Amino Acid Sequence Similarity Between Rabies Virus Glycoprotein and Snake Venom Curaremimetic Neurotoxins

Abstract. *Evidence was presented earlier that a host-cell receptor for the highly neurotropic rabies virus might be the acetylcholine receptor. The amino acid sequence of the glycoprotein of rabies virus was compared by computer analysis with that of snake venom curaremimetic neurotoxins, potent ligands of the acetylcholine receptor. A statistically significant sequence relation was found between a segment of the rabies glycoprotein and the entire sequence of long neurotoxins. The greatest identity occurs with residues considered most important in neurotoxicity, including those interacting with the acetylcholine binding site of the acetylcholine receptor. Because of the similarity between the glycoprotein and the receptor-binding region of the neurotoxins, this region of the viral glycoprotein may function as a recognition site for the acetylcholine receptor. Direct binding of the rabies virus glycoprotein to the acetylcholine receptor could contribute to the neurotropism of this virus.*

Rabies virus (RV) is an enveloped, negative-strand RNA virus belonging to the rhabdovirus family, *Lyssavirus* genus. The glycoprotein comprising the surface spikes of enveloped viruses attaches to the cell surface (*1*) by binding to normal cellular constituents, which act as viral receptors (*2*). The amino acid sequence of the glycoprotein from two strains of RV has been deduced from the nucleotide sequence of cloned complementary DNA (*3, 4*). The glycoprotein is composed of 505 amino acids, with three carbohydrate attachment sites on the ectodomain. After replicating in muscle cells, RV gains access through peripheral nerves to the central nervous system (*5*) where it selectively infects certain neuronal populations (*2*). We showed earlier that RV was localized in regions containing a high density of acetylcholine receptors (AChR) on mouse diaphragms and cultured chick myotubes (*6*). Also, α-bungarotoxin, a snake venom neurotoxin, and *d*-tubocurarine greatly reduced the titer of RV strain 1820-B necessary to infect myotubes, and α-bungarotoxin reduced attachment of [^{3}H]uridine-labeled standard challenge virus strain (CVS) to myotubes. On the basis of these observations, we propose that a host-cell receptor for RV might be the AChR.

Neurotoxins from snakes belonging to the families Elapidae (cobras, kraits, mambas, and others) and Hydrophidae (sea snakes) are polypeptides that bind with high affinity to the nicotinic AChR and, like curare, competitively block the depolarizing action of acetylcholine (ACh) (*7, 8*). The snake venom neurotoxins (molecular weights, 7000 to 8000) are of two distinct size groups: short neurotoxins containing 60 to 62 amino acids, and long neurotoxins containing 71 to 74 residues (*8*). More than 60 of these neurotoxins have been sequenced (*8, 9*) and have been aligned and enumerated by Karlsson (*8*) to characterize their maximum homology. Comparative sequence data, chemical modification, and determination of the three-dimensional structure of the neurotoxins has provided considerable information on structure-function relations. Several amino acids that are highly conserved or invariant among all the neurotoxins are considered to be important in the binding of the toxin to the ACh binding site on the AChR (*8, 10–14*). These residues are located at the end of loop 2 (the "toxic" loop), a long central loop protruding from the toxin molecule (*10–15*) (see Fig. 3). The guanidinium group of Arg^{37} is the only cationic group common to all the neurotoxins and may be the counterpart of the quaternary ammonium group of ACh (*8, 10, 11*). A hydrogen-bonded ion

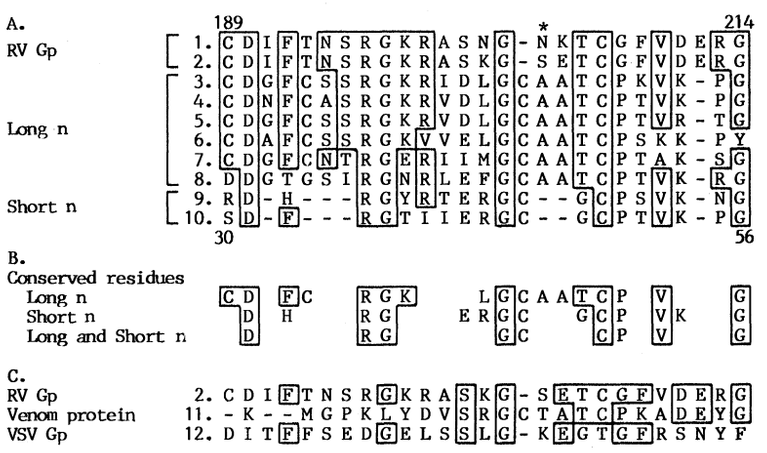

Fig. 1. Comparison of amino acid sequences of rabies virus glycoprotein (RV Gp) (residues 189 to 214) with (A) long and short neurotoxins (n); (B) amino acids highly conserved or invariant among n (based on all sequenced n); and (C) a nonneurotoxic venom protein and vesicular stomatitis virus glycoprotein (VSV Gp). Rabies virus is aligned with the Karlsson homology alignment positions 30 to 56 for n and venom proteins (*8*) and with VSV as described (*20*). Amino acid identities between RV Gp and other sequences are enclosed in boxes. A possible site of *N*-linked glycosylation in rabies glycoprotein (CVS strain) is indicated by an asterisk. (1) RV Gp (CVS strain) (*4*); (2) RV Gp (ERA strain) (*3*); (3) *Ophiophagus hannah* (king cobra), toxin b; (4) *Naja melanoleuca* (forest cobra), toxin b; (5) *Naja naja naja* (Indian cobra), toxin 4 (toxin b); (6) *Bungarus multicinctus* (Formosan banded krait), α-bungarotoxin; (7) *Astrotia stokesii* (Stoke's sea snake), toxin b; (8) *Laticauda colubrina* (yellow-lipped sea krait), toxin b; (9) *Naja naja atra* (Formosan cobra), cobratoxin; (10) *Laticauda semifasciata* (broad-banded blue sea snake), erabutoxin; (11) *Dendroaspis angusticeps* (eastern green mamba), toxin C_9S_3; (12) VSV Gp (*22*). Toxin sequences from (*8, 9*).

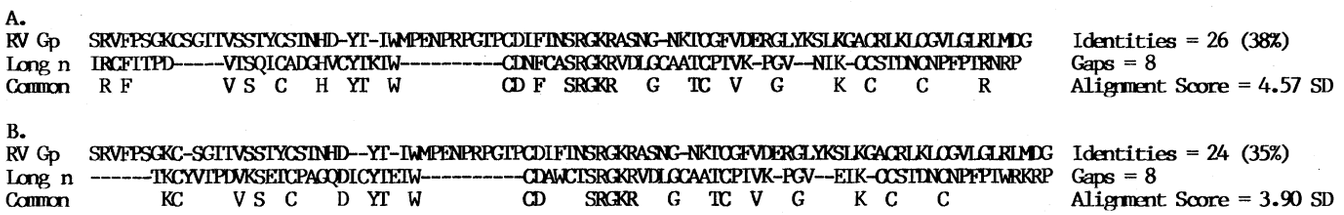

Fig. 2. Alignment of long neurotoxin (n) sequences with a segment (residues 151 to 238) of rabies virus glycoprotein (RV Gp) by means of the computer program ALIGN. The mutational data matrix for 250 PAM's (point accepted mutations) (*17*) was used. The above alignments were produced by means of a matrix bias of +3 and a gap penalty of 6. Alignment scores were derived from standard deviations obtained with alignments made with at least 100 scrambled sequences having the same lengths and compositions as the RV Gp and n. (A) Alignment of RV Gp (CVS strain) with long toxin b of *Naja melanoleuca*; (B) alignment of RV Gp with long toxin a of *Ophiophagus hannah*.

pair between the guanidinium group of Arg^{37} and the side-chain carboxylate of Asp^{31} stereochemically resembles ACh (*11*).

Because of the possibility that both RV and the neurotoxins bind at the ACh binding site of the AChR, we performed a comparison of the amino acid sequences of the RV glycoprotein with the sequences of the neurotoxins. The amino acid sequence of RV glycoprotein was manually aligned with the Karlsson homology alignment of neurotoxins. There was a high degree of identity (50 percent) between residues 189 to 214 of the glycoprotein and between the alignment positions 30 to 56 of the neurotoxins (Fig. 1A). The full length of the long neurotoxins was compared to the first 250 residues of RV glycoprotein with the computer program ALIGN (*16, 17*). These analyses revealed a significant structural similarity between the entire long neurotoxin sequence and a segment of the RV glycoprotein (residues 151 to 238) (Fig. 2). Alignment scores for both the neurotoxin segment and the entire neurotoxin ranged between 3 and 5 standard deviation (S.D.) units (normalized standard deviations). Scores above 3 S.D. units indicate significant sequence relation and a possible common evolutionary origin (*18*). Analysis of 20 long neurotoxins (*19*) revealed ten long neurotoxins with alignment scores above 3 S.D. units.

Of even more significance, the greatest similarity between the RV glycoprotein and neurotoxins occurred with the amino acids that are highly conserved or invariant among all the neurotoxins, including those residues that bind to the ACh binding site on the AChR (Fig. 1B). All the highly conserved amino acids at the end of the long neurotoxin loop 2 interacting with the ACh binding site (Asp^{31}, Phe^{33} [His or Trp], Arg^{37}, Gly^{38}, Lys^{39}) (*13*) were matched in the RV glycoprotein (Figs. 1B and 3). In contrast, other nonneurotoxic venom proteins whose entire sequences are related to the neurotoxins, as well as vesicular stomatitis virus glycoprotein which is homologous to the entire RV glycoprotein sequence (*20*), showed little similarity to RV glycoprotein and neurotoxins in this particular region and lack Asp^{31} and Arg^{37} (Fig. 1C).

The RV glycoprotein showed identity with all the residues in loop 2 that are highly conserved or invariant in both the long and short neurotoxins (Tyr^{25}, Trp^{29}, Asp^{31}, Arg^{37}, Gly^{38}, Gly^{44}) (Fig. 2). As a result, the structurally similar segment of the RV glycoprotein sequence can be superimposed over models of neurotoxin loop 2 (*9–13, 15*) (Fig. 3). The alignment given in Fig. 2 would place an insertion of ten residues in the glycoprotein (Fig. 3, boxed region) between Trp^{29} and the adjacent Cys^{30} of loop 2 in the long neurotoxins. Such a structure appears plausible because the insertion includes four prolines capable of producing a random-coil loop. Furthermore, this region of the glycoprotein is probably exposed because a site for *N*-linked glycosylation is present at position 204 of the CVS strain (*4*) (Fig. 1).

The observation that a portion of a viral glycoprotein closely resembles the structure of a specific ligand for a normal cell receptor has important general implications for viral diseases. The regions of virus molecules involved in binding to the cellular components acting as viral receptors may, in general, be structurally similar to the binding domains of cellular proteins and molecules that normally bind to the cell components. Such similarity between parts of viruses and ligands of normal cell constituents may provide a basis for the pathogenesis of some autoimmune diseases because viral infection could be followed by production of antibodies to the idiotypes, which would react with the cellular structures (*21*). Finally, identification of the viral domains that act in binding to cells should be important in the treatment of viral diseases. These regions could be used as immunogenic agents in the development of safe and specific subunit vaccines or as the targets of chemical agents that might prove effective in inhibiting viral binding.

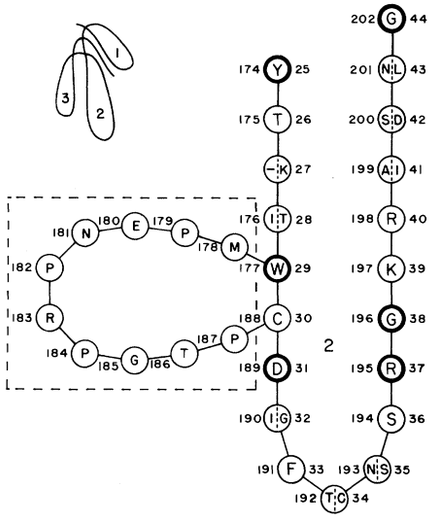

Fig. 3. A model showing homology of rabies glycoprotein with the "toxic" loop of the neurotoxins. The segment of the glycoprotein (residues 174 to 202) corresponding to loop 2 of the long neurotoxins (Karlsson positions 25 to 44) as determined by computer modeling is positioned relative to a schematic representation of loop 2. Toxin scheme based on (*12*). Within circles, residues or gaps in the glycoprotein are shown on the left and those in the neurotoxin on the right. One letter is shown where the glycoprotein and toxin are identical. Bold circles are residues highly conserved or invariant among all the neurotoxins. A ten-residue insertion in the glycoprotein is enclosed in the box. The rabies virus sequence is of the CVS strain. Neurotoxin sequence is *Ophiophagus hannah*, toxin b. (Inset) Schematic of neurotoxin structure showing positions of loops 1, 2, and 3.

THOMAS L. LENTZ
PAUL T. WILSON
Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06510
EDWARD HAWROT
Department of Pharmacology, Yale University School of Medicine
DAVID W. SPEICHER
Protein Chemistry Laboratory, Department of Pathology, Yale University School of Medicine

References and Notes

1. J. White, M. Kielian, A. Helenius, *Q. Rev. Biophys.* **16**, 151 (1983).
2. R. T. Johnson, *Viral Infections of the Nervous System* (Raven, New York, 1982).
3. A. Anilionis, W. H. Wunner, P. J. Curtis, *Nature (London)* **294**, 275 (1981).
4. E. Yelverton, S. Norton, J. F. Obijeski, D. V. Goeddel, *Science* **219**, 614 (1983).
5. F. A. Murphy, *Arch. Virol.* **54**, 279 (1977).
6. T. L. Lentz, T. G. Burrage, A. L. Smith, J. Crick, G. H. Tignor, *Science* **215**, 182 (1982).
7. C. Y. Lee, *Annu. Rev. Pharmacol.* **12**, 265 (1972); R. Chicheportiche, J.-P. Vincent, C. Kopeyan, H. Schweitz, M. Lazdunski, *Biochemistry* **14**, 2081 (1975).
8. E. Karlsson, *Handb. Exp. Pharmacol.* **52**, 159 (1979).
9. M. J. Dufton and R. C. Hider, *CRC Crit. Rev. Biochem.* **14**, 113 (1983).
10. B. W. Low *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2991 (1976); C. Y. Lee, *Adv. Cytopharmacol.* **3**, 1 (1979); B. W. Low, *ibid.*, p. 141; *Handb. Exp. Pharmacol.* **52**, 213 (1979).
11. D. Tsernoglou, G. A. Petsko, R. A. Hudson, *Mol. Pharmacol.* **14**, 710 (1978).
12. R. M. Stroud, *Neurosci. Comment.* **1**, 124 (1983).
13. R. H. Fairclough *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **48**, 9 (1983).
14. M. A. Juillerat, B. Schwendimann, J. Hauert, B. W. Fulpius, J. P. Bargetzi, *J. Biol. Chem.* **257**, 2901 (1982).
15. M. R. Kimball, A. Sato, J. S. Richardson, L. S. Rosen, B. W. Low, *Biochem. Biophys. Res. Commun.* **88**, 950 (1979); M. D. Walkinshaw, W. Saenger, A. Maelicke, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2400 (1980).
16. M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1978), vol. 5, suppl. 3, pp. 1–8.
17. R. M. Schwartz and M. O. Dayhoff, *ibid.*, pp. 353–358.
18. W. C. Barker and M. O. Dayhoff, *ibid.* (1972), vol. 5, pp. 101–110; R. F. Doolittle, *Science* **214**, 149 (1981).
19. W. C. Barker *et al.*, *Protein Sequence Database* from *The Atlas of Protein Sequence and Structure* (Version 8.1, 21 January 1984; 2511 sequences, 470,158 residues) (National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C., 1984).
20. J. K. Rose, R. F. Doolittle, A. Anilionis, P. J. Curtis, W. H. Wunner, *J. Virol.* **43**, 361 (1982).
21. K. Sege and P. A. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2443 (1978); N. H. Wassermann *et al.*, *ibid.* **79**, 4810 (1982); P. H. Plotz, *Lancet* **1983-II**, 824 (1983).

22. J. K. Rose and C. J. Gallione, *J. Virol.* **39**, 519 (1981).
23. We thank A. Helenius and C. F. Stevens for reviewing the manuscript. Supported by NSF grant 82-03825 and NIH grant GM 32629.

17 May 1984; accepted 2 July 1984