JOURNAL OF VIROLOGY, July 1982, p. 361-364
0022-538X/82/070361-04$02.00/0

Vol. 43, No. 1

Homology Between the Glycoproteins of Vesicular Stomatitis Virus and Rabies Virus

J. K. ROSE,[1]* R. F. DOOLITTLE,[2] A. ANILIONIS,[3]† P. J. CURTIS,[3] AND W. H. WUNNER[3]

Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138[1]; Department of Chemistry, University of California, San Diego, La Jolla, California 92093[2]; and The Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania 19104[3]

Received 7 December 1981/Accepted 2 March 1982

We compared the predicted amino acid sequences of the vesicular stomatitis virus and rabies virus glycoproteins by using a computer program which provides an optimal alignment and a statistical significance for the match. Highly significant homology between these two proteins was detected, including identical positioning of one glycosylation site. A significant homology between the predicted amino acid sequences of vesicular stomatitis virus and influenza virus matrix proteins was also found.

Vesicular stomatitis virus (VSV) and rabies virus were originally classified as members of the rhabdovirus family because they both have the bullet-shaped morphology characteristic of members of this family (12). Subsequent biochemical characterization has shown numerous other similarities, including the presence in both viruses of single, negative-stranded RNA genomes from which five separate mRNA species are transcribed. These mRNAs encode five structural proteins of similar sizes in both viruses (3, 5, 8, 14). The single glycoprotein species (G protein) found in virions of both VSV and rabies virus protrudes from the viral envelope, forming visible "spikes." This protein is required for binding of the virus to the host cell (12).

Recent determination of the nucleotide sequences of the mRNAs encoding a VSV G protein (11) and a rabies virus G protein (1) has made it possible for us to examine the extent of sequence relatedness between the predicted amino acid sequences of these two rhabdovirus proteins. We have also examined the extent of homology between the matrix (M) protein of VSV (11) and the M protein of an influenza virus (13), an enveloped, negative-stranded RNA virus from a separate family, the orthomyxoviruses.

The VSV and rabies virus G proteins are approximately the same size, being composed of 511 and 524 amino acids, respectively. The two sequences were compared by a computer program with an algorithm similar to that described by Needleman and Wunsch (7). The computer essentially considers all possible alignments of the two sequences. An alignment score is generated such that a penalty is imposed for every gap that is equal to 2.5 times the score given for each identity (4). The statistical significance of the alignment score is then gauged by generating a set of scrambled sequences that have the same lengths and compositions as the two sequences being compared. A large number (36 in this case) of alignments are then made with the scrambled sequences, scores are tabulated, and a mean and standard deviation are calculated. The significance of the authentic comparison can then be expressed in standard deviations above (or below) the mean score of the scrambled comparisons. A score of 3.0 standard deviations above the mean has been accepted as evidence for a significant relationship implying common ancestry (2, 4). The VSV and rabies virus sequences have 102 identities (20%) in the optimum alignment; there are seven gaps (Fig. 1). The alignment score was 12.6 standard deviations above the mean obtained from the randomized comparisons, a highly significant result.

Certain regions of the proteins show greater homology. For example, in the 24 amino acids preceding and including the putative glycosylation sites nearest the COOH terminus of both proteins (arrowed), the identity is greater than 50% and includes exact alignment of the putative glycosylation sites. The similar positioning of cysteine residues in both proteins is also especially notable, because cysteine residues are known to be highly conserved in protein evolution (2, 4). Of the 15 cysteine residues in VSV G protein, 8 are aligned with 8 of the 17 cysteine residues in rabies virus G protein. Both proteins have hydrophobic signal peptides at their amino termini (Fig. 1, shaded), and although they differ in length, the alignment places the sites of pro-

† Present address: Southern Biotech, Inc., Tampa, FL 33612.

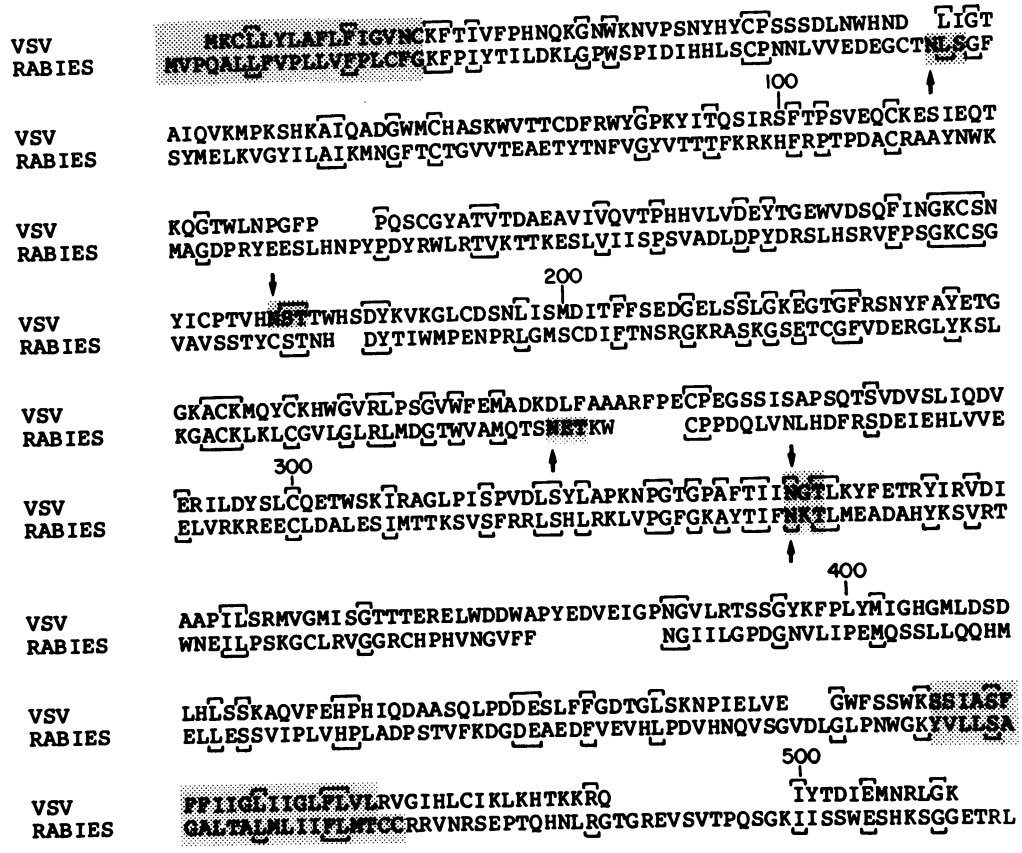

FIG. 1. Comparison between the predicted sequences of VSV (San Juan strain) G protein and rabies virus (ERA strain) G protein. The amino acid sequences were aligned for optimum identity as described by Doolittle (4). The entire sequence of each protein, including the signal peptide, is shown. Areas of identity are bracketed. The NH_2-terminal signal sequences are shaded, as are the putative transmembrane segments of each protein near their COOH termini. Putative glycosylation signals, Asn-X-Ser and Asn-X-Thr, are also shaded. Arrows indicate the three putative glycosylation sites in rabies virus G protein and two putative glycosylation sites in VSV G protein. The letter codes for the amino acids are as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine.

teolytic cleavage in register. Similarly, the hydrophobic transmembrane segments near the COOH terminus (Fig. 1, shaded) are aligned exactly on one side. It is notable that the signal sequences and transmembrane segments are not highly homologous, presumably because the major constraint is only that a hydrophobic character be maintained in these regions. Our unpublished work (C. J. Gallione and J. K. Rose) indicates that these regions are also highly variable among VSV strains and serotypes.

Taken together, these similarities provide strong evidence for related three-dimensional structures in both G proteins and provide a quantitative measure of the evolutionary relatedness of these two viruses. Future comparisons of rhabdovirus G protein sequences will allow us to determine whether these homologies are conserved throughout the rhabdovirus family. In addition to important structural roles, such conserved domains could function as signals for transport of these proteins through the cellular membrane system to the cell surface or in virus binding and entry into cells.

The degree of relatedness between rabies virus and VSV is not as great as that seen for the structural proteins of Sindbis virus and Semliki Forest virus, two members of the alphavirus family (10). These viruses show approximately 47% identity, making the homology obvious without computer analysis. The relatedness of the rabies virus and VSV G proteins is similar to that found for the β-chain of human hemoglobin and human myoglobin (4).

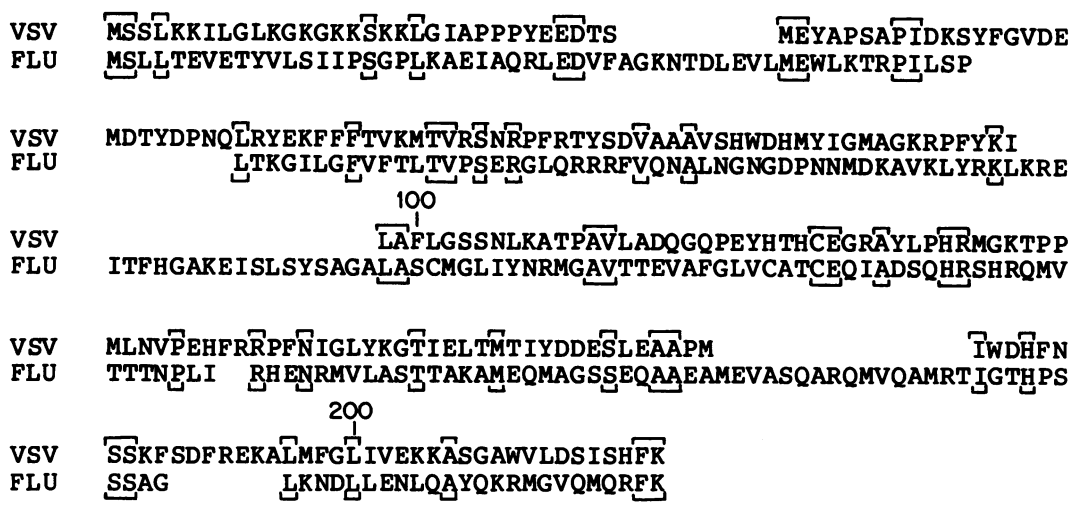

FIG. 2. Comparison between the predicted sequences of VSV (San Juan strain) M protein and influenza (A/PR/8/34 strain) M protein. The optimum alignment was determined as described by Doolittle (4). Identities are bracketed.

It should be noted that even in regions of high amino acid identity between VSV G protein and rabies virus G protein, the extent of nucleotide sequence identity is less than 50% because of degeneracy in the genetic code. No sequence identities longer than six nucleotides are found in regions of high amino acid homology, indicating that nucleic acid homology should not be detectable by nucleic acid hybridization. In fact, even among different serotypes of VSV it is well established that there is little nucleic acid homology detectable by nucleic acid hybridization (9) and strong conservation of protein sequences (6).

Although VSV and influenza virus are members of different virus families, they both contain matrix (M) proteins with similar molecular weights and apparently similar functions. In each virus family, the M protein is thought to form a shell beneath the viral envelope (12). Because the M gene sequences have been determined recently for both VSV (11) and an influenza virus strain (13), we examined the possible relatedness of these proteins as described above. The maximum alignment is shown in Fig. 2. Although the overall identity is 20% for this comparison (as it is in the case of the VSV and rabies virus G proteins), the alignment score is lower, 3.3 standard deviations above the mean of 36 comparisons of randomized sequences of the M proteins. This lower significance results mainly from the comparison of much shorter amino acid sequences and the introduction of six gaps to obtain optimal alignment. This is the same degree of significance as is found for the haptoglobin β-chain and chymotrypsinogen, a well-established case of common ancestry for proteins in the same size range as these M proteins.

It is notable that the single cysteine residue of VSV M protein is aligned with a cysteine residue in influenza M protein and that both proteins begin with the sequence Met-Ser-X-Lys and end with identical dipeptides. Given the presumably similar structural roles of the two proteins, we believe that the significant relatedness indicates common ancestry. It should be noted, however, that in comparisons of influenza virus and VSV G protein sequences and VSV nucleocapsid and influenza virus nucleocapsid protein sequences we have not detected significant relatedness. It seems possible, however, that greater evolutionary constraints might have been applied to the M protein structure so that it fulfills a crucial role in virus maturation. Because the degree of relatedness of the M proteins stands out from comparisons of a large number of proteins of similar sizes (4), we are confident that the relatedness is genuine.

We thank Bart Sefton, Kaaren Janssen, and Roberto Weinmann for comments on the manuscript and Maureen Brennan and Carolyn Goller for typing the manuscript.

This work was supported by grants from the Public Health Service (AI15481 and HE18576) and the National Cancer Institute (CA14195).

LITERATURE CITED

1. **Anilionis, A., W. H. Wunner, and P. Curtis.** 1981. Structure of the glycoprotein gene in rabies virus. Nature (London) **294**:275–278.
2. **Barker, W. C., and M. O. Dayhoff.** 1972. Atlas of protein sequence and structure, vol. 5, p. 101–110. National Biochemical Research Foundation, Washington, D.C.
3. **Both, G. W., S. A. Moyer, and A. K. Banerjee.** 1975.

Translation and identification of the mRNA species synthesized in vitro by the RNA polymerase of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. **72**:274–278.

4. **Doolittle, R. F.** 1981. Similar amino acid sequences: chance or common ancestry. Science **214**:149–159.
5. **Knipe, D., J. K. Rose, and H. F. Lodish.** 1975. Translation of individual species of vesicular stomatitis viral mRNA. J Virol. **15**:1004–1011.
6. **McGeoch, D. J., A. Dolan, and C. R. Pringle.** 1980. Comparisons of nucleotide sequences in the genomes of the New Jersey and Indiana serotypes of vesicular stomatitis virus. J. Virol. **33**:69–77.
7. **Needleman, S. B., and D. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–458.
8. **Pennica, D., B. Holloway, J. Heyward, and J. Obijeski.** 1980. In vitro translation of rabies virus messenger RNAs. Virology **103**:517–521.
9. **Repik, P., A. Flamand, H. F. Clark, J. F. Obijeski, P. Roy, and D. H. L. Bishop.** 1974. Detection of homologous RNA sequences among six rhabdovirus genomes. J. Virol. **13**:250–252.
10. **Rice, C. M., and J. H. Strauss.** 1981. Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. U.S.A. **78**:2062–2066.
11. **Rose, J. K., and C. J. Gallione.** 1981. Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. J. Virol. **39**:519–528.
12. **Wagner, R. R.** 1975. Reproduction of rhabdoviruses. p. 1–93. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology. Plenum Publishing Corp., New York.
13. **Winter, G., and S. Fields.** 1980. Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. Nucleic Acids Res. **8**:1965–1974.
14. **Wunner, W. H., P. J. Curtis, and T. J. Wiktor.** 1980. Rabies mRNA translation in *Xenopus laevis* oocytes. J. Virol. **36**:133–142.