., K. Mizumoto, and A. Ishihama. 1983. RNA polymerase of influenza virus. IV. Catalytic properties of the capped RNA endonuclease associated with the RNA polymerase. Nucl. Acids Res. 11:3637-3649.

Kemdirim, S., J. Palefsky, and D. J. Briedis. 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.

Kendal, A. P. 1975. A comparison of influenza C with prototype myxoviruses: Receptor- destroying activity (neuraminidase) and structural polypeptides. Virology 65:87-99.

Kendal, A. P., M. Goldfield, G. R. Noble, and W. R. Dowdle. 1977. Identification and preliminary antigenic analysis of swine influenza-like virus isolated during an influenza outbreak at Fort Dix, New Jersey. J. Infect. Dis. 136:381-385.

Kendal, A. P., G. R. Noble, J. J. Skehel, and W. R. Dowdle. 1978. Antigenic similarity of influenza A (H1N1) viruses from epidemics in 1977-1978 to "Scandinavian" strains isolated in epidemics of 1950-1951. Virology 89:632-636.

Khorana, H. G. 1979. Total synthesis of a gene. Science 203:614-625.

Kilbourne, E. D., W. G. Laver, J. L. Schulman, and R. G. Webster. 1968. Antiviral activity of antiserum specific for an influenza virus neuraminidase. J. Virol. 2:281-288.

Kimura, M. 1983. The neutral theory of molecular evolution. In: Evolution of Genes and Proteins, M. Nei and R. K. Koehn, eds., Sinauer Associates, Sunderland, MA, pp. 208-233.

Kistner, O., H. Muller, H. Becht, and C. Scholtissek. 1985. Phosphopeptide fingerprints of nucleoproteins of various influenza A strains grown in different host cells. J. gen. Virol. 66:465-472.

Klenk, H.-D., R. W. Compans, and P. W. Choppin. 1970. An electron microscopic study of the presence or absence of neuraminic acid in enveloped viruses. Virology 42:1158-1162.

Koennecke, I., C. B. Boschek, and C. Scholtissek. 1981. Isolation and properties of a temperature-sensitive mutant (ts412) of an influenza A virus recombinant with a ts lesion in the gene coding for the nonstructural protein. Virology 110:16-25.

Krug, R. M. 1981. Priming of influenza viral RNA transcription by capped heterologous RNAs. Curr. Topics Microbiol. Immunol. 93:125-150.

Krug, R. M., B. A. Broni, A. J. LaFiandra, M. A. Morgan, and A. J. Shatkin. 1980. Priming and inhibitory activities of RNAs for the influenza viral transcriptase do not require base-pairing with the virion RNA template. Proc. Natl. Acad. Sci. USA 77:5874-5878.

Krug, R. M., M. Shaw, B. Broni, G. Shapiro, and O. Haller. 1985. Inhibition of influenza viral mRNA synthesis in cells expressing the interferon-induced Mx gene product. J. Virol. 56:201-206.

Krug, R. M., and R. Soeiro. 1975. Studies on the intranuclear localization of influenza virus-specific proteins. Virology 64:378-387.

Krug, R. M., M. Ueda, and P. Palese. 1975. Temperature-sensitive mutants of influenza WSN virus defective in virus-specific RNA synthesis. J. Virol. 16:790-796.

Krystal, M., D. A. Buonagurio, J. F. Young, and P. Palese. 1983a. Sequential mutations in the NS genes of influenza virus field strains. J. Virol. 45:547-554.

Krystal, M., R. M. Elliott, E. W. Benz, J. F. Young, and P. Palese. 1982. Evolution of influenza A and B viruses: conservation of structural features in the hemagglutinin gene. Proc. Natl. Acad. Sci. USA 79:4800-4804.

Krystal, M., S. Nakada, D. A. Buonagurio, D. C. DeBorde, H. F. Maassab, and P. Palese. 1984. The nonstructural gene segment of influenza A virus: Expression of NS1 protein in mammalian cells; Analysis of a deletion mutant. In: Proceedings of the 5th International Symposium on Negative Strand Viruses, D. H. L. Bishop and R. W. Compans, eds., Elsevier, New York, pp. 147-157.

Krystal, M., J. F. Young, P. Palese, I. A. Wilson, J. J. Skehel, and D. C. Wiley. 1983b. Sequential mutations in the hemagglutinins of influenza B virus isolates: Definition of antigenic domains. Proc. Natl. Acad. Sci. USA 80:4527-4531.

Lamb, R. A. 1983. The influenza virus RNA segments and their encoded proteins. In: Genetics of Influenza Viruses, P. Palese and D. W. Kingsbury, eds., Springer-Verlag, New York, pp. 21-69.

Lamb, R. A., and P. W. Choppin. 1976. Synthesis of influenza virus proteins in infected cells: Translation of viral polypeptides, including three P polypeptides, from RNA produced by primary transcription. Virology 74:504-519.

Lamb, R. A., and P. W. Choppin. 1977. Synthesis of influenza virus polypeptides in cells resistant to alpha-amanitin: Evidence for the involvement of cellular RNA polymerase II in virus replication. J. Virol. 23:816-819.

Lamb, R. A., and P. W. Choppin. 1979. Segment 8 of the influenza virus genome is unique in coding for two polypeptides. Proc. Natl. Acad. Sci. USA 76:4908-4912.

Lamb, R. A., and P. W. Choppin. 1981. Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. Virology 112:729-737.

Lamb, R. A., P. W. Choppin, R. M. Chanock, and C.-J. Lai. 1980. Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of the influenza virus genome. Proc. Natl. Acad. Sci. USA 77:1857-1861.

Lamb, R. A., and C.-J. Lai. 1980. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell 21:475-485.

Lamb, R. A., and C.-J. Lai. 1981. Conservation of the influenza virus membrane protein (M1) amino acid sequence and an open reading frame of RNA segment 7 encoding a second protein (M2) in H1N1 and H3N2 strains. Virology 112:746-751.

Lamb, R. A., and C.-J. Lai. 1982. Spliced and unspliced messenger RNAs synthesized from cloned influenza virus M DNA in an SV40 vector: Expression of the influenza virus membrane protein (M1). Virology 123:237-256.

Lamb, R. A., and C.-J. Lai. 1984. Expression of unspliced NS1 mRNA, spliced NS2 mRNA, and a spliced chimera mRNA from cloned influenza virus NS DNA in an SV40 vector. Virology 135:139-147.

Lamb, R. A., C.-J. Lai, and P. W. Choppin. 1981. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: Colinear and interrupted mRNAs code for overlapping proteins. Proc. Natl. Acad. Sci. USA 78:4170-4174.

Lamb, R. A., S. L. Zebedee, and C. D. Richardson. 1985. Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. Cell 40:627-633.

Laver, W. G., G. M. Air, T. A. Dopheide, and C. W. Ward. 1980. Amino acid sequence changes in the hemagglutinin of A/Hong Kong (H3N2) influenza virus during the period 1968-1977. Nature 283:454-457.

Laver, W. G., G. M. Air, and R. G. Webster. 1981. The mechanisms of antigenic drift in influenza virus. Amino acid sequence changes in an antigenically active region of Hong Kong (H3N2) influenza virus hemagglutinin. J. Mol. Biol. 145:339-361.

Laver, W. G., G. M. Air, R. G. Webster, and L. J. Markoff. 1982. Amino acid sequence changes in antigenic variants of type A influenza virus N2 neuraminidase. Virology 122:450-460.

Laver, W. G., and R. C. Valentine. 1969. Morphology of the isolated hemagglutinin and neuraminidase subunits of influenza virus. Virology 38:105-119.

Laver, W. G., and R. G. Webster. 1973. Studies on the origin of pandemic influenza. III. Evidence implicating duck and equine influenza viruses as possible progenitors of the Hong Kong strain of human influenza. Virology 51:383-391.

Lazarowitz, S. G., and P. W. Choppin. 1975. Enhancement of infectivity of influenza A and B viruses by proteolytic cleavage of the hemagglutinin polypeptide. Virology 68:440-454.

Lazarowitz, S. G., R. W. Compans, and P. W. Choppin. 1971. Influenza virus structural and nonstructural proteins in infected cells and their plasma membranes. Virology 46:830-843.

Lazarowitz, S. G., R. W. Compans, and P. W. Choppin. 1973. Proteolytic cleavage of the hemagglutinin polypeptide of influenza virus: function of the uncleaved polypeptide HA. Virology 52:199-212.

Lazdins, I., E. A. Haslam, and D. O. White. 1972. The polypeptides of influenza virus. VI. Composition of the neuraminidase. Virology 49:758-765.

Lentz, M. R., G. M. Air, W. G. Laver, and R. G. Webster. 1984. Sequence of the neuraminidase gene of influenza virus A/Tokyo/3/67. Virology 135:257-265.

Li, W.-H., and T. Gojobori. 1983. Rapid evolution of goat and sheep globin genes following gene duplication. Mol. Biol. Evol. 1:94-108.

Li, W.-H, C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative liklihood of nucleotide and codon changes. Mol. Biol. Evol. 2:150-174.

Lohmeyer, J., L. T. Talens, and H.-D. Klenk. 1979. Biosynthesis of the influenza virus envelope in abortive infection. J. gen. Virol. 42:73-88.

Londo, D. R., A. R. Davis, and D. P. Nayak. 1983. Complete nucleotide sequence of the nucleoprotein gene of influenza B virus. J. Virol. 47:642-648.

Lubeck, M. D., J. L. Schulman, and P. Palese. 1980. Antigenic variants of influenza viruses: Marked differences in the frequencies of variants selected with different monoclonal antibodies. Virology 102:458-462.

Maassab, H. F., and D. C. DeBorde. 1983. Characterization of an influenza A host range mutant. Virology 130:342-350.

Maassab, H. F., C. W. Smitka, A. M. Donabedian, A. S. Monto, N. J. Cox, and A. P. Kendal. 1981. Characterization of influenza virus "cold" recombinants derived at the non-permissive temperature (38°). In: The Replication of Negative Strand Viruses, D. H. L. Bishop and R. W. Compans, eds., Elsevier/North Holland, New York, pp. 395-404.

Mahy, B. W. J., T. Barrett, S. T. Nichol, C. R. Penn, and A. J. Wolstenholme. 1981. Analysis of the functions of influenza virus genome RNA segments by use of temperature-sensitive mutants of fowl plague virus. In: The Replication of Negative Strand Viruses, D. H. L. Bishop and R. W. Compans, eds., Elsevier, New York, pp. 379-387.

Maizel, J. V., Jr. 1971. Polyacrylamide gel electrophoresis of viral proteins. In: Methods in Virology, K. Maramorosch and H. Koprowski, eds., Academic Press, New York, pp. 179-245.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York, p. 93.

Mark, G. E., J. M. Taylor, B. Broni, and R. M. Krug. 1979. Nuclear accumulation of influenza viral RNA transcripts and the effects of cycloheximide, actinomycin D, and alpha-amanitin. J. Virol. 29:744-752.

Markoff, L., and C.-J. Lai. 1982. Sequence of the influenza A/Udorn/72 (H3N2) virus neuraminidase gene as determined from cloned full-length DNA. Virology 119:288-297.

Markoff, L., B.-C. Lin, M. M. Sveda, and C.-J. Lai. 1984. Glycosylation and surface expression of the influenza virus neuraminidase requires the N-terminal hydrophobic region. Mol. Cell. Biol. 4:8-16.

Martinez, C., L. Del Rio, A. Portela, E. Domingo, and J. Ortin. 1983. Evolution of the influenza virus neuraminidase gene during drift of the N2 subtype. Virology 130:539-545.

Masurel, N. 1969. Relation between Hong Kong virus and former human A2 isolates and the A/equi2 virus in human sera collected before 1957. Lancet 1:907-910.

Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. 65:499-560.

McCauley, J. W., and B. W. J. Mahy. 1983. Structure and function of the influenza virus genome. Biochem. J. 211:281-294.

McCauley, J. W., B. W. J. Mahy, and S. C. Inglis. 1982. Nucleotide sequence of fowl plague virus RNA segment 7. J. gen. Virol. 58:211-215.

Meier-Ewert, H., A. Nagele, G. Herrler, S. Basak, and R. W. Compans. 1981a. Analysis of influenza C virus structural proteins and identification of a virion RNA polymerase. In: The Replication of Negative Strand Viruses, D. H. L. Bishop and R. W. Compans, eds., Elsevier, New York, pp. 173-180.

Meier-Ewert, H., T. Petri, and D. H. L. Bishop. 1981b. Oligonucleotide fingerprint analyses of influenza C virion RNA recovered from five different isolates. Arch. Virol. 67:141-147.

Merril, C. R., D. Goldman, S. A. Sedman, and M. H. Ebert. 1981. Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. Science 211:1437-1438.

Meyer, T., and M. A. Horisberger. 1984. Combined action of mouse alpha and beta interferons in influenza virus-infected macrophages carrying the resistance gene Mx. J. Virol. 49:709-716.

Min Jou, W., M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, W. Fiers, G. Threlfall, C. Barber, N. Carey, and S. Emtage. 1980. Complete structure of the hemagglutinin gene from the human A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell 19:683-696.

Minor, P. D., G. C. Schild, J. Bootman, D. M. A. Evans, M. Ferguson, P. Reeve, M. Spitz, G. Stanway, A. J. Cann, R. Hauptmann, L. D. Clarke, R. C. Mountford, and J. W. Almond. 1983. Location and primary structure of a major antigenic site for poliovirus neutralization. Nature 310:674-679.

Miyata, T., H. Hayashida, R. Kikuno, M. Hasegawa, M. Kobayashi, and K. Koike. 1982. Molecular clock of silent substitution: At least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. 19:28-35.

Mount, S. M. 1982. A catalogue of splice junction sequences. Nucl. Acids Res. 10:459-472.

Murti, K. G., and R. G. Webster. 1986. Distribution of hemagglutinin and neuraminidase on influenza virions as revealed by immunoelectron microscopy. Virology 149:36-43.

Nakada, S., R. S. Creager, M. Krystal, R. P. Aaronson, and P. Palese. 1984a. Influenza C virus hemagglutinin: Comparison with influenza A and B virus hemagglutinins. J. Virol. 50:118-124.

Nakada, S., R. S. Creager, M. Krystal, and P. Palese. 1984b. Complete nucleotide sequence of the influenza C/California/78 virus nucleoprotein gene. Virus Res. 1:433-441.

Nakada, S., P. N. Graves, U. Desselberger, R. S. Creager, M. Krystal, and P. Palese. 1985. Influenza C virus RNA 7 codes for a nonstructural protein. J. Virol. 56:221-226.

Nakada, S., P. N. Graves, and P. Palese. 1986. The influenza C virus NS gene: Evidence for a spliced mRNA and a second NS gene product (NS2 protein). Virus Res. 4:263-273.

Nakajima, K., U. Desselberger, and P. Palese. 1978. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. Nature 274:334-339.

Nakajima, K., E. Nobusawa, and S. Nakajima. 1984. Genetic relatedness between A/Swine/Iowa/15/30 (H1N1) and human influenza viruses. Virology 139:194-198.

Nakamura, K., G. Herrler, T. Petri, H. Meier-Ewert, and R. W. Compans. 1979. Carbohydrate components of influenza C virions. J. Virol. 29:997-1005.

Nei, M. 1983. Genetic polymorphism and the role of mutation in evolution. In: Evolution of Genes and Proteins, M. Nei and R. K. Koehn, eds., Sinauer Associates, Sunderland, MA, pp. 165-190.

Newton, S. E., G. M. Air, R. G. Webster, and W. G. Laver. 1983. Sequence of the hemagglutinin gene of influenza virus A/Memphis/1/71 and previously uncharacterized monoclonal antibody- derived variants. Virology 128:495-501.

Norton, G. P., T. Tanaka, K. Tobita, S. Nakada, D. A. Buonagurio, D. Greenspan, M. Krystal, and P. Palese. Infectious influenza A and B virus variants with long carboxyl terminal deletions in the NS1 polypeptide. Virology (submitted).

O'Callaghan, R. J., R. S. Gohd, and D. D. Labat. 1980. Human antibody to influenza C virus: Its age-related distribution and distinction from receptor analogs. Infect. Immun. 30:500-505.

O'Hara, P. J., S. T. Nichol, F. M. Horodyski, and J. J. Holland. 1984. Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. Cell 36:915-924.

Odagiri, T., D. C. DeBorde, and H. F. Maassab. 1982. Cold-adapted recombinants of influenza A virus in MDCK cells. I. Development and characterization of A/Ann Arbor/6/60 x A/Alaska/6/77 recombinant viruses. Virology 119:82-95.

Ohuchi, M., R. Ohuchi, and K. Mifune. 1982. Demonstration of hemolytic and fusion activities of influenza C virus. J. Virol. 42:1076-1079.

Ortin, J., C. Martinez, L. Del Rio, M. Davila, C. Lopez-Galindez, N. Villanueva, and E. Domingo. 1983. Evolution of the nucleotide sequence of influenza virus RNA segment 7 during drift of the H3N2 subtype. Gene 23:233-239.

Ortin, J., R. Najera, C. Lopez, M. Davila, and E. Domingo. 1980. Genetic variability of Hong Kong (H3N2) influenza viruses: spontaneous mutations and their location in the viral genome. Gene 11:319-331.

Oxford, J. S., T. Corcoran, and G. C. Schild. 1981. Intratypic electrophoretic variation of structural and non-structural polypeptides of human influenza A viruses. J. gen. Virol. 56:431-436.

Oxford, J. S., A. I. Klimov, T. Corcoran, Y. Z. Ghendon, and G. C. Schild. 1984. Biochemical and serological studies of influenza B viruses: Comparisons of historical and recent isolates. Virus Res. 1:241-258.

Palese, P. 1977. The genes of influenza virus. Cell 10:1-10.

Palese, P., R. M. Elliott, M. Baez, J. J. Zazra, and J. F. Young. 1981. Genome diversity among influenza A, B and C viruses and genetic structure of RNA 7 and 8 of influenza A viruses. In: Genetic Variation among Influenza Viruses, D. P. Nayak, ed., Academic Press, New York, pp. 127-140.

Palese, P., V. R. Racaniello, U. Desselberger, J. F. Young, and M. Baez. 1980. Genetic structure and genetic variation of influenza viruses. Phil. Trans. Roy. Soc. London B 288:299-305.

Palese, P., M. B. Ritchey, and J. L. Schulman. 1977a. P1 and P3 proteins of influenza virus are required for complementary RNA synthesis. J. Virol. 21:1187-1195.

Palese, P., M. B. Ritchey, and J. L. Schulman. 1977b. Mapping of the influenza virus genome. II. Identification of the P1, P2, and P3 genes. Virology 76:114-121.

Palese, P., and J. L. Schulman. 1974. Isolation and characterization of influenza virus recombinants with high and low neuraminidase activity: Use of 2-(3'-methoxyphenyl)-N- acetylneuraminic acid to identify cloned populations. Virology 57:227-237.

Palese, P., and J. F. Young. 1982. Variation of influenza A, B, and C viruses. Science 215:1468-1474.

Palese, P., and J. F. Young. 1983. Molecular epidemiology of influenza virus. In: Genetics of Influenza Viruses, P. Palese and D. W. Kingsbury, eds., Springer-Verlag, New York, pp. 321-336.

Parvin, J. D., A. Moscona, W. T. Pan, J. M. Leider, and P. Palese. 1986. Measurement of the mutation rates of animal viruses: Influenza A virus and poliovirus type 1. J. Virol. 59:(in press).

Parvin, J. D., J. F. Young, and P. Palese. 1983. Nonsense mutations affecting the lengths of the NS1 nonstructural proteins of influenza A virus isolates. Virology 128:512-517.

Peacock, A. C., and C. W. Dingman. 1967. Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. Biochem. 6:1818-1827.

Penn, C. R., D. Blaas, E. Kuechler, and B. W. J. Mahy. 1982. Identification of the cap-binding protein of two strains of influenza A/FPV. J. gen. Virol. 62:177-180.

Petri, T., and N. J. Dimmock. 1981. Phosphorylation of influenza virus nucleoprotein *in vivo*. J. gen. Virol. 57:185-190.

Petri, T., G. Herrler, R. W. Compans, and H. Meier-Ewert. 1980. Gene products of influenza C virus. FEMS Microbiol. Lett. 9:43-47.

Petri, T., S. Patterson, and N. Dimmock. 1982. Polymorphism of the NS1 proteins of type A influenza virus. J. gen. Virol. 61:217-231.

Pfeifer, J. B., and R. W. Compans. 1984. Structure of the influenza C glycoprotein gene as determined from cloned DNA. Virus Res. 1:281-296.

Plotch, S. J., M. Bouloy, I. Ulmanen, and R. M. Krug. 1981. Initiation of influenza viral RNA transcription by capped RNA primers: a unique cap ($m^7GpppXm$)-dependent virion endonuclease generates 5' terminal RNA fragments that prime transcription. Cell 23:847-858.

Porter, A. G., J. C. Smith, and J. S. Emtage. 1980. Nucleotide sequence of influenza virus RNA segment 8 indicates that coding regions for NS1 and NS2 proteins overlap. Proc. Natl. Acad. Sci. USA 77:1857-1861.

Portner, A., R. G. Webster, and W. J. Bean. 1980. Similar frequencies of antigenic variants in Sendai, vesicular stomatitis, and influenza A viruses. Virology 104:235-238.

Prabhakar, B. S., M. V. Haspel, P. R. McClintock, and A. L. Notkins. 1982. High frequency of antigenic variants among naturally occurring human Coxsackie B4 virus isolates identified by monoclonal antibodies. Nature 300:374-376.

Privalsky, M. L., and E. E. Penhoet. 1981. Structure and synthesis of influenza virus phosphoproteins. J. Biol. Chem. 256:5368-5376.

Racaniello, V. R., and P. Palese. 1979. Influenza B virus genome: assignment of viral polypeptides to RNA segments. J. Virol. 29:361-373.

Ransohoff, R. M., P. A. Maroney, D. P. Nayak, T. M. Chambers, and T. W. Nilsen. 1985. Effect of human alpha A interferon on influenza virus replication in MDBK cells. J. Virol. 56:1049-1052.

Raymond, F. L., A. J. Caton, N. J. Cox, A. P. Kendal, and G. G. Brownlee. 1986. The antigenicity and evolution of influenza H1 haemagglutinin, from 1950-1957 and 1977-1983: Two pathways from one gene. Virology 148:275-287.

Raymond, F. L., A. J. Caton, N. J. Cox, A. P. Kendal, and G. G. Brownlee. 1983. Antigenicity and evolution amongst recent influenza viruses of H1N1 subtypes. Nucl. Acids Res. 11:7191-7203.

Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113:237-251.

Ritchey, M. B., and P. Palese. 1977. Identification of defective genes in three mutant groups of influenza virus. J. Virol. 21:1196-1204.

Ritchey, M. B., P. Palese, and E. D. Kilbourne. 1976a. RNAs of influenza A, B, and C viruses. J. Virol. 18:738-744.

Ritchey, M. B., P. Palese, and J. L. Schulman. 1976b. Mapping of the influenza virus genome. III. Identification of genes coding for nucleoprotein, membrane protein and nonstructural protein. J. Virol. 20:307-313.

Ritchey, M. B., P. Palese, and J. L. Schulman. 1977. Differences in protein patterns of influenza A viruses. Virology 76:122-128.

Robertson, J. S., E. Robertson, I. Roditi, J. W. Almond, and S. C. Inglis. 1983. Sequence analysis of fowl plague virus mutant ts47 reveals a nonsense mutation in the NS1 gene. Virology 126:391-394.

Robertson, J. S., M. Schubert, and R. A. Lazzarini. 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. 38:157-163.

Rogers, G. N., G. Herrler, J. C. Paulson, and H.-D. Klenk. 1986. Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. J. Biol. Chem. 261:5947-5951.

Romanos, M. A., and A. J. Hay. 1984. Identification of the influenza virus transcriptase by affinity- labeling with pyridoxal 5'-phosphate. Virology 132:110-117.

Rott, R., M. Reinacher, M. Orlich, and H.-D. Klenk. 1980. Cleavability of hemagglutinin determines spread of avian influenza viruses in chorioallantoic membrane of chicken embryo. Arch. Virol. 65:123-133.

Saitou, N., and M. Nei. 1986. Polymorphism and evolution of influenza A virus genes. Mol. Biol. Evol. 3:57-74.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schild, G. C., J. S. Oxford, and R. W. Newman. 1979. Evidence for antigenic variation in influenza A nucleoprotein. Virology 93:569-573.

Scholtissek, C., and A. L. Bowles. 1975. Isolation and characterization of temperature-sensitive mutants of fowl plague virus. Virology 67:576-587.

Scholtissek, C., W. Rohde, V. Von Hoyningen, and R. Rott. 1978a. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology 87:13-20.

Scholtissek, C., V. Von Hoyningen, and R. Rott. 1978b. Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). Virology 89:613-617.

Schulman, J. L., M. Khakpour, and E. D. Kilbourne. 1968. Protective effects of specific immunity to viral neuraminidase on influenza virus infection of mice. J. Virol. 2:778-786.

Schulman, J. L., and P. Palese. 1977. Virulence factors of influenza A viruses: WSN virus neuraminidase required for productive infection in MDBK cells. J. Virol. 24:170-176.

Sekikawa, K., and C.-J. Lai. 1983. Defects in functional expression of an influenza virus hemagglutinin lacking the signal peptide sequences. Proc. Natl. Acad. Sci. USA 80:3563-3567.

Selimova, L. M., V. M. Zaides, and V. M. Zhadanov. 1982. Disulfide bonding in influenza virus proteins as revealed by polyacrylamide gel electrophoresis. J. Virol. 44:450-457.

Sever, J. L. 1962. Application of a microtechnique to viral serological investigations. J. Immunol. 88:320-329.

Shaw, M. W., and P. W. Choppin. 1984. Studies on the synthesis of the influenza B virus NB glycoprotein. Virology 139:178-184.

Shaw, M. W., P. W. Choppin, and R. A. Lamb. 1983. A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminidase. Proc. Natl. Acad. Sci. USA 80:4879-4883.

Shaw, M. W., and R. W. Compans. 1978. Isolation and characterization of cytoplasmic inclusions from influenza A virus-infected cells. J. Virol. 25:608-615.

Shaw, M. W., and R. A. Lamb. 1984. A specific sub-set of host cell mRNAs prime influenza virus mRNA synthesis. Virus Res. 1:455-467.

Shaw, M. W., R. A. Lamb, B. W. Erickson, D. J. Briedis, and P. W. Choppin. 1982a. Complete nucleotide sequence of the neuraminidase gene of influenza B virus. Proc. Natl. Acad. Sci. USA 79:6817-6821.

Shaw, M. W., E. W. Lamon, and R. W. Compans. 1982b. Immunologic studies on the influenza A virus nonstructural protein NS1. J. Exp. Med. 156:243-254.

Skehel, J. J., P. M. Bayley, E. B. Brown, S. R. Martin, M. D. Waterfield, J. M. White, I. A. Wilson, and D. C. Wiley. 1982. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc. Natl. Acad. Sci. USA 79:968-972.

Skehel, J. J., and A. J. Hay. 1978a. Nucleotide sequences of the 5' termini of influenza virus RNAs and their transcripts. Nucl. Acids Res. 5:1207-1219.

Skehel, J. J., and A. J. Hay. 1978b. Influenza virus transcription. J. gen. Virol. 39:1-8.

Skehel, J. J., and G. C. Schild. 1971. The polypeptide composition of influenza A viruses. Virology 44:396-408.

Sleigh, M. J., G. W. Both, P. A. Underwood, and V. J. Bender. 1981. Antigenic drift in the hemagglutinin of the Hong Kong influenza subtype: Correlation of amino acid changes with alterations in viral antigenicity. J. Virol. 37:845-853.

Smith, D. B., and S. C. Inglis. 1985. Regulated production of an influenza virus spliced mRNA mediated by virus-specific products. EMBO J. 4:2313-2319.

Smith, G. L., and A. J. Hay. 1982. Replication of the influenza virus genome. Virology 118:96-108.

Sobrino, F., M. Davila, J. Ortin, and E. Domingo. 1983. Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture. Virology 128:310-318.

Spindler, K. R., F. M. Horodyski, and J. J. Holland. 1982. High multiplicities of infection favor rapid and random evolution of vesicular stomatitis virus. Virology 119:96-108.

Spooner, L. L. R., and R. D. Barry. 1977. Participation of DNA-dependent RNA polymerase II in replication of influenza viruses. Nature 268:650-652.

Staden, R. 1977. Sequence data handling by computer. Nucl. Acids Res. 4:4037-4051.

Staden, R. 1978. Further procedures for sequence analysis by computer. Nucl. Acids Res. 5:1013-1015.

Staeheli, P., O. Haller, W. Boll, J. Lindenmann, and C. Weissmann. 1986. Mx protein: Constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. Cell 44:147-158.

Staeheli, P., and O. Haller. 1985. Interferon-induced human protein with homology to protein Mx of influenza virus-resistant mice. Mol. Cell. Biol. 5:2150-2153.

Steinhauer, D. A., and J. J. Holland. 1986. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. J. Virol. 57:219-228.

Sugawara, K., K. Nakamura, and M. Homma. 1983. Analyses of structural polypeptides of seven different isolates of influenza C virus. J. gen. Virol. 64:579-587.

Sveda, M., L. J. Markoff, and C.-J. Lai. 1982. Cell surface expression of the influenza virus hemagglutinin requires the hydrophobic carboxy-terminal sequences. Cell 30:649-656.

Tanaka, T., M. Urabe, H. Goto, and K. Tobita. 1984. Isolation and preliminary characterization of a highly cytolytic influenza B virus variant with an aberrant NS gene. Virology 135:515-523.

Tanimura, M., K. Miyamura, and N. Takeda. 1985. Construction of a phylogenetic tree of enterovirus 70. Jpn. J. Genet. 60:137-150.

Thierry, F., and O. Danos. 1982. Use of specific single stranded DNA probes cloned in M13 to study the RNA synthesis of four temperature-sensitive mutants of HK/68 influenza virus. Nucl. Acids Res. 10:2925-2937.

Tobita, K., A. Sugiura, C. Enomoto, and M. Furuyama. 1975. Plaque assay and primary isolation of influenza A viruses in an established line of canine kidney cells (MDCK) in the presence of trypsin. Med. Microbiol. Immunol. 162:9-14.

Ueda, M., A. Tobita, A. Suguira, and C. Enomoto. 1978. Identification of hemagglutinin and neuraminidase genes of influenza B virus. J. Virol. 25:685-686.

Ulmanen, I., B. Broni, and R. M. Krug. 1983. Influenza virus temperature-sensitive cap (m^7GpppNm)-dependent endonuclease. J. Virol. 45:27-35.

Ulmanen, I., B. A. Broni, and R. M. Krug. 1981. Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m^7GpppNm) on RNAs and in initiating viral RNA transcription. Proc. Natl. Acad. Sci. USA 78:7355-7359.

Van Rompuy, L., W. Min Jou, D. Huylebroeck, and W. Fiers. 1982. Complete nucleotide sequence of a human influenza neuraminidase gene of subtype N2 (A/Victoria/3/75). J. Mol. Biol. 161:1-11.

Van Wyke, K. L., W. J. Bean, Jr., and R. G. Webster. 1981. Monoclonal antibodies to the influenza A virus nucleoprotein affecting RNA transcription. J. Virol. 39:313-317.

Van Wyke, K. L., J. W. Yewdell, L. J. Reck, and B. R. Murphy. 1984. Antigenic characterization of influenza A virus matrix protein with monoclonal antibodies. J. Virol. 49:248-252.

Van Wyke, K. L., V. S. Hinshaw, W. J. Bean, and R. G. Webster. 1980. Antigenic variation of influenza A virus nucleoprotein detected with monoclonal antibodies. J. Virol. 35:24-30.

Varghese, J. N., W. G. Laver, and P. M. Colman. 1983. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. Nature 303:35-40.

Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers. 1980. Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature 286:771-776.

Verhoeyen, M., L. Van Rompuy, W. Min Jou, D. Huylebroeck, and W. Fiers. 1983. Complete nucleotide sequence of the influenza B/Singapore/222/79 virus hemagglutinin gene and comparison with the B/Lee/40 hemagglutinin. Nucl. Acids Res. 11:4703-4712.

Ward, C. W., and T. A. Dopheide. 1981a. Amino acid sequence and oligosaccharide distribution of the hemagglutinin from an early Hong Kong influenza virus variant A/Aichi/2/68 (X-31). Biochem. J. 193:953-962.

Ward, C. W., and T. A. Dopheide. 1981b. Evolution of the Hong Kong influenza A subtype. Biochem. J. 195:337-340.

Waterfield, M., G. Scrace, and J. J. Skehel. 1981. Disulphide bonds of haemagglutinin of Asian influenza virus. Nature 289:422-424.

Waterfield, M. D., M.-J. Gething, G. Scrace, and J. J. Skehel. 1980. The carbohydrate side chains and disulfide bonds of the hemagglutinin of the influenza virus A/Japan/305/57 (H2N2). In: Structure and Variation in Influenza Virus, G. Laver and G. M. Air, eds., Elsevier/North Holland, New York, pp. 11-20.

Webster, R. G., L. E. Brown, and W. G. Laver. 1984. Antigenic and biological characterization of influenza virus neuraminidase (N2) with monoclonal antibodies. Virology 135:30-42.

Webster, R. G., C. H. Campbell, and A. Granoff. 1971. The in vivo production of "new" influenza viruses. I. Genetic recombination between avian and mammalian influenza viruses. Virology 44:317-328.

Webster, R. G., V. S. Hinshaw, and W. G. Laver. 1982a. Selection and analysis of antigenic variants of the neuraminidase of N2 influenza viruses with monoclonal antibodies. Virology 117:93-104.

Webster, R. G., A. P. Kendal, and W. Gerhard. 1979. Analysis of antigenic drift in recently isolated influenza A (H1N1) viruses using monoclonal antibody preparations. Virology 96:258-264.

Webster, R. G., and W. G. Laver. 1980. Determination of the number of non-overlapping antigenic areas of Hong Kong (H3N2) influenza virus hemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance. Virology 104:139-148.

Webster, R. G., W. G. Laver, G. M. Air, and G. C. Schild. 1982b. Molecular mechanisms of variation in influenza viruses. Nature 296:115-121.

Wharton, S. A., J. J. Skehel, and D. C. Wiley. 1986. Studies of influenza haemagglutinin-mediated membrane fusion. Virology 149:27-35.

White, J., J. Kartenbeck, and A. Helenius. 1982. Membrane fusion of influenza virus. EMBO J. 1:217-222.

White, J., M. Kielian, and A. Helenius. 1983. Membrane fusion proteins of enveloped animal viruses. Q. Rev. Biophys. 16:151-195.

Whitton, J. L., F. Hundley, B. O'Donnell, and U. Desselberger. 1983. Silver staining of nucleic acids. Applications in virus research and in diagnostic virology. J. Virol. Met. 7:185-198.

Wiley, D. C., I. A. Wilson, and J. J. Skehel. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289:373-378.

Wilson, A. C., S. S. Carlson, and T. J. White. 1977. Biochemical evolution. Ann. Rev. Biochem. 46:573-639.

Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289:366-373.

Winter, G., and S. Fields. 1980. Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. Nucl. Acids Res. 8:1965-1974.

Winter, G., and S. Fields. 1981. The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. Virology 114:423-428.

Winter, G., and S. Fields. 1982. Nucleotide sequence of human influenza A/PR/8/34 segment 2. Nucl. Acids Res. 10:2135-2143.

Winter, G., S. Fields, and G. G. Brownlee. 1981a. Nucleotide sequence of the haemagglutinin of a human influenza virus H1 subtype. Nature 292:72-75.

Winter, G., S. Fields, M. J. Gait, and G. G. Brownlee. 1981b. The use of synthetic oligonucleotide primers in cloning and sequencing segment 8 of influenza virus (A/PR/8/34). Nucl. Acids Res. 9:237-245.

Wolstenholme, A. J., T. Barrett, S. T. Nichol, and B. W. J. Mahy. 1980. Influenza virus-specific RNA and protein syntheses in cells infected with temperature-sensitive mutants defective in the genome segment encoding nonstructural proteins. J. Virol. 35:1-7.

Wrigley, N. G. 1979. Electron microscopy of influenza virus. Brit. Med. Bull. 35:35-38.

Yewdell, J. W., R. G. Webster, and W. Gerhard. 1979. Antigenic variation in three distinct determinants of an influenza type A hemagglutinin molecule. Nature 279:246-248.

Yokota, M., K. Nakamura, K. Sugawara, and M. Homma. 1983. The synthesis of polypeptides in influenza C virus-infected cells. Virology 130:105-117.

Young, J. F., U. Desselberger, and P. Palese. 1979. Evolution of human influenza A viruses in nature: Sequential mutations in the genomes of new H1N1 isolates. Cell 18:73-83.

Young, J. F., U. Desselberger, P. Palese, B. Ferguson, A. R. Shatzman, and M. Rosenberg. 1983. Efficient expression of influenza virus NS1 nonstructural proteins in *E. coli*. Proc. Natl. Acad. Sci. USA 80:6105-6109.

Young, J. F., and P. Palese. 1979. Evolution of human influenza A viruses in nature: Recombination contributes to genetic variation of H1N1 strains. Proc. Natl. Acad. Sci. USA 76:6547-6551.

Zarling, D. A., and H. M. Temin. 1976. High spontaneous mutation rate of an avian sarcoma virus. J. Virol. 17:74-84.

Zebedee, S., C. D. Richardson, and R. A. Lamb. 1985. Characterization of the influenza virus M2 integral membrane protein and expression at the infected-cell surface from cloned cDNA. J. Virol. 56:502-511.

Zhirnov, O., and A. G. Bukrinskaya. 1984. Nucleoproteins of animal influenza viruses, in contrast to those of human strains, are not cleaved in infected cells. J. gen. Virol. 65:1127-1134.

Zhirnov, O. P., and A. G. Bukrinskaya. 1981. Two forms of influenza virus nucleoprotein in infected cells and virions. Virology 109:174-179.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. In: Evolving Genes and Proteins, V. Bryson and H. J. Vogel, eds., Academic Press, New York, pp. 97-166.