Order Number 8801747

Analysis of mutation rates of RNA viruses

Parvin, Jeffrey David, Ph.D.

City University of New York, 1987

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

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 _____

_____

_____

ANALYSIS OF MUTATION RATES OF RNA VIRUSES

by

Jeffrey D. Parvin

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.

1987

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.

6/19/87
Date

Peter Palese
Peter Palese, Ph.D.
Chair of Examining Committee

7/15/87
Date

Terry A. Krulwich, Ph.D.
Executive Officer

Ming-Ta Hsu, Ph.D.

Jerome L. Schulman, M.D.

Carter Bancroft, Ph.D.

Lu-Hai Wang, Ph.D.

Supervisory Committee

The City University of New York

ABSTRACT

ANALYSIS OF MUTATION RATES OF RNA VIRUSES

by

Jeffrey D. Parvin

Advisor: Dr. Peter Palese

Epidemiologic and genetic evidence suggests that influenza A viruses evolve more rapidly than other viruses in man. In this thesis the evolution of the influenza A virus nonstructural (NS) gene was studied by two approaches. First, the evolution of the gene derived from viruses isolated over a 50 year period was delineated by sequence analysis. We found that the NS gene evolved at a very high rate of 2×10^{-3} substitutions/nucleotide/year and that this evolution was along a primary lineage. The lack of significant branching in the evolutionary tree suggested that selective pressures are involved in shaping the evolution. Second, we determined the mutation rate of this gene in tissue culture. Although high mutation rates are often implicated in generating rapid evolution, direct measurement of this parameter has not been previously determined for any animal virus. In this study, the rates of mutation of the NS gene of the influenza A virus and for the VP1 gene in poliovirus type 1 were measured by direct sequence analysis. Each gene was repeatedly sequenced in over one hundred viral clones which were descended from a single virion in one plaque generation. 108 NS genes of influenza virus were sequenced, and in the 91,708 nucleotides analyzed seven point changes were observed. 105 VP1 genes of poliovirus were sequenced and in the 95,688 nucleotides analyzed zero

mutations were observed. We then calculated mutation rates of 1.5 X 10^{-5} and less than 2.1 X 10^{-6} mutations/nucleotide/infectious cycle for influenza virus and poliovirus, respectively. We suggest that the higher mutation rate of influenza A virus may promote the rapid evolution of this virus in nature.

Also described is the development of technologies which should facilitate future mutation rate analyses. Single point mutations in the NS gene of influenza virus were detected by electrophoresis of double stranded RNA heteroduplexes in denaturing gradient gels. The heteroduplex RNAs were made by hybridization of virion RNA with SP6-derived RNA probes of varying length. Mutations located at different positions along the NS gene (890 nucleotides long) were all detected in a predictable fashion. The method of heteroduplex analysis was also successfully used in detecting single point mismatches in DNA-RNA hybrids.

Further, a method is described for directly sequencing an RNA probe using SP6 RNA polymerase in the presence of 3' deoxynucleotide triphosphates.

Finally, the genome of the viral RNA was found to be held in a circular conformation by a terminal panhandle which is present in the ribonucleoprotein complex in virions and in influenza A virus infected cell. We suggest a regulatory function in this panhandle conformation in promoting synthesis of virus specific message RNA. We also suggest that the structure may be involved in packaging the genome in progeny virus.

FORMAT OF THESIS

This thesis is prepared according to the new guidelines of the City University of New York which permit the direct incorporation of published research articles as chapters. The thesis has a general introduction, and chapters may have specific introductory statements. Materials and methods and results sections are in each individual chapter. Also, each chapter has a specific discussion section, and there is a general discussion as the final chapter of the thesis. The references for all chapters are pooled, in order to avoid redundancy.

Copyright permission for each chapter has been obtained from the publishers.

It should be noted that publications 4 through 8 (listed on page 8) form the basis of this thesis. Publications 1 and 3 were from studies completed during apprenticeships in the laboratories of Dr. P. Palese and of Dr. L.-H. Wang. Publication 2 is derived from research performed while I was an undergraduate student.

TABLE OF CONTENTS

PUBLICATIONS

Parvin, J.D., Young, J.F. and Palese, P. (1983). Nonsense mutations affecting the lengths of the NS1 nonstructural proteins of influenza A virus isolates, Virology 128, 512-517.

Boyer, K.M., Papierniak, C.K., Gadzala, C.A., Parvin, J.D. and Gotoff, S.P. (1984). Transplacental passage of IgG antibody to group B streptococcus serotype Ia. The Journal of Pediatrics 104, 618-620.

Parvin, J.D. and Wang, L.-H. (1984). Mechanisms for the generation of src-deletion mutants and recovered sarcoma viruses: Identification of viral sequences involved in src deletions and in recombination with c-src sequences, Virology 138, 236-245.

Parvin, J.D., Smith, F.I. and Palese, P. (1986). Rapid RNA sequencing using double-stranded template DNA, SP6 polymerase and 3'-deoxy nucleotide triphosphates, DNA 5, 167-171.

Smith, F.I., Parvin, J.D. and Palese, P. (1986). Detection of single base substitutions in influenza virus RNA molecules by denaturing gradient gel electrophoresis of RNA-RNA or DNA-RNA heteroduplexes, Virology 150, 55-64.

Buonagurio, D.A., Nakada, S., Parvin, J.D., Krystal, M., Palese, P. and Fitch, W.M. (1986). Rapid evolution of influenza A viruses in man: A model for a molecular clock, Science 232, 980- 982.

Parvin, J.D., Moscona, A., Pan, W.T., Leider, J.M. and Palese, P. (1986). The mutation rate of animal viruses: Influenza A virus and poliovirus type 1, J. Virol. 59, 377-383.

Hsu, M.-T., Parvin, J.D., Gupta, S., Krystal, M. and Palese, P. The genome RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proc. Natl. Acad. Sci. U.S.A. (in press).

TABLES

FIGURES

I. INTRODUCTION

A. The influenza A virus causes respiratory infections in man and in animals. It is most noted for periodic pandemics and epidemics which affect large segments of the population. These pandemics and epidemics arise due to the unusual evolutionary characteristics of the virus. Its very rapid evolution, with cumulative substitutions, has confounded vaccination programs (for reviews: Palese and Kingsbury, 1983; Webster, et. al., 1982). Immunization against virus from only a few years ago is not protective against virus which may infect today or in the future. Thus in a day of widespread vaccination, influenza A virus remains a significant viral pathogen. A major thrust of this thesis is to understand whether there is a fundamental property of the influenza A virus which permits it to have this unusually rapid evolution. The significance of this question is highlighted by the emergence of an epidemic caused by human immunodeficiency virus. Although the evolutionary pattern of this virus is less well characterized, it appears that this virus also varies dramatically in nature (Hahn, et. al., 1986).

Another emphasis in the thesis is the development of methodology which may facilitate experimental findings regarding mutation rates of viruses. The first of these findings is the development of a technique for detecting point mutations without sequencing. The second is the development of a method which allows one to sequence directly the probes used for point mutation detection.

A final theme is the secondary structure of the viral genome. We observed that the termini of the single-stranded RNAs of the influenza viruses are base-paired, and we propose that these panhandle structures provide a mechanism to regulate primary transcription and to package genomic RNAs into

virus.

B. STRUCTURE OF INFLUENZA A VIRION

The influenza virion is observed in the electron microscope as a roughly spherical particle about 80 to 120 nm in diameter containing a lipid envelope (Horne, et. al., 1960; Wrigley, 1979), seven structural proteins and the RNA genome (reviewed in Lamb, 1983). The viral RNA consists of eight single stranded segments of negative polarity (Ritchey, et. al., 1976; Palese and Schulman, 1977; Palese, 1977).

There are ten known protein products encoded by the viral RNA; a summary of the physical properties and functions of each can be seen in Table 1. The three polymerase proteins function in a complex (Rochovansky, 1976; Krystal, et. al., 1986; Detjen, et. al., 1987) in association with the nucleocapsid protein (NP) on the RNA template. The major surface proteins, and thus the major antigens, are the hemagglutinin (HA) and neuraminidase (NA) proteins. The matrix protein (M1), the predominant protein of the virion, underlies the lipid membrane. The three nonstructural proteins, M2, NS1 and NS2 have undetermined functions. The last four proteins are encoded by two genome RNAs. Two of the translation products result from the unspliced form of their respective mRNAs, the M1 and NS1 proteins, and the M2 and NS2 each are translated from spliced mRNAs (see Fig. 1). (See Table 1 and Fig. 1 for references.)

Table 1.

INFLUENZA A/PR/8/34 VIRUS GENOME AND CODING ASSIGNMENTS*

Segment	vRNA length (nucleotides)	encoded polypeptide	function (references)
1	2,341	PB2	cap endonuclease, polymerase complex (Plotch, et. al., 1981, Ulmanen, et al., 1981)
2	2,341	PB1	initiation and elongation of transcription, polymerase complex (Ulmanen, et. al., 1981; Palese, et. al., 1977)
3	2,233	PA	polymerase complex (Braam, et. al., 1983)
4	1,778	HA	major surface antigen, receptor binding, fusion (Wilson, et. al., 1981)
5	1,565	NP	nucleocapsid protein, structural component of polymerase complex (Compans, et. al., 1972)
6	1,413	NA	minor surface antigen, receptor destroying enzyme (Colman, et. al., 1983)
7	1,027	M1	major structural protein, underlies lipid envelope (Compans, et. al., 1970)
		M2	spliced mRNA, nonstructural glycoprotein, membrane bound (Lamb and Choppin, 1981)
8	890	NS1	nonstructural protein (Skehel, 1972)
		NS2	spliced mRNA, nonstructural protein (Lamb and Lai, 1980)

* Modified from Lamb (1983).

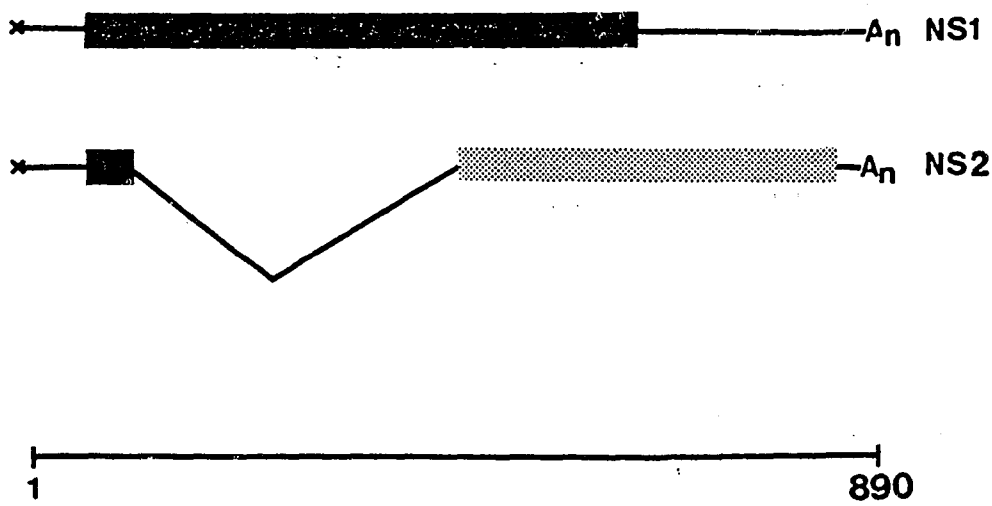


Figure 1. SPLICED AND UNSPLICED MESSAGE RNAS OF SEGMENT 8.

The mRNAs encoded by viral RNA segment 8 are depicted with the associated polypeptide indicated at the right of each. The NS1 protein is translated from unspliced mRNA with a single open reading frame. The spliced mRNA interrupts the reading frame and at the splice junction forms a new reading frame. The NS1 and NS2 reading frames overlap for the first 10 codons, and the NS2 polypeptide is translated in a +1 reading frame (relative to NS1) downstream of the splice junction (Lamb and Lai, 1980). The size of the second overlap region is variable among strains due to nonsense mutations which occur in the NS1 coding region (Parvin, et. al., 1983; Krystal, et. al., 1983). A similar coding strategy is observed with segment 7 (Lamb, et. al., 1981).

C. TRANSCRIPTION AND REPLICATION

The virus penetrates the host cell via the endosome pathway and by fusing its envelope with the host membrane is able to deposit its RNP core into the cytosol (White, et. al., 1982). The core complex is transported to the nucleus where the viral transcription occurs. The input viral RNA is of negative polarity with regard to its coding strand (Pons, 1971; Baltimore, 1971). Encapsidating the RNA in the core are the nucleocapsid protein (NP) and the viral polymerase complex consisting of the PB2, PB1 and PA proteins (Kato, et. al., 1985). The viral polymerase transcribes from the negative strand two classes of RNA: message and template (Hay, et. al., 1977; Beaton and Krug, 1986). The primary transcription (mRNA synthesis) is initiated by an unusual process. It has been shown that the PB2 protein binds to host-derived capped mRNA and that the protein cleaves the RNA ten to twelve nucleotides downstream of the guanosine cap. The viral polymerase then initiates mRNA synthesis using this capped RNA oligomer as a primer (Plotch, et. al., 1979). Synthesis continues until about 15 to 20 nucleotides from the 5' end of the vRNA template where termination apparently occurs and the nascent RNA is polyadenylated (Robertson, et. al., 1981; Beaton and Krug, 1986). This mRNA product may or may not be spliced and then is transported to the cytoplasm where translation occurs. The second kind of synthesis produces complementary template RNA. This cRNA is an exact copy of the input vRNA, and it thus serves as template for synthesizing viral RNA. In <u>in vitro</u> studies cRNA synthesis has been initiated with a dinucleotide primer (McGeogh and Kitron, 1975; Beaton and Krug 1986), but it is not known how it is primed <u>in vivo.</u> In chapter VI of the thesis, evidence supporting the role of secondary structure of the vRNA (which serves as template for both kinds of positive strand RNA) in

regulating the switch from message RNA synthesis to template RNA synthesis will be described.

D. EPIDEMIOLOGY

The epidemiology of influenza A viruses is unique among viruses because of the very high rate of evolution and because of the singular shape of the evolutionary tree. Pandemics of influenza occur when the genes encoding the surface antigens, the HA and NA molecules, become exchanged (through reassortment) with the genes derived from a nonhuman influenza virus resulting in a virus which is completely novel to the human population with respect to its antigens. This process is called antigenic shift. In addition, during the periods between pandemics, the genes encoding the major antigen mutate, thus generating subtly different antigens against which the host has reduced immune protection. This latter process is called antigenic drift. These mutations have been shown to accumulate in subsequent strains, and thus a very rapid evolution along a single lineage is observed with these genes (reviewed in: Webster, et. al., 1982).

These two properties of shift and drift have greatly hampered vaccination programs against influenza (Bull. WHO, 65, 479-484, 1985). It is impossible to predict when antigenic shift will occur or what new antigen it will present, thus making vaccination impossible. Antigenic drift also limits the efficaciousness of vaccination in periods between pandemics. The changes accumulate in the HA with sufficient rapidity to limit the protection conferred by vaccination against virus only several years old. Not only do the vaccine strains have to be frequently updated, but frequent revaccination is also required.

Evolutionary studies of the HA gene indicate that the changes in the protein structure are reflected by substitutions in the coding sequence. The two key observations are that the substitutions are sequential in time and that many mutations correlate with regions that have been determined to encode the antibody binding sites (Wiley, et. al., 1981; Webster, et. al., 1982; Raymond, et. al., 1986). These findings strongly suggest that there is immunological selection acting on the viral population. The immune surveillance of the host population may limit the growth of viruses which do not have mutated antibody binding sites. Thus, a very potent selection appears to be driving the evolution of these viruses by forcing the virus to change in order to survive.

E. EVOLUTION OF THE NS GENE OF INFLUENZA A VIRUSES

In our laboratory, we have been studying the evolution of the nonstructural gene. It has been found that this gene tolerates a great deal of point mutations, particularly in the NS1 reading frame. For example, in analysis performed before commencement of the thesis project we demonstrated that the NS1 is a highly plastic protein in which the carboxy terminus can vary greatly in length (Parvin, et. al., 1983). We found point mutations which generate nonsense codons in the NS1 reading frame while maintaining the NS2 reading frame. These changes generated NS1 proteins whose migrations by polyacrylamide gel electrophoresis were highly variable but were relatively consistent with regard to size. Subsequent studies further demonstrated the highly variable nature of the NS gene sequence (Krystal, et. al.,1983; Buonagurio, et. al., 1984). The second chapter of this thesis is adapted from a publication in which the principle investigator was D.A. Buonagurio in which the evolutionary tree of the NS gene from viruses isolated over 50 years is

described (Buonagurio, et. al., 1986b). We found that mutations accumulated in the NS gene in a sequential manner, as was found for the HA gene, and that the rate of evolution was high, 2 X 10^{-3} substitutions/ nucleotide/year, but not as high as with the HA gene (about 7 X 10^{-3} substitutions/nucleotide/year; D.A. Buonagurio PhD. Thesis). Since the NS gene has not reassorted in human strains, as has the HA, we were able to study genes isolated over much greater times. We found a single lineage for the NS gene, even across pandemics with the exception of the post-1977 H1N1 isolates. In this latter case a virus apparently derived from 1950 re-emerged in 1977, and continued as a single separate lineage (Nakajima, et. al., 1978; Kendal, et. al., 1978). The shape of the evolutionary tree of the NS, with a single lineage, is indicative of a selective pressure acting on the virus. We develop in chapter II the concept that mutant NS genes "hitchhike" with an immune selected HA variant. Thus, selection seems to act indirectly on the NS gene, via the HA gene, in driving and shaping its evolution.

F. MUTATION RATE ANALYSIS

1. POSSIBLE ROLE OF MUTATION RATE IN EVOLUTION

Why is the influenza A virus apparently uniquely suited to escape the immune surveillance of the host population? There are many vaccination programs against viral pathogens which are enormously successful. For example, every child in the United States receives a series of immunizations against poliovirus using a vaccine strain derived from a virus isolated 50 years ago. Except for rare cases of vaccine-related poliomyelitis, the vaccine has virtually eliminated the disease. The selective pressures acting on the poliovirus may be somewhat different than is the case with influenza A virus since it is an

enteric virus, but both classes of virus elicit IgA antibody-mediated immune response. In each case this immune response is protective against future infections by the virus. In chapter III a study is described in which the mutation rates of influenza A virus and of poliovirus type 1 are measured. A comparison of this fundamental property of each virus might indicate whether high mutation rate is responsible for the rapid evolution in influenza A viruses.

2. STRUCTURE AND EPIDEMIOLOGY OF POLIOVIRUS

Poliovirus belongs to the class of picornaviruses. There is no viral envelope, but rather there is a tightly packed shell consisting of four structural proteins encapsidating the genome RNA. The three-dimensional structure of the Mahoney strain of poliovirus type 1 has been solved (Hogle, et. al., 1985) and demonstrates that three of the proteins, VP1, VP2 and VP3 form the external shell of the virus, and VP4 is on the inside of the shell. Antibodies against VP1, VP2 and VP3 are protective with VP1 being a major target of the immune response (Emini, et. al., 1983).

In the experimental settings, the poliovirus has demonstrated an extensive capability for change. Point mutants can be selected using monoclonal antibodies, and these antigenic variants arise with frequencies similar to most other viruses (Emini, et. al., 1983; Holland, et. al., 1982). However, antigenic variants do not seem to appear in nature. The Sabin type 1 vaccine strain has been sequenced in its entirety (Nomoto, et. al., 1982) as has its parent strain, the Mahoney virus (Racaniello and Baltimore, 1981; Kitamura, et. al., 1981). Sequence comparison between these two viruses reveals that the VP1 gene contains the most point changes and the highest number of coding changes in comparison to all the other genes. Variation in wildtype and vaccine-related polioviruses has been observed in nature (Nottay, et. al., 1981), and a portion of

the VP1 gene sequence has been shown to vary to some extent (Rico-Hesse, et. al., 1986). There is no indication, however, that extensive antigenic variation of polioviruses is common in nature.

3. The question we address in chapter three is whether there is a fundamental property of the influenza A virus which permits it to escape the immune surveillance and thus grow despite the presence of antibodies specific for influenza viruses. We measured a higher mutation rate (mutations per nucleotide per infectious cycle) for influenza A virus than for poliovirus. We believe that this high mutation provides a mechanism for influenza A virus to escape the immune surveillance of the host. Therefore, the ability to mutate rapidly coupled with an immune selective pressure results in a high evolution rate.

G. NEW METHODOLOGY -- POINT MUTATION DETECTION

1. The extensive sequence analysis described in chapter III was found to be a cumbersome procedure. Using the point mutants detected by sequencing we sought to develop procedures which would simplify and speed up the analysis to detect single point mutations. Methods developed by other laboratories utilize heteroduplexes in which a point mutation becomes a mismatch. The mismatch can then be detected by disruption of the double helix with subsequent cleavage by ribonuclease A (Winter, et. al., 1986; Myers, et. al., 1986) or by binding the mismatch with carbodiimide (Novack, et. al., 1986) or by destabilization of the double strand in denaturing gels(Lerman, et. al., 1984; Myers, et. al., 1985).

The destablizing effect of a mismatch in a DNA heteroduplex has been successfully used by Lerman and colleagues to detect point mutations in DNA molecules. These workers developed methods for determining sequence-specific

melting domains in DNA. A mismatch in a domain lowers the melting temperature (Tm) for that domain. Therefore a mismatch in the heteroduplex causes the domain to melt at a lower temperature, or under milder conditions, than in the perfectly complementary hybrid. If the point mutation is in a domain which melts at a lower temperature than the adjacent domain, then denaturation of the lower domain results in a molecule which is partially single stranded (the low melting domain) and partially double stranded (the higher melting domain). This partially denatured duplex migrates very slowly in polyacrylamide gels as compared to the non-melted duplex. These characteristics have been used to develop a detection system by which the hybrids are electrophoresed under conditions close to, but just below, the Tm for the domain. The gel contains a gradient of denaturant (formamide and urea), and as the sample migrates, the denaturant increases to a point where the mismatched duplex melts and has greatly slowed gel migration while the non-mismatched duplex continues to migrate. Following electrophoresis, the gel is autoradiographed, and the mismatched hybrid is observed as a band which has not migrated as far as the sample without mismatches (Lerman, et. al., 1984; Myers, et. al., 1985).

2. In chapter IV we describe an extension of this form of analysis. In experiments, in which the primary investigator was F.I. Smith, we determined that this method of denaturant-gradient gel electrophoresis could be extended for use with RNA-RNA and RNA-DNA heteroduplexes. Also, this finding permits usage of SP6 derived probes which can be obtained easily and in high quantity. In addition, RNA genomes can be investigated.

H. NEW METHODOLOGY -- SP6 SEQUENCING

Current sequencing systems of general application use one of two strategies. The first strategy is that of specifically cleaving end-labelled DNAs with chemical reagents (Maxam and Gilbert, 1977). The other strategy is that of specific termination of synthesis during polymerization (Sanger, et. al., 1977). By this latter method, a single stranded DNA is the template for reactions which use the Klenow fragment of DNA polymerase I and dideoxynucleotide triphosphate analogs as chain terminators (Sanger, et. al., 1977; Messing, et. al., 1981). This latter strategy has been extended to using a template which is a double stranded, supercoiled plasmid (Chen and Seeburg, 1985). Additionally, RNAs could be sequenced by this strategy using reverse transcriptase (Zimmern and Kaesberg, 1978; chapter III of this thesis). In chapter V we describe the first generally applicable sequencing method which uses the Sanger strategy, but uses an RNA polymerase and ribonucleotide triphosphate analogs as chain terminators. The sequencing reactions do not require a primer binding step, but, rather, utilize the promoter present in the plasmid to initiate RNA synthesis. In chapter V we discuss the advantages and disadvantages of this method.

I. GENOME STRUCTURE -- THE PANHANDLE

1. Detergent disrupted virus releases a ribonucleoprotein (RNP) core which contains the genome RNA ensheathed by a large number of nucleoprotein (NP) molecules and the three polymerase proteins in approximately stoichiometric amounts (Kato, et. al., 1985). The RNP complexes are in different size classes, and it was shown that the size depends on the individual RNA segment present (Duesberg, 1969). In the electron microscope a higher order structure was observed as a large double helix at one end and a loop at the other end

(Compans, et. al., 1972). The function of these structures is unknown. Sequence information of the genome segments revealed that the ends of each RNA are constant with 12 bases conserved at the 3' end and 13 bases conserved at the 5' end (Skehel and Hay, 1978; Robertson, 1979; Desselberger, et. al., 1980). In addition, it was observed that these regions are partially complementary, and that they were predicted to base pair and form a panhandle (Robertson, 1979; Desselberger, et. al. 1980). Adjacent to the common 12 to 13 nucleotide termini there are in each RNA three additional nucleotides which may contribute to the stability of the panhandle since they are complementary. These three nucleotides are specific for an RNA segment.

Juxtaposed to the hypothetical panhandle on the 5' terminus of the viral RNA are five to six uridine residues which are thought to function as the signal for termination and polyadenylation of message RNAs. This U_5 or U_6 sequence occurs many times in the genome but will terminate and polyadenylate only in this location. It was therefore suggested that its proximity to the panhandle structure may be critical to its recognition as a signal (Robertson, et. al., 1981). It should be added, however, that the panhandle structure has never previously been shown to exist, and that such a demonstration may be difficult since the stability of the secondary structure may be very low due to the low level of base pairing.

2. In chapter VI, in a project in which the primary investigator was M.-T. Hsu, we demonstrate that the panhandle structure does in fact exist. Using a psoralen derivative which can crosslink double stranded RNA, we observed the panhandle was present in intact virus and in infected cells at both early and late time points. At earlier times of infection, a greater percent of the viral RNA was in the panhandle conformation than at later time points. The time

course of panhandle formation approximately coincided with the regulation of message RNA versus replicative RNA synthesis. These data led to the conclusion that the panhandle conformation does in fact have a regulatory function in the transcription of mRNA, and since it was observed in virus, we believe that it also may be necessary for packaging of the RNA into virus.

SPECIFIC AIMS

A. The antigenic variation of influenza A viruses is a well recognized phenomenon (e.g. Webster, et. al., 1982) and is a characteristic which has curtailed vaccination programs (Bull. WHO, 63, 479-484, 1985). The antigenic drift within a subtype has been extensively studied for the hemagglutinin protein (e.g. Both, et. al., 1983). We sought to first identify the evolutionary pattern of the nonstructural gene, a gene which might not be subject to immune selection. Second, we sought to develop a model which would explain the evolution of the influenza A virus.

B. We sought to measure the mutation rate of the influenza A virus in tissue culture and also measure the same parameter in the evolutionarily more stable poliovirus type 1. The determination of this fundamental parameter of each virus should help in elucidating the role of the mutation rate in the evolution of viruses.

C. We sought to develop new technologies which would facilitate future mutation rate studies. These included a method for detecting point mutations in the RNA genomes of viruses and a method for sequencing the RNA probes used in the analyses.

D. Finally, we wished to determine the structure of the influenza virus genome with regard to secondary structure. A panhandle structure of the RNAs had been predicted based on sequence analysis. We indeed provided proof for the existence of such panhandle structures in virus and in virus infected cells, and we suggested possible functions for these structures.

II.

EVOLUTION OF HUMAN INFLUENZA A VIRUSES OVER 50 YEARS: RAPID AND UNIFORM RATE OF CHANGE IN THE NS GENE

Deborah A. Buonagurio, Susumu Nakada, Jeffrey D. Parvin,

Mark Krystal, Peter Palese and Walter M. Fitch*

Dept. of Microbiology, Mt. Sinai School of Medicine of CUNY
Fifth Avenue and 100th Street, New York, N.Y. 10029

*Dept. of Physiological Chemistry, University of Wisconsin, Center for Health Sciences, 1300 University Avenue, Madison, WI 53706

Originally published in Science 232, 980-982 (1986). Reprinted with permission from the American Association for the Advancement of Science.

A. INTRODUCTION

Variation of influenza A viruses was examined by comparison of nucleotide sequences of the NS gene (890 bases) of fifteen human viruses isolated over fifty-three years (1933-1985). Changes in the genes accumulate with time and an evolutionary tree based on the maximum parsimony method can be constructed. The evolutionary rate is approximately 2×10^{-3} substitutions per site per year in the NS genes, which is about 10^6 times faster than the evolutionary rate of germ-line genes in mammals. This uniform and rapid rate of evolution in the NS gene thus makes it a good molecular clock and is compatible with the hypothesis that positive selection is operating on the hemagglutinin (or perhaps some other viral genes) to preserve random mutations in the NS gene.

B. MATERIALS AND METHODS

1. VIRUSES

Influenza viruses A/Puerto Rico/8/34, A/Bellamy/42, A/Fort Monmouth/1/47, A/Fort Warren/1/50, A/Denver/1/57, A/Ann Arbor/6/60, A/Berkeley/1/68, A/Alaska/6/77, A/Houston/24269/85, A/Maryland/2/80, A/USSR/90/77, A/Houston/18515/84 and A/Houston/23284/85, are abbreviated PR34, BEL42, FM47, FW50, DEN57, AA60, BERK68, ALA77, (H3)HT85, MD80, USSR77, (H1)HT84 and (H1)HT85), respectively. Virus stocks were grown and RNA was prepared using standard procedures (Ritchey and Palese, 1976).

2. SEQUENCE ANALYSIS

The NS sequences of the A/WSN/33, 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 direct RNA sequencing of purified vRNA by the dideoxy chain terminator method

(Sanger, et. al., 1977; see chapter III for a thorough description). Sequencing primers were oligonucleotides 15-20 bases in length synthesized using either the phosphoramidite or phosphotriester chemistry on the Biosearch SAM One automated DNA synthesizer (Biosearch, Inc., CA) (Buonagurio, et. al., 1984). The 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. Nucleotide sequences were stored, edited and compared in an IBM 370 computer at the City University of New York using published programs (Staden, 1978).

C. RESULTS

Fig. 2 shows the nucleotide sequences of the NS genes of 15 human influenza A virus strains. The viruses were isolated over a 53 year period and represent all three human hemagglutinin serotypes (H1, H2 and H3). Except for the three Houston isolates, the strains were obtained from diverse geographical locations. The 15 sequences are easily aligned for analysis because of the size conservation of the NS gene segment of 890 bases. Nucleotide substitutions occur at 149 positions scattered throughout the gene and usually, once a base change is observed in a virus isolate, it is found in subsequent strains. The sequence information as presented in Fig. 2 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. 3. The parsimony method also yielded four alternative trees containing 187 substitutions. These alternative trees contain only minor branch perturbations of the best tree.

Figure 4 shows the number of nucleotide substitutions between the origin of the best tree and the tip of each branch (Fig. 3) plotted against the date of isolation of the viruses whose NS gene is represented by that tip. The major

Figure 2. NUCLEOTIDE SEQUENCES OF 15 NS GENES OF HUMAN INFLUENZA A VIRUSES.

The NS sequences of the A/WSN/33, 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 in this study. The sequences of the remaining NS genes have been published (Krystal, et. al., 1983; Buonagurio, et. al., 1984; 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 cDNA clone and the universal primer.

Figure 2

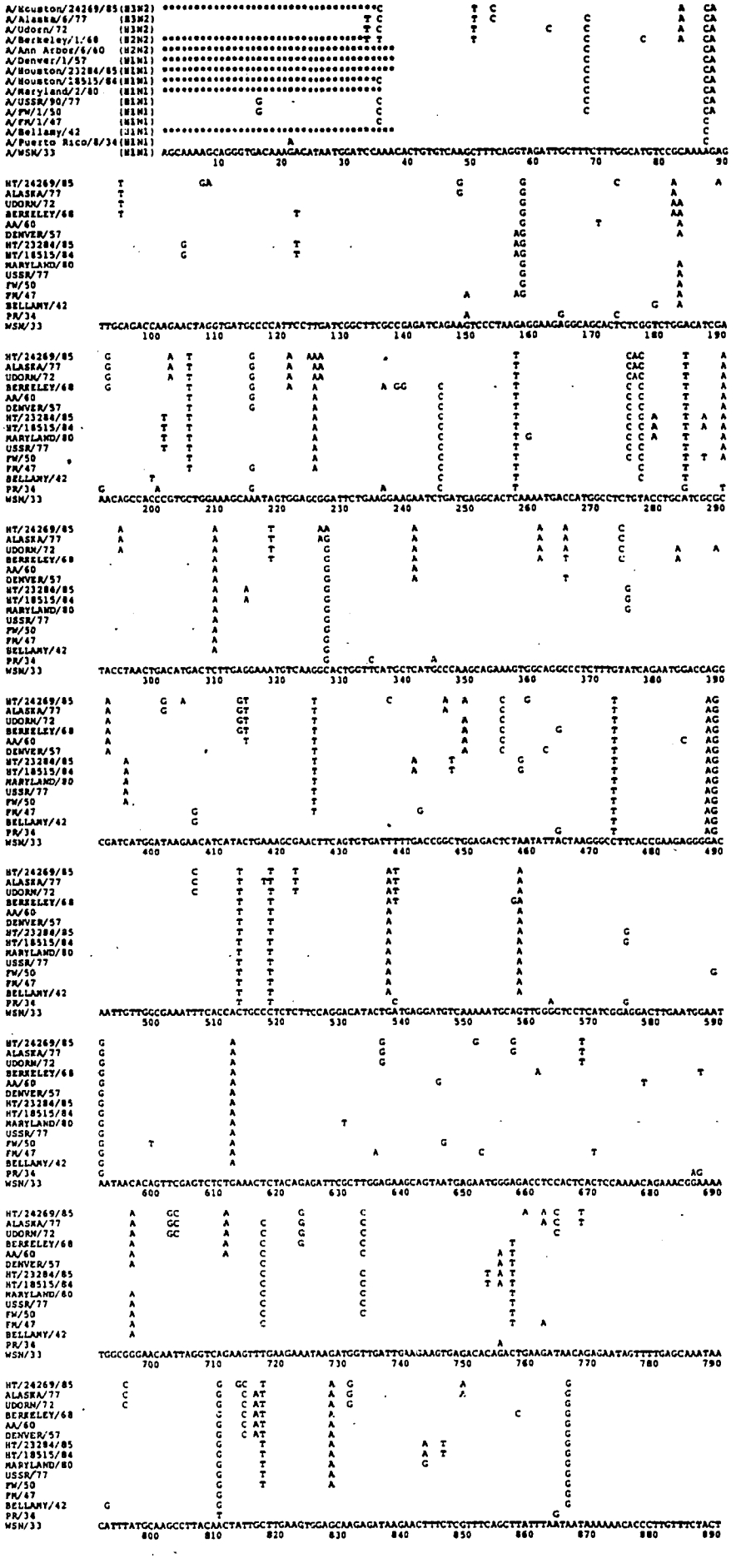

A/Kcuston/24269/85 (H3N2)
A/Alaska/6/77 (H3N2)
A/Udorn/72 (H3N2)
A/Berkeley/1/68 (H2N2)
A/Ann Arbor/6/60 (H2N2)
A/Denver/1/57 (H1N1)
A/Houston/23284/85 (H1N1)
A/Houston/18515/84 (H1N1)
A/Maryland/2/80 (H1N1)
A/USSR/90/77 (H1N1)
A/FW/1/50 (H1N1)
A/FM/1/47 (H1N1)
A/Bellamy/42 (H1N1)
A/Puerto Rico/8/34 (H1N1)
A/WSN/33 (H1N1)
AGCAAAAGCAGGGTGACAAAGACATAATGCATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCTTTGGCATGTCCGCAAAAGAG
10 20 30 40 50 60 70 80 90
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
TTGCAGACCAAGAACTAGGTGATGCCCCATTCCTTGATCGGCTTCGCCGAGATCAGAAGTCCCTAAGAGGAAGAGGCAGCACTCTCGGTCTGGACATCGA
100 110 120 130 140 150 160 170 180 190
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
AACAGCCACCCGTGCTGGAAAGCAAATAGTGGAGCGGATTCTGAAGGAAGAATCTGATGAGGCACTCAAAATGACCATGGCCTCTGTACCTGCATCGCGC
200 210 220 230 240 250 260 270 280 290
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
TACCTAACTGACATGACTCTTGAGGAAATGTCAAGGCACTGGTTCATGCCTCATGCCCAAGCAGAAAGTGGCAGGCCCTCTTTGTATCAGAATGGACCAGG
300 310 320 330 340 350 360 370 380 390
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
CGATCATGGATAAGAACATCATACTGAAAGCGAACTTCAGTGTGATTTTTGACCGGCTGGAGACTCTAATATTACTAAGGGCCTTCACCGAAGAGGGGAC
400 410 420 430 440 450 460 470 480 490
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
AATTGTTGGCGAAATTTCACCACTGCCCTCTCTTCCAGGACATACTGATGAGGATGTCAAAAATGCAGTTGGGGTCCTCATCGGAGGACTTGAATGGAAT
500 510 520 530 540 550 560 570 580 590
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
AATAACACAGTTCGAGTCTCTGAAACTCTACAGAGATTCGCTTGGAGAAGCAGTAATGAGAATGGGAGACCTCCACTCACTCCAAAACAGAAACGGAAAA
600 610 620 630 640 650 660 670 680 690
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
TGGCGGGAACAATTAGGTCAGAAGTTTGAAGAAATAAGATGGTTGATTGAAGAAGTGAGACACAGACTGAAGATAACAGAGAATAGTTTTTGAGCAAATAA
700 710 720 730 740 750 760 770 780 790
NT/24269/85
ALASKA/77
UDORN/72
BERKELEY/68
AA/60
DENVER/57
NT/23284/85
NT/18515/84
MARYLAND/80
USSR/77
FW/50
FM/47
BELLAMY/42
PR/34
WSN/33
CATTTATGCAAGCCTTACAACTATTGCTTGAAGTGGAGCAAGAGATAAGAACTTTCTCGTTTCAGCTTATTTAATAATAAAAAACACCCTTGTTTCTACT
800 810 820 830 840 850 860 870 880 890

Fig. 3. MOST PARSIMONIOUS EVOLUTIONARY TREE FOR 15 INFLUENZA A VIRUS NS GENES.

The nucleotide sequences shown in Fig. 2 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 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).

Figure 4. 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. 3 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.

Figure 3

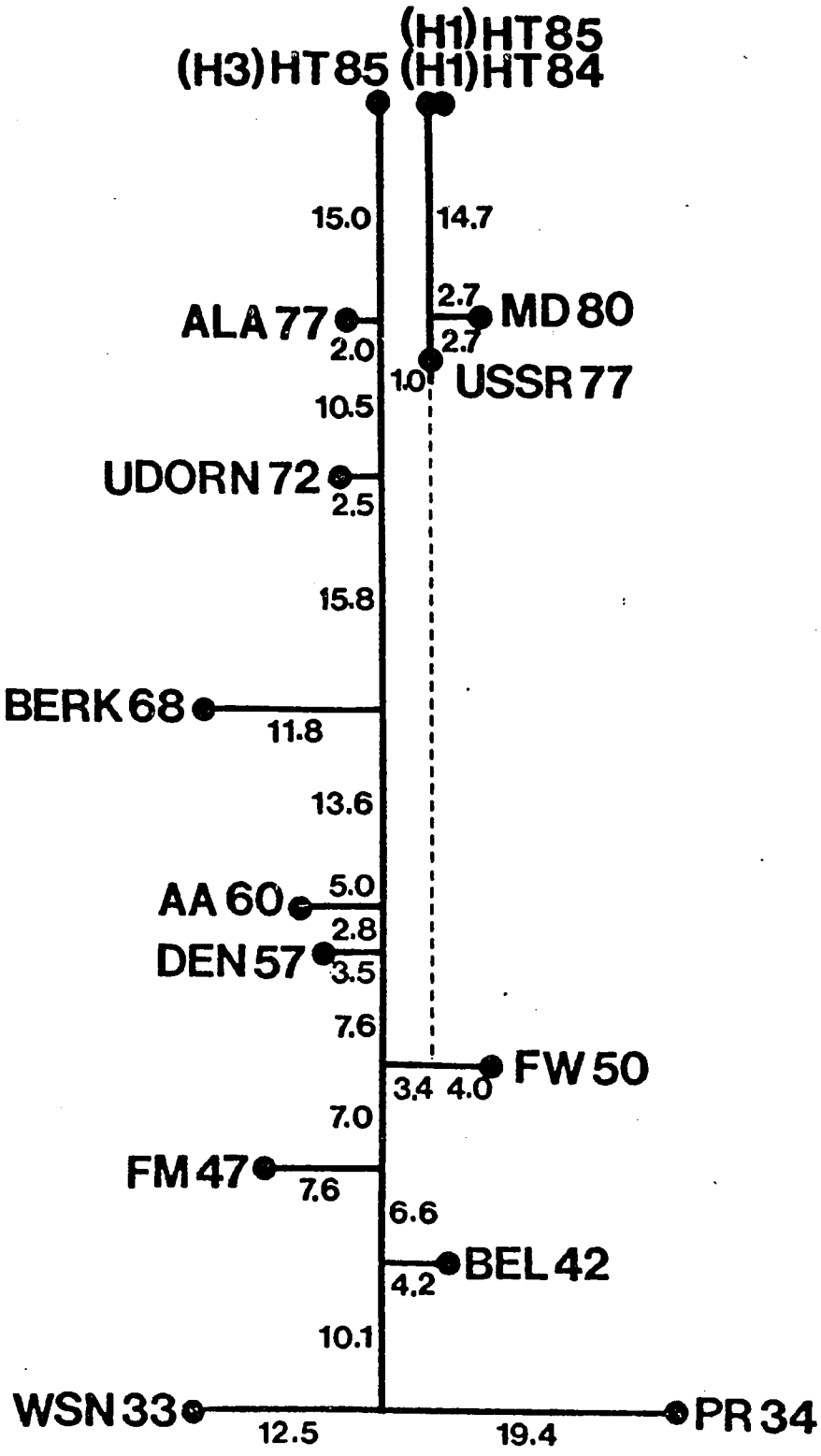
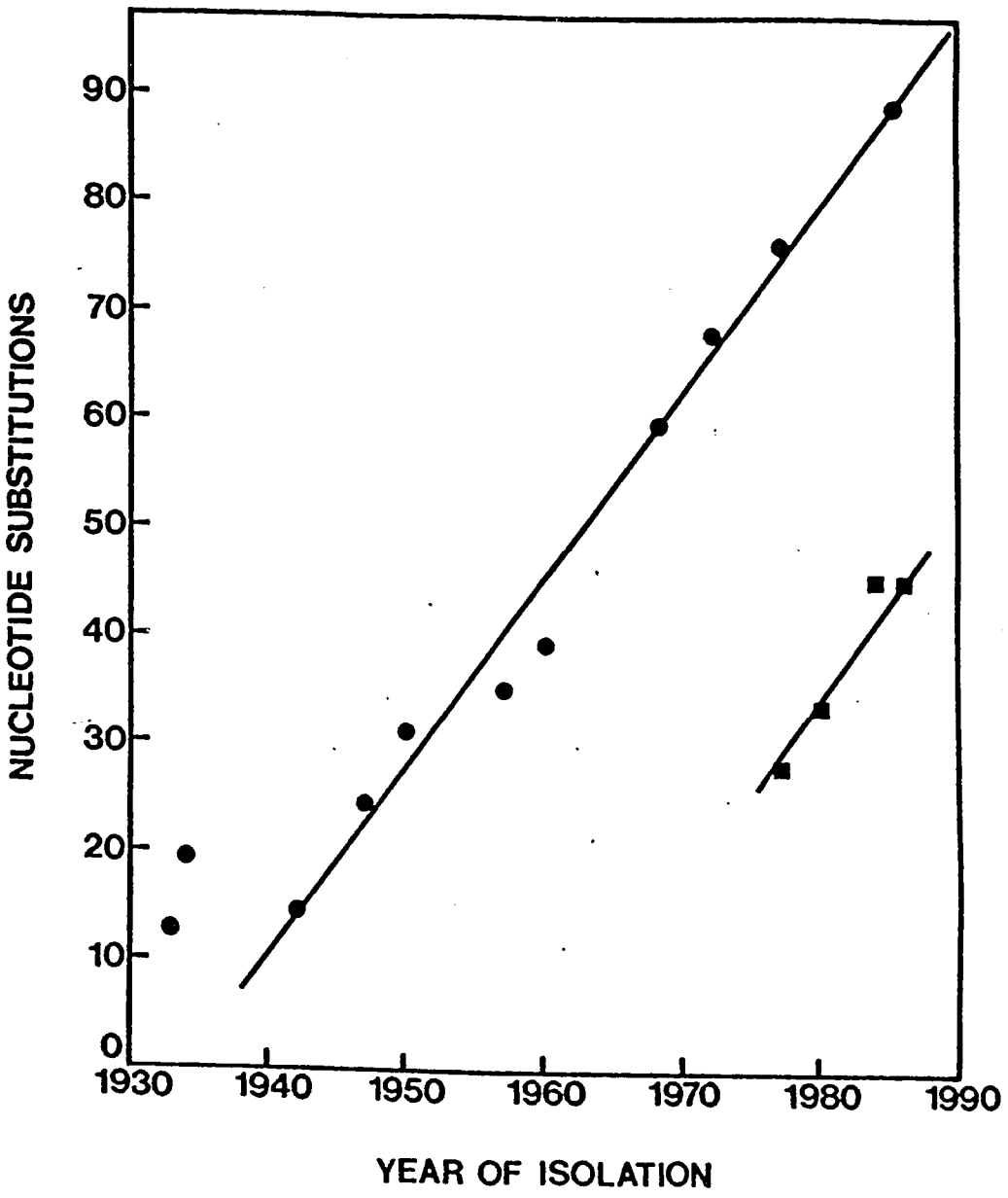

(H1)HT85
(H3)HT85
(H1)HT84
15.0
14.7
2.7
ALA77
MD80
2.0
2.7
1.0
USSR77
10.5
UDORN72
2.5
15.8
BERK68
11.8
13.6
5.0
AA60
2.8
DEN57
3.5
7.6
FW50
3.4
4.0
7.0
FM47
7.6
6.6
BEL42
4.2
10.1
WSN33
PR34
12.5
19.4

Figure 4


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

line, derived by linear regression analysis, shows that these sequences are evolving at the steady rate of 1.73 + 0.08 nucleotide substitutions per year or 1.94 + 0.09 x 10^{-3} substitutions per nucleotide site per year. The WSN/33 and 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.4 also shows that the group of H1N1 subtype strains, which reemerged in the human population in 1977 after a 27-year absence (Kendal, et. al., 1978; Nakajima, et. al., 1978), is evolving at the same rate. These "new" H1N1 viruses have been cocirculating with the H3N2 viruses since 1977 and form a separate evolutionary branch (Fig. 3). In reality, the H1N1 branch should be directly connected to the FW/50 branch of the main tree, since there are only five nucleotide differences between the FW/50 and 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 USSR/77 (represented by the broken line in Fig. 3). The observed data, thus suggest a unique epidemiology of the new H1N1 isolates.

D. DISCUSSION

1. Several points can be made following the analysis of the data. First, calibration of the molecular clock is not affected by inaccurate paleontological dates since the time of fossilization (isolation) of these strains is recorded. This may partly explain why the NS gene of influenza A viruses behaves as an

accurate molecular clock (for review, Wilson, et. al., 1977). Thus, given only the NS gene sequence of a main line isolate, one can closely estimate the year of its isolation (Fig. 4). Although fewer points are available for measuring the rate in the new H1N1 strains (1977-1985), the data (Fig. 4, solid squares) are compatible with a molecular clock ticking at the same evolutionary rate for these NS genes. The mutations seen in the NS genes of the new H1N1 strains (1977-1985) are different from those seen in the 1950-1957 H1N1 strains. Second, the evolutionary rate of $1.94 + 0.09 \times 10^{-3}$ 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.) Third, we have concentrated on an analysis of the NS genes because there was evidence that the NS gene was not exchanged (Scholtissek, et. al., 1978) during the reassortment of influenza viruses leading to new A virus subtypes (Webster, et. al., 1982). Furthermore, preliminary data from several different laboratories had suggested that substitutions in influenza virus NS genes (as well as other genes) are generally retained when strains obtained at later times are compared with earlier isolates (Air, 1981; Both, et. al., 1983; Hayashida, et. al., 1983; Nei, 1983; Raymond, et. al., 1983; Skehel, et. al., 1983; Martinez, et. al., 1983; Ortin, et. al., 1983; Krystal, et. al., 1983; Buonaguric, et. al., 1984). We have compared NS gene sequences of nine new strains with six previously published sequences (Krystal, et. al., 1983; Buonagurio, et. al.,1984). Thus, we can rely upon a much larger data base and on isolates obtained over a longer period of time than was previously possible.

2. Another point concerns the long, slender nature of the evolutionary tree of the NS genes. 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 (the average age of the side branches is only three years). 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 (Buonagurio, et. al., 1985). 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 vaccine strain goes back to Pasteur's time, whereas the influenza A viruses used for vaccine manufacture are changed every 2-3 years (even in the absence of a subtype shift) to accomodate changes in the prevalent strains.

3. What model, then, could explain the rapid evolutionary change in influenza virus genes? One speculation suggests that the unusually high variation of influenza A viruses in nature is the result of a high mutation rate of the virus (Epidemiology of Influenza--Summary of Influenza Workshop IV., J. Infect. Dis. 128, 361, 1973). 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; Gojobori

and Yokoyama, 1985) and herpesviruses (Holland, et. al., 1983), have also been shown to possess the potential to undergo rapid genetic change. (Precise comparative data on the mutation rates of influenza A and other human viruses are not available at this time.) Moreover, a high rate of change does not explain the slender nature of the geneology.

An explanation for the unusual pattern 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 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" (Kimura, 1983) for the A virus NS gene from these data. We suggest that positive selection represents a significant factor in the unusual pattern and high rate of influenza A virus evolution in man.

III.

MEASUREMENT OF THE MUTATION RATE OF ANIMAL VIRUSES: INFLUENZA A VIRUS AND POLIOVIRUS TYPE 1

Jeffrey D. Parvin, Anne Moscona, Wayne T. Pan,

Jason M. Leider and Peter Palese

Department of Microbiology
Mount Sinai School of Medicine of CUNY
One Gustave L. Levy Place
New York, N.Y. 10029

Originally published in J. Virol. 59, 377-383 (1986). Reprinted with permission from the American Society for Microbiology.

A. INTRODUCTION

The variability of animal viruses is a well recognized phenomenon (reviewed in Holland, et. al., 1982). The best studied example is the influenza A virus in which the antigenic nature is continually changing by reassortment (antigenic shift) and by mutation (antigenic drift). The property of antigenic drift allows virus from a single subtype to persist in the human population in spite of immunity to strains from previous years (Webster, et. al., 1982). It has been observed from sequence analyses that the genes coding for surface as well as non-surface proteins of the influenza A virus evolve at a rate which is approximately a million-fold greater than that of eukaryotic genes in nature (Buonagurio, et. al., 1986b; Palese, 1985; Saitou and Nei, 1986). Although a high mutation rate of influenza A viruses is most often implicated as the cause of the extensive variation (Hayashida, et. al., 1985; Saitou and Nei, 1986), a direct measurement of this parameter has never been obtained in vivo.

Mutation rates have been estimated for several bacteriophages. Based on the growth kinetics of spontaneous revertants from a deleterious point mutation a mutation rate was indirectly calculated for the bacteriophage QB (Batschelet, et. al., 1976) and mutation frequencies determined by growth under selective conditions have been estimated for bacteriophage lambda and bacteriophage T4 (Drake, 1969). Attempts have also been made to assess the mutability of animal viruses. For example, the substitution frequency of a specific 5' terminal nucleotide in the vesicular stomatitis virus (VSV) genome was determined (Steinhauer and Holland, 1986). Similarly, the frequency of variants which have lost the recognition site for a neutralizing monoclonal antibody has been used as a measure of mutation frequency. This method has been used to analyze mutability of RNA and DNA containing viruses including influenza viruses,

poliovirus, VSV, and herpes viruses (Emini, et. al., 1983; Holland, et. al., 1983; Lubeck, et. al., 1980; Portner, et. al., 1980). Alternatively, reversion frequencies of viral mutants have been used to assess mutation frequencies. For example, the accumulation of thymidine kinase deficient mutants of herpes simplex virus was employed to analyze the role of mutator and anti-mutator phenotypes of this virus (Hall, et. al., 1984).

In this study we attempt to measure directly the mutation rate of two animal viruses by sequencing genes in randomly selected viral clones. These viral clones are all descended from a single virion after only one plaque generation, confining the time over which mutations can occur to about five infectious cycles. The mutation rate we measure is a neutral mutation rate, since all lethal and deleterious mutations are not observed. We find that the mutation rate of influenza A virus is higher than that of the poliovirus, and we speculate this difference correlates with the speed of evolution and the lack of success of vaccination against influenza viruses.

Accurate in vivo measurements of mutation rates should allow a greater understanding of the evolution of viruses in nature and may also provide a new way to test different evolutionary theories with respect to animal viruses. Knowledge of the inherent mutability of a virus may also clarify the mechanism by which variants emerge that are resistant to the immunity conferred by vaccination. The probability of successful vaccination against a particular virus may be determined by the mutation rate of that virus, and thus knowledge of this parameter may influence the strategy of vaccine design. Finally, an understanding of the mutability of a virus may help in studying the development of drug resistance as well as the changes in virulence observed with different RNA and DNA containing viruses.

B. MATERIALS AND METHODS

1. VIRUSES AND CELLS

The influenza virus A/WSN/33 was derived from an uncloned viral stock preparation grown in Madin-Darby Canine Kidney (MDCK) cells in the presence of Eagle's minimal essential medium containing 1 ug/ml trypsin (Brand and Palese, 1980). Virus was then plaque passaged twice in MDCK cells, and virus from the second plaque passage was directly employed for the experiment described. The agar overlay for plaquing contained 0.6% agar (Oxoid Ltd.), minimal essential medium, 0.2% bovine albumin, 0.01% DEAE dextran, and 1 ug/ml trypsin.

The Mahoney strain of poliovirus type 1 and the HeLa cell line used for its passage were kindly provided by V. Racaniello. Again, the virus was plaque passaged twice before being used in the experiment. Viral passage was done as described in (Bernstein, et. al., 1985).

2. PREPARATION OF INFLUENZA A VIRUS CLONES AND PURIFICATION OF VIRAL RNA

Confluent MDCK cells were infected with serial ten-fold dilutions of the plaque-purified virus preparation. One hour post infection (pi), the innoculum was removed by aspiration and the standard agar overlay was added. Forty-eight hours pi a well isolated plaque was identified and the overlay above the plaque was gently aspirated with a Pasteur pipette. Virus was eluted from the agar plug into 0.5 ml phosphate buffered saline containing 0.2% bovine albumin (PBS+BA). The cell monolayer from which the plaque was picked was stained with 0.1% crystal violet in 20% methanol to demonstrate that the plaque was indeed discrete.

In order to clone the individual virions in the plaque, a second plaque passage was done using the virus yield from the first plaque. This time many dishes were prepared to allow the isolation of several hundred discrete plaques. The total PFU in each plaque was about 1 X 10^6. One fifth of the yield from each plaque was used to infect 2 X 10^6 MDCK cells with a liquid overlay which contained the same components as the agar overlay except the agar and the dextran. The supernatant from the completely lysed monolayer was harvested 24 to 30 hours pi. The supernatant was then diluted in PBS+BA to infect at an MOI of about 0.2 seven or eight dishes of MDCK cells, each containing 2 X 10^7 cells. The medium (65 ml) was harvested approximately 30 hours pi after complete lysis of the cell monolayer.

The medium containing virus harvested from lysed cells was pre-cleared by centrifugation at 8000 g for 30 min. The supernatant was then layered over a 3 ml 20% sucrose cushion, and virus was pelleted by centrifuging at 25,000 RPM for 2 h in a SW27 rotor. The virus pellet was resuspended in 4 ml of 100 mM NaCl, 10 mM tris-HCl pH 7.4, and 1 mM Na_2EDTA (NTE buffer). The RNA from the resuspended virus was extracted (approximately 100 ug) (Palese and Schulman, 1976).

3. PREPARATION OF POLIOVIRUS TYPE 1 CLONES AND OF VIRAL RNA

Confluent HeLa cell monolayers were infected with serial dilutions of the plaque-purified poliovirus preparation. One hour pi the virus innoculum was removed by aspiration and the standard agar overlay containing 1% agar, Dulbecco's modified Eagle medium, 5% horse serum and 0.01% DEAE dextran was added. Thirty-six hours pi a well isolated plaque was identified and the agar above it was aspirated as before. The monolayer was stained with the crystal violet solution to confirm that the plaque was indeed discrete. Virus was eluted

from the agar plug into PBS containing 0.2% horse serum. The total yield from the plaque was 5.2 X 10^6 PFU.

In order to clone individual virions from the virus yield of the plaque, a second round of plaque passage was done as before except that the time allowed for plaque formation was 48 hours. One fifth of the virus yield from each plaque was used to infect 2 X 10^6 HeLa cells with a liquid overlay which contained the same components as the agar overlay except the agar and the dextran. Twenty-four hours pi the medium above the lysed cells was harvested. This supernatant was then diluted for the infection (at an MOI of 0.5) of five large dishes containing 2 X 10^7 HeLa cells each. Twenty-four hours pi the yield from the lysed cells was harvested.

The medium containing virus (40 ml) was pre-cleared by centrifugation at 9000 g for 30 min. The supernatant was centrifuged at 130,000 g for 90 min. The pelleted virus was resuspended in 4 ml NTE buffer, and the RNA (approximately 100 ug) was extracted as was done for the influenza virus.

4. SEQUENCING OF VIRAL RNA

The viral RNA was sequenced directly using the strategy of Sanger, et. al. (1977). The hybridization mixture contained 10 ug of viral RNA and 200 ng of a specific oligonucleotide as primer. Three primers were used to sequence each gene. The primers used for the NS gene were complementary to the viral RNA at positions 10 to 29, 293 to 312 and 593 to 612 (Buonagurio, et. al., 1986b). The VP1 gene primers were complementary to positions 3422 to 3403, 3136 to 3117, and 2812 to 2793 on the viral RNA (Kitamura, et. al., 1981). The hybridization mix was heated in boiling water for five min., cooled to room temperature, and aliquoted to four reaction tubes. The A reaction contained 250 uM dGTP and dTTP, 100 uM dATP, 12.5 uM dideoxy ATP and 10 uM dCTP.

The G reaction contained 250 uM dATP and dTTP, 100 uM dGTP, 12.5 uM dideoxy GTP, and 10 uM dCTP. Both the A and G reactions had 15 uCi of alpha-^{32}P-dCTP (3000 Ci/mmole) or alpha-^{35}S-dCTP (600 Ci/mmole) added to each reaction. The C reaction contained 250 uM dGTP and dTTP, 100 uM dCTP, 12.5 uM dideoxy CTP, and 10 uM dATP. The T reaction contained 250 uM dCTP and dGTP, 100 uM dTTP, 12.5 uM dideoxy TTP, and 10 uM dATP. Both the C and T reactions contained 15 uCi of alpha-^{32}P-dATP (3000 Ci/mmole) or alpha-^{35}S-dATP (600 Ci/mmole) added to each reaction. The reactions were started by the addition of 5 U of AMV reverse transcriptase (Molecular Genetics Resources, Inc.) and incubated at 42^{o}C for 30 min. The total reaction volume was 10 microliters. A cold-chase solution containing 2mM concentrations of each nucleotide was added (1.5 ul). Following a second 30 min incubation at 42^{o}C, the reactions were terminated by the addition of 11 ul of a formamide dye mix. Prior to gel electrophoresis, the reactions were heated in boiling water for 5 min and then rapidly cooled in an ice/water bath.

As a standard procedure two different gel systems were used to analyze the sequencing results. The first gel was a 6% polyacrylamide buffer-gradient gel (Biggin, et. al., 1983) or a 6% polyacrylamide wedge-shaped gel (CBS Scientific Company, Inc.) which generally permitted the resolution of sequence ten nucleotides from the primer to at least 180 nucleotides from the primer. A second gel, which was the standard 6% polyacrylamide gel, was used for a long electrophoresis of the reaction products. This second gel generally permitted the resolution of nucleotides from 160 to 320 or more bases away from the primer. Gels were fixed in a 10% acetic acid, 10% methanol solution, dried, and exposed to Cronex 4 film for three to ten days.

Using this method, bands occasionally appeared in more than one lane at a

given position. These secondary bands proved not to be troublesome since their pattern was uniform for a position among the different RNAs.

For the influenza virus NS gene, 91,708 nucleotides were sequenced. Of these, 50 nucleotides could not be positively identified because the bands in the expected lane were too faint. However, since no other band appeared in any of the other three lanes at the same level, it was assumed that no changes had occurred at these positions. Of the 95,688 sequenced nucleotides for the VP1 gene, 323 could not be positively identified for the same technical reason. Again, none of these positions was scored for mutations because bands in any of the other three lanes were not observed.

5. COMPARISON OF GROWTH KINETICS OF NS GENE MUTANTS

In order to determine the relative growth kinetics of each influenza virus NS gene variant, a multicycle infection was done for each mutant, the parental virus, and two randomly picked controls whose NS genes had the wild-type sequence. Aliquots of each virus were titered in advance, and then each was used to infect a dish of MDCK cells at a multiplicity of 1000 PFU per 3 X 10^6 cells. One hour pi the innocula were removed from the dish, each monolayer was washed with PBS, and the standard liquid overlay was added. At each time point, 10, 19, 24, 28, 32, 37, 42, and 48 hours pi, 0.3 ml of medium was removed from each dish and was used for plaque titrations. Cytopathic effect was first observed at the 37 hour time point.

6. CONVERSION OF THE NS GENE MUTATION RATE FOR COMPARISON

The mutation rate of the NS gene as defined in this paper was calculated as the amount of change which would occur with each cell burst. Drake (1969) and Koch and Drake (1973) calculated mutation rates from mutation frequencies. By this definition of mutation rate, the number of replications

reflected the number of times the genome was copied. Replication number was approximately equal to the population size. The formula was:

mutation rate=.4343 X mutation frequency /log(population size).

Using this formula, the NS gene mutation frequency in a plaque was seven changes per 91,708 nucleotides, and the population size was equal to the total nucleotide pool (1.2×10^6 PFU X 850 nucleotides analyzed per NS gene), which resulted in the converted mutation rate of 3.7×10^{-6} mutations/nucleotide/ replication. Alternatively, a correction factor substituting population size by population size X mutation rate could be made since no mutants would be expected until the population size equaled the reciprocal of the mutation rate (J.W. Drake, personal communication). The corrected mutation rate was then 8.4×10^{-6} mutations/ nucleotide/replication. If the mutation rate ceiling for the poliovirus were similarly converted, it would be 4.7×10^{-7} mutations/nucleotide/replication or, including the correction factor, 1.2×10^{-6} mutations/nucleotide/ replication. Using this procedure the values for the mutation rates of bacteriophage lambda, bacteriophage T4, Salmonella typhimurium, Escherichia coli and Neurospora crassa were 2.4×10^{-8}, 1.7×10^{-8}, 2.0×10^{-10}, 2.0×10^{-10} and 0.7×10^{-11}, respectively (Drake, 1969).

C. RESULTS

1. DETERMINATION OF THE NS GENE MUTATION RATE

The influenza virus clones used to determine the mutation rate in the NS gene were derived from a single plaquing experiment (Fig. 5). Plaque formation occurred under standard tissue culture conditions in the absence of any new

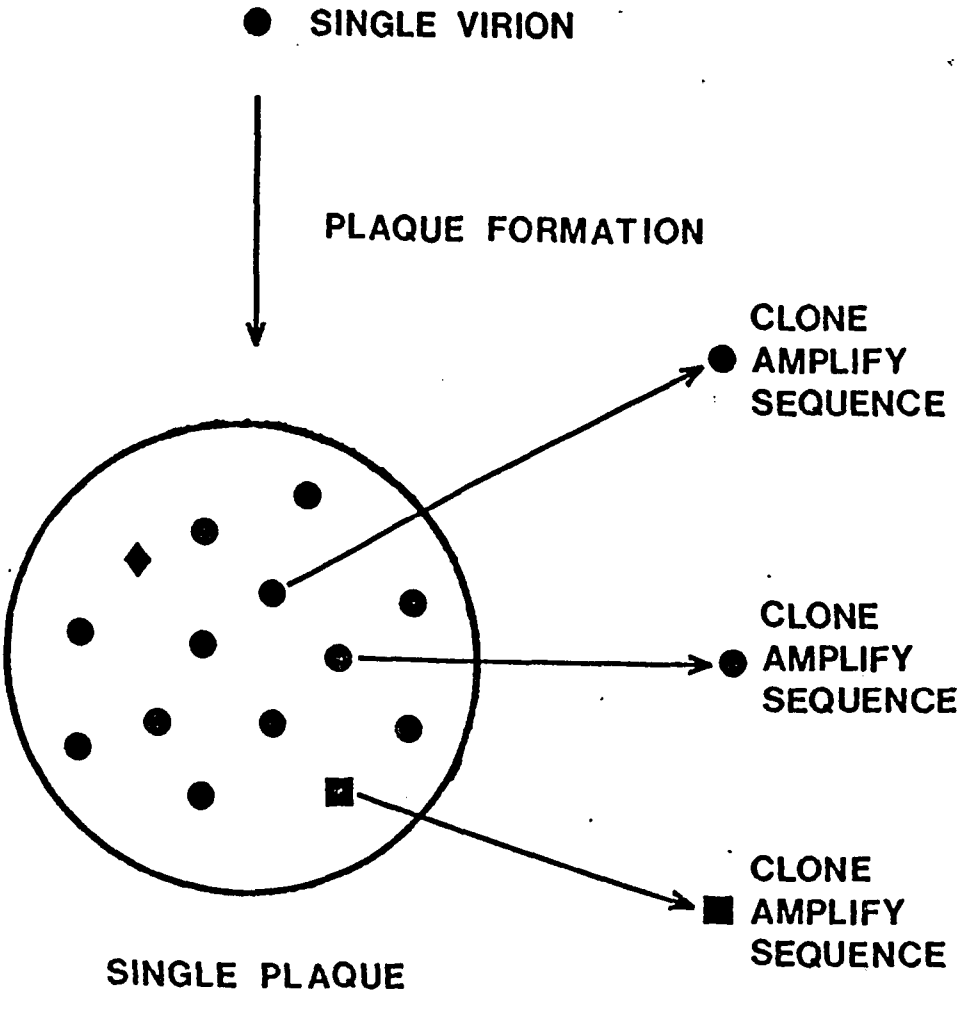
selective pressure. Cell monolayers were infected at high dilution with plaque-purified A/WSN/33 virus, and 48 hours post infection (pi) a single, well isolated plaque was picked. This plaque, descended from a single virion, contained 1.2 X 10^6 infectious virus particles (PFU). In order to determine the number of mutant NS genes among these viral particles, individual clones were obtained and viral RNA was prepared following only two amplification steps. The multiplicity of infection (MOI) during the amplification steps was maintained at 0.1 for the first step and 0.2 for the second to minimize the enrichment of new mutants. Since RNA obtained from amplified viral preparations was sequenced using the primer extension protocol, the sequence represents a "consensus" sequence which should be identical to that of the individual clone. Any mutants which may arise in the amplification protocol would not be detected.

The NS genes of 108 different clones were sequenced producing 91,708 nucleotides of information. Comparison of the NS gene sequences revealed seven mutants. Figure 6 shows the relevant sequencing gels of two of these mutants. In the top half of the figure, the sequences for five different clones are presented for positions 151 to 181. As indicated by the arrow, a G to T transversion in clone 014 was observed at position 169. In the bottom half of the figure, the sequences for five other clones are shown for positions 751 to 788. On this gel the G reactions of the five clones were loaded side-by-side as were the A, T, and C reactions. This latter arrangement facilitated the rapid detection of mutants. The A to G transition at position 772 in clone 024 can easily be observed in this gel (arrows). All seven point mutants are listed in table 1 along with the predicted amino acid coding changes. As can be seen, the mutations were evenly distributed along the gene, and there was only one

Figure 5. EXPERIMENTAL DESIGN FOR MEASURING MUTATION RATES IN VIRAL GENES.

A single virion (the parental virus) formed a plaque after sufficient time for about five infectious cycles. The many progeny virions contained within the plaque carried genes with the parental sequence (represented by filled circles), and a fraction of the virions present carried genes with a point mutation (represented by square and diamond). The individual virions were cloned by plaquing and amplified under conditions that minimize the effect of new mutations on the consensus sequence of the gene in the clone. RNA obtained from purified virus was directly sequenced. The mutation rate was determined by dividing the number of observed point mutations by the number of nucleotides analyzed and by the number of infectious cycles.

Figure 5


SINGLE VIRION
PLAQUE FORMATION
CLONE
AMPLIFY
SEQUENCE
CLONE
AMPLIFY
SEQUENCE
CLONE
AMPLIFY
SEQUENCE
SINGLE PLAQUE

mutation per variant gene. Two mutants, 050 and 069, which were derived from clones on separate dishes were found to have the same point change. This finding was not unexpected since point mutations which occur early in the development of the plaque would represent a higher fraction of the variants than ones which occur late. Six of the seven variants were found to encode amino acid changes in the NS gene protein products. This observed frequency of coding changes is in agreement with the predicted frequency of coding changes in a randomly mutated coding for overlapping NS1 and NS2 proteins.

It is assumed that back-mutations are negligible and that the mutants have similar growth fitness to wild-type (see below). Based on these assumptions, dividing the number of variants by the number of infectious cycles then accounts for the effect of mutations occurring over several replications and of mutations being sampled twice.

The mutation rate was calculated using the formula:

mutation rate=7 mutations/91,708 nucleotides/5 infectious cycles

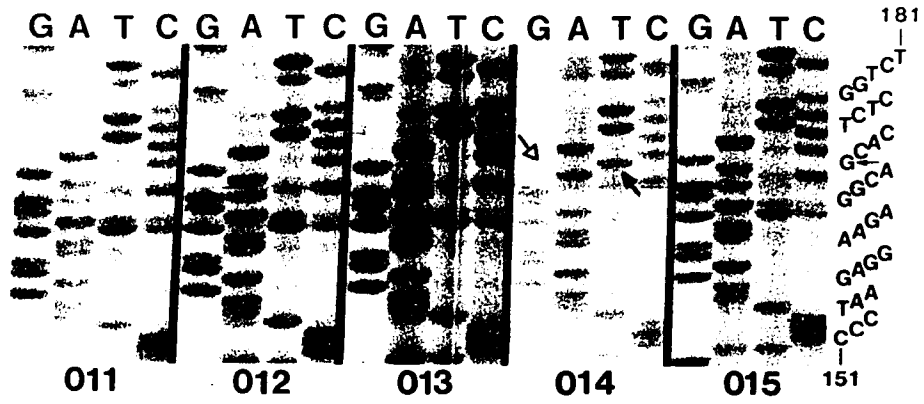
$=1.5 \times 10^{-5}$ mutations/nucleotide/infectious cycle.

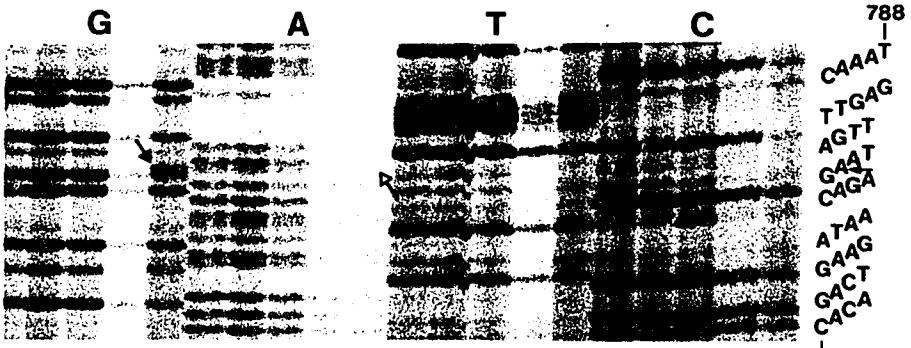
Although plaque growth dynamics are complex, and although synchrony in burst cycles may be lost, the number of growth cycles was most probably five. If all of the sampled virus clones were derived from only four infectious cycles, then the mutation rate would be underestimated by 26%. Similarly, if all sampled clones resulted from six bursts, the mutation rate would be overestimated by 17% in our calculations. Thus, the error contributed by clones that had different numbers of growth cycles was small.

Figure 6. DETECTION OF POINT MUTATIONS IN NS GENES OF INFLUENZA VIRUS CLONES 014 AND 024.

In the top half of the figure, the sequencing reactions of the NS genes of clones 011, 012, 013, 014 and 015, were elctrophoresed in a 6% polyacrylamide buffer-gradient gel. Positions 151 through 181 are shown and the G to T change at position 169 in clone 014 is indicated by the arrows. In the bottom half of the figure, the sequencing reactions for the NS genes of clones 039, 040, 041, 042 and 024 were electrophoresed in a buffer-gradient gel. The G reactions for five clones, 039, 040, 041, 042 and 024 (left to right) were loaded side-by-side as were the A, T and C reactions. Positions 751 to 788 are shown, and the A to G transition at position 772 in clone 024 is indicated by arrows.

Figure 6





CLONES 039, 040, 041, 042, AND 024

Table 2. POINT MUTATIONS DETECTED AMONG NS GENES OF 108 INFLUENZA A/WSN/33 VIRUS CLONES

Mutant	Mutation*	Amino acid change	Position
014	G --> U	Ser --> Ile in NS1	169
024	A --> G	Asn --> Asp in NS2	772
031	U --> C	Trp --> Arg in NS1	72
049	U --> C	Silent	68
050	G --> A	Arg --> Lys in NS1	379
069	G --> A	Arg --> Lys in NS1	379
082	C --> U	Leu --> Phe in NS2	856

*Mutations detected by RNA sequencing are presented in the message orientation. The parent NS gene sequence is according to Buonagurio, et. al., 1986b.

2. CHARACTERIZATION OF THE NS GENE VARIANTS

It was assumed in the calculation of the mutation rate that the mutants had no significant selective advantage or disadvantage during plaque formation. If these mutations were phenotypically neutral, then the percentage of a given mutant would remain constant as the total virus population expanded in the plaque.

In order to test whether the NS gene mutations were neutral, the rate of virus production was assayed in a multicycle infection. The time course of virus production in multicycle infections is very sensitive to small changes in virus fitness. The results are diagrammed in figure 7. Included in the study

were the parent clone and two randomly selected virus clones whose NS genes did not contain mutations. Samples from the liquid overlay taken at the indicated times were titered for the concentration of PFU. In the diagram, lines connect the time points in the controls. As can be seen in the figure, there was no significant difference between the wild-type and variant clones. One virus appeared to lag behind the others at some intermediate time points, but it was the variant carrying the silent mutation, and it would be very unlikely that this NS gene mutation would be responsible for the small lag. All seven variants were therefore indistinguishable from the wild-type and thus the mutations appeared to be neutral.

3. DETERMINATION OF THE VP1 GENE MUTATION RATE

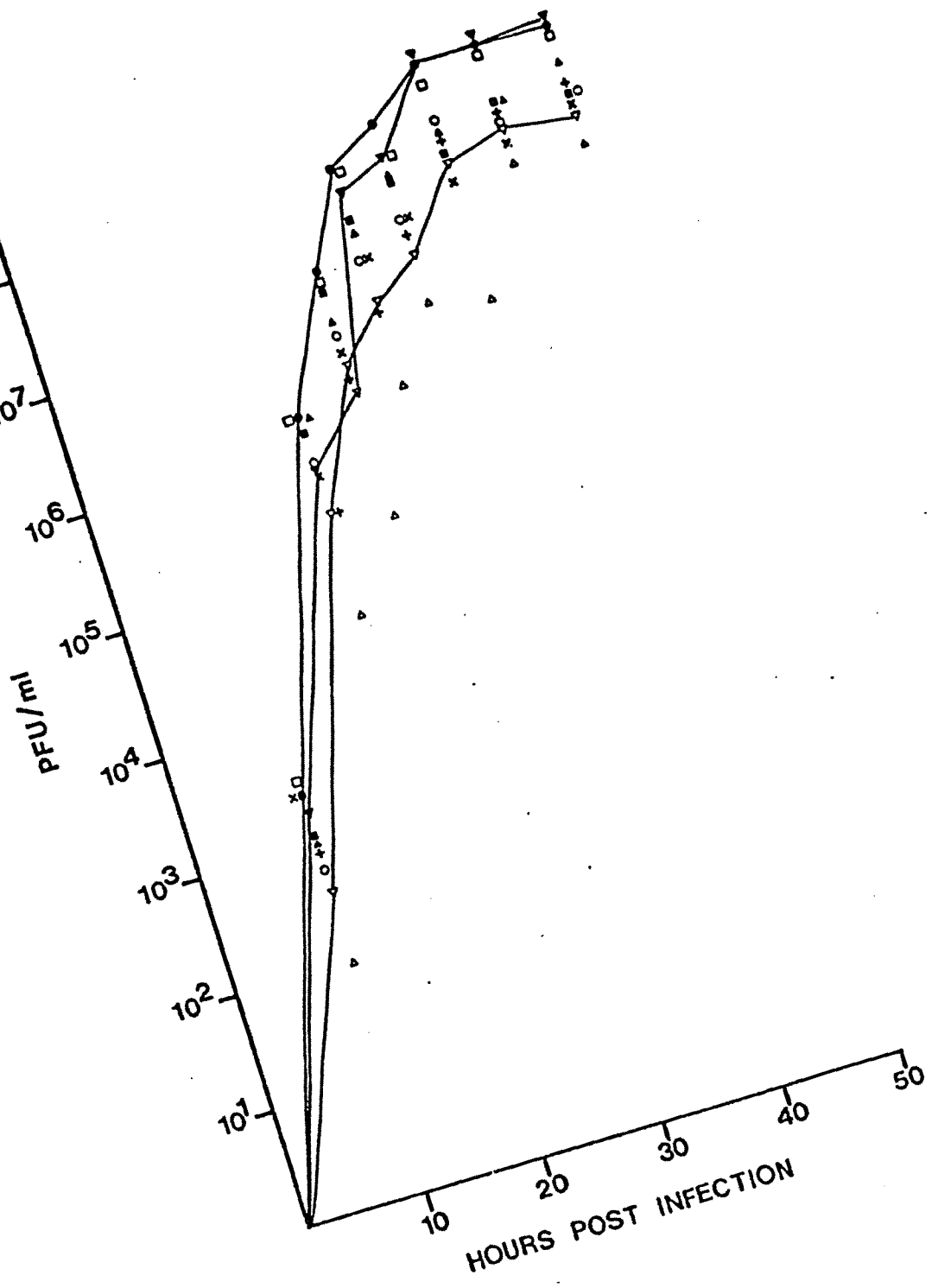
The plaque passage of the Mahoney strain of poliovirus type 1 was done according to the same scheme as followed for the influenza virus experiment. Virus which had been twice plaque purified was infected onto cell monolayers, and 36 hours pi a well isolated plaque containing 5.2 X 10^6 infectious particles was picked. Since the generation time of poliovirus is about 7 hours, this time of plaque formation allowed for about five infectious cycles. Again, individual clones from among the 5.2 X 10^6 PFU were obtained from this plaque and viral RNA was prepared following a two step amplification process.

The sequence of the Mahoney strain VP1 gene was the same as presented in Racaniello and Baltimore (1981) except for three point changes at positions: 2837 (A to G), 3139 (C to T), and 3151 (A to T). (Numbering is according to Kitamura, et al., 1981). These three positions also differ from the sequence of Kitamura, et. al., and probably represented the changes which randomly occur after repeated passage and were only seen after cloning. Interestingly, two of the three changes were associated with amino acid changes in the VP1 protein

Figure 7. COMPARISON OF GROWTH KINETICS OF INFLUENZA VIRUS NS GENE VARIANTS IN MDCK CELLS.

The time course of production of virus following infection at an MOI of 3 X 10^{-4} was determined for the parental virus and two clones which had the parental NS gene sequence and for the seven variants. The PFU per ml of the parental virus (•) and clones: 007 (▼), 014 (■), 024 (□), 031 (o), 049 (△), 050 (▲), 069 (×), 082 (+) and 086 (▽). The data points for the viruses containing the parental NS gene sequence (the parental virus and clones 007 and 086) were connected by lines for clarity. The infectious virus per ml of supernatant medium was determined for different time points by plaque titration on MDCK cells.

Figure 7


PFU/ml
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10
20
30
40
50
HOURS POST INFECTION

in comparison to the sequence reported by Racaniello and Baltimore (1981). When the sequences of the 105 VP1 genes were compared, zero point mutations were observed. Since no mutations occurred in the VP1 genes of the clones analyzed, the upper limit of the mutation rate can be calculated as follows:

mutation rate <1 mutation/95,688 nucleotides/5 infectious cycles

$<2.1 \times 10^{-6}$ mutations/nucleotide/infectious cycle.

Analysis of the statistical significance of the difference in mutation rate between the two experiments was performed by a chi-square test using a two-by-two contingency table of the data. It was found that the mutation rate of the VP1 gene was significantly lower than that of the NS gene ($p < 0.025$). In addition, the difference was statistically significant using a t-test applied to the binomial standard error of the influenza virus mutation rate ($p < 0.01$). However, since each mutation rate was only measured once, we cannot predict the variation in mutation rates had more than one mutation experiment been performed for each virus.

D. DISCUSSION

1. COMPARISON OF INFLUENZA A VIRUS AND POLIOVIRUS TYPE 1 MUTATION RATES

The present study was undertaken to determine the precise mutation rate of two different RNA viruses. The specific strains were chosen because of technical ease of plaque passaging at $37^{o}C$, and because of excellent growth characteristics which facilitated sequencing

template preparations. The genes were selected based on prior demonstrations that they were capable of variation in nature or in the laboratory. The NS gene of influenza A virus has been shown to evolve very rapidly in nature (Buonagurio, et. al., 1986b, Krystal, et. al., 1983) and the VP1 gene of poliovirus type 1 was shown to tolerate mutations (Emini, et. al., 1983; Nomoto, et. al., 1982). The influenza A virus mutation rate was found to be significantly higher than that of poliovirus type 1. This result was surprising in the light of earlier data involving the selection of antigenic variants with neutralizing monoclonal antibody preparations. The frequency of variants, as measured by the plaque-reduction assay, was as high for poliovirus type 1 as for influenza A virus in many cases (Emini, et. al., 1983; Lubeck, et. al., 1980). However, this assay may not reflect the true mutation frequency of the viral genomes, since the result may be affected by various other parameters. For example, the avidity and/or discriminating capabilities of the monoclonal antibody preparations may differ. Also, the genomic target size for the antibody combining sites is unknown. Further, it has been shown with poliovirus that even mutations outside of the antibody binding site can inhibit neutralization by the antibody (Blondel, et. al., 1986). It is thus difficult to predict a precise mutation rate from data on the frequency of antigenic variants.

Is the lower mutation rate in the poliovirus VP1 gene the result of high replicase fidelity or due to very high constraints against lethal or deleterious mutations? The data on the selection of antigenic variants (Emini, et. al., 1983) demonstrates that there must be regions along the gene which have relaxed constraint, and comparison of the VP1 sequences of the Mahoney and the Sabin 1 strains reveals that many coding and many silent changes are allowed in the VP1 gene. The VP1 gene, which represents 12% of the genome,

contains nine of the 57 total point changes and seven of the 19 coding differences between the Mahoney and Sabin strains (Nomoto, et. al., 1982). It is thus suggested that the absence of poliovirus VP1 mutants among the 105 clones sequenced may not be due to constraints alone, but may also be the result of increased fidelity of the viral replicase.

Based on the calculated mutation rates, the following prediction can be made regarding the frequencies of mutants present in influenza and polio virus populations. If all of the influenza A virus genes are under the same constraints as the NS gene and thus have the same mutation rate, then every virus in a plaque would average one point change per genome (mutation rate X 13.6 Kb genome X five cycles). An influenza viral population therefore represents a quasispecies (Eigen, 1971; Eigen and Schuster, 1977) in which there is a consensus sequence, but each individual is unique. A similar calculation suggests that a poliovirus plaque would contain less than 8% mutant infectious particles. Therefore, the diversity in the poliovirus population is not as great as that predicted in the influenza virus population.

2. COMPARISONS OF MUTATION RATE ESTIMATES IN OTHER SYSTEMS

Mutation rates and mutation frequencies have been estimated in a variety of other systems. Recently, a study of the VSV substitution frequency revealed a surprisingly high <u>in vitro</u> rate of misincorporation of an extracistronic nucleotide (Steinhauer and Holland, 1986). The frequency of misincorporation was about 10^{-4} substitutions per base incorporated at the site. The same site was analyzed <u>in vivo,</u> and the substitution frequency was again found to be about 10^{-4}. Caution must be exercised in comparing this substitution frequency to the mutation rate determined for influenza and polio viruses since this specific site may not be characteristic for the genome. In

this respect, it should be noted that mutation frequencies of specific single bases are often idiosyncratic. In the phage T4rII locus, 10^4-fold differences in mutation frequency have been observed for single nucleotides at different sites in the gene (Salts and Rosen, 1971). The indication that VSV might have a high mutation rate would suggest that a direct comparison to the mutation rates measured in this study using the same methodology would be very interesting. The direct analysis of clones derived from a single plaque passage, as is done in the present study, examines the mutation rate of an entire gene and thus appears to give more reliable data.

Mutation frequencies have also been estimated for virus populations after serial uncloned tissue culture passage using T1 oligonucleotide mapping analysis. After 28 passages FMDV was shown to accumulate virus that consists of particles with two to eight mutations per genome (Sobrino, et. al., 1983), and bacteriophage QB was similarly shown to have one to two changes per genome after 30 or more passages (Domingo, et. al., 1978). These estimates, based on multiple passages with an unknown number of replication cycles, do not allow us to calculate mutation rates (mutation frequency per infectious cycle) for these viruses and thus these values cannot be directly compared with the data obtained for influenza and polio viruses during a single plaque generation.

Other estimates of mutation rates for bacteriophages, bacteria, and a fungus were derived from experiments in which the yield of test organisms grown under selective conditions was compared to the yield grown under permissive conditions. Although the molecular basis of the mutation allowing growth under selective conditions was unknown, it was assumed to be a point mutation, and the gene was assumed to have a standard length if the length

was unknown (Drake, 1969). To compare our data to these results, the NS gene mutation rate must be converted to fit the definitions of Drake (1969). The mutation rate of the NS gene after this conversion is approximately 10^2 times higher than those of bacteriophages lambda and T4, approximately 10^4 times higher than those of the bacteria Salmonella typhimurium and Escherichia coli, and approximately 10^5 times higher than that of the fungus Neurospora crassa. (For detailed calculations see Materials and Methods.)

3. THE ROLE OF MUTATION RATE IN VIRAL EVOLUTION

There is epidemiologic and genetic evidence that influenza A viruses evolve more rapidly than other viruses in man (Webster, et. al., 1982; Bull. WHO, 1985, 63, 479-484). Specifically, vaccine strains used against influenza A virus have to be changed frequently (at least every two to three years) in order to protect against an evolving virus population. In contrast, vaccines against poliovirus (and most other human viruses) are based on strains which have been used for the last several decades, ostensibly without loss in efficacy. One might therefore speculate that the higher mutation rate found for influenza A viruses could provide a molecular basis for this difference. High mutation rates would be necessary to generate the great amount of diversity required for the extraordinarily rare event of producing a beneficial mutant which would be selected in the host environment. It should be noted, however, that extensive variation has been observed in poliovirus populations circulating in nature (Nottay, et. al., 1981) and thus the above model for the role of mutation rate must remain but one of the possible hypotheses to explain the phenomenon of rapid evolution of influenza A virus.

Furthermore, one would like to know whether other viruses, for example, the AIDS virus or herpes viruses show comparably high mutation rates which

may then affect the successful use of vaccines. The precise measurements of mutation rates of other viruses may thus help to dissect the factors which determine the complex genetic interactions of viruses with the natural environment.

IV.

DETECTION OF SINGLE BASE SUBSTITUTIONS IN INFLUENZA VIRUS RNA MOLECULES BY DENATURING GRADIENT GEL ELECTROPHORESIS OF RNA-RNA OR DNA-RNA HETERODUPLEXES

Frances I. Smith, Jeffrey D. Parvin and Peter Palese

Department of Microbiology
Mount Sinai School of Medicine of CUNY
Fifth Avenue and 100th Street
New York, N.Y. 10029

Originally published in Virology 150, 55-64 (1986). Reprinted with permission from Academic Press, Inc.

A. INTRODUCTION

Electrophoretic methods have often been used to separate RNA molecules on the basis of sequence differences. For example, the origin of the eight different genomic ssRNA molecules of influenza A virus reassortants can often be assigned to one of the parental viruses by analysis of their electrophoretic migration in polyacrylamide gels (Palese and Ritchey, 1977). However, there are many instances in which ssRNA influenza genes known to contain differences migrate identically. Using reovirus, which has a segmented dsRNA genome, Ito and Joklik (1972) showed that sequence differences not detectable by the migration of homologous dsRNA molecules sometimes could be detected in the migration of heterologous RNA-RNA hybrid molecules (e.g., hybrids made using the plus strand of the mutant and the minus strand of the wildtype virus). The presence of urea was noted to enhance the migrational differences (Schuerch and Joklik, 1973), indicating that these differences are caused by a change in the secondary structure of these hybrids. Ito and Joklik (1972) suggested that under these conditions of electrophoresis (6-7 M urea, low voltage, room temperature), the homologous ds RNA hybrids were present as a linear duplex, whereas the mismatched base in the heterologous hybrids caused a structural change, possibly a looping out or formation of a flower configuration. Whatever the reason, these authors noted that the heteroduplex was not rendered sensitive to ribonuclease A degradation. Hay et. al. (1979) extended Ito and Joklik's system to look at mutations in the influenza A virus hemagglutinin (HA) gene. By hybridizing cRNA (plus sense ssRNA obtained from influenza virus-infected cells) with virion RNA (minus sense ssRNA), they showed that mutations which resulted in a single amino acid change could be

detected by this method. However, the basis of mutant detection was not clear, and no prediction could be made of whether a certain mutation would be detectable by this method.

In recent years, considerable work has been done by Lerman and colleagues (Lerman et al., 1984; Myers et al., 1985) on the conditions required to detect single base substitutions in DNA molecules. By using a gradient of denaturing solvent in a polyacrylamide gel maintained at the temperature of incipient DNA melting, they determined that single base mismatches may alter the melting behaviour of the heteroduplex. Partially melted molecules consisting of both double helix and disordered single-stranded sections move much more slowly than complete double helices, and thus the destabilized heteroduplexes are selectively retarded. DNA melting proceeds under equilibrium conditions as a series of relatively abrupt transitions of regions (domains) of the molecule from helix to random chain. By analysis of the sequence of a DNA molecule, it can be predicted under which conditions the different domains will melt. This knowledge enables a prediction of whether a certain mutation will be detectable using this method.

In the present paper, we have investigated whether the methods developed by Lerman et al. (1984) for DNA hybrids may be used to predict the behaviour of RNA hybrids. As a model system, we have utilized a series of mutants of the influenza A/WSN/33 virus that contain a single base substitution in the NS gene. The results presented here show that this method is indeed capable of detecting single base substitutions in RNA molecules in a predictable fashion.

B. MATERIALS AND METHODS

1. PREPARATION OF VIRAL RNA (MINUS SENSE)

Influenza viruses A/PR/8/34 (PR8) and A/USSR/90/77 (USSR) were grown in embryonated chicken eggs. Virus purification through sucrose gradient centrifugation and RNA extraction were done as described previously (Ritchey and Palese, 1976). Influenza virus A/WSN/33 (WSN) and mutants derived from this strain were grown in MDCK cells. Thirty-six hours after infection, culture supernatant was clarified by centrifugation at 10 K for 15 minutes, and the virus was pelleted through a 20% sucrose cushion. Subsequent RNA extractions were done as described previously (Ritchey and Palese, 1977). NS mutants of WSN virus were obtained in an experiment unrelated to the ones described here and were detected by directed RNA sequencing of the NS genes (chapter III). Mutants are named according to the position of the single base substitution. For example, the NS gene of mutant M068 differs from the wildtype gene only at position 68.

2. CLONING INTO SP6 AND M13 PLASMIDS

SP6 (Promega Biotec) plasmids were used as cloning vectors for viral cDNA. cDNA was made from WSN viral RNA by oligonucleotide priming of reverse transcriptase. XhoI linkers were added to the blunt-ended DNA, which was then ligated into the SalI site of SP64. After transfection of E. coli MM293 with this DNA, colonies which hybridized with a radiolabelled nick-translated cDNA copy of the NS gene of PR8 (Baez et al., 1980) were recloned, expanded, and the plasmid DNA purified by the alkaline lysis method (Maniatis et al., 1982). Clones having a full-length copy of the NS gene in the m-sense (plus sense) orientation (pSP64-WSN-NSM) were selected by analysis of restriction enzyme digestion patterns. The plasmid pSP64-WSN-NSM was then doubly digested with PstI and XbaI restriction enzymes, and the fragment

containing the cDNA insert was ligated into the PstI/XbaI site of the ds DNA replicative form of the phage M13mp18 (New England Biolabs). After transfection of E. coli JM109 with this DNA, phage producing colourless plaques on indicator medium were recloned, amplified, and the ds DNA replicative form was purified by the alkaline lysis method (Maniatis et al., 1982). Clones having a full-length copy of the NS gene in the v-sense (minus sense) orientation (M18-WSN-NSV) were selected by analysis of restriction enzyme digestion patterns.

A cDNA copy of the PR8 NS gene cloned into plasmid pBR322 (Baez et al., 1980) was recloned into the EcoRI site of plasmid SP65 by P. Gottlieb (pSP65-PR8-NSM).

3. PREPARATION OF PLUS-SENSE RNA

For production of ^{32}P-RNA probes carrying a full-length copy of the NS gene, plasmid DNA was doubly digested using either SmaI and SalI (pSP65-PR8-NSM) or SmaI and EcoRI (pSP64-WSN-NSM) restriction enzymes. (Plasmid DNA was doubly digested to insure against incomplete digestion.) For production of probes carrying truncated copies of the WSN NS gene, pSP64-WSN-NSM was singly digested with the enzymes DdeI, RsaI, BstNI, HpaII or FokI. Restriction enzymes which yielded DNA templates with blunt or 5'-protruding ends were chosen because the use of such templates has been reported to reduce the production of RNA which is specific for the non-coding strand of the template (Schenborn and Mierendorf, 1985). Linear DNA templates (100 ug/ml) were transcribed by SP6 RNA polymerase (7.5 units/ug template DNA; Promega Biotec) in the presence of 40 uCi alpha-^{32}P-CTP in a reaction volume of 20 ul at 40^{0}C for 1 hour, as described previously by Melton et. al., (1984). The DNA template was then removed by DNase 1 digestion,

followed by phenol:chloroform (1:1) extraction, and the ssRNA was ethanol precipitated (Melton et. al., 1984). Typically, 3 to 5 ug of ssRNA were produced per ug of input plasmid DNA.

4. PREPARATION OF PLUS-SENSE DNA

M18-WSN-NSV phage were grown to high titer in cultures of E. coli JM109, the phage was precipitated from the supernatant by polyethylene glycol and the ssDNA was purified by phenol extraction and isopropanol precipitation (Schreier and Cortese, 1979). The ss template DNA was incubated with the Klenow fragment of DNA polymerase I and an oligonucleotide primer specific for phage sequences adjacent to the site of cDNA insertion under conditions allowing chain elongation (New England Biolabs Laboratory Manual). The resulting dsDNA was separated from unincorporated alpha-^{32}P-dCTP by column chromatography using Sephadex G75 and digested with restriction enzymes PstI and XbaI to liberate the cDNA insert from the phage sequences. Following phenol:chloroform (1:1) extraction, the ds DNA was ethanol precipitated.

5. PREPARATION OF DOUBLE STRANDED HYBRIDS

^{32}P-labelled plus-sense RNA or DNA was resuspended in 30 ul of hybridization buffer (80% deionized formamide, 40 mM Pipes, pH 6.7, 0.4 M NaCl, 1 mM EDTA), mixed with 1-2 ug virion RNA (minus-sense), denatured for 10 minutes at 85^{o}C and incubated for 14-16 hours at 45^{o}C. To remove unhybridized species, 300 ul of either RNase digestion buffer (0.3 M NaCl, 10mM Tris-HCl, pH 7.5, 5 mM EDTA, containing 2 ug/ml of RNase T1, Sigma) or S1 digestion buffer (30 mM sodium acetate, pH 4.4, 280 mM NaCl, 5 mM $ZnCl_2$, containing 1000 unit/ml of nuclease S1, Sigma) was then added, and the reaction was incubated at 37^{o}C for 1 hour. This step was followed by a phenol:chloroform (1:1) extraction, and the ^{32}P-RNA was ethanol precipitated

with 12 ug carrier tRNA. The precipitate was washed with 70% ethanol, dissolved in 20-50 ul of sterile water, and stored at -20^{o}C.

6. ELECTROPHORESIS

The submerged gel apparatus used is that described by Fischer and Lerman (1979). Gels are 6.5% polyacrylamide (acrylamide/bisacrylamide = 30/0.8), containing a denaturant gradient either perpendicular or parallel to the direction of electrophoresis (100% denaturant = 7 M urea and 40% formamide, as defined by Lerman et al., 1984). Parallel gradient gels had a short stacking gel (3.8% acrylamide, no denaturant). The gels are run at 200V submerged in an aquarium containing TAE buffer (40 mM Tris-acetate, pH7.4, 20 mM sodium acetate, and 1 mM EDTA) that was maintained at a constant temperature of 60^{o}C by a Thermomix (B. Braun Melsungen, AG) heater and stirrer. After electrophoresis, the gel was dried and autoradiographed at -70^{o}C.

7. MEASUREMENT OF INTERNAL GEL TEMPERATURE

Temperature measurements were made using a Luxtron Fluoroptic Thermometer (Model 1000B, Luxtron, Mountain View, CA). The optical fibre probe was inserted into the center of the gel, and the temperature of the gel was monitored during electrophoresis using a BBC Goerz Metrawatt (Model SE 120) chart recorder.

8. NOTATION OF HYBRIDS

The constitution of ds hybrid molecules is designated as plus strand/minus strand. Thus, PR8/WSN denotes a hybrid possessing the plus strand of PR8 and the minus strand of WSN. Ds hybrids made between plus strand of wildtype WSN virus and minus strand of mutant WSN viruses were designated as follows:

H1=wildtype/M169; H2=wildtype/M772; H3=wildtype/M072; H4=wildtype/M068; H5=wildtype/M379; H6=wildtype/M856; C=wildtype/wildtype.

C. RESULTS

1. DETECTION OF MULTIPLE SUBSTITUTIONS IN RNA MOLECULES

Ds RNA hybrids were made by hybridizing PR8, USSR, and WSN virion RNA with an SP6 probe containing a plus-sense copy of the PR8 NS gene. The PR8 NS gene is known to differ from the USSR NS gene at 45 positions, and from the WSN NS gene at 31 positions (Krystal et. al., 1983; Buonagurio et. al., 1986b). After hybridization, hybrids were treated with RNase T1 to remove non- paired single-stranded RNA. On electrophoresis in 6.5% acrylamide gels in the absence of denaturant at room temperature, PR8/PR8 hybrids (Fig. 8A, lane 2) were easily distinguishable from PR8/USSR (lane 1) and PR8/WSN (lane 3) hybrids. The presence of urea and formamide altered the relative migration rates of these three hybrids, resulting in different patterns of migration at different depths in a denaturing gradient gel (Figs. 8B, C, D). Therefore, electrophoresis in both the presence and absence of denaturant enabled the detection of hybrid ds RNA molecules containing multiple mismatches.

2. DETECTION OF SINGLE BASE SUBSTITUTIONS IN RNA MOLECULES

a. Low temperature gels.

Ds RNA hybrids were made by hybridizing a full-length plus sense copy of the WSN NS gene with viral RNA purified from wildtype virus or from WSN mutants which are known to differ from wildtype by only a single base change. Five of these hybrids (C, H1, H2, H3 and H5) were analyzed on low temperature gels. No difference in electrophoretic migration of these hybrids was observed in the absence of urea and formamide (data not shown), and

only one of the four hybrids containing a mismatch (H5) migrated anomalously in gels containing a 0-100% denaturing gradient (figure 8E). Temperature measurements showed that a 6.5% acrylamide gel containing 50% denaturant run at 200 V at room temperature reached a maximum temperature of approximately 37^{o}C.

b. High temperature gels.

(a) Full length SP6 probes.

To analyze the electrophoretic behaviour of ds RNA hybrids under conditions of incipient melting, gels were made containing a 0-100% or a 20-60% denaturant gradient perpendicular to the direction of electrophoresis. All hybrids tested gave biphasic curves when the gels were run at 60^{o}C (Fig. 9). There are two obvious transitions, with midpoints at approximately 27% and 50% of denaturant (see arrows in Fig. 9). At denaturant concentrations above 55%, the hybrids are practically stationary in the gel matrix. We therefore chose to examine the behaviour of the hybrids in parallel gradients of 0-60% denaturant, run at 60^{o}C. Fig. 10 shows that three of the six heterologous hybrids were distinguishable from wildtype hybrids using these conditions. The migration of hybrid H6 was significantly retarded at concentrations of denaturant as low as 20% (Fig. 10A, lane 7), and remained retarded throughout the gel (Fig. 10C, lane 7). Hybrid H5 showed significant retardation only between 30-40% denaturant (Fig. 10B, lane 6). However, longer electrophoresis times resulting in migration of the bands into higher concentrations of denaturant eliminated this difference (Fig. 10C, lane 6). Hybrid H2 migrated more slowly than the wildtype control at denaturant concentrations above 47% (Fig. 10C, lane 3). The single base mismatch present in hybrids H2 and H6 is located at positions 772 and 856, respectively. This

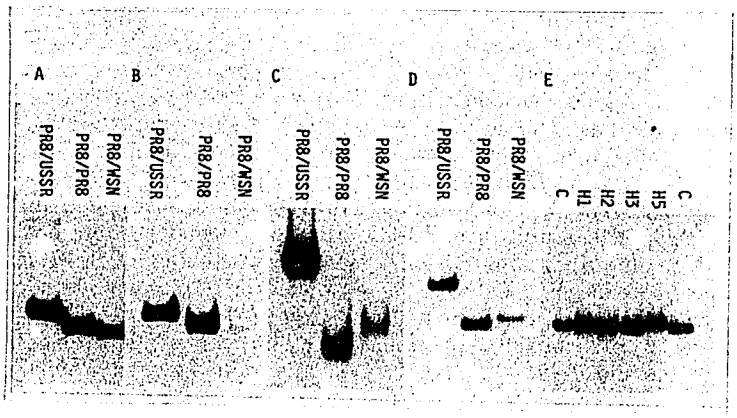


Figure 8. IDENTIFICATION OF MISMATCHES IN DS RNA HYBRIDS BY ELECTROPHORESIS ON 6.5% ACRYLAMIDE GELS AT ROOM TEMPERATURE.

A. Gel contains no denaturant. B-E. Gels contain a 0-100% gradient of denaturant parallel to the electric field and the patterns of migration shown were obtained when the wildtype/wildtype hybrids reached the following concentrations of denaturant: 40% (B), 75% (C), 95% (D) and 90% (E). PR8/PR8, PR8/USSR and PR8/WSN represent hybrids derived from radioactively labeled (plus sense) PR8 RNA and virion RNA of PR8, USSR and WSN, respectively. For explanation of hybrid notation (C, H1, H2, H4 and H5) see Materials and Methods. Results show that ds RNA hybrids containing multiple mismatches can be separated under these conditions, whereas only one out of four of the hybrids containing a single base mismatch is distinguishable.

region has the lowest G-C content of the NS gene (Buonagurio, et. al., 1986b) and thus may be expected to be the domain which melts at the lowest temperature (or denaturant concentration). Hybrids H1, H3 and H4 with mutations in regions of higher G-C content (see below) were not distinguishable from the control hybrid C at any concentrations of denaturant tested.

(b). Truncated SP6 probes.

Analysis of the results above suggested that melting of one region of the RNA-RNA heteroduplex (including positions 772 and 856) caused a drastic reduction of mobility. Therefore, we predicted that if we removed this domain, the shorter hybrid would continue to migrate into the denaturing gradient, until the next lowest melting domain became disordered. Mutations within this domain should then become detectable. Therefore, we made a series of shorter SP6 probes by digesting the plasmid SP64-WSN- NSM with restriction enzymes that cut at various sites within the NS gene. Hybridization of these short probes with minus-sense RNA, followed by S1 nuclease digestion of the unpaired bases, resulted in ds RNA hybrids of length 153 bp ("153-hybrids"), 277 bp ("277-hybrids), 387 bp ("387-hybrids"), 443 bp ("443-hybrids") and 560 bp ("560-hybrids"). Obviously, hybrids made with these short probes cannot be used to detect mutations which occur in the minus-sense RNA at positions beyond the length of the probe. For example, hybrids H2, H5 and H6 in the 277-hybrid series should not differ from wildtype hybrids.

Hybrids were electrophoresed on parallel gradients of 0-100% denaturant run at 60^{o}C. All hybrids ran quickly through the low concentrations of denaturant without showing any migrational differences. However, at

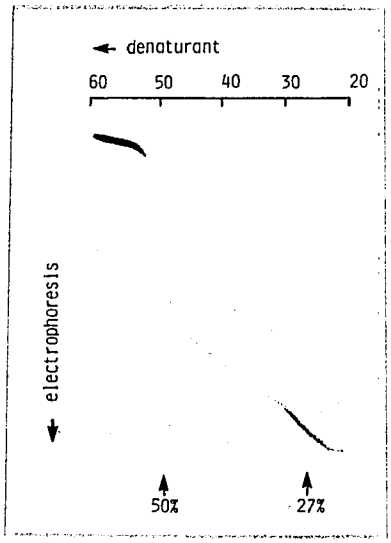


Figure 9. EXPERIMENTAL MELTING PATTERN OF WILDTYPE/WILDTYPE DS RNA HYBRID.

The mobility of a ds RNA hybrid containing a radioactively labelled plus-sense strand was examined on a denaturing 20-60% gradient perpendicular to the electric field. The sample (hybrid C) was applied uniformly across the top of the gel. The arrows indicate the midpoint of two transitions observed.

concentrations of denaturant between 70% and 80%, hybrid H5 (which has a mismatch at position 379) became retarded in 560-, 443-, and 387-hybrids, and hybrids H3 and H4 (which have mismatches at positions 68 and 72, respectively) became retarded in all series of short hybrids (Fig. 11). However, hybrid H1, which has a mismatch at position 169, migrated indistinguishably from the wildtype hybrid C under these conditions. These results suggested that position 169 is located in a domain which melts at even higher concentrations of denaturant. To examine the behaviour of hybrids at very high temperature equivalents, a gradient gel was made containing higher denaturant concentrations and the temperature of the buffer-filled aquarium used for electrophoresis was increased. 277- and 387-hybrids were run under these conditions. As samples migrated into high concentrations of denaturant, bands became diffuse and it was noted in both series of hybrids that the band derived from hybrid H1, containing a mismatch at position 169, ran faster than the bands seen in the other lanes (Fig. 12A and 12B, lane 2; data for 277-hybrids not shown). These results indicate that we are observing melting of the highest temperature melting domain under these conditions, and that a mismatch in this domain causes an earlier complete strand separation. Analysis of ss RNA and ds RNA on denaturant gradients perpendicular to the direction of electrophoresis confirms that ss RNA migrates faster than the ds hybrids at high denaturant concentrations (data not shown).

(c). Full-length SP6 probe made with ITP.

The stability and thus the melting behaviour of RNA-RNA molecules is dependent on the G-C content of the duplexes. We investigated whether replacing GMP with IMP in the plus strand of the hybrids changed their migrational behaviour. Analysis by electrophoresis at 60^{o}C on a gel containing

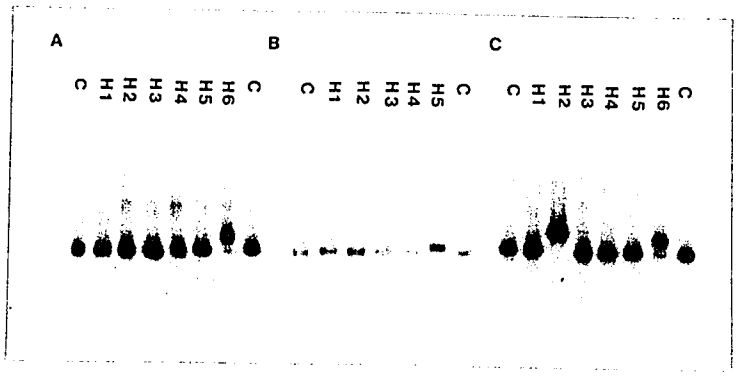


Figure 10. IDENTIFICATION OF SINGLE BASE MISMATCHES IN DS RNA HYBRIDS BY ELECTROPHORESIS ON 6.5% ACRYLAMIDE GELS CONTAINING A DENATURING GRADIENT.

The gradient (0-60%) was parallel to the electric field. The patterns of migration shown were obtained when hybrid C reached the following approximate concentrations of denaturant: 20% (A), 35% (B), 50% (C). Hybrid notations are as described in Materials and Methods. Results show that 3 out of 6 of the hybrids containing a single base mismatch are distinguishable. (Note: hybrid H6 is not analyzed in panel B).

a 0-60% denaturant gradient perpendicular to the direction of electrophoresis showed that the hybrids made with ITP melt at low concentrations (0-5%) of denaturant (data not shown). Therefore, these hybrids were electrophoresed at a lower temperature, to enable examination of their behaviour over a wide range of denaturing conditions. Parallel denaturing gradient gels run at 40^{o}C showed that hybrid H5 is selectively retarded at concentrations of denaturant between 20% and 40% (Fig. 13, A and B, lane 6), whereas hybrids H2 and H6 lag behind at concentrations of denaturant greater than 50% (Fig. 13C, lanes 3 and 7). Therefore, the same mismatches were detected using an IMP-containing or a GMP-containing probe.

(d). Full length M13-DNA probe

^{32}P-DNA plus-sense probes were hybridized with minus-sense RNA and unpaired single stranded DNA was digested with S1 nuclease. Analysis of DNA-RNA hybrids on a gel run at 60^{o}C and containing a 0-50% denaturing gradient perpendicular to the direction of electrophoresis showed that the DNA-RNA hybrid undergoes a sharp decrease in migration at a concentration of approximately 15-18% denaturant (data not shown). DNA-RNA hybrids were therefore analysed on gels containing a 0-20% denaturant gradient, parallel to the direction of electrophoresis. Fig. 14 shows that hybrids H2 and H5 were selectively retarded as was seen using RNA-RNA heteroduplexes made with GTP or ITP-containing probes.

3. MELTING MAP OF THE DS DNA COPY OF THE WSN VIRUS NS GENE

The expected melting progression along the ds DNA copy of the NS gene of WSN virus was calculated as described by Lerman et. al. (1984; Fig. 15). The ordinate in figure 15 shows the temperature at which each base pair has an equal probability of helix or random chain configuration.

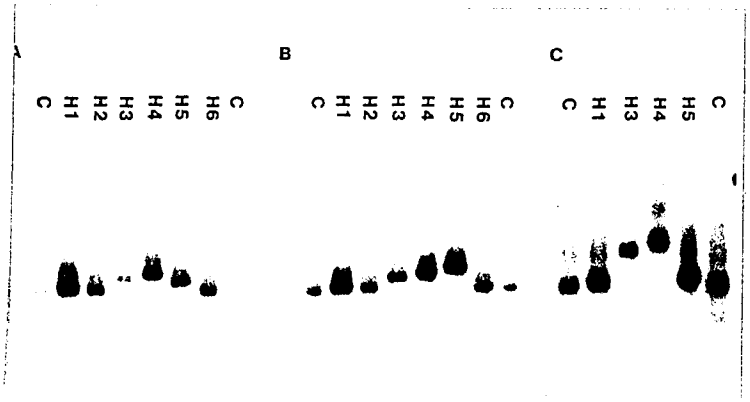


Figure 11. IDENTIFICATION OF SINGLE BASE MISMATCHES IN DS RNA HYBRIDS PREPARED WITH TRUNCATED RADIOACTIVELY LABELED PLUS-SENSE RNA PROBES.

Electrophoresis was on 6.5% acrylamide gels containing a denaturant (0-100%) gradient parallel to the electric field and the run was stopped when the control hybrid C reached a denaturant concentration of approximately 75-80%. A. 560 bp length hybrids, B. 443 bp length hybrids, C. 387 bp length hybrids. Results show that hybrids H3, H4 and H5 are distinguishable using truncated RNA probes.

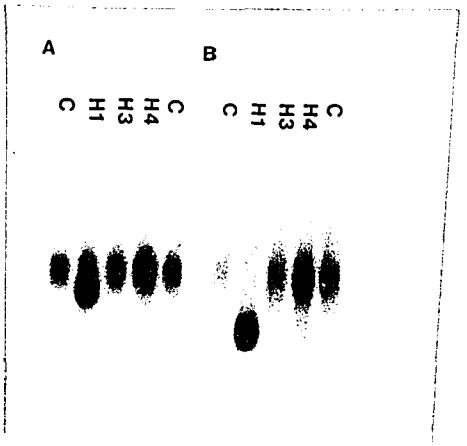


Fig. 12. IDENTIFICATION OF A SINGLE BASE MISMATCH LOCATED IN A HIGH MELTING DOMAIN OF DS RNA HYBRIDS.

387 bp length hybrids were electrophoresed at 64^{o}C on 6.5% acrylamide gels containing a denaturant (63-100%) gradient parallel to the electric field. In order to establish denaturing conditions which allow complete strand separation of the ds RNA hybrids, a gradient was formed using solutions containing 4.4 M urea and 38% formamide and 7 M urea and 7 M urea and 60% formamide, respectively. Patterns of migration shown were observed when hybrid C reached a denaturant concentration of approximately 90% (A) or 100% (B).

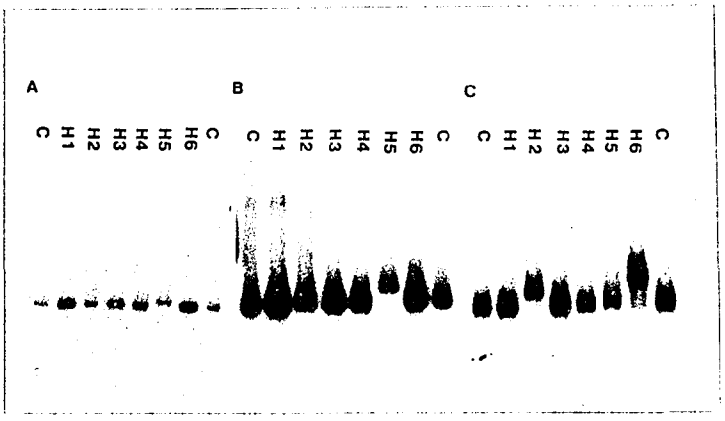


Figure 13. ELECTROPHORESIS OF DS RNA HYBRIDS CONTAINING IMP IN THE RADIOACTIVELY LABELLED PLUS-SENSE STRAND.

Hybrids were electrophoresed at 40°C on 6.5% acrylamide gels containing a denaturing (0-100%) gradient parallel to the electric field. Patterns of migration shown were obtained when hybrid C reached a denaturant concentration of 22% (A), 40% (B) or 50% (C).

The calculation shows that melting is expected to proceed stepwise as the temperature is raised. The effect of cooperativity is strong enough that fairly long blocks of contiguous helix melt within narrow temperature intervals. At least five different melting domains can be seen.

D. DISCUSSION

Ito and Joklik (1972) showed that differences in electrophoretic migration of ds RNA heteroduplexes can be used to detect sequence differences in RNA molecules. Also, Lerman and colleagues (Lerman et al., 1984; Myers et al., 1985) reported that single point mutations in ds DNA molecules can be detected predictably by analysis of their electrophoretic behavior under conditions of incipient melting. We have used these observations as a basis to establish conditions whereby single point mutations in RNA molecules can be detected easily, reliably and predictably.

The results presented in this paper show that six point mutations, located at different positions along the NS gene (890 nucleotides in length) of influenza WSN virus, were all distinguishable by electrophoresis of ds RNA duplexes on denaturing gradient gels. By using the plasmid SP6 as a cloning vector, we were able to produce large amounts of plus-sense ss RNA labelled to high specific activity. This removes the need for working with cRNA obtained from infected cells (Ito and Joklik, 1972; Hay et al., 1979) and allows for the separate analysis of individual genes.

Although ds RNA duplexes are more stable than their ds DNA counterparts, we assume that the relative stability of the various domains will be the same. Analysis of the thermodynamics of several stacking interactions of ds DNA (Gotoh and Tagashira, 1981) and ds RNA (Borer et. al., 1974) show similarities in the hierarchy of stabilities of nearest neighbour doublets.

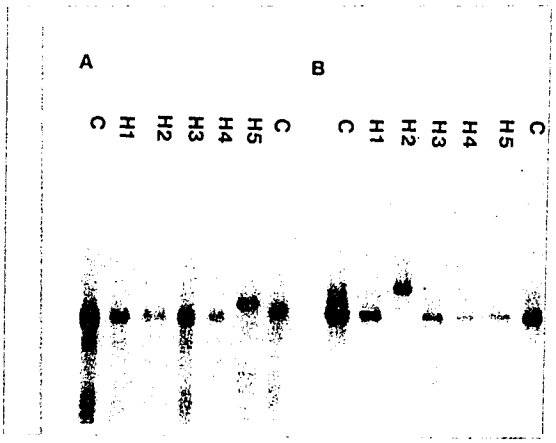


Figure 14. IDENTIFICATION OF SINGLE BASE MISMATCHES IN DS DNA-RNA HYBRIDS PREPARED WITH RADIOACTIVELY LABELLED PLUS-SENSE DNA.

Hybrids were electrophoresed at 60°C on 6.5%polyacrylamide gels containing a denaturant gradient (0-20%) parallel to the electric field. Patterns of migration shown were obtained when hybrid C reached a denaturant concentration of 12% (A) and 17% (B).

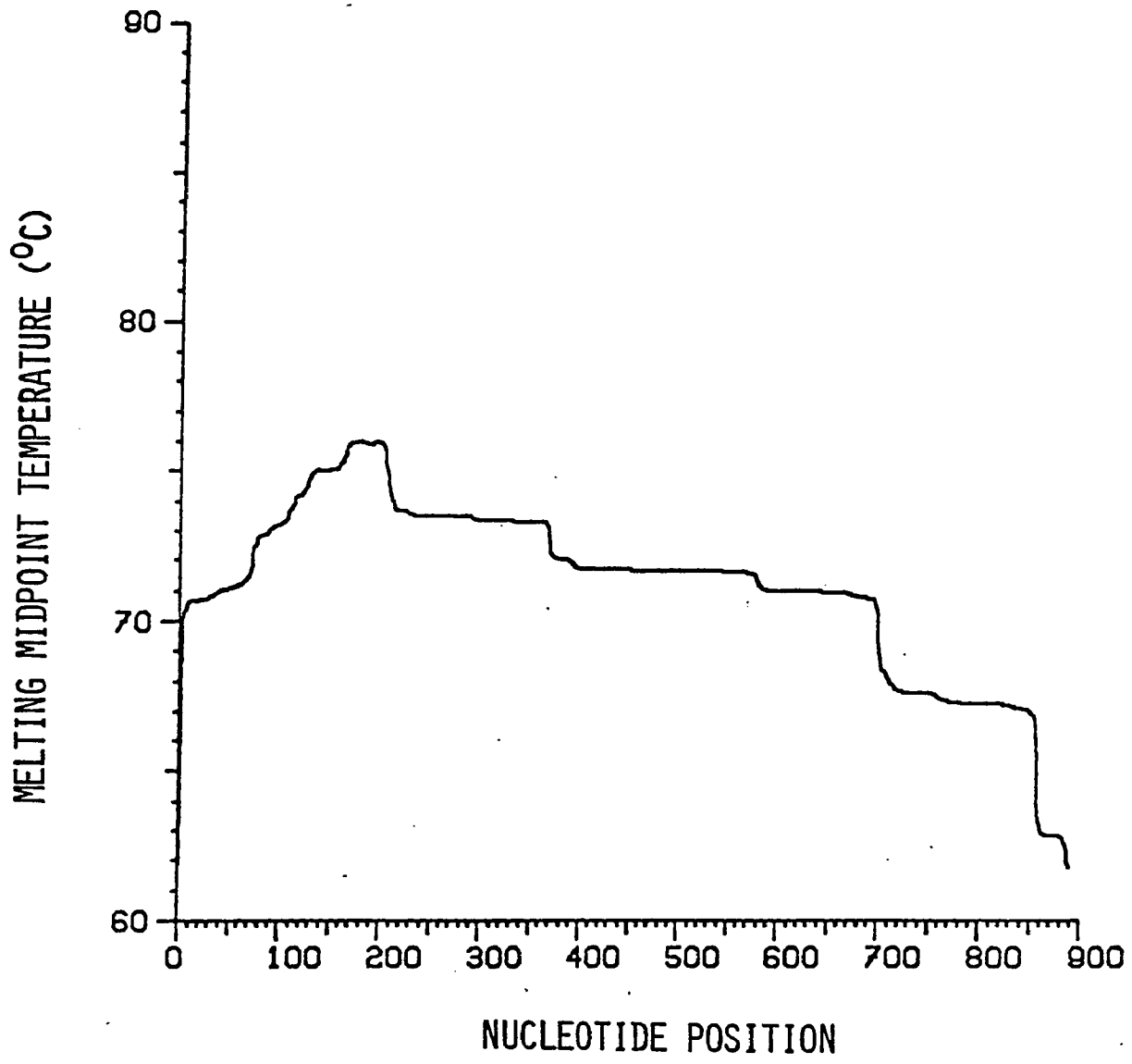


Figure 15. MELTING MAP OF THE DS DNA COPY OF THE WSN VIRUS NS GENE.

The abscissa represents positions along the sequence of the NS gene. The ordinate shows the temperature at which each base has an equal probability of helix or random chain configuration.

Therefore, we have compared experimental results obtained with ds RNA duplexes (Figs. 9, 10, 11 and 12) with those predicted by the theoretical melting map (Fig. 15) for the ds DNA duplexes. The first domain expected to melt is relatively short, and consists of bases between 875 and 890. The melting of this domain may correspond to the first transition seen at 27% denaturant in the perpendicular gradient gel run at 60°C (Fig. 9). The next domain to melt is much longer (between nucleotides 705 and 875) and may correspond to the very sharp mobility transition seen at 50% denaturant (Fig. 9). These domains are likely to melt earlier in heteroduplexes H2 and H6 with mismatches in positions 772 and 856, respectively, than in the control homoduplex. This partial melting of H2 and H6 RNA hybrids should result in a selective retardation on denaturing gradient gels because the melted arms of the molecules decrease gel mobility. Results shown in Fig. 10C bear out this prediction. It is interesting to note that hybrid H6 is detectable at lower denaturant concentrations than is H2. Because the mismatch in H6 is located close to the border between the first two melting domains, it apparently influences the melting of the first domain, whereas the mismatch in H2 only affects melting of the second domain.

Melting map analysis predicts that mutations located in domains between positions 15 to 80 and between 375 and 705 would be detected at higher denaturant concentrations. The use of probes eliminating the lowest melting domains allows the hybrids to migrate deeper into the gradient gel to attain these conditions. Results obtained using truncated probes (Fig. 11) show that mutations at positions 68, 72 and 379 caused selective retardation of hybrid migration at high denaturant concentrations, thus confirming this prediction. It is interesting that hybrid H5 (mutation at position 379) can be detected in a

387 bp length heteroduplex, where the mismatch is only 8 bases from the end. It should be noted that the mutation at position 379 also caused a transient change in mobility of ds RNA hybrids made with full length probes when hybrids were electrophoresed at low concentrations of denaturant (Fig. 10). This effect is probably not due to an alteration in melting behaviour but rather the result of a conformational change.

Melting of the highest melting domain of a duplex molecule leads to complete strand separation. Any strand liberated by dissociation from a partially melted molecule will migrate more rapidly. At high denaturant concentrations in gels run at high temperature, we observed that the hybrid containing a mismatch at position 169 ran faster than the other hybrids (Fig. 12). Therefore, we are able to detect mutations in the highest melting domain by subjecting the ds RNA hybrids to conditions promoting strand separation.

Replacement of GMP with IMP in the radioactively labelled full length plus-sense probe also enabled the detection of single base mismatches. This modification may be useful if it is more convenient to work at lower temperature, or when the replacement alters the melting map in such a way as to increase the number of mutations detectable with a full-length probe. Although in this study the same mutations were detectable using IMP probes as were seen using probes containing GMP, a different distribution of G residues in another gene may make this a method with broader applications.

Results obtained using an M13-derived probe showed that denaturing gradient gel electrophoresis can also be used effectively to distinguish single point mismatches in DNA-RNA hybrids. Therefore, the detection of point mutations in DNA viruses is also feasible by denaturing gel electrophoresis, in which the highly efficient production of ss RNA probes by SP6 polymerase may

be exploited for the production of DNA-RNA hybrids.

Another important potential application of this method would be in the diagnosis of genetic diseases. Mutations in genes do not always cause an alteration in a restriction enzyme cleavage site. In such cases, diagnosis of the presence of the mutant gene (in either affected individuals or symptomless carriers) is dependent on finding linkage to a restriction fragment length polymorphism (RFLP). Tightly-linked RFLP's have not yet been found for every gene, and when found are not informative in every family. The ability to detect directly any single point mutation, in either genomic DNA or in mRNA, by denaturing gel electrophoresis of DNA-RNA or RNA-RNA hybrids would overcome these problems.

V.

RAPID RNA SEQUENCING USING DOUBLE-STRANDED TEMPLATE DNA, SP6 POLYMERASE AND 3'-DEOXY NUCLEOTIDE TRIPHOSPHATES

Jeffrey D. Parvin, Frances I. Smith and Peter Palese

Department of Microbiology
Mount Sinai School of Medicine of CUNY
Fifth Avenue and 100th Street
New York, N.Y. 10029

Originally published in DNA 5, 167-171 (1986). Reprinted with permission from Mary Ann Liebert, Inc.

A. INTRODUCTION

Many methods now exist for sequencing RNA and DNA. In addition to the Maxam and Gilbert technique of sequencing (Maxam and Gilbert, 1977), the Sanger method of specific termination of nascent polynucleotide chains by incorporation of 2',3'-dideoxy nucleotides (Sanger, et. al., 1977) has been widely utilized. Based on this strategy, single-stranded RNA can be sequenced using reverse transcriptase (Zimmern and Kaesberg, 1978), and effective sequencing systems using single-stranded bacteriophage M13 DNA (Messing, et. al., 1981) or denatured plasmid DNA (Chen and Seeburg, 1985) have also been developed. Another application of the Sanger strategy was the sequencing of RNA by QB replicase using 3'-deoxy nucleotides (Kramer and Mills, 1978). However, this latter technique has limited application due to the template specificity of the polymerase (Blumenthal, 1980) and due to problems in resolving a sequencing ladder caused by the secondary structure present in the RNA molecules (Kramer and Mills, 1978).

The recent availability of bacteriophage SP6 polymerase has provided a relatively simple and efficient method for preparing RNA transcripts from plasmid DNA (Butler and Chamberlin, 1982; Kassavetis et. al., 1982; Melton et. al., 1984). This transcription system can also be utilized to provide a means of sequencing by chain termination.

B. MATERIALS AND METHODS

1. PREPARATION OF TEMPLATE DNA

Two templates were used in this study. One template (pSP64- WSN-NS) is derived from RNA segment 8 (NS gene) of influenza virus A/WSN/33 inserted into the SalI site of the polylinker region of pSP64 using XhoI linkers

ligated to the full length ds cDNA copy. The second template (pSPR1-801-d6) is a genetically engineered deletion mutant of a cDNA encoding RNA segment 8 of influenza virus A/PR/8/34 inserted into the EcoR1 site of a modified pSP64 plasmid (Baez, et. al., 1980; Greenspan, et. al., unpublished). Both inserts are in message orientation relative to the SP6 polymerase promoter. The plasmids were prepared using standard procedures and were either passed over an AG-150 sizing column or banded on a CsCl gradient.

2. NUCLEOTIDES AND ENZYME

The SP6 polymerase was obtained from Promega Biotec. ATP, CTP and UTP were purchased from Pharmacia, Inc. and ITP from Sigma Chemical Co. The 3'-deoxy ATP (cordycepin triphosphate) was obtained from Sigma, the 3'-deoxy GTP, 3'-deoxy CTP and 3'-deoxy UTP from Pharmacia and the radioactive label, alpha-^{32}P-CTP from New England Nuclear.

3. SEQUENCING REACTIONS

The final composition of the A,G,C and U reactions are outlined in Table 1. Each of the four reactions was carried out in 10 ul volumes containing pre-mixed nucleotide solutions, 2 ug template DNA, 20-30 uCi alpha-^{32}P-CTP and 4 U of SP6 polymerase. RNasin, as frequently used in SP6 transcription (Melton et. al., 1984), can be omitted. The reactions were incubated at 42^{o}C for one hour. One microliter of a cold chase mix containing 3.5 mM ATP, 3.5 mM UTP, 3.5 mM ITP, 5.5 mM CTP and 1X transcription buffer was added and incubation at 42^{o}C was continued for 30 minutes. The reactions were terminated by the addition of 11 ul running mix containing 78% deionized formamide, 10 mM EDTA pH 8, 0.1% xylene cyanol and 0.05% bromophenol blue. No further treatment of the reactions was necessary before electrophoresis.

TABLE 3. SP6 SEQUENCING REACTION MIXTURES.

Transcription buffer*	G	A	U	C
CTP	12.5 uM	12.5 uM	12.5 uM	25.0 uM
UTP	500 uM	500 uM	100 uM	500 uM
ATP	500 uM	100 uM	500 uM	500 uM
ITP	400 uM	1000 uM	1000 uM	1000 uM
3'-dNTP	20 uM	33 uM	600 uM	50 uM
alpha-^{32}P-CTP**	20 uCi	20 uCi	20 uCi	30 uCi

*Transcription buffer contains 40 mM tris-HCl pH 7.5, 6 mM $MgCl_2$, 2mM spermidine, 100 mM NaCl, and 10 mM DTT.

**The specific activity of alpha-^{32}P-CTP was 800 Ci/mmole.

4. GEL ELECTROPHORESIS OF SEQUENCING REACTION PRODUCTS

Immediately before loading gels, samples were heated in boiling water for five minutes and rapidly cooled in an ice water bath. Samples were stable after repeated boilings. One to three microliters of each reaction sample was applied to gels containing 6% or 8% polyacrylamide, 50 mM tris-borate pH 8.3, 2.5 mM Na_2EDTA and 7.7 M urea. Gels were run at approximately 50^{o}C on a Hoefer Scientific Instruments or an International Biotechnologies, Inc. sequencing apparatus. The gel temperature was determined using a Pacific Transducer Corp. surface thermometer. After electrophoresis gels were fixed in a 10% (v/v) acetic acid, 10% (v/v) methanol solution, dried and then exposed to Cronex 4 film without an intensifying screen for one to seven days.

C. RESULTS

1. SEQUENCING RESULTS

The gel depicted in figure 16 shows the partial sequence of the genetically engineered deletion mutant pSPR1-801-d6. The first 24 bases are transcribed from the polylinker region of pSP64 followed by six bases of linker DNA and 72 bases of viral sequence. The viral sequence can be read in the expanded region of the figure. Figure 17 shows sequencing results of plasmid pSP64-WSN-NS. The corresponding viral RNA is represented next to the gel and the expanded region is from nucleotides 226 to 261.

Accurate reading of sequencing ladders begins about 20 nucleotides from the transcription initiation site and, as can be seen in Figure 17, continues beyond 280 nucleotides. It is our experience that these reaction conditions often enable well over 500 nucleotides to be read, but the resolution of RNAs of this length is poorer (data not shown). The ratios of NTP:3'-dNTP in our reaction mixes are as follows: ITP:3'-dGTP = 20:1, ATP:3'- dATP = 3:1, UTP:3'-dUTP = 1:6 and CTP:3'-dCTP = 1:2. When the 3'-dNTP concentration is decreased (raising the ratios), the appearance of non-specific bands increased, and when the chain-terminating nucleotide concentration is increased, the length of the sequencing ladder is decreased without any concomitant improvement of specificity. However, increasing the 3'-dNTP concentrations allowed sequences as close as 12 nucleotides from the transcription initiation to be read when the sample was electrophoresed on 20% polyacrylamide gels (data not shown).

2. SPECIFICITY OF SEQUENCING REACTIONS

The NS genes from A/WSN/33 and A/PR/8/34 have been sequenced in our laboratory by Maxam and Gilbert sequencing of a cDNA clone (A/PR/8/34) (Baez, et. al., 1980) and by reverse transcription of viral RNA in the presence

Figure 16. SEQUENCE DETERMINATION OF pSPR1-801-d6.

The autoradiographs of two loadings of an 8% polyacrylamide gel are shown with the accompanying sequence. The first 24 bases are derived from the pSP64 polylinker, the next six nucleotides are derived from an EcoR1/BamH1 linker and the first virus-specific nucleotide is indicated as position 48. The expanded region, representing a longer time of electrophoresis, more clearly resolves the higher molecular weight molecules from viral positions 51 to 119.

Figure 17. SEQUENCE DETERMINATION OF pSP64-WSN-NS.

The autoradiographs of an 8% polyacrylamide sequencing gel and of a 6% gel electrophoresed about four times longer are shown. The sequences correlating to viral positions 50 to 119 and 226 to 261 are depicted. Since the transcription initiation is at position -28, molecules of 289 nucleotides are shown to be resolved from their nearest neighbor.

Figure 16

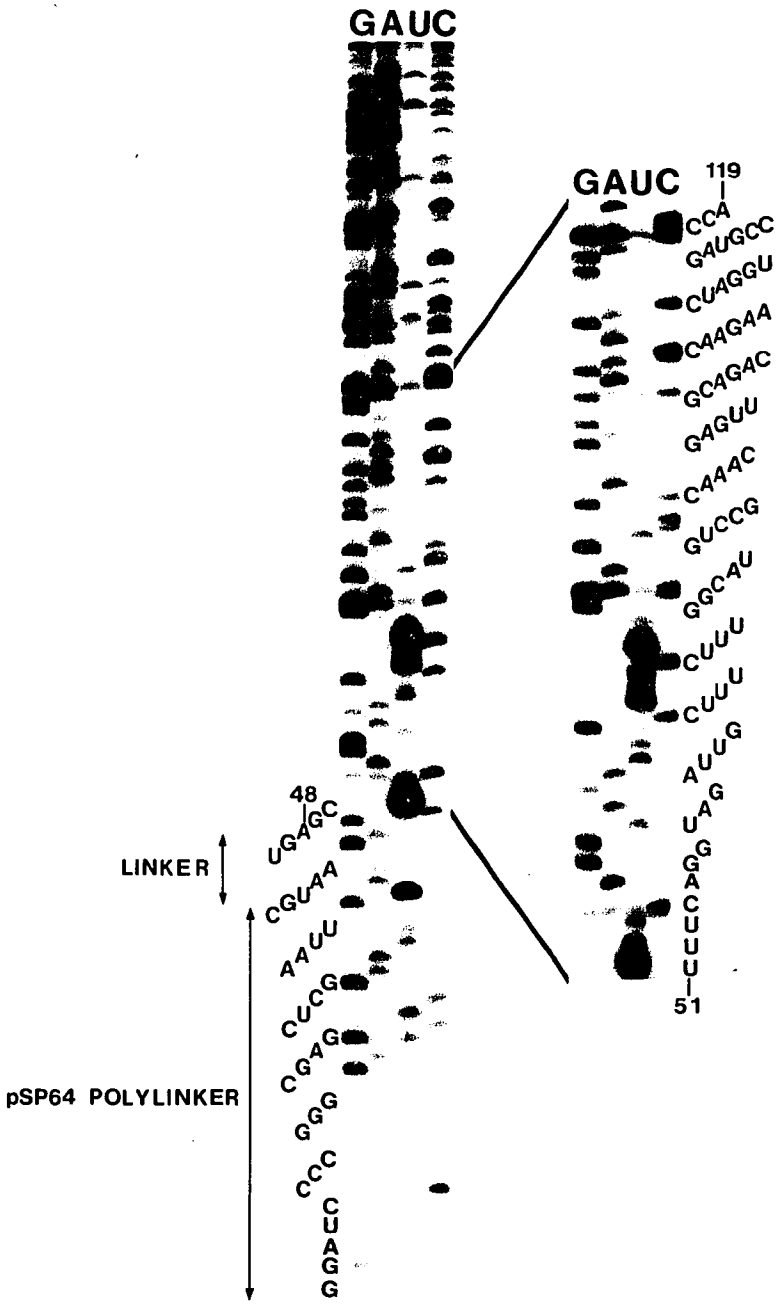
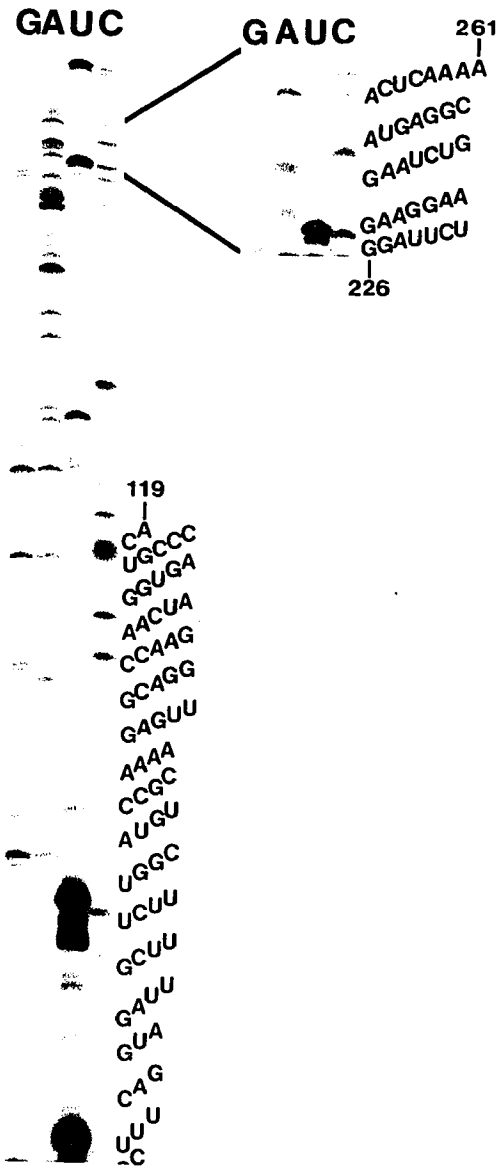


Figure 17


GAUC
GAUC
261
ACUCAAAA
AUGAGGC
GAAUCUG
GAAGGAA
GGAUUCU
226
119
CA
UGCCC
GGUGA
AACUA
CCAAG
GCAGG
GAGUU
AAAA
CCGC
AUGU
UGGC
UCUU
GCUU
GAUU
GUA
CAG
UUU
UC

of 2',3'-dideoxy nucleotides (A/WSN/33, Buonagurio, et. al., 1986b). The sequence of the A/WSN/33 cDNA clone determined using SP6 polymerase agreed with the previously determined sequence of the viral RNA with the exception that an A to G change was observed at position 97. We assume that this nucleotide change represents a true mutant, possibly the result of the cloning procedure involving reverse transcription. The sequence of the A/PR/8/34 clone obtained using SP6 polymerase is identical with that of the same cDNA insert previously established by Maxam and Gilbert analysis with the exception that the first 47 nucleotides have been deleted. This deletion seen in the pSPR1-801-d6 plasmid was predicted on the basis of its manner of preparation (Greenspan, et. al., unpublished).

The sequencing gels show that the reactions are specific, generally free of extra bands and low in background. There are several sites where the most intense band must be read, and instances where a G is represented by a band in all four lanes. It is possible that the SP6 polymerase has decreased affinity for ITP and thus occasionally halts synthesis in all four reactions. The most difficult specificity problem is the variable intensity of bands in the U track. There is sometimes very heavy labelling of the first U in a run of three or more uridines. Conversely, in several instances the expected band in the U lane is very faint and perhaps even absent. The variable band intensity observed in the U lane may be related to the observation made by Melton, et. al. (1984) who reported that chain elongation by SP6 polymerase is apparently more sensitive to changes in UTP concentration than to changes in the other three nucleotides when working at low nucleotide concentrations.

THE EFFECT OF TEMPERATURE. It was hoped that by altering the reaction conditions the specificity of the U reaction would improve. In addition to

altering the nucleotide concentrations as described earlier, we tested several reaction temperatures: 37°C, 42°C, 45°C and 50°C. The intensity of bands in all four lanes was markedly decreased at 50°C, and reactions at 45°C had decreased band intensity in the A and U lanes. The appearance of consecutive uridines was improved on comparing reactions at 42°C with those at 37°C, while the band intensity was comparable.

3. GEL ELECTROPHORESIS

Secondary structures in RNA molecules are more stable than those in analogous single-stranded DNA molecules (Kallenbach, 1968) and therefore resolution of a sequencing ladder of transcribed RNAs is more difficult. The replacement of guanosine by inosine decreases the stability of secondary structures (Mills and Kramer, 1979), and combined with electrophoresis temperatures of about 50°C, almost all compressions are eliminated.

D. DISCUSSION

APPLICATIONS AND ADVANTAGES OF THE SEQUENCING METHOD

SP6 polymerase is widely used for the synthesis of single-stranded RNA, which may be used for a variety of biochemical techniques including studies on RNA splicing (Green, et. al., 1983; Krainer, et. al., 1984). In such studies, it is often useful to generate new constructs and/or mutants in order to ascertain their effect in a given system. The structure of the novel plasmid and/or the presence of expected mutations could most easily be verified by using the sequencing system outlined in this paper. For example, we are currently using this system for analysis of the deletions in insert influenza viral DNAs generated by exonuclease Bal 31 digestion. It may then be possible in the future to directly sequence minipreparations of plasmid DNAs using a modified

alkaline lysis method (Maniatis, et. al., 1982) which improves the purity of the plasmid DNA.

We believe that this sequencing method presents several advantages, both in ease of technical manipulations and in scope of applications. The template is untreated plasmid prepared by standard procedures. Since the plasmid already contains a specific SP6 promoter, tedious preparation and hybridization of a primer is obviated. To set up the four reactions, we have premixed solutions containing buffer, nucleotides and terminators; we simply add DNA, alpha-labeled nucleotide and polymerase to these solutions. After termination of the reaction using the formamide-dye mix, caution regarding RNase contamination need not be exercised. The gel electrophoresis conditions are not difficult since many commercially available sequencing apparatuses are designed to run gels at high temperatures. We can expect about 300 nucleotides to be determined from a reaction set. Although this number is not as high as is often achieved by sequencing DNA from an M13 template, the fact that the template is plasmid DNA paves the way for new and simplified applications of sequence analysis.

The possibility that T7 RNA polymerase (Davanloo, et. al., 1984) may similarly be used to sequence insert DNA may increase the versatility of this technique. Special plasmids are commercially available which have the SP6 and T7 promoters present in opposite orientations on either end of the polylinker region. With such a plasmid it may be possible to sequence insert DNA from both ends, therefore improving the speed with which sequences may be determined. A single plasmid preparation may thus allow the determination of the complete sequence of short inserts. In addition, a shotgun strategy, as is used in the M13 system (Messing, et. al., 1981), may be applied in order to

determine the complete sequence of large genes.

VI.

THE GENOMIC RNAS OF INFLUENZA VIRUSES ARE HELD IN A CIRCULAR CONFORMATION IN VIRIONS AND IN INFECTED CELLS BY A TERMINAL PANHANDLE

by

Ming-Ta Hsu, Jeffrey D. Parvin, Suresh Gupta, Mark Krystal and Peter Palese

Department of Microbiology
Mount Sinai School of Medicine of CUNY
Fifth Avenue and 100 th Street
New York, New York 10029

This chapter will be published in Proc. Natl. Acad. Sci. U.S.A. (in press).

A. INTRODUCTION

Influenza A virus contains eight negatively stranded RNA segments which code for at least 10 distinct proteins (for review: Palese and Kingsbury, 1983). The core of the virus consists of eight individual nucleocapsids which are made up of the RNA segments, the three polymerase (P) proteins and nucleoprotein molecules (NP). Early analyses revealed that these nucleocapsids have different molecular weights (Duesberg, 1969) and that they appear as strands (Compans, et. al., 1972) which show loops at one end. The tertiary structure is formed by folding back on itself and coiling, which results in a large double helix (Compans, et. al., 1972; Jennings, et. al., 1983). The lengths of the nucleocapsid strands can be correlated with the molecular weights of the viral RNA segments. Also, the presence of multiple nucleocapsid segments correlates with the finding that the mRNAs are transcribed independently (Pons and Rochavansky, 1979; Hay and Skehel, 1979). Each of the nucleocapsids directs the synthesis of complementary RNA starting from the 3' ends of the virion RNAs (Braam, et. al., 1983). Although sequencing studies of the 3' and 5' termini of influenza A, B and C virus RNAs revealed the presence of inverted complementary sequences (Skehel and Hay, 1978; Robertson, 1979; Desselberger, et. al., 1980), little was known about the tertiary structure of the termini of influenza virion RNAs.

In the present study we show that UV crosslinking of virion RNAs in the presence of 4'-substituted psoralen leads to circular RNA structures. This finding supports the proposition that stable hairpin structures involving the RNA termini are present in infectious influenza virus particles. In addition, we find that the viral RNA in infected cells is predominantly circular during the phase when message RNA synthesis occurs. We therefore suggest that these

structures serve as cis regulatory elements for the transcription of influenza virus RNAs and possibly also for the packaging of these RNAs into virions.

B. MATERIALS AND METHODS

1. CROSSLINKING OF INFLUENZA VIRUS RNA IN PURIFIED VIRIONS AND IN INFLUENZA VIRUS INFECTED CELLS

For crosslinking of virion RNA, sucrose gradient-purified influenza virus A/PR/8/34 was incubated with 12 ug/ml of AMT (4'-aminomethyltrioxsalen, Calbiochem) for one minute and then irradiated with a 366 nm ultraviolet light (Ultroviolet Light Products, Inc.) for one hour at a dosage of approximately 1000 $J/m^2/sec$. The drug was replenished twice at 20 minute intervals. At the end of the crosslinking reaction, the solution was made 1% in SDS (sodium dodecyl sulfate) and digested with 200 ug/ml of proteinase K (Boehringer Mannheim Biochemicals) at 37^oC for 3 hours. Viral RNA was purified by extraction with phenol and chloroform as described previously (chapter III of thesis).

Intracellular RNA from infected cells was crosslinked directly in the tissue culture dish as described above. At the end of the crosslinking reaction, the cells (approximately 4×10^7 per sample) were lysed using a solution containing 1% SDS, 100 mM NaCl, 10 mM Tris-HCl, 1mM EDTA, pH 7.4, 200 ug/ml of tRNA, and 200 ug/ml proteinase K. After incubation at 37^oC for 2 hours, the lysates were filtered through a thick layer of sterile glass wool packed in a 10 ml disposable syringe. Viscous cell DNA was trapped in the glass wool and the RNA molecules were recovered in the filtrates (details of the latter method are to be published elsewhere). RNA was then purified by phenol and chloroform extractions.

2. ELECTRON MICROSCOPY OF CROSSLINKED VIRAL RNA

The procedures for denaturation of crosslinked nucleic acid and for spreading were described previously (Hsu, 1985). Samples were examined in a Zeiss 10 CA electron microscope.

3. NORTHERN TRANSFER OF CROSSLINKED RNAS

Virions of the A/PR/8/34 strain were treated with psoralen and the RNA was extracted as described. Psoralen-treated and control RNAs (500 ng) were denatured by treatment at 50^{o}C for 30 minutes with 1 M glyoxal, 50% dimethyl sulfoxide, 10 mM sodium phosphate buffer pH 7.0 and 0.1 mM EDTA. Samples were electrophoresed at 80 V for 64 hours on a 4% polyacrylamide gel containing 7.7M urea, 90 mM tris-borate pH 8.3 and 2.5 mM EDTA. The gel was then treated in 3 mM NaOH for 15 minutes at room temperature and then 15 minutes in chilled 10 mM tris, 5 mM sodium acetate, 0.5 mM EDTA, pH 7.8 (TAE) buffer. The RNA was then electrophoretically transferred to a Zeta-Bind (AMF Cuno Laboratories) nylon membrane using a Trans-Blot Cell apparatus (Bio-Rad Laboratories) filled with TAE buffer. The membrane was dried in an 80^{o}C vacuum oven and then treated with pre-hybridization buffer as described in the Bio-Rad protocol.

Gene-specific probes were prepared by nick translation of gel-purified insert DNAs of the cloned A/WSN/33 strain NS, HA or NA genes following standard protocols. The plasmid containing influenza-specific DNA were prepared by reverse transcription of A/WSN/33 viral RNA and insertion of the cDNAs into pSP65 or pGEM-3 plasmids by standard procedures (chapter IV of thesis). The membrane was probed first using an NS gene and then leaving the first probe still hybridized, an HA probe was employed for hybridization. Finally, without stripping off the old probes, an NA specific probe was used for analysis of the membranes.

4. DETERMINING THE SIZE OF THE RNA PANHANDLE BY NUCLEASE S1 PROTECTION OF TERMINAL PROBES.

A probe specific for the 5' terminus of the NS vRNA was prepared by reverse transcription of A/PR/8/34 vRNA using a specific oligonucleotide primer complementary to the NS gene at base 820 to 834 generating a run-off product of 71 nucleotides in length (bases 820 to 890). Specifically, 100 ng primer plus 2 to 5 ug vRNA in 9 ul total volume were denatured by heating in boiling water and slowly cooled to room temperature. The remaining components were added to final concentrations: 50 mM tris-HCl pH 8.3, 60 mM NaCl, 20 mM DTT, 12 mM $MgCl_2$, 0.5 mM of each dCTP, dGTP and dTTP and 10 uM dATP. 40 uCi alpha-^{32}P-dATP (New England Nuclear) was added, and the reaction was begun by addition of 25 U of reverse transcriptase (Molecular Genetics Resources, Inc.). The mixture was incubated at 25^{o}C for 30 minutes and chased for another 30 minutes following addition of dATP to 0.4 mM. The reaction product was gel purified on a 4% polyacrylamide gel containg 7 M urea.

A probe for the 3' end of the NS vRNA was generated by SP6 RNA polymerase (Promega Biotec Corp.) using a plasmid containing a cDNA insert of the NS gene of A/PR/8/34 virus downstream from the SP6 promoter (chapter IV of thesis). The template plasmid was digested with HindIII which cleaved after base 52 of the NS gene. The run-off transcription product of the polymerase reaction (Melton, et. al., 1984) thus contained 19 nucleotides derived from the vector and 52 nucleotides complementary to the 3' end of the viral RNA. The transcription product unexpectedly yielded two major species of RNA, a 71 and a 63 nucleotide molecule, which is most likely the result of differential termination.

The 5' and 3' specific probes were each hybridized to 1 ug of RNA

obtained either from psoralen-treated virions or control virions. Specifically, the probe and 1 ug vRNA in 25 ul containing 50% formamide, 40 mM Pipes pH 6.7, 0.4 M NaCl and 1 mM EDTA were denatured at 85°C for 5 minutes and incubated overnight at 42°C following the protocol of Melton, et. al. (1984). The hybridization mix was aliquoted to several tubes and 300 ul of buffer was added containing 300 mM NaCl, 30 mM sodium acetate pH 5.0, 5 mM $ZnSO_4$ and nuclease S1 (Sigma Chemical Company) at concentrations varying from 20 U/ml to 1000 U/ml. The digestion mix was incubated at 37°C for 60 minutes. Ten micrograms of tRNA were added to each tube and the sample was ethanol precipitated. The pellet was resuspended in a dye-mix containing 80% formamide and denatured by heating in boiling water. The sample was then electrophoresed on an 8% polyacrylamide sequencing gel containing 7.7 M urea. A set of sequencing reactions was also loaded adjacent to the lane so that the molecular weights of the resultant bands could be determined.

5. STRAND SPECIFIC PROBES FOR NORTHERN ANALYSIS

A probe complementary to the viral sense RNA of the NS gene was prepared by synthesis of a radioactive RNA derived from a plasmid insert of the A/WSN/33 cDNA cloned between the SP6 and T7 promoters of pGEM-3 (Promega Biotec). Hybridizations were performed using the standard procedures.

C. RESULTS

1. OBSERVATION IN THE ELECTRON MICROSCOPE OF CIRCULAR INFLUENZA VIRION RNA CROSSLINKED WITH AMT

The secondary structure of influenza virus RNAs in viral particles was investigated by crosslinking the RNA with AMT (4'-aminomethyltrioxsalen). Viral RNA extracted from AMT crosslinked virus particles was denatured using glyoxal and examined in the electron microscope. As shown in Fig. 18 about

30-50% of the viral RNA was found to have circular conformation. No circular RNA was observed in the control uncrosslinked viral RNA. Under the crosslinking conditions used, influenza virus morphology remained intact as judged by negative staining (data not shown). The histogram of the length distribution of the 325 circular RNA molecules examined is shown in Fig.19. Single strand fd phage DNA present in the same spreading solution was used as internal length standard. Major peaks were found around 1, 1.5, 1.7 and 2 kb respectively. Less abundant RNA species of 2.1, 2.3, 2.5 and 2.7 kb length were also observed. Because of the limitation in resolution of measuring the lengths of small RNA molecules in the electron microscope (standard deviation approximately 10-15%) the two smallest RNA subunits with lengths of 890 and 1027 nucleotides can not be well differentiated by electron microscopic analysis. The peak at 1 kb is broad and probably contains the two smallest RNA subunits, NS and M, of influenza virus. Similarly the peak at 1.5 kb probably consists of two RNAs which represent the NP and NA segments. The length distribution of the circles with peaks at these lower molecular weights and at 1.7 kb, at 2.1 kb and around 2.5 kb therefore suggests that all RNA subunits in the virus particles can exist in circular form. The lower abundance of the larger RNA subunits in Fig. 19 most likely reflects the lower distribution of these subunits in influenza virus particles (Palese and Schulman, 1977).

2. ELECTROPHORETIC EVIDENCE OF CIRCULAR INFLUENZA VIRION RNA

In order to corroborate the observation of circular forms of influenza virus RNAs found in the electron microscope we analyzed the crosslinked RNA preparation by denaturation gel electrophoresis. As seen in Fig. 20 the crosslinked RNA shows two bands when probed with nick-translated DNA

Figure 18. ELECTRON MICROGRAPH OF AMT CROSSLINKED INFLUENZA VIRION RNA.

Purified influenza virus was treated with the crosslinking reagent AMT as described in Materials and Methods. After isolation, the virion RNA was denatured in glyoxal and mounted on electron microscope grids as described previously (Hsu, 1985). Circular RNA molecules ranging from 900 to 2300 nucleotides are present in this micrograph. The bar represents 0.5 um.

Figure 19. HISTOGRAM OF THE LENGTH DISTRIBUTION OF CIRCULAR INFLUENZA VIRION RNA AS OBSERVED IN THE ELECTRON MICROSCOPE.

Following experiments described in Fig. 18, a total of 325 molecules were measured. Circular single stranded fd phage DNA was used as internal length standard. Lengths of molecules were measured using the Sigmascan (Jandel Scientific, California) computer program, and the result was plotted using the Sigmaplot (Jandel Scientific) program.

Figure 18

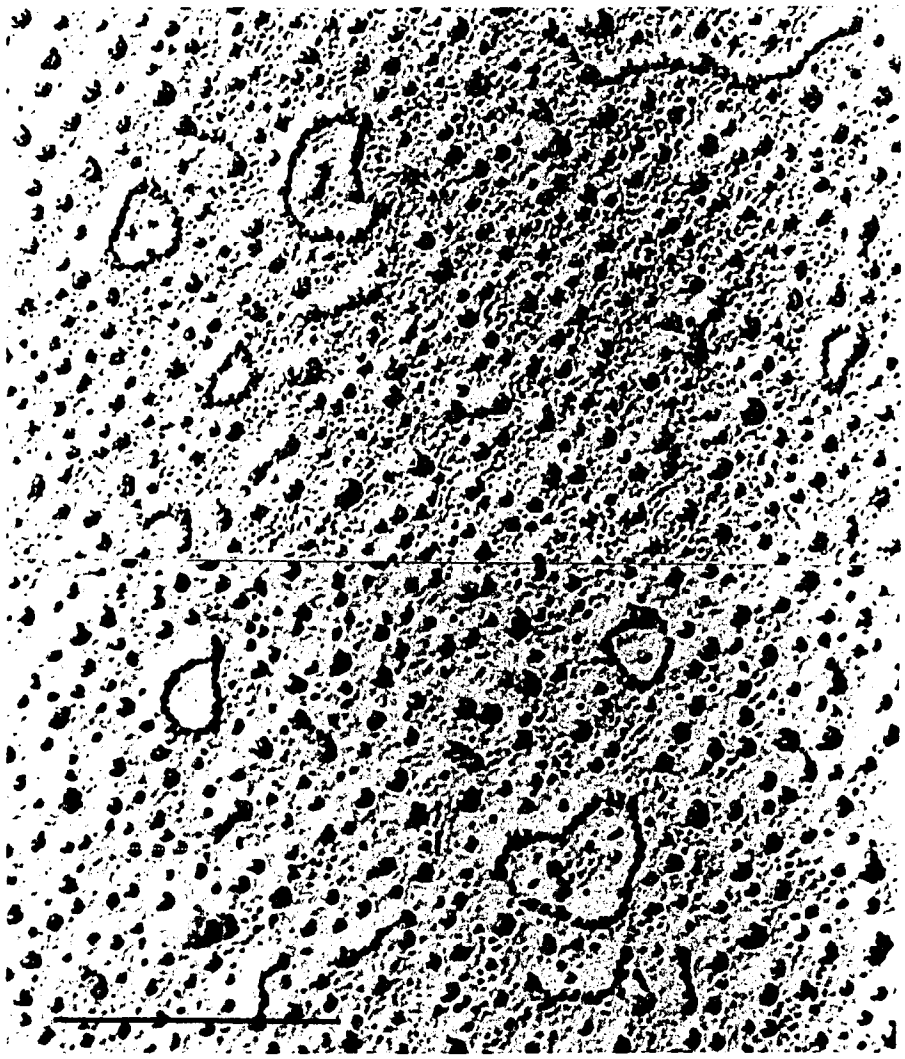
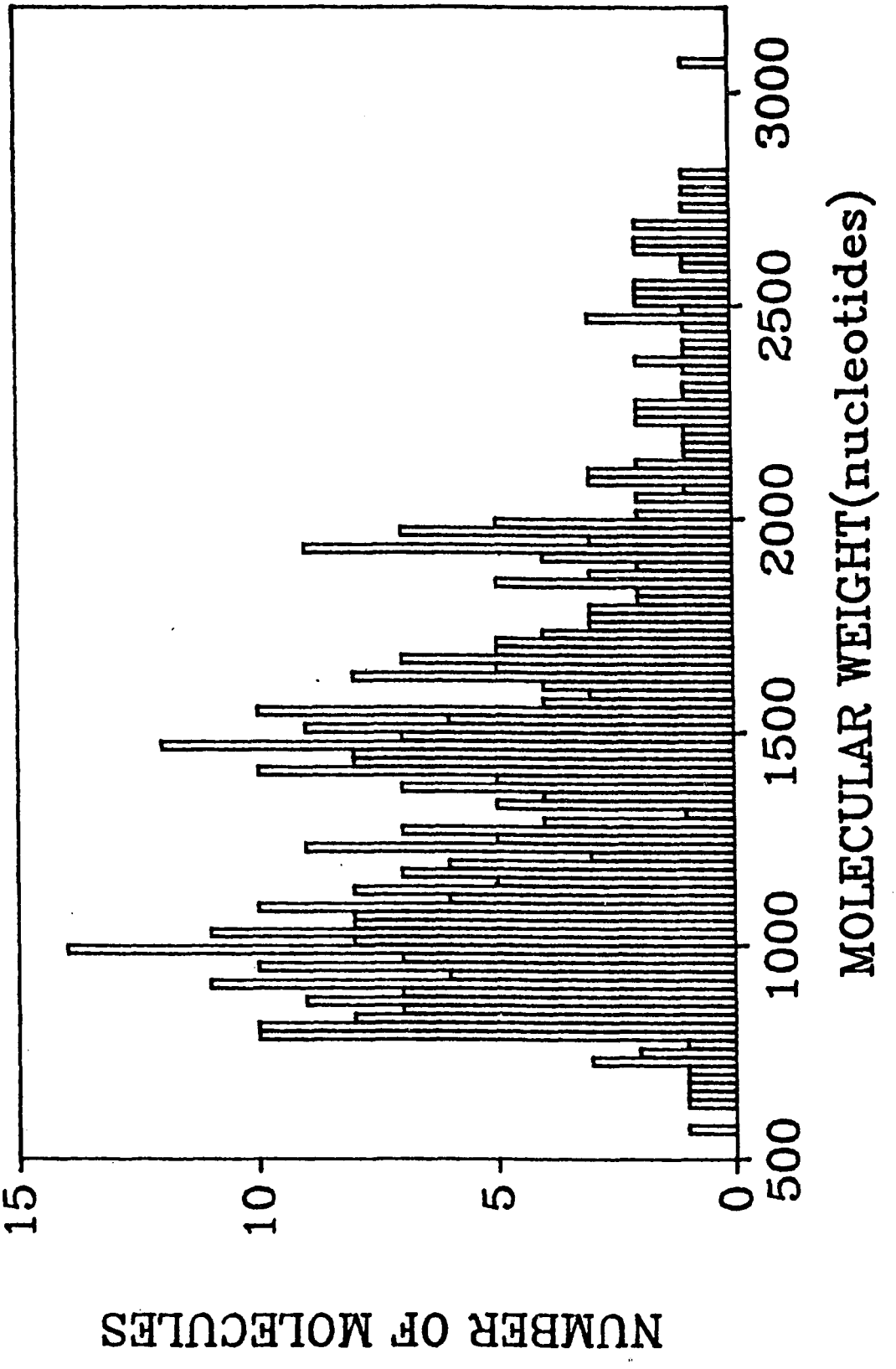


Figure 19


NUMBER OF MOLECULES
15
10
5
0
500
1000
1500
2000
2500
3000
MOLECULAR WEIGHT(nucleotides)

specific for either the NS, NP or HA gene. In each case one of the bands co-migrates with the RNA of the control preparation while the other band has a much slower gel mobility. Since circular RNA is known to have a slower gel mobility in denaturing gels than the linear form of the same molecular weight (Kos, et. al., 1986), these data are consistent with the electron microscopic observation that there are circular forms of influenza virus RNAs following the crosslinking with AMT. In addition, the migration of the upper bands (the circular RNAs) are consistent with the sizes of the circles. However, in comparison to the electron microscopic data, most of the RNA, 60 to 75%, is in circular conformation.

3. NUCLEASE S1 MAPPING OF PANHANDLE SEQUENCE IN THE CIRCULAR INFLUENZA VIRUS RNA

The observation of circular RNA molecules in the electron microscope suggests that inside the virus particles viral RNA segments exist in circular conformation as the result of the complementarity of the terminal RNA sequences. If this is indeed the case, then the terminal sequences in the crosslinked circular RNA should not be available for hybridization with exogenous DNA or RNA probes specific for either of the ends of the viral RNA (Fig. 21a). A probe of 71 nuleotides complementary to the 5' end of the NS RNA segment was synthesized by reverse transcription of NS viral RNA, purified by gel electrophoresis and hybridized to the crosslinked viral RNA. Digestion of the hybrid between the probe and the circular RNA using single strand specific S1 nuclease generated a shortened probe about 56 nucleotides long (Fig. 21b). This result thus provides strong evidence that the circular conformation of RNA in the virus particles is the consequence of a 15 base pair long panhandle

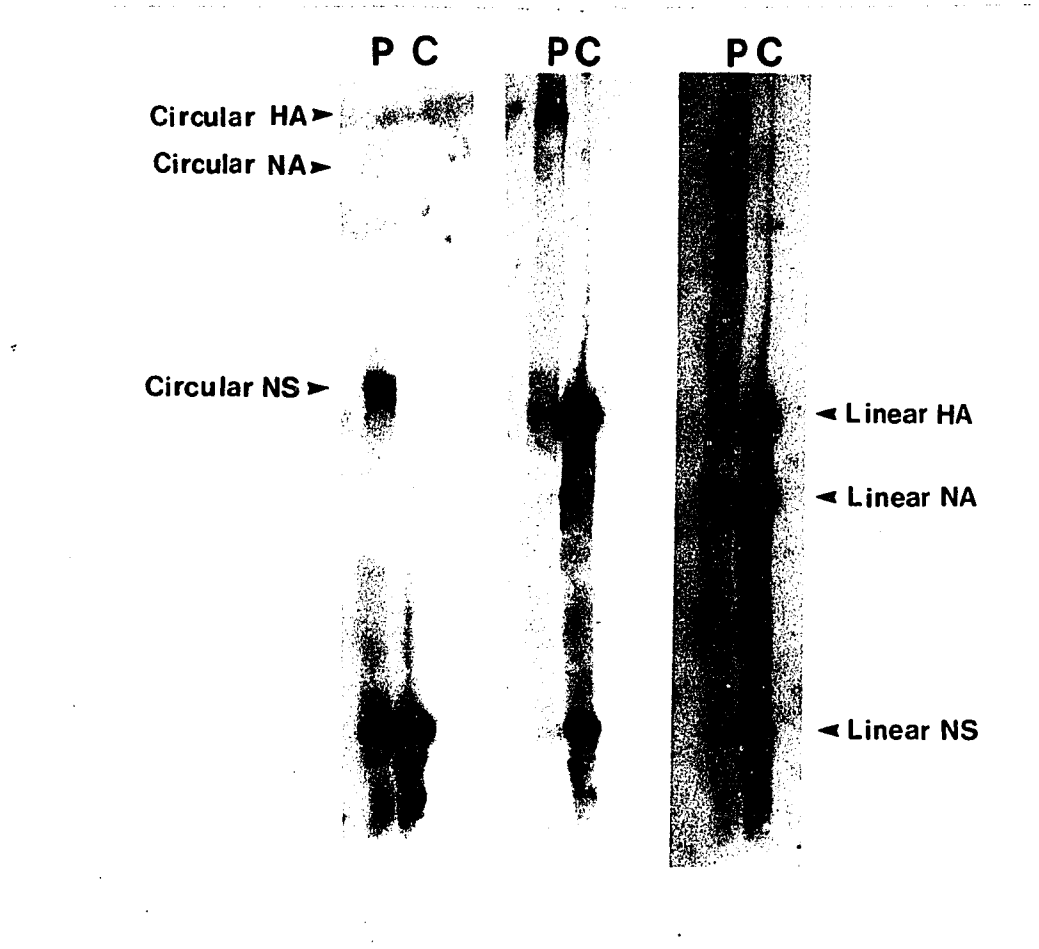


FIGURE 20

RETARDED GEL MIGRATION OF CROSS-LINKED VIRAL RNAS.

RNA extracted from virions treated with the psoralen derivative (P) or RNA from control virions (C) was glyoxalated and then electrophoresed on a 4% polyacrylamide gel containing 7.7 M urea. Following electro-transfer to a nylon membrane, the membrane was sequentially probed with nick-translated DNA inserts of the NS (left), HA (center) and NA (right) genes without stripping off the previous probe.

FIGURE 21

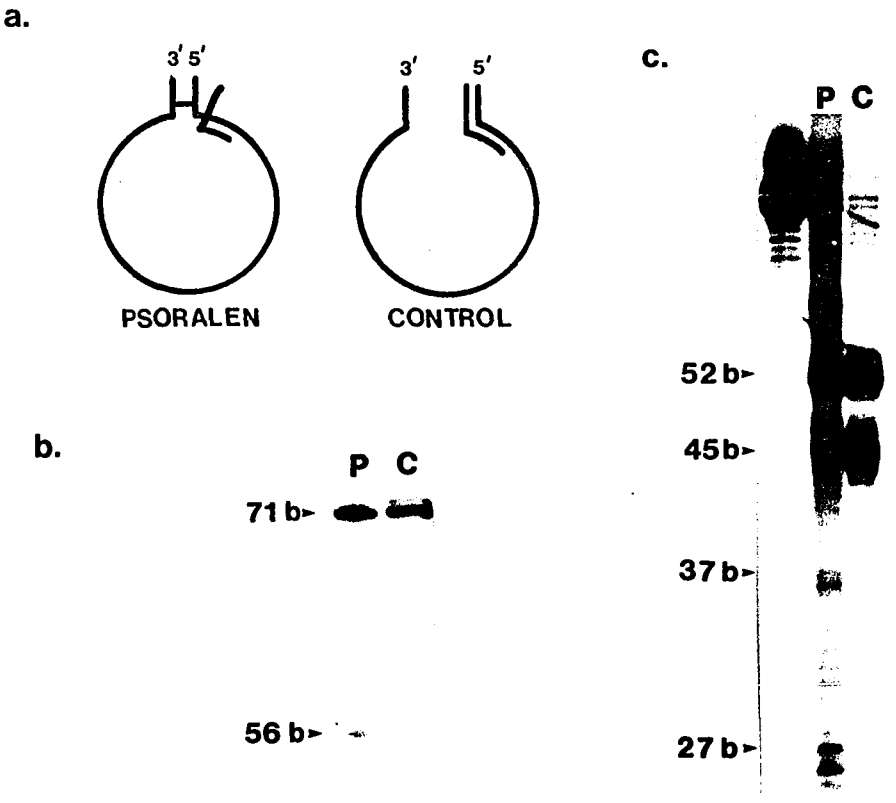
DETERMINATION OF THE SIZE OF THE PANHANDLES BY NUCLEASE S1 PROTECTION OF 5' AND 3' SPECIFIC PROBES.

a. The strategy for determining the number of base pairs in the panhandle is depicted. The psoralen-treated vRNA contains a covalently closed panhandle and will not be opened by the branch migration of a probe for the vRNA terminus, but the control vRNA will. Nuclease S1 digestion will remove probe which is not base-paired, and thus electrophoresis will resolve the difference in lengths.

b. A probe complementary to the 5' terminal 71 nucleotides of the NS gene viral RNA was prepared by reverse transcription using a specific primer. Following hybridization to RNA from psoralen-treated (P) or control (C) virions the sample was digested with nuclease S1 and electrophoresed on a 8% polyacrylamide gel containing 7.7 M urea. Sizes of the resultant DNAs were determined by an adjacent sequencing ladder (not shown).

c. A 71 base RNA was derived from a plasmid with the NS gene located downstream of an SP6 RNA polymerase promoter. This 71 b RNA product had 52 nucleotides complementary to the 3' terminus of the NS gene viral RNA. The first lane shows the probe prior to hybridization to psoralen-treated or control vRNAs. The probe was hybridized to psoralen-treated (lane 2) or control (lane 3) vRNAs and treated as in (b).

Figure 21



structure. The intact 71 nucleotide band is also present in the crosslinked sample after nuclease S1 digestion. This could be due in part to undigested probe. Also, the linear RNA would form a more stable hybrid with the probe than would the circular RNA. Therefore, even though the linear RNA was shown to be present in lower abundance in the crosslinked sample (see Fig. 20), it may outcompete the circular RNA for the radioactive probe and thus produce a 71 nucleotide band.

In the S1 digestion experiments we found that the short RNA- DNA duplex is amenable to S1 nicking. Therefore the optimum concentration of S1 nuclease was determined by titration using a fixed amount of RNA-DNA duplex. At the optimum concentration of S1 nuclease many of the excess probe molecules remained undigested. Therefore, the ratio of the 71 nucleotide band to the 56 nucleotide band as seen in Fig. 21b does not reflect the relative amount of linear and circular conformation in the crosslinked RNA population.

The results presented above confirm the presence of a 15 base region at the 5' terminus of the NS RNA which is protected from hybridization with labelled probe. To confirm the presence of the crosslinked panhandle structure at the 3' terminus a probe of 71 nucleotide length (with 52 nucleotides complementary to the 3' end of the NS virion RNA) was synthesized from the NS gene cloned in an SP6 vector. To synthesize the probe the plasmid containing NS gene was cut with HindIII endonuclease and the linearized DNA transcribed with SP6 polymerase. In repeated experiments synthesized probe was found to contain not only full length, 71 nucleotides long RNA, but also an RNA species 7-8 nucleotides shorter than the full length probe (Fig. 21c, lane 1). This is most likely due to the fact that DNA cut by HindIII has a 5' overhang of four bases and SP6 polymerase sometimes terminates before

reaching the end of the DNA (Schenborn and Mierendorf, 1985). When the probe was hybridized with the control uncrosslinked viral RNA and digested with S1 nuclease two prominent size classes of S1 resistant probe were observed. The 52 nucleotides band and the bands around 45 nucleotides were derived from heteroduplexes between the linear viral RNA and full length probe and shortened probe, respectively. In contrast, in the S1 nuclease experiments using crosslinked viral RNA four promiment classes of S1 resistant RNA species were observed. Two of the four S1 resistant RNA species comigrate with those in the control samples. They represent the probes that hybridized with linear viral RNA present in the sample. In addition to these two classes, S1 resistant RNA molecules 36-37 and 26-27 nucleotides long were observed in the crosslinked samples. These correspond to the full length and shortened probes that hybridized to the circular influenza viral RNA with crosslinks in the panhandle. This result is consistent with the data obtained using the 5' specific probe and suggests that RNA panhandle structures exist inside influenza virus particles. There is a faint band in the control lane corresponding to 37 nucleotides, but since it is present in much lower abundance than in the crosslinked sample, we believe this results from a low amount of panhandle forms present in the control.

We also examined the panhandle structure in the HA segment using the S1 nuclease digestion technique. A DNA probe of 75 nucleotides complementary to the 5' end of the HA viral RNA was prepared and used in hybridization experiments with crosslinked virion RNA. Nuclease S1 resistant bands with length ranging from 53 to 59 nucleotides were observed confirming the existence of a panhandle structure with the hemagglutinin specific RNA segment (data not shown).

4. THE OCCURRENCE OF CIRCULAR GENOME RNAS IN INFECTED CELLS

Influenza virus infected MDCK cells were treated with saturating amounts of AMT, total RNA was extracted and subjected to northern gel analysis as was done for purified virion RNA. Strand specific probes were generated using SP6 polymerase so that only genomic RNA would be detected. We found that the viral-sense RNA was predominantly in the circular conformation early in infection (2 and 4 hours post infection) and fewer than 50% circular structures were observed late in infection (8.5 hpi) (Fig. 22).

D. DISCUSSION

Using psoralen crosslinking techniques we have shown that the negative strands of influenza viral RNA are organized as circular structures inside the virus particles and in the infected cells during lytic infection. The S1 nuclease analysis of the psoralen crosslinked viral RNA strongly suggests that the circular conformation of influenza viral RNA is the result of a terminal panhandle resulting from sequence complementarity between the two ends of each of the viral RNA segments. Inspection of the sequences of the influenza virus RNA show that the sequences at the termini of each of the eight influenza viral RNAs are complementary and are highly conserved (Fig. 23). Twelve of the fifteen nucleotides at the 3' ends are common among all the eight segments of PR8 virus. At the 5' ends thirteen out of sixteen are conserved (Skehel and Hay, 1978; Robertson, 1979; Desselberger, et. al., 1980). The other three nucleotides at each end of an RNA are segment specific and since they are complementary, they may also contribute to the stability of the panhandle. Sequence analysis of different influenza viral RNAs suggests that conservation of the panhandle sequence is necessary for viability. For example,

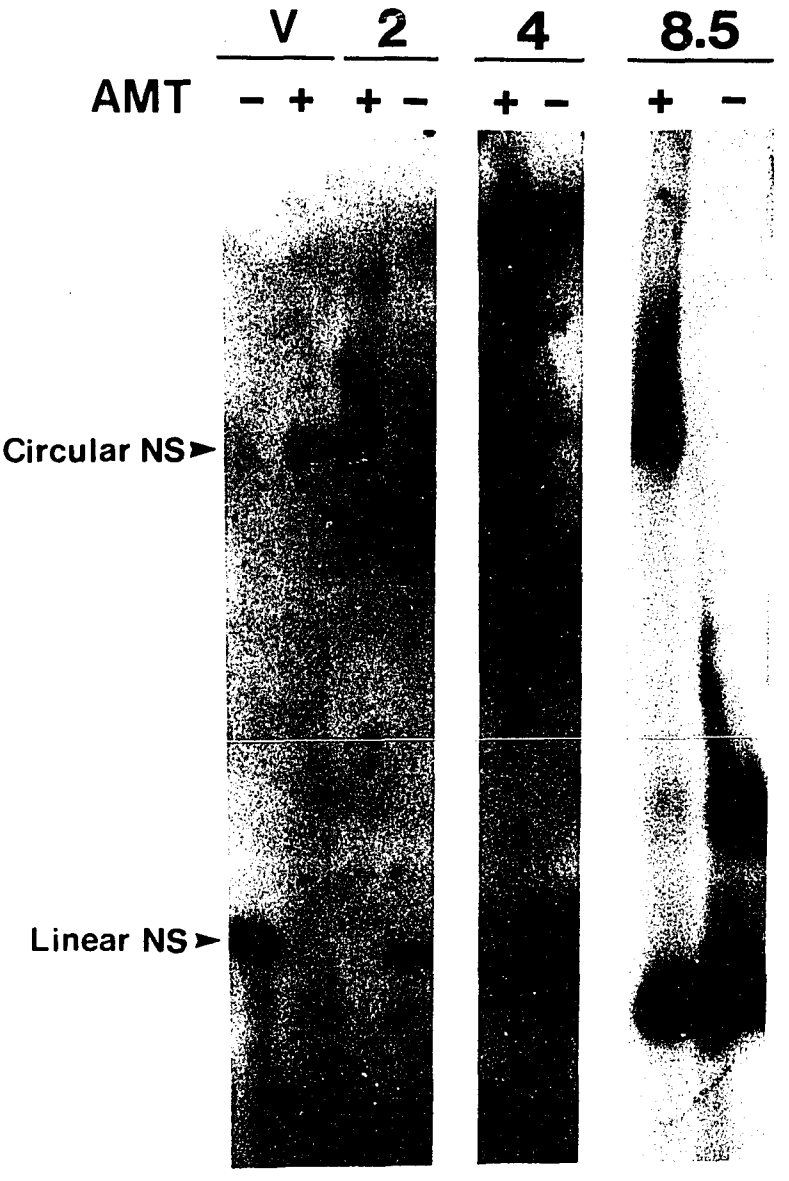
FIGURE 22

DETECTION OF CIRCULAR VIRAL RNAS IN INFECTED CELLS.

MDCK cells were infected with influenza A/WSN/33 virus and at 2, 4, and 8.5 hours post infection cells were treated with the psoralen cross-linking reagent. The RNA was extracted and subjected to electrophoresis on a 4% polyacrylamide gel containing 7.7 M urea. Following electro-transfer to a nylon membrane, the membrane was probed using an SP6-derived RNA complementary to the virion sense RNA of the NS gene. RNA samples from purified virus (V) or from cells 2, 4 or 8.5 h.p.i. which were untreated (-) psoralen treated (+) were examined.

The 8.5 h.p.i. samples were run on a separate gel.

Figure 22



in the defective interfering particles of influenza virus the sequences at the two termini of the viral RNA are preserved (Jennings, et. al., 1983; Nayak, et. al., 1982) in contrast to the VSV DI particles which conserve only the 5' end sequence of the viral RNA (Keene, et. al., 1979). Additionally, evolutionary studies of influenza virus variants showed that the sequences at the termini of viral RNA are invariant despite frequent mutations in viral RNA (Buonagurio, et. al., 1986b). These facts argue for an important role of the panhandle sequences in the replication cycle of influenza viruses.

Inverted terminal repeat sequences have been found in the genome of other negative strand RNA viruses including arenaviruses (Auperin, et. al., 1984) and bunyaviruses (Obijeski, et. al., 1980; Cabradilla, et. al., 1983). The termini of the S RNA of the Pichinde strain of arenavirus is predicted to form a 23 base pair panhandle with 13 of these being very stable G-C base pairs (Auperin, et. al.,1984) and the S segments of two bunyaviruses, La Crosse and Snow Shoe Hare viruses, have 25 of the terminal 26 to 28 nucleotides base-pairing with eight G-C pairs (Obijeski, et. al., 1980; Cabradilla, et. al., 1983). The panhandle of arenavirus vRNAs is stable enough that circular viral RNAs were observed in the electron microscope (Vezza, et. al., 1978). For both arenaviruses (Palmer, et. al., 1977) and bunyaviruses (Petterson and von Bonsdorff, 1975) circular nucleocapsids were observed by electron microscopy. In contrast to these viruses, the influenza A virus NS gene contains only 14 base-pairs (counting G-U mismatches as base-pairs) in the terminal 15 to 16 nucleotides, with six G-C pairs (Desselberger, et. al., 1980). Furthermore, we only observe the terminal panhandle in the ribonucleoprotein (RNP) complex of influenza viruses and not with purified RNA as in the case of the arenavirus RNAs. We

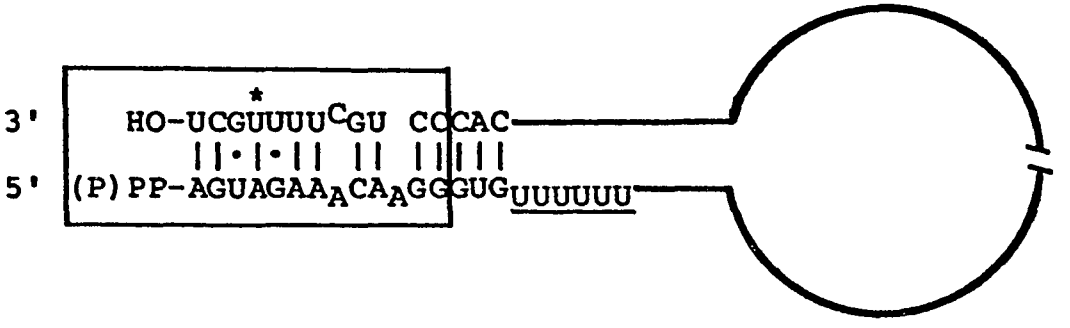


FIGURE 23

MODEL FOR THE STRUCTURE OF THE PANHANDLE AND ITS ROLE IN TRANSCRIPTION

The predicted panhandle sequence for the NS gene is diagrammed with the predicted base-pairing indicated with lines. G-U pairs, which are less stable than conventional base pairs, are indicated with dots. The boxed region is common to all eight segments with the exception of the asterisked U at position 4 which is a C in RNA segments 1, 2, 3 and 6. The additional three nucleotide pairs are specific to the NS gene, however, such pairing can be found in the other seven RNA segments. The U_5 to U_6 termination/polyadenylation signal is underlined, and is shown to be immediately adjacent to the panhandle structure.

thus suggest that protein-RNA interactions are required to stabilize the panhandle conformation of influenza virus RNAs. The possible requirement of a protein could indicate a means of regulating whether the panhandle is open or closed.

The present data complement the observations of Compans, et. al. (1972) who observed the influenza virus nucleocapsid as a large RNP double helix with a loop on one end. We would predict that the other end is held closed by the terminal panhandle. In addition, we find that the panhandle is present in the intracellular pool at times when transcription is occurring, and, ostensibly, no viral assembly is occurring. We interpret this observation as indicating that the panhandle of influenza virus RNAs may serve as a regulatory signal in the control of mRNA synthesis.

The control of switch between mRNA and positive strand template synthesis during influenza virus infection is an intriguing question which remains unresolved. The panhandle structure observed in the present study could serve as a regulatory signal for the switch. Because we observed a high level of circular viral RNA during early phase of infection when the majority of viral RNA synthesized is mRNA, the template for the synthesis of viral mRNA is likely to be in the circular conformation. The panhandle structure could function as the initiation signal for mRNA transcription. Alternatively, or in addition, it could serve as the signal for transcription termination (Robertson, et. al., 1981). It is interesting that the U rich transcription termination sequence occurs 36 times in the A/PR/8/34 genome (Lamb, 1983), but only when this sequence directly abuts the panhandle sequence does polyadenylation and termination occur. The presence of the panhandle structure directly behind the termination site of viral mRNA suggests that the termination of transcription

could be the result of the sequestration of the last 15 nucleotides by the panhandle structure.

If the panhandle structure is indeed a signal for the switch between mRNA and genomic RNA synthesis then one would predict that the RNA templates for the genomic RNA synthesis would lack the panhandle structure. In our preliminary crosslinking analysis, a circular form of the positive strand genomic RNA was not observed in the infected cells. However because of the low concentration of positive strand genomic RNA in the infected cells we could not completely rule out the existence of panhandle structures in the positive strand genomic RNAs (data not shown). Further analysis would be required to resolve this issue.

Another possible function of the panhandle structure in influenza virus RNAs is to serve as a signal for packaging of viral RNA in the virus particles. We observed a higher percent of circular vRNA in virus than we did in infected cells late in infection when packaging is beginning to occur, suggesting that circular vRNAs are selectively packaged. Plus sense RNA is not packaged into virus and we were unable to observe any panhandle structures of plus sense RNA in infected cells. This interpretation is also consistent with the conservation of the panhandle sequences in all the defective interfering viruses (Nayak, et. al., 1982) which do not code for any polypeptides but are packaged in mature virus particles. In conclusion it is possible that the panhandle must be present to package the RNP complex, and then upon infection the complex is in the correct conformation for primary transcription to occur.

VII. DISCUSSION

A. EVOLUTION OF INFLUENZA A VIRUSES

The data presented in chapter II indicate that the influenza A virus evolves at a very high rate and along a single lineage. Other studies (Webster, et. al., 1982; Raymond, et. al.; 1986; Saitou and Nei, 1986) have also demonstrated these points. In comparing the epidemiology of influenza A and C viruses, Buonagurio, et. al. (1985) have proposed a model which is diagrammed in figure 24. The influenza A virus tree in this diagram is a tall, slender tree with long internodal distances, and there are no side branches of any significance. In contrast, the influenza C virus was found to be more like other viruses with respect to its evolution (Buonagurio, et. al., 1985; Buonagurio, et. al., 1986a). It shows much shorter internodal distances and many more branchings which persist. It appears that once a successful variant appears in the influenza A virus population, it outcompetes all other viruses and becomes the progenitor of all future isolates. This model then explains the observation of sequential mutants and is thus suggestive that selection is shaping the evolutionary tree. It is most likely that this is immunological selection since point changes accumulate most in antibody binding sites of the hemagglutinin protein (Both, et. al., 1983; Wiley, et. al., 1981; Webster, et. al., 1982). The cumulative point changes in the NS gene then probably are not directly selected but occur due to "hitchhiking" with the selected marker. When a variant HA gene is selected it will carry with it an NS gene which may, or may not, contain mutations. The NS gene mutations then become fixed in the population via this indirect selection acting on the HA.

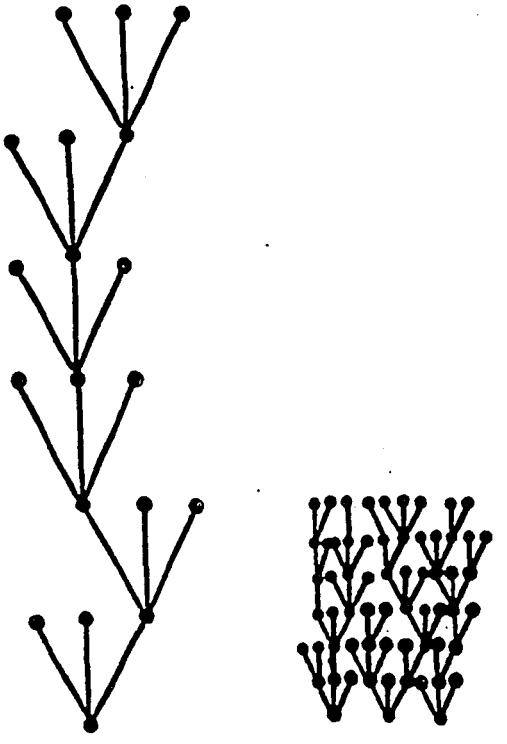


Figure 24. MODEL FOR THE EVOLUTIONARY PATTERN OF INFLUENZA A VIRUS

The influenza A viruses have been shown to evolve along a single primary lineage with only minor side-branching. In comparison, most other viruses have evolutionary patterns which is depicted on the right. This latter model was derived for influenza C viruses (Buonagurio, et. al., 1985). There are many side-branches which persist in the population and thus generate many co-circulating lineages. In addition, the distances between nodes in the tree for influenza A viruses is greater than in the tree on the right. The difference in internodal distances is indicative of the rapid rate of evolution of influenza A viruses. (Modified from Buonagurio, et. al., 1985.)

B. MUTATION RATE ANALYSIS

In chapter III we measure the mutation rates of the influenza A virus NS gene and of the VP1 gene of poliovirus type 1. As is discussed in chapter III, mutation rates have been estimated in a number of systems, but the assay usually involves a phenotype for which the associated nucleotide change is not known, or for which the number of replications over which mutations could occur and accumulate is unknown. These estimates of mutation frequencies for viruses were generally around 10^{-3} to 10^{-4} mutations/nucleotide for many markers in many viruses (Holland, et. al., 1982). The mutation rates we measure, 1.5 X 10^{-5} and less than 2.1 X 10^{-6} mutations per nucleotide per infectious cycle for influenza virus and poliovirus, respectively, are much lower than these previously determined estimates. Since the study was published, two other studies have yielded estimates of mutation rates. In these cases, a phenotype was measured, but what was new was that an attempt was made to know the number of replications. Dougherty and Temin (1986) estimated the mutation rate of a spleen necrosis virus-based retrovirus vector, which uses viral proteins for replication, to be about 0.5% mutations per cycle. This number can be compared to those measured for influenza and polioviruses by considering that the possible target size for mutation in this vector was at most 1850 nucleotides. Therefore the mutation rate is at least 0.5%/1850 nucleotides, or about 2.6 X 10^{-6} mutations per nucleotide per cycle. Since not all 1850 nucleotides are likely to be involved in the phenotype being measured, the mutation rate is probably higher for this retroviral vector. Durbin and Stollar (1986) attempted to measure the mutation rate of sindbis virus using a

phenotype for which they had isolated all possible revertant mutations. They failed to detect any mutants in a single plaque passage, and estimated then that the mutation rate could not be higher than 5 X 10^{-7} mutations per nucleotide per infectious cycle. Therefore, data has been accumulating that the original estimates of mutation rates may have been several orders of magnitude too high and that mutation rates for most viruses may be around 10^{-6} or lower. The mutation rate of influenza A virus is still the highest among these more recent mutation rate determinations (with the possible exception of the retrovirus), thus supporting the role of mutation rate in driving the rapid evolution.

C. EVOLUTIONARY CONCEPTS

There are three major evolutionary concepts which may be useful for considering the data. The first is exemplified by Darwinian selection. By this theory, variants can be positively selected by virtue of superior growth characteristics. The strength of the selective pressures is the critical parameter in driving the evolution. In the case of viruses growing in animals, the principal selective pressure would be exerted by the host's immune system. The immune system can actually promote the evolution, because only those viruses which change and are no longer recognized can grow. The second concept is that of neutral mutation and random drift (Kimura, 1968; Kimura, 1983). Kimura emphasizes mutations which are neutral with respect to phenotype. These are mutations which randomly appear and which randomly become fixed in the population. The mutation rate is the critical parameter along with generation time in driving the evolution rate. It is possible that the polymerases which replicate the viral genome are error-prone and thus generate a high rate of random mutations. These mutations may become fixed by an

entirely random process. A third theory, the quasispecies concept, in effect merges the other two theories into a single framework (Eigen, 1971; Eigen and Schuster, 1977). A species is thought to actually consist of extensive genetic diversity generated by random mutation. Mutations become fixed in the population by stochastic mechanisms (as according to the neutral theory) or by positive selection of variants in the population of the quasispecies.

The quasispecies principle provides a useful framework for interpreting our data of the evolution rate and mutation rate.

It was suggested in chapter III that the higher mutation rate of influenza A virus generates greater diversity in the viral quasispecies. This additional diversity allows very rare beneficial mutants arise. Thus a virus with a high mutation rate may escape the highly selective immunological pressures by having present in its quasispecies the exceedingly rare mutant which is not recognized by the immune system of the host (Fig. 25a). The authors of the quasispecies theory predict that a high mutation rate in combination with a potent selection would generate the highest evolutionary rate. This prediction appears to be borne out by the data of the influenza A virus. In addition, a prediction based on the theory is that if only a single variant in a population can replicate efficiently, the diversity of the species will be temporarily eliminated. All further variation will then be generated from that single variant (Eigen, 1971) (Fig. 25b). This prediction is in strong agreement with the finding of a single lineage of virus observed in chapter II.

D. SEQUENCE ANALYSIS METHODOLOGY

In chapters IV and V new techniques of sequence analysis are described. The first, detection of single mismatches in RNA heteroduplexes, provides a

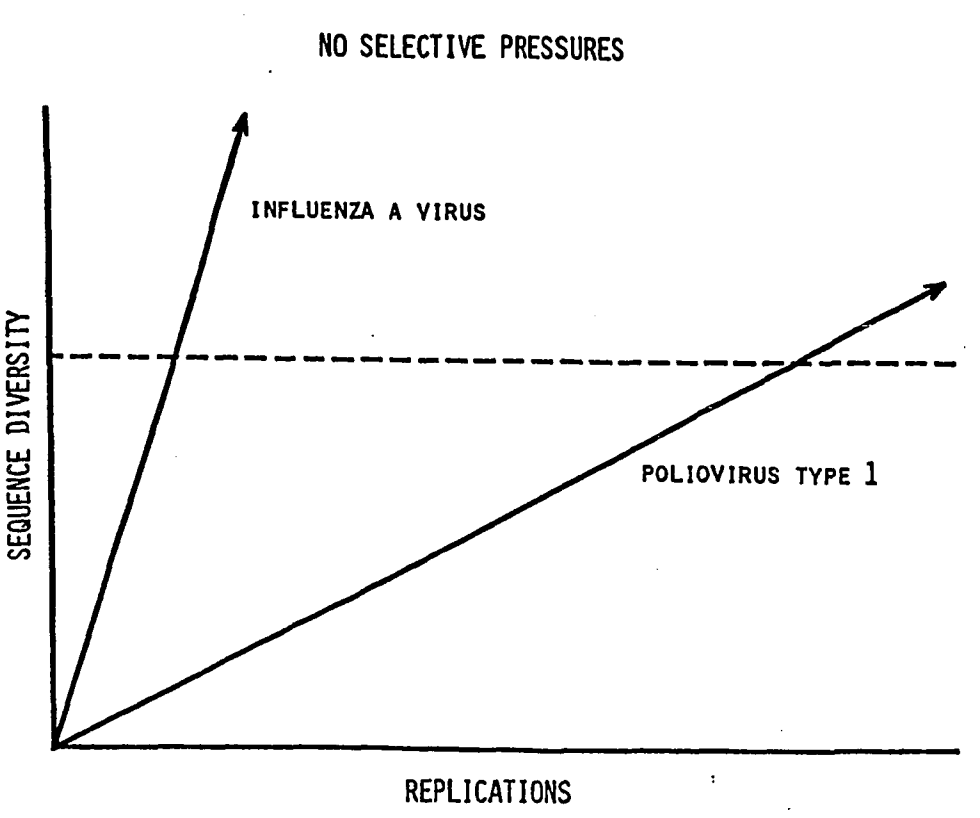


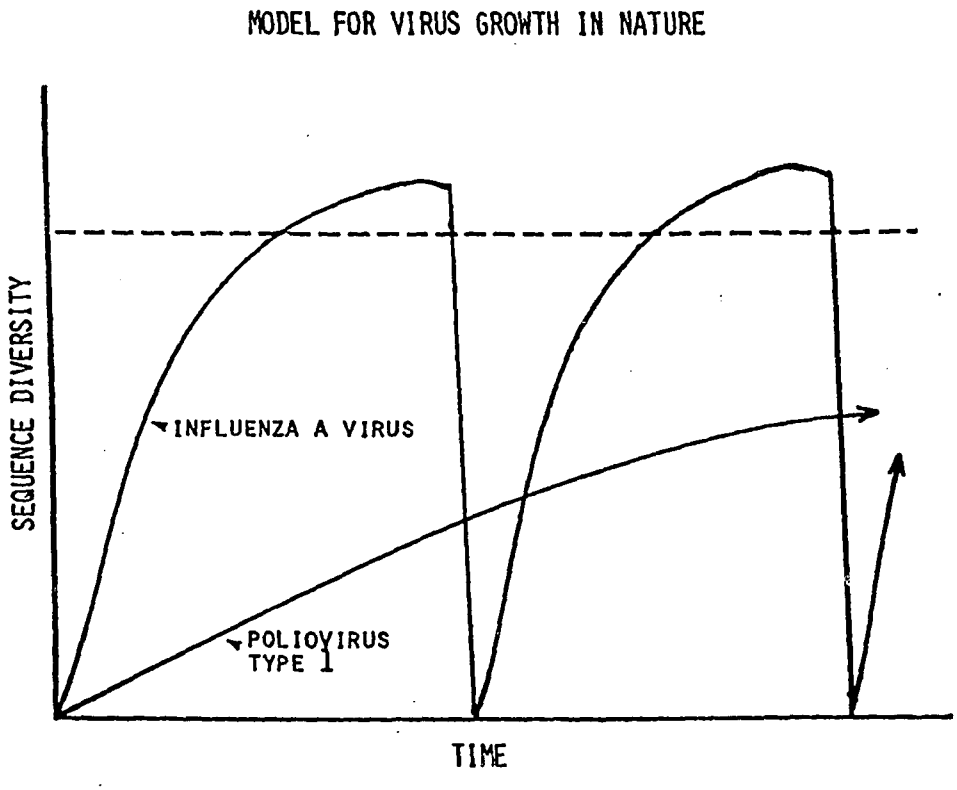
Figure 25. MODEL FOR THE GENERATION OF THE QUASISPECIES AND SELECTION OF THE RARE VARIANT

A. The influenza A virus replicates with a mutation rate which is approximately seven-fold higher than the upper limit of the mutation rate observed with poliovirus type 1. This finding is indicated by the slopes of the lines for each virus. The dashed line indicates a level of sequence diversity in which it becomes probable that a very rare selectable phenotype is present in the quasispecies population. In the case of viruses replicating in an immune environment, this level may reflect the failure of the host's antibodies to recognize the virus. Since this may require several mutations in the same gene, such a mutant may be exceedingly rare. Because of its higher mutation rate, the influenza A virus would be predicted to achieve this level more rapidly than would the poliovirus.

B. The model in this diagram suggests that a high mutation rate of the virus may result in the formation of variants with sequential mutations. Depicted is the introduction of two new viruses into an immunologically naive population. The viruses initially grow unrestricted by the hosts' antibodies. With time, the host population obtains immunity to these viruses, and thus the growth of the viruses is limited. As growth slows, so does the rate of change of the sequence diversity in the viral quasispecies. Since the mutation rate of the influenza A virus is higher than in poliovirus, it may still achieve the level of diversity at which the very rare selectable variant is present in the population (dashed line). This variant may then grow unrestricted and outcompete all other members of the quasispecies. Since this variant would become the progenitor of future isolates, during the time of transition the sequence diversity would drop to zero. A prediction based on this model is that influenza A virus isolates would accumulate point mutations at a high rate along a single lineage, while poliovirus (and other viruses) may have significant variation in nature, but not enough to generate the antigenic drift observed in influenza A viruses.

Figure 25b.

B.


MODEL FOR VIRUS GROWTH IN NATURE
SEQUENCE DIVERSITY
INFLUENZA A VIRUS
POLIOVIRUS TYPE 1
TIME

very convenient method for detecting point mutations in a viral genome. The amount of viral RNA needed will be less than would be needed for sequencing, and the techniques involved are much simpler and require less difficultanalysis. The major drawback is that not all sequences can be analyzed, but only those which are in low melting domains. This limitation, however, is not at all crippling since a high percent of most genomes is low melting domain.

In chapter V a companion project is described in which the SP6 probe used for detecting point mutations can be directly sequenced. The chief advantage of the method is its simplicity. There is no need for primer hybridization prior to sequencing, and the template can be prepared using routine techniques. A similar protocol using the T7 polymerase has now been established (Axelrod and Kramer, 1985). The major application of this method may not be the original sequencing of a newly isolated sequence. There are other methods which are superior in accuracy, and require special single stranded DNA templates (Sanger, et. al., 1977; Messing, et. al., 1981). The SP6 sequencing is far more appropriate when confirming that a special construct which uses the SP6 promoter has the desired sequence. Another comparable procedure for such plasmids is primer directed sequencing using a denatured DNA template (Chen and Seeburg, 1985). This procedure has the advantage that the synthetic product is DNA which is more stable and the mixes used for sequencing are essentially the same as are used for the M13 sequencing system (Messing, et. al., 1981). The major disadvantage of the double stranded DNA sequencing method is that preparation of a primer is required. The SP6 sequencing system is thus a fine complement to analysis of plasmid constructs designed to synthesize RNA.

E. PANHANDLE CONFORMATION OF INFLUENZA VIRAL RNA

In chapter VI we describe observations which indicate that the genomic RNA in influenza virions is held in a circular conformation with a terminal panhandle. In addition, we find that the viral sense RNA in influenza A virus infected cells is predominantly circular at early time points. At later times, the circular form is present, but at less than 50% of the total viral RNA. Early in the infection, mRNA is the primary transcript, and later replicative synthesis occurs. We suggest that the conformation of the viral RNA template regulates the RNA synthesis which occurs on that template. A circular template will make message RNA, while a linear template makes replicative RNA. Inspection of the sequence of the RNA termini reveals several points (see Lamb, 1983, for the sequence of the entire genome of A/PR/8/34 virus). First, the predicted panhandle structure is 15 to 16 base pairs long, which is in strong agreement with our nuclease S1 data. Second, the panhandle sequence is present in all eight RNA segments with 12 or 13 of these being common to all segments and 3 being specific to a segment. Third, we found that the circular form is the predominant vRNA species in infected cells at times when mRNA is the predominant transcription product. Fourth, the U_5 termination signal occurs 36 times in the genome of A/PR/8/34, but only when it is abutting the panhandle sequence does termination and polyadenylation occur. For these reasons, we believe that our data indicate that the transcript from the circular viral RNA template is message RNA and that linear viral RNA is the template for replicative RNA synthesis. Further, since we find the RNA in the virion is circular, we predict that the panhandle acts as a packaging signal as well. Therefore, upon infection of a cell, the viral genome is in the appropriate conformation for synthesizing its primary transcript.

SIGNIFICANCE

Vaccination against influenza A virus fails to confer lasting immunity to the virus due to its antigenic variation. In developing vaccination programs against influenza A virus or any other new viral pathogen one must consider how the variation occurs and what properties of the virus promote the antigenic drift. Sequence analysis reveals that this variation is occurring by a rapid evolution along a single lineage. We have determined the evolution rate of a nonstructural gene of the virus to be about 2×10^{-3} substitutions/nucleotide/year. This rate is less than a third of that of the hemagglutinin gene, the major antigenic protein. We have determined the mutation rate of the NS gene to be 1.5×10^{-5} mutations/nucleotide/infectious cycle and the rate of mutation of the VP1 gene of poliovirus to be less than 2.1×10^{-6} mutations/ nucleotide/infectious cycle. The difference in the mutation rates of the two viruses is significant and thus may indicate a fundamental property of the influenza A virus which permits the high degree of variation observed in nature. The high mutation rate of influenza virus creates increased diversity in the viral quasispecies allowing the selection of the very rare variant. Thus, it is a combination of a high mutation rate and potent selective pressures which drive the antigenic drift of influenza A virus.

We have developed two new methods of sequence analysis which should facilitate future mutation rate analyses. First, a method for detecting single point mutations in RNA molecules is described which uses denaturing gradient gel electrophoresis of RNA heteroduplexes. Second, a simple method for sequencing the RNA probes is described, improving the ease with which one uses these SP6 probes.

Finally, we have demonstrated that the genome RNA in viruses and in

ribonucleoprotein complexes in infected cells is circular in conformation. The termini of each segment base-pair in a panhandle. We show that the time points at which the circular RNA is the predominant form are during phases in which the major transcription product is message RNA. We thus propose a novel regulatory mechanism in which the secondary structure of the RNA template determines the kind of transcript synthesized.

BIBLIOGRAPHY

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

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

Axelrod, V.D. and Kramer, F.R. (1985). Transcription from bacteriophage T7 and SP6 RNA polymerase promoters in the presence of 3'-deoxyribonucleoside 5'triphosphate chain terminators. Biochemistry 24, 5716-5723.

Auperin, D.D., Romanowski, V., Galinski, M. and Bishop, D.H.L. (1984). Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, and ambisense viral S RNA. J. Virol., 52, 897-904.

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

Baltimore, D. (1971). Expression of animal virus genomes. Bacteriol. Rev. 35, 235-241.

Batschelet, E., Domingo, E., and Weissmann, C. (1976). The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. Gene 1, 27-32.

Beaton, A.R. and Krug, R.M. (1986). Transcription antitermination during influenza viral template RNA synthesis repuires the nucleocapsid protein and the absence of a 5' capped end. Proc. Natl. Acad. Sci. U.S.A. 83, 6282-6286.

Bernstein, H.D., Sonenberg, N. and Baltimore, D. (1985). Poliovirus mutant that does not selectively inhibit host cell protein synthesis. Mol. Cell. Biol. 5, 2913-22923.

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

Blondel, B., Crainic, R., Fichot, O., Dufraisse, G., Candrea, A., Diamond, D.,Girard, M., and Horaud, F. (1986). Mutations conferring resistance to neutralization with monoclonal antibodies in type 1 poliovirus can be located outside or inside the antibody-binding site. J. Virol. 57, 81-90.

Blumenthal, T. (1980). QB replicase template specificity: Different templates require different GTP concentrations for initiation. Proc. Natl. Acad. Sci. USA 77, 2601-2605.

Borer, P.N., Dengler, B., Tinoco, I., Jr. and Uhlenbeck, O.C. (1974). Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. 86, 843-853.

Both, G.W., Sleigh, M.J., Cox, N.J. 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., Ulmanen, I. and Krug, R.M. (1983). Molecular model of a eukaryotic transcription complex: Functions and movements of influenza P proteins during capped RNA-primed transcription. Cell 34, 609-618.

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

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

Buonagurio, D.A., Nakada, S., Fitch, W.M. and Palese, P. (1986a). Epidemiology of influenza C virus in man: Multiple evolutionary lineages and low rate of change. Virology 153, 12-21.

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

Butler, E.T. and Chamberlin, M.J. (1982). Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. J. Biol. Chem. 257, 5772-5778.

Cabradilla, C.D., Holloway, B.P. and Obijeski, J.F. (1983). Molecular cloning and sequencing of the La Crosse virus S RNA. Virology, 128, 463-468.

Chen, E.Y. and Seeburg, P.H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4, 165- 170.

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

Compans, R.W., Content, J. & Duesberg, P.H. (1972). Structure of the ribonucleoprotein of influenza virus. J. Virol., 10, 795-800.

Compans, R.W., Klenk, H.D., Caliguiri, L.A. and Choppin, P.W. (1970). Influenza virus proteins. I. Analysis of polypeptides of the virion and identification of spike glycoproteins. Virology 42, 880-889.

Davanloo, P., Rosenburg, A.H., Dunn, J.J. and Studier, F.W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Proc. Natl.

Acad. Sci. USA 81, 2035-2039.

Desselberger, U., Racaniello, V.R., Zazra, J.J. & Palese, P. (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.

Detjen, B.M., St. Angelo, C., Katze, M.G. and Krug, R.M. (1987). The three influenza virus polymerase (P) proteins not associated with viral nucleocapsids in the infected cell are in the form of a complex. J. Virol. 61, 16-22.

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

Drake, J.W. (1969). Comparative rates of spontaneous mutation. Nature 221, 1132.

Duesberg, P. (1969). Distinct subunits of the ribonucleoprotein of influenza virus. J. Mol. Biol., 42, 485-499.

Durbin, R.K. and Stollar, V. (1986). Sequence analysis of the E2 gene of a hyperglycosylated, host restricted mutant of Sindbis virus and estimation of mutation rate from frequency of revertants. Virology 154, 135-143.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 465-523.

Eigen, M. and Schuster, P. (1977). The hypercycle, a principle of natural self-organization, Part A: Emergence of the hypercycle. Naturwissenschaften 64, 541-565.

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

Fisher, S.G. and Lerman, L.S. (1979). Two-dimensional electrophoretic separation of restriction fragments of DNA. Methods in Enzymology 68, 183-191.

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

Gojobori, T. and Yokoyama, S. (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.

Gotoh, O. and Tagashira, Y. (1981). Stabilities of nearest neighbour doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. Biopolymers 20, 1033-1042.

Green, M.R., Maniatis, T. and Melton, D.A. (1983). Human-globin pre-mRNA synthesized in vitro is accurately spliced in Xenopus oocyte nuclei. Cell 32, 681-694.

Hahn, B.H., Shaw, G.M., Taylor, M.E., Redfield, R.R., Markham, P.D., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C., Parks, E.S. and Parks, W.P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science 232, 1548-1553.

Hall, J.D., Coen, D.M., Fisher, B.L., Weisslitz, M., Randall, S., Almy, R.E., Gelep, P.T., and Schaffer, P.A. (1984). Generation of genetic diversity in herpes simplex virus: An antimutator phenotype maps to the DNA polymerase locus. Virology 132, 26-37.

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

Hay, A.J. and Skehel, J.J. (1979). Characterization of influenza virus RNA transcripts synthesized in vitro. J. gen. Virol., 44, 599-608.

Hay, A.J., Skehel, J.J. and Webster, R.G. (1979). Differentiation of the hemagglutinin genes of variant influenza viruses by RNA-RNA hybridization. J. gen. Virol. 45, 245-248.

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

Hogle, J.M., Chow, M. and Filman, D.J. (1985). Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229, 1358-1365.

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

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

Horne, R.W., Waterson, A.P., Wildy, P. and Farnham, A.E. (1960). The structure and composition of the myxoviruses. I. Electron microscope studies of the structure of myxovirus particles by negative staining techniques. Virology 11, 79-98.

Hsu, M.-T. (1985). Electron microscopic evidence for the cruciform structure in intracellular SV40 DNA. Virology, 143, 617-621.

Hsu, M.-T., Parvin, J.D., Gupta, S., Krystal, M. and Palese, P. The genome RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proc. Natl. Acad. Sci. U.S.A. (in press).

Ito, Y. and Joklik, W.K. (1972). Temperature-sensitive mutants of reovirus. II. Anomalous electrophoretic migration of certain hybrid RNA molecules composed of mutant plus strands and wild-type minus strands. Virology 50, 202-208.

Jennings, P.A., Finch, J.T., Winter, G. and Robertson, J.S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence

rearrangements in influenza viral RNA? Cell, 34, 619-627.

Kallenbach, N.R. (1968). Theory of thermal transitions in low molecular weight RNA chains. J. Mol. Biol. 37, 445-466.

Kassavetis, G.A., Butler, E.T., Roulland, D. and Chamberlin, M.J. (1982). Bacteriophage SP6-specific DNA polymerase II. Mapping of SP6 DNA and selective in vitro transcription. J. Biol. Chem. 257, 5779-5788.

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

Keene, J.D., Schubert, M. and Lazzarini, R.A. (1979). Terminal sequences of vesicular stomatitis virus RNA are both complementary and conserved. J. Virol., 32, 167-174.

Kendal, A.P., Noble, G.R., Skehel, J.J. and Dowdle, W.R. (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.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature (London) 217, 624-626.

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

Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecaf, R., Lee, J.J., van der Werf, S., Anderson, C.W., and Wimmer, E. (1981). Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature 291, 547-553.

Koch, R.E. and Drake, J.W. (1973). Lipase-defective bactiophage T4. I. Effects on mutation rates. J. Virol. 11, 35- 40.

Kos, A., Dijkema, R., Arnberg, A.C., van der Meide, P.H. & Schellekens, H. (1986). The hepatitis delta virus possesses a circular RNA. Nature, 323, 558-560.

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

Krystal, M., Ruan, L., Lyles, D., Pavlakis, G. and Palese, P. (1986). Expression of the three influenza virus polymerase proteins in a single cell allows growth complementation of viral mutants. Proc. Natl. Acad. Sci. U.S.A. 83, 2709-2713.

Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984). Normal and mutant human alpha-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell 36, 993-1005.

Kramer, F.R. and Mills, D.R. (1978). RNA sequencing with radioactive chain-terminating ribonucleotides. Proc. Natl. Acad. Sci. USA 75, 5334-5338.

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, Wien, New York, pp. 21-69.

Lamb, R.A. and Choppin, P.W. (1981). Identification of a second protein (M_2) encoded by RNA segment 7 of influenza virus. Virology 112, 729-737.

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

Lamb, R.A., Lai, C.-J. and Choppin, P.W. (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. U.S.A. 78, 4170-4174.

Lerman, L.S., Fischer, S.G., Hurley, I., Silverstein, K. and Lumelsky, N. (1984). Sequence-determined DNA separations. Ann. Rev. Biophys Bioeng. 13, 399-423.

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

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

McGeoch, D. and Kitron, N. (1975). Influenza virion RNA-dependent RNA polymerase: Stimulation by guanosine and related compounds. J. Virol. 15, 686-695.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory , 368-369.

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

Maxam, A.M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560-564.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12, 7035-7056.

Messing, J., Crea, R. and Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucl. Acids Res. 9, 309-321.

Mills, D.R. and Kramer, F.R. (1979). Structure-independent nucleotide sequence

analysis. Proc. Natl. Acad. Sci. USA 76, 2232-2235.

Myers, R.M., Fischer, S.G., Maniatis, T. and Lerman, L.S. (1985). Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. Nucl. Acids Res. 13, 3111-3129.

Myers, R.M., Larin, Z. and Maniatis, T. (1985). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA-DNA duplexes. Science 230, 1242-1246.

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

Nayak, D.P., Sivasubramanian, N., Davis, A.R., Cortini, R. and Sung, J. (1982). Complete sequence analyses show that two defective interfering influenza viral RNAs contain a single internal deletion of a polymerase gene. Proc. Natl. Acad. Sci. U.S.A., 79, 2216-2220.

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

Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Kataoka, Y., Genba, Y., Nakano, Y. and Imura, N. (1982). Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome. Proc. Natl. Acad. Sci. USA 79, 5793-5797.

Nottay, B.K., Kew, O.M., Hatch, M.H., Heyward, J.T., and Obijeski, J.F. (1981). Molecular variation of type 1 vaccine- related and wild polioviruses during replication in humans. Virology 108, 405-423.

Novack, D.F., Casna, N.J., Fischer, S.G. and Ford, J.P. (1986). Detection of single base-pair mismatches in DNA by chemical modification followed by electrophoresis in 15% polyacrylamide gel. Proc. Natl. Acad. Sci. U.S.A. 83, 586-590.

Obijeski, J.F., McCauley, J. and Skehel, J.J. (1980). Nucleotide sequences at the termini of La Crosse virus RNAs. Nuc. Acids Res., 8, 2431-2438.

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

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.

Palese, P. (1977). The genes of influenza virus. Cell 10, 1-10.

Palese, P. (1986). Rapid evolution of human influenza viruses. In: Evolutionary

Processes and Theory (S. Karlin and E. Novo, eds.) Academic Press, New York, pp. 53-68.

Palese, P. and Kingsbury, D.W., eds. Genetics of Influenza Viruses, (Springer-Verlag, Wien, New York) 1983.

Palese, P. and Ritchey, M.B. (1977). Live attenuated influenza virus vaccines. Strains with temperature-sensitive defects in P3 protein and nucleoprotein. Virology 78, 183-191.

Palese, P., Ritchey, M.B. and Schulman, J.L. (1977). P1 and P3 proteins of influenza virus are required for complementary RNA synthesis. J. Virol. 21, 1187-1195.

Palese, P. and Schulman, J.L. (1976). Mapping of the influenza virus genome: Identification of the hemagglutinin and the neuraminidase genes. Proc. Natl. Acad. Sci. USA 73, 2142- 2146.

Palese, P. and Schulman, J.L. (1977). Differences in RNA patterns of influenza A viruses. J. Virol., 17, 876-884.

Palmer, E.L., Obijeski, J.F., Webb, P.A. and Johnson, K.M. (1977). The circular, segmented mucleocapsid of an arenavirus-Tacaribe virus. J. gen. Virol., 36, 541-545.

Parvin, J.D., Moscona, A., Pan, W.T., Leider, J.M. and Palese, P. (1986). The mutation rate of animal viruses: Influenza A virus and poliovirus type 1, J. Virol. 59, 377-383.

Parvin, J.D., Smith, F.I. and Palese, P. (1986). Rapid RNA sequencing using double-stranded template DNA, SP6 polymerase and 3'-deoxy nucleotide triphosphates, DNA 5, 167-171.

Parvin, J.D., Young, J.F. and Palese, P. (1983). Nonsense mutations affecting the lengths of the NS1 nonstructural proteins of influenza A virus isolates, Virology 128, 512-517.

Petterson, R.F. and von Bonsdorff, C.-H. (1975). Ribonucleoproteins of Uukuniemi virus are circular. J. Virol., 15, 386-392.

Plotch, S.J., Bouloy, M., Krug, R.M.. (1979). A unique cap (m^7GppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23, 847-858.

Pons, M.W. (1971). Isolation of influenza virus ribonucleoprotein from infected cells. Demonstration of the presence of negative stranded RNA in viral RNP. Virology 46, 149-160.

Pons, M.W. and Rochavansky, O.M. (1979). Ultraviolet inactivation of influenza virus RNA in vitro and in vivo. Virology, 97, 183-189.

Portner, A., Webster, R.G., and Bean, W.J. (1980). Similar frequencies of

antigenic variants in Sendai, vesicular stomatitis and influenza A viruses. Virology 104, 235-238.

Racaniello, V.R. and Baltimore, D. (1981). Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. Proc. Natl. Acad. Sci. USA 78, 4887-4891.

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.

Raymond, F.L., Caton, A.J., Cox, N.J., Kendal, A.P. and Brownlee, G.G. (1986). The antigenicity and evolution of influenza H1 haemagglutinin, from 1950-1957 and 1977-1983: Two pathways from one gene. Virology 148, 275-287.

Rico-Hesse, R., Pallansch, M.A., Nottay, B.K. and Kew, O. (1986). Natural distribution of wild type 1 poliovirus genotypes. In: Positive Strand RNA Viruses, UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 54. Brinton, M.A. and Rueckert, R., eds. Alan R. Liss, Inc., New York, NY.

Ritchey, M.B., Palese, P. and Schulman, J.L. (1976). Mapping of the influenza virus genome. III. Identification of genes coding for nucleoprotein, membrane protein and nonstructural protein. J. Virol. 20, 307-313.

Robertson, J.S. (1979). 5' and 3' terminal nucleotide sequences of the RNA genome segments of influenza virus. Nuc. Acids Res., 6, 3745-3757.

Robertson, J.S., Schubert, M. and Lazzarini, R.A. (1981). Polyadenylation sites for influenza virus mRNA. J. Virol., 38, 157-163.

Rochovansky, O.M. (1976). RNA synthesis by ribonucleoprotein-polymerase complexes isolated from influenza virus. Virology 73, 327-338.

Saitou, N. and Nei, M. (1986). Polymorphism and evolution of influenza A virus genes. Mol. Biol. Evol. 3, 57-74.

Salts, Y. and Rosen, A. (1971). Neighbor effects in the mutation of ochre triplets in the T4rII gene. Mutation Res. 13, 109-113.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schenborn, E.T. and Mierendorf, R.C. (1985). A novel transcription property of SP6 and T7 RNA polymerases: Dependence on template structure. Nucl. Acids Res. 13, 6223-6236.

Scholtissek, C., W. Rohde, V. von Hoyningen and R. Rott (1978). On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology 87, 13-20.

Schreier, P.H. and Cortese, R. (1979). A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. 129, 169-172.

Schuerch, A.R. and Joklik, W.K. (1973). Temperature-sensitive mutants of reovirus. IV. Evidence that anomalous electrophoretic migration behaviour of certain double-stranded RNA hybrid species is mutant group-specific. Virology 56, 218- 229.

Skehel, J.J. (1972). Polypeptide synthesis in influenza virus-infected cells. Virology 49, 23-36.

Skehel, J.J., R.S. Daniels, A.R. Douglas and D.C. Wiley (1983). Antigenic and amino acid sequence variations in the haemagglutinins of type A influenza viruses recently isolated from human subjects. Bull. WHO 61, 671-676.

Skehel, J.J. and Hay, A.J. (1978). Nucleotide sequences of the 5' termini of influenza virus RNAs and their transcripts. Nuc. Acids Res., 5, 1207-1219.

Smith, F.I., Parvin, J.D. and Palese, P. (1986). Detection of single base substitutions in influenza virus RNA molecules by denaturing gradient gel electrophoresis of RNA-RNA or DNA-RNA heteroduplexes, Virology 150, 55-64.

Sobrino, F., Davila, M., Ortin, J., and Domingo, E. (1983). Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture. Virology 128, 310- 318.

R. Staden. (1978). Further procedures for sequence analysis by computer. Nucl. Acids Res. 5, 1013-1015.

Steinhauer, D.A. and Holland, J.J. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. J. Virol. 57, 219-228.

Tanimura, M., Miyamura, K. and Takeda, N. (1985). Construction of a phylogenetic tree of enterovirus 70. Jpn. J. Genet. 60, 137 (1985).

Ulmanen, I., Broni, B.A., Krug, R.M. (1981). The role of two of the influenza virus core P proteins in recognizing cap 1 structures (m^7GpppNm) on RNAs and in initiation viral RNA transcription. Proc. Natl. Acad. Sci. U.S.A. 78, 7355-7359.

Vezza, A.C., Clewley, J.P., Gard, G.P., Abraham, N.Z., Compans, R.W. and Bishop, D.H.L. (1978). Virion RNA species of the arenaviruses Pichinde, Tacaribe, and Tamiami. J. Virol., 26, 485-497.

Webster, R.G., Laver, W.G., Air, G.M. and Schild, G.C. (1982). Molecular mechanisms of variation in influenza viruses. Nature 296, 115-121.

White, J., Kartenbeck, J. and Helenius, A. (1982). Membrane fusion activity of influenza virus. The EMBO J. 1, 217-222.

Wiley, D.C., Wilson, I.A. and Skehel, J.J. (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., Carlson, S.S. and White, T.J. (1977). Biochemical evolution. Ann. Rev. Biochem. 46, 573-639.

Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981). Structure of the Haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289, 366-373.

Winter, E., Yamamoto, F., Almoguera, C. and Perucho, M. (1985). A method to detect and characterize point mutations in transcribed genes: Amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells. Proc. Natl. Acad. Sci. U.S.A. 82, 7575-7579.

Wrigley, N.G. (1979). Electron microscopy of influenza virus. Brit. Med. Bull. 35, 35-38.

Zarling, D.A. and Temin, H. M. (1976). High spontaneous mutation rate of an avian sarcoma virus. J. Virol. 17, 74-84.

Zimmern, D. and Kaesberg, P. (1978). 3'-Terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc. Natl. Acad. Sci. 75, 4257-4261.