JOURNAL OF VIROLOGY, Jan. 1984, p. 276–278
0022-538X/84/010276-03$02.00/0

Vol. 49, No. 1

Nucleotide Sequence of the Influenza Virus A/USSR/90/77 Hemagglutinin Gene

PATRICK CONCANNON,[1]* IAN W. CUMMINGS,[2]† AND WINSTON A. SALSER[1,2]

Departments of Biology[1] and Molecular Biology,[2] University of California, Los Angeles, California 90024

Received 9 June 1983/Accepted 19 September 1983

The complete nucleotide sequence of the hemagglutinin gene of influenza virus A/USSR/90/77 was determined. Comparison of hemagglutinin amino acid sequences from H1 field strains revealed five potential antigenic sites. Four of these sites correspond to those observed for H3 hemagglutinins, whereas the fifth apparently derives from differences in the glycosylation patterns between subtypes.

Influenza A viruses direct the synthesis of two surface glycoproteins, hemagglutinin (HA), the major antigen, and neuraminidase. Variation in these antigens has regularly caused major epidemics and has hampered efforts to create useful vaccines. In attempts to understand this variation, a number of amino acid and nucleic acid sequences of HAs from field strain viruses and in vitro-generated variants have been determined (11). Previous analysis has focused on the HA of the H3 subtype for which the three-dimensional structure has been solved (13). To examine the variation in another human subtype, H1, we determined the nucleotide sequence of the HA gene of influenza virus A/USSR/90/77. Comparison of this sequence with the only other existing complete H1 HA sequences (2, 7) derived from H1N1 viruses isolated in 1933 and 1934 allowed us to identify clusters of amino acid substitutions which suggest the location of the important antigenic sites in the H1 HAs.

The amino acid sequence reported here bears less than 40% homology with the corresponding sequences of H3 strains (Table 1). Consequently, it seemed plausible that the separate evolution of these strains might have created at least some changes in the three-dimensional folding of the protein which could aid in avoiding immune surveillance by covering up some antigenic sites and exposing others. On the contrary, our data indicate either that antigenic sites are in the same locations in the H1 and H3 subtypes or, if not, that the differences can more readily be explained by masking of an antigenic site owing to changes in glycosylation rather than changes in polypeptide folding.

We have previously reported the construction of a bank of cDNA clones derived from the reverse transcription of the viral RNA segments of the recombinant virus A/USSR/90/77 NBA2 (4). Identification of plasmids bearing HA gene sequences was made by colony hybridization (6) and Northern blotting (1). A number of plasmids containing the complete coding sequence of the A/USSR/90/77 HA gene and various amounts of the untranslated regions were isolated. The nucleotide sequence of the inserted DNA from one of these plasmids, pD49, was determined by the method of Maxam and Gilbert (8).

The coding sequence of the A/USSR/90/77 HA gene contains 1,698 base pairs encoding 566 amino acids, corresponding to 326 amino acids in the mature HA1 protein, 222 amino acids in the mature HA2, and 17 amino acids in the signal peptide. This sequence is shown in Fig. 1, along with representative sequences of HA genes (2, 5, 7, 9) from other human subtypes. The alignment of sequences in Fig. 1 is complicated by the lack of homology between the HA genes of different subtypes.

The low levels of amino acid homology noted in Table 1 suggest that there could be substantial differences in the three-dimensional structures of HAs of different subtypes. Antigenic selection would have favored any viable mutations which could have changed the folding of these molecules so as to cover up or change the location of antigenic sites. Such changes could help explain the complete lack of antigenic cross-reactivity between subtypes.

To test this hypothesis we have compared the amino acid sequences of the early H1 strains with that of A/USSR/90/77 HA to locate clusters of amino acid substitutions which might indicate the location of antigenic sites. The substitutions observed tend to cluster into five distinct regions as detailed in Table 2. Four of these regions correspond exactly in location with the antigenic sites predicted from sequencing the HA genes of field strains and laboratory variants of the H3 subtype (11, 12). There is a fifth significant cluster, labeled E in Table 2, which does not correspond to any antigenic site previously identified in the H3 HAs. This could indicate that a novel antigenic site has been exposed by a difference in the three-dimensional foldings of the H1 and H3 molecules. We have noticed, however, that all H3 proteins analyzed thus far share a glycosylation site at residue 81 which is uniformly absent in the H1 strains. It has been shown that the position 81 site is glycosylated in H3 strains (10). It had been postulated earlier that this carbohydrate side chain would hinder antibody attachment to this exposed region of the protein (12). Therefore, we think that the many changes we see at site E are more importantly due to this difference in glycosylation rather than to changes in protein folding.

Like all proteins, HAs accumulate a "background" level

TABLE 1. Percent homology shared between the HA gene of A/USSR/90/77 and the HA genes of other influenza viruses[a]

Virus	Amino acid	Nucleic acid
A/USSR/90/77 (H1)	100	100
A/PR/8/34 (H1)	91	92.2
A/WSN/33 (H1)	88.2	91.1
A/Jap/305/57 (H2)	66.4	63.7
A/Aichi/2/68 (H3)	37.1	47.7

[a] Percent homology has been calculated as the total number of identical positions in sequences aligned at conserved cysteine residues divided by the total number of comparable positions. Insertions or deletions are calculated as positions of nonhomology. These comparisons include the signal peptide.

* Corresponding author.
† Present address: School of Medicine, University of California, San Francisco, CA 94143.

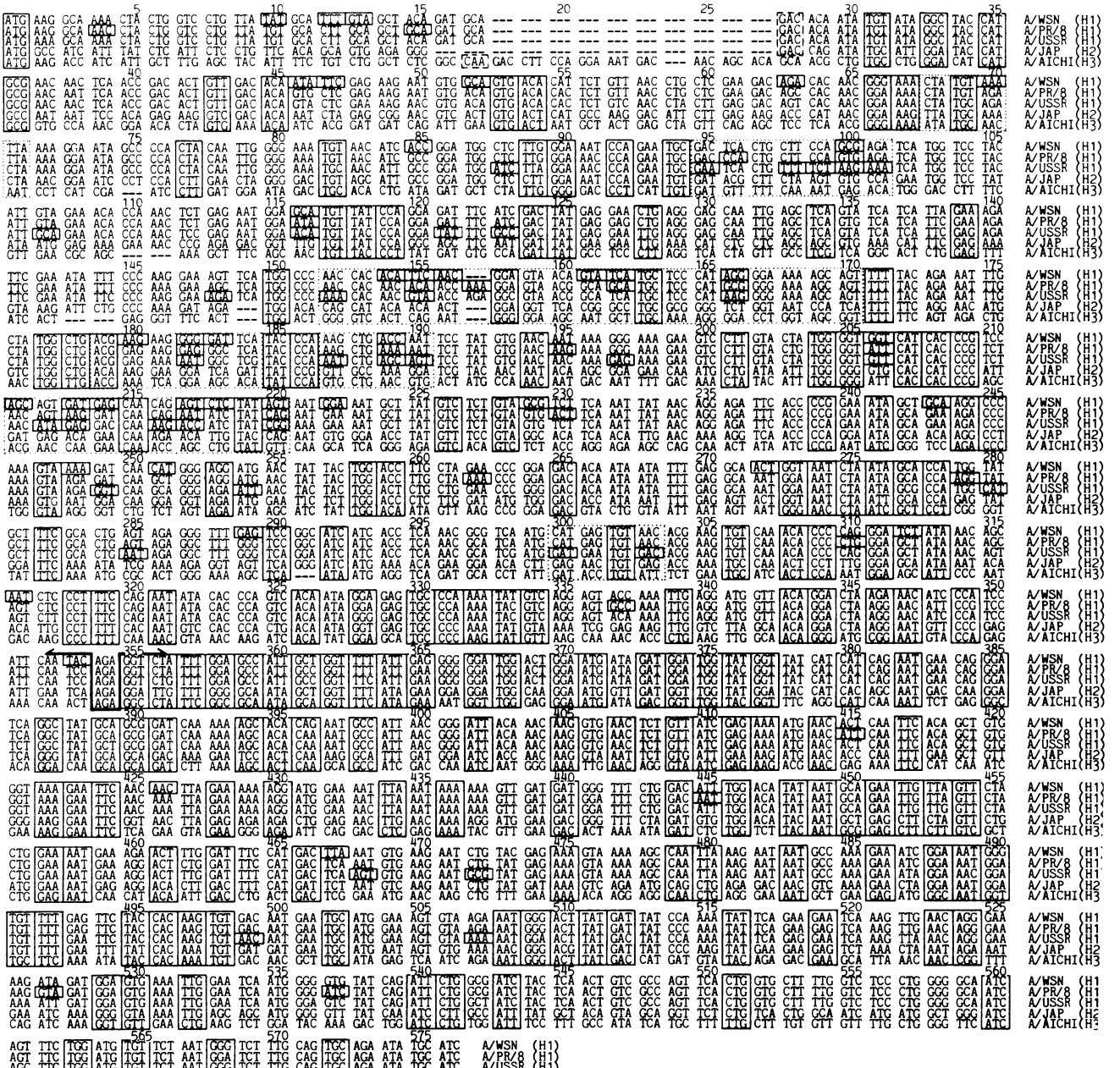

FIG. 1. The nucleotide sequence of the A/USSR/90/77 H1 gene. The nucleotides are arranged in triplets corresponding to the correct translational reading frame. Solid line boxes surrounding single triplets indicate amino acid differences relative to the other H1 sequences. Solid line boxes enclosing a triplet in all five sequences indicate an amino acid homology at that position in all five sequences. Dashed line boxes at positions 17 and 28 indicate the probable initial amino acid of the mature protein. The dark arrows surrounding position 354 indicate the HA1/HA2 boundary. Dotted boxes enclose the proposed antigenic sites as described in the text. These sites were identified by a moving average analysis as regions that displayed clustering of amino acid substitutions at levels significantly higher than background. The highest "background signals" which we ignored in this analysis were changes at only single amino acid residues, positions 89 and 179. These correspond to positions in the interior of the H3 protein.

of amino acid substitutions which may be neutral, since they appear to be randomly placed and tend to be conservative in nature. Nevertheless, the clustering and the nonconservative nature of the amino acid substitutions which we observed was significantly greater than this background level and appeared to accurately indicate the locations of antigenic sites. For instance, in a comparison of the A/PR/8/34 and A/USSR/90/77 HAs, the regions designated as probable antigenic sites in Fig. 1 and Table 1 encompass only 6.7% of the protein but contain 26% of the amino acid differences, a level about fivefold greater than in the remainder of the molecule. As mentioned earlier, four of the five regions used in the above calculation correspond to the location of sites predicted from sequencing the HA genes of field strains of the H3 subtype as well as laboratory variants selected by growth in the presence of antibodies directed against HA

TABLE 2. Tabulations of positions of amino acid substitutions that fall within proposed antigenic sites[a]

Site and position	Amino acid substitutions for the following strains:		
	A/WSN/33	A/PR/8/34	A/USSR/90/77
Site A			
125	Asn	Asn	Lys
127	Thr	Asn	Asn
128	Phe	Thr	Val
129	Asn	Thr	Thr
130	—[b]	Lys	Arg
134	Val	Ala	Ala
135	Ser	Ala	Ser
139	Arg	Ala	Lys
Site B			
153	Lys	Glu	Glu
155	Gly	Glu	Asn
156	Asp	Lys	Glu
184	Ser	Asn	Asn
185	Ser	Ser	Ile
186	Asp	Lys	Glu
187	Glu	Asp	Asp
189	Gln	Gln	Lys
190	Ser	Asn	Thr
191	Leu	Ile	Ile
193	Ser	Gln	Arg
Site C			
43	Lys	Arg	Arg
273	His	His	Asp
276	Asn	Asn	Asp
Site D			
160	Lys	Lys	Asn
162	Thr	Lys	Ser
163	Asn	Asn	Ser
Site E			
68	Asp	Asp	Glu
69	Ser	Pro	Ser
71	Leu	Leu	Phe
72	Pro	Pro	Ser
73	Ala	Val	Lys
74	Arg	Arg	Lys

[a] Note that because of insertions and deletions our numbering system is different from that of Wiley et al. (12). Otherwise, our sites A through D correspond to theirs, except that we have included in site A several substitutions which might extend new side chains into the originally designated site A or affect it indirectly by altering the protruding loop. These areas have been included in site A in more recent reviews of the H3 HA (e.g., Fig. 1 of reference 11).

[b] —, Position 130 of A/WSN/33 is a deletion in our sequence alignment.

(11, 12). Our identification of antigenic sites also corresponds with those located by analysis of antibody-selected laboratory variants of A/PR/8/34 (H1) (3) and confirms that these sites identified in vitro are also important for circulation in vivo.

Host immune responses mounted against influenza virus HA should have strongly selected for any viable alterations in folding which would have covered up some antigenic sites or caused antigenic sites to move to new locations. Although the evolutionary divergence of the H1 and H3 proteins is so great that they share only about 40% of the overall amino acid sequence, our data offer no evidence for any significant alterations in folding. Instead, the only significant change in the localization of antigenic sites appears to be due to the absence of a shielding carbohydrate side chain in the H1 proteins. This suggests a surprising degree of long-term conservation of the detailed three-dimensional folding of the HA protein.

We thank Jocyndra Wright and Dean Gilbert for valuable technical assistance in plasmid DNA isolation and purification and sequencing gel preparation and Susan Salser for expert assistance in manuscript preparation. In addition, we thank D. P. Nayak and G. Brownlee for providing us with nucleotide sequences of the A/WSN/33 and the A/PR/8/34 HA genes, respectively, before their publication.

This work was supported by Public Health Service grants AI-16348 and GM-18586 from the National Institutes of Health.

LITERATURE CITED

1. **Alwine, J. C., D. J. Kemp, B. A. Parker, J. Reiser, J. Renart, G. R. Stark, and G. M. Wahl.** 1979. Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Methods Enzymol. **68**:220–242.
2. **Brownlee, G. G., G. Winter, and S. Fields.** 1981. The hemagglutinin gene of influenza A/PR/8/34, p. 65–75. *In* D. Nayak and C. F. Fox (ed.), ICN-UCLA symposia on molecular and cellular biology. Academic Press, Inc., New York.
3. **Caton, A. J., G. G. Brownlee, J. W. Yewdell, and W. Gerhard.** 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell **31**:417–427.
4. **Cummings, I., and W. A. Salser.** 1980. The synthesis and cloning of large influenza A cDNAs using synthetic DNA primers, p. 147–155. *In* G. Laver and G. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
5. **Gething, M.-J., J. Bye, J. Skehel, and M. Waterfield.** 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 (London) **287**:301–306.
6. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3972.
7. **Hiti, A. L., A. R. Davis, and D. P. Nayak.** 1981. Complete sequence analysis of the influenza A/WSN/33 (H0 subtype) hemagglutinin and its relationship to the A/Japan/305/57 (H2 subtype) hemagglutinin, p. 217–223. *In* D. H. L. Bishop and R. W. Compans (ed.), The replication of negative strand viruses. Elsevier/North-Holland Publishing Co., New York.
8. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
9. **Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers.** 1980. Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature (London) **286**:771–776.
10. **Ward, C. W., and T. A. Dopheide.** 1980. The Hong Kong (H3) hemagglutinin. Complete amino acid sequence and oligosaccharide distribution for the heavy chain of A/Memphis/102/72, p. 147–155. *In* G. Laver and G. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
11. **Webster, R. G., W. G. Laver, G. M. Air, and G. G. Schild.** 1982. Molecular mechanisms of variation in influenza viruses. Nature (London) **296**:115–121.
12. **Wiley, D. C., I. A. Wilson, and J. J. Skehel.** 1981. Structural identification of the antibody binding sites of the Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature (London) **289**:373–378.
13. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 angstrom resolution. Nature (London) **289**:366–373.