Acta virol. **26** : 295—300, 1982

INFLUENZA VIRUS RNA

GHENDON Y. Z.

Moscow Research Institute for Viral Preparations, 109088 Moscow, U.S.S.R.

Received July 1, 1981; revised December 20, 1981

Summary. — Recent data are presented concerning molecular weights of influenza virus RNA segments, nucleotide sequences of individual segments, their homology in different influenza virus serotypes and changes of RNA sequences observed in the course of antigenic drift.

Key words: *influenza virus*; *RNA segments*; *RNA sequence analysis*

According to recent data influenza A and B virus genomes contain 8 RNA fragments differing in molecular weights and encoded genetic information (Palese, 1977; Racaniello and Palese, 1979). Seven RNA fragments of influenza C virus have been detected so far. Studies of RNA preparations denaturated with glyoxal showed that the total molecular weight (m. w.) of influenza A virus (A/PR/8/34) genome represents 4.9×10^6 daltons (Desselberger and Palese, 1978). The total m. w. of influenza B virus RNA (v RNA) turned out to be slightly higher (5.3×10^6 daltons) than that of influenza A virus (Desselberger and Palese, 1978). Each RNA fragment of influenza B virus appeared a little larger than the corresponding fragment of influenza A vRNA (Racaniello and Palese, 1979). The analysis of DNA copies of influenza vRNA fragments allowed to make rather accurate estimates of the sizes of each RNA fragment ranging from 2390 bases (fragments 1 and 2 of strain A/NT/60/68) to 890 ones (fragment 8) Sleigh *et al.*, 1979). It has been also calculated that nearly all the information present in each genome fragment is translated when the functional proteins are synthesized.

Each of the influenza A, B and C virus vRNA fragments contains a common sequence of 12 nucleotides at its 3′-terminus. The 5′-terminus of each fragment reveals a common sequence of 11-13 nucleotides. In addition, there are three unique sequences and, beginning from nucleotide 17, all the fragments contain 5-7 uridine residues. Of special interest are the data that 5′ and 3′ends of influenza A, B and C vRNA fragments contain small inverted complementary sequences (Skehel and Hay, 1978; Robertson, 1979; Desselberger *et al.* 1980). Unfortunately, the biological significance of common sequences in each vRNA fragment as well as that of the terminal complementary sequences is still unclear.

At present new data have allowed to determine proteins coded for by each vRNA fragment (Palese, 1977; Scholtissek, 1978; Racaniello and Palese, 1979). High molecular weight fragments 1, 2 and 3 were shown to code for

P proteins. Gene 4 is known to code for haemagglutinin, gene 5 in the majority of strains codes for NP protein, gene 6 — for neuraminidase and gene 7 codes for M proteins. Gene 8 codes for two proteins — NS1 and NS2 (Inglis *et al.*, 1979; Lamb and Choppin, 1979). However, the capacity of genetic information contained in gene 8 is insufficient to encode for both proteins. This gene appeared to contain a region of 144-159 nucleotides from which a common part of both mRNAs is transcribed (Inglis *et al.*, 1980; Lamb *et al.*, 1980). Bearing in mind that NS1 and NS2 proteins share no common amino-acid sequences, one can assume that the transcription of both mRNAs from the common region of vRNA is accomplished by a frame shift. Recently it has been reported (Allen *et al.*, 1980; Winter and Fields, 1980) that the gene 7 can also code for two proteins. The first one has a m. w. of about 27,000 daltons (this is apparently the classical M protein); the second protein (m. w. of about 11,000—15,000 daltons) is usually not seen in gels. Similarly to gene 8, the sequences coding for these two proteins are overlapping. Lamb and Choppin (1981) have recently confirmed the existence of the M2 protein which, like the M1 protein (27,000) is encoded by gene 7.

The majority of investigators agree that influenza A and B virus genomes contain 8 RNA fragments. However, sometimes larger as well as smaller RNA fragments are detected in orthomyxovirus virions. Harms *et al.* (1978) revealed an additional nucleic acid fragment in highly-purified fowl plague virions; it had a m. w. of about 1.5×10^6 daltons. That high m. w. fragment contained RNA sequences homologous on one hand to the fragment 2, which codes for one of the P proteins, on other hand to the fragment 7, which codes for M proteins, as well as varying amounts of DNA apparently of cellular origin.

In virions of orthomyxoviruses (especially in populations containing defective virions) several low m. w. RNA fragments were detected. The analysis of these small RNAs showed sequences belonging to the high m. w. fragments 1, 2 and 3, which apparently arose from internal deletion of high m. w. fragments of vRNAs (Davis, and Nayak, 1979; Nakajima *et al.*, 1979; Davis *et al.*, 1980; Ueda *et al.*, 1980).

The development of the methods of RNA investigation (namely, preparation of complementary DNA copies) allowed a comparison of the same gene in different types and strains of influenza virus. Both and Air (1979) studied the sequences of the first 150 nucleotides of the DNA complementary to fragment 7 in a number of influenza A virus strains and showed that the nucleotide sequences of this gene region were rather similar. The first 80 nucleotide sequences in fragment 8 of two influenza A strains differed only in three bases; however, the first 130 nucleotides of this gene in A and B strains had no common sequences at all (Air and Hackett, 1980). Recent analysis of nucleotide sequences of complementary (c) DNA to genes 7 and 8 of a number of influenza A virus strains isolated during the last 43 years, has revealed a high conservatism of these genes. In the course of 23 years only 5-7 out of 230 analysed nucleotides were substituted in the gene 7; alternatively only 12-13 nucleotides were substituted in the gene 8. This

resembles to drift changes in genes coding for haemagglutinin and neuraminidase (Hall and Air, 1981).

Studying the NS gene homology of different influenza A virus strains by hybridization of labelled vRNA fragments with unlabelled cRNA Scholtissek and von Hoyningen-Huene (1980) found that all tested strains could be divided into three groups. Within each group the homology was 90% or more but between the groups it was about 40%. The first group included all tested human influenza A viruses, the avian influenza viruses were divided into two groups. It should be noted, however, that according to Baez *et al.* (1980) who studied the complete nucleotide sequence of the NS gene of two human influenza virus strains and that of fowl plague virus, the nucleotide sequence of the NS gene turned out to be rather similar in all human and avian influenza virus strains that had been studied.

Analysis of the first 200 nucleotides of the gene coding for neuraminidase has shown (Block and Air, 1980) that this gene derived from the same neuraminidase subtype differs in a few bases amounting not more than 20-26 ones. At the same time, comparison of genes coding for N1 and N2 neuraminidases has shown that from the first 200 nucleotides only 100 turned out to be similar. The N-terminal region of N1 nad N2 neuraminidase contained conserved sequences. It is worth mentioning that not all nucleotide changes wer followed by anmino acid changes. These results are of great importance since they show that not all changes of the genome detected by modern methods can result in alteration of amino acid sequences and, correspondingly, in alternation of biological properties of viruses. At present, complete analysis of sequences of 1413 nucleotides of the gene coding for the influenza virus neuraminidase has been carried out (Fields *et al.*, 1981). The authors came to conclusion that the neuraminidase, unlike to haemagglutinin, is oriented with its N-terminus buried into the viral membrane.

A complete analysis of nucleotide sequences of DNA complementary to the gene coding for influenza virus nucleoportein has revealed 1565 nucleotides encoding 498 amino acids (Winter and Fields, 1981).

A complete analysis of 1765–1768 nucleotides of the gene coding for haemagglutinin allowed to detect the regions coding for a hydrophobic leader peptide (16 residues), for JA1 (328–329 residues), for HA2 (221 residue), as well as for arginine which joins the haemagglutinin subunits (Both and Sleigh, 1980; Jou *et al.*, 1980).

Untranslated regions containing 32 nucleotides at 5′end and 48 nucleotides at 3′end were found in the influenza virus gene coding for haemagglutinin (Hiti *et al.*, 1981). The comparison of nucleotide sequences of a gene region coding for HA1 showed that the drift of the haemagglutinin antigenic specificity might occur due to a single base mutation (Jou *et al.*, 1980; Verhoeyen *et al.*, 1980). The comparison of genes coding for haemagglutinins of different influenza virus subtypes showed that nearly 40% of HA2 amino acid sequences and only 20% of that of HA1 was conserved (Gething *et al.*, 1980).

The development of accurate methods of influenza virus genome RNA analysis allowed to carry out experiments on the study of possible origin of pandemc influenza virus strains. By hybridization of separate fragments of labelled vRNA with unlabelled cRNA Scholtissek *et al.* (1978) found that fragments 1, 5, 7 and 8 of strain A/Singapore/1/57 (H2N2) were completely homologous to the corresponding fragments of strain A/FM/1/47 (H1N1); the homology of ther fragments was 24—76%. The comparison of A/Singapore/1/57 (H2N2) and A/Hong Kong/1/68 (H3N2) strains showed that al fragments, with the exception of fragment 4, coding for haemagglutinin, appeared to be completely homologous. The data obtained led to the conclusion that H2N2 and H3N2 pandemic influenza virus strains arose from recombination of a previous pandemic strain with an unknown parent. According to Hiti *et al.* (1981) and Winter *et al.* (1981) who analysed the complete nucleotide sequences of the haemagglutinin gene, the homology between H1(H0) and H2 haemagglutinin of a number of influenza virus serotypes turned out to be significantly greater than between H1(H0) and H3. According to Hiti *et al.* (1981) influenza viruses of H1(H0) and H2 serotypes have only recently diverged one from another.

Using the method of vRNA nucleotide analysis Young *et al.* (1979) revealed differences in several oligonucleotides in the genes of H1N1 influenza virus strains isolated in 1977, the number of these differences having been variabl for different strains. Proceeding from these data and assuming an accumulation of mutation points in the course of the virus evolution, the authors drew a scheme of the generation of H1N1 influenza virus strains isolated in 1977 based on the fact that all strains had had a common ancestor. Recent experiments of Block and Airt (1980) who studied the sequence of the first 200 nucleotides of the neuraminidase coding have shown that appearance and accumulation of mutations in that gene is not always linear by time. Comparing they studied H0N1 strains isolated in 1933, 1934 and 1942; the number and location of mutations in the neuraminidase gene was more similar between the strains from 1933 and 1942 than between those isolated in 1933 and 1934. Apparently, the notion that a greater number of mutations in a gene indicates its later appearance needs further investigation.

Young *et al.* (1979) showed by oligonucleotide analysis that mutation changes of H1N1 influenza virus strains isolated in 1977 had not been restricted only to the genes coding for haemagglutinin and neuraminidase only, but involved other genes as well. Similar data were obtained by Hay *et al.* (1977) who analysed a number of H3N2 influenza virus strains isolated in different years. They applied a method that allowed to detect even point mutations in individual genes of the influenza virus (hybridization of vRNA with cRNA followed by nuclease treatment and polyacrylamide gel electrophoresis). In our studies (Ghendon *et al.*, 1981) of H3N2 influenza virus strains isolated during the epidemic of 1979—80 using the method of Hay *et al.* (1977), we found H3N2 strains that circulated throughout the epidemic but differed in 6-8 homologous genes. All these data indicate that appearance of epidemic influenza virus strains may be accompanied by changes not only in genes coding for haemagglutinin and neuraminidase, but also in other

genes coding for nonglycosylated proteins. Their role in the emergence of highly virulent epidemic influenza strains may be rather significant.

The data presented in this short review have shown that modern investigation of influenza vRNAs allowed a highly accurate analysis of properties represented by individual genes of different virus strains. Further progress in this direction can help to solve many important problems, namely, the cardinal question of the emergence of epidemic influenza strains.

References

Air, G. M., and Hackett, J. A. (1980): Gene 8 of influenza virus: Sequences of cDNA transcribed from the 3′end of viral RNA of influenza A and B strains. *Virology* **103**, 291—298.

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

Baez, M., Taursig, R., Zazra, J. J., Young, F. F., Palese, P., Reifeld, A., and Skalka, A. M. (1980): Complete nucleotide sequence of the influenza A/PR/8/34 virus NS gene and comparison with the A/Udorn/74/ and A/FPV/Rostock/37 strains. *Nucleis Acids Res.* **8**, 5845.

Block, J., and Air, G. M. (1980): Comparative nucleotide sequences at the 3′end of the neuraminidase gene from eleven influenza type A viruses. *Virology* **107**, 50—60.

Both, G. W., and Air, G. M. (1979): Nucleotide sequence coding for the N-terminal region of the matrix protein of influenza virus. *Eur. J. Biochem.* **96**, 363—372.

Both, G. W., and Sleigh, M. J. (1980): Complete nucleotide sequence of the haemagglutinin gene from a human influenza virus of the Hong Kong subtype. *Nucleic Acids Res.* **8**, 2561——2575.

Davis, A. R., and Nayak, D. P. (1979): Sequence relationships among defective interfering influenza viral RNAs. Proc. natn. *Acad. Sci. U.S.A.*, **76**, 3092—3096.

Davis, A. R., Hiti, A. L., and Nayak, D. P. (1980): Influenza defective interfering viral RNA is formed by internal deletion of genomic RNA. *Proc. natn. Acad. Sci. U.S.A.* **77**, 215—219.

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

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

Fields, S., Winter, G., and Brownlee, G. G. (1981): Structure of the neuraminidase gene in human influenza virus A/PR/34. *Nature* (*Lond.*) **290**, 213—217.

Gething, M. J., Bye, J., Skehel, J., and Waterfield, M. (1980): Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. *Nature* (*Lond.*) **287**, 301—302.

Ghendon, Y., Klimov, A., Gorodkova, N., and Döhner, L. (1981): Genome analysis of influenza A virus strains isolated during an epidemic of 1979—1980. *J. gen. Virol.*, in press.

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

Harms, E., Rohde, W., and Scholtissek, C. (1978): Minor nucleic acids in influenza virus. *J. gen. Virol.* **40**, 691—694.

Hay, A. J., Bellamy, A. R., Abraham, G., Skehel, J. J., Brand, C. M., and Webster, R.G. (1977): Procedures for characterization of the genetic material of candidate vaccine strain s. *Development in Biological Standardization* **39**, 15—24.

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

Inglis, S. C., Barrett, Th., Brown, C. M., and Almond, J. W. (1979): The smallest genome RNA segment of influenza virus contains two genes that may overlap. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3790.

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

Jou, W. M., Verhoeyen, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., and Fiers, W. (1980): Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. *Cell* **19**, 683–696.

Lamb, R. A., and Choppin, P. W. (1979): Segment 8 of the influenza virus genome is unique in coding for two polypeptides. *Proc. natn. Acad. Sci. U.S.A* **76**, 4908–4912.

Lamb. R. A., Choppin, R. M., and Lai, Ching-Juh (1980): Mapping of the two overlapping genes for polypeptides NS_1 and NS_2 on RNA segment 8 of influenza virus genome. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1857–1861.

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

Nakajima, K., Ueda, M., and Sugiura, A. (1979): Origin of small RNA in von Magnus particles of influenza virus. *J. Virol.* **29**, 1142–1148.

Palese, P. (1977): The genes of influenza virus. *Cell* **10**, 1–10.

Racaniello, V. R., and Palese, P. (1979): Influenza B virus genome: assignment of viral polypeptides to RNA segments. *J. Viro* . **29**, 361–373.

Robertson, J. S. (1979): 5′ and 3′ terminal nucleotide sequences of the RNA genome segments of influenza virus. *Nucleic Acids Res.* **6**, 3745–3758.

Scholtissek, C. (1978): The genome of the influenza virus.
Current Topics in Microbiology and Immunology **80**, 139–163.

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

Scholtissek, C., and von Hoyiningen-Huene, V. (1980): Genetic relatedness of the gene which codes for the nonstructural (NS) protein of different influenza A strains. *Virology* **102**, 13–20.

Skehel, J. J., and Hay, A. J. (1978): Nucleotide sequences at the 5′ termini of influenza virus RNAs and their transcripts. *Nucleic Acids Res.* **5**, 1207–1220.

Sleigh, M. J., Both, G. W., and Brownlee, G. G. (1979): A new method for the size estimation of the RNA genome segments of influenza virus. *Nucleic Acids. Res.* **6**, 1309–1322.

Ueda, M., Nakajima, K., and Sugiura, A. (1980): Extra RNAs of von Magnus particles of influenza virus cause reduction of particular polymerase genes. *J. Virol.* **34**, 1–8.

Verhoeyen, M., Fang, R., Jou, W. M., Devos, R., Huylebroeck, D., Saman, E., and Fiers, W. (1980): Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. *Nature* (*Lond.*) **286**, 771–776.

Winter, G., and Fields, S. (1980): Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic Acids Res.* **8**, 1965–1974.

Winter, G., and Fields, S. (1981): The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. *Virology* **114**, 423–428.

Winter, G., Fields, S., adn Brownlee, G. G. (1981): Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature* (*Lond.*) **292**, 72–75.

Young, J. F., Desselberger, U., and Palese, P. (1979): Evolution of human influenza A viruses in nature: sequential mutations in the genomes of new H1N1 isolates. *Cell* **18**, 73–83.