JOURNAL OF VIROLOGY, Nov. 1983, p. 335–339
0022-538X/83/110335-05$02.00/0

Vol. 48, No. 2

Coding Assignment and Nucleotide Sequence of Simian Rotavirus SA11 Gene Segment 10: Location of Glycosylation Sites Suggests That the Signal Peptide Is not Cleaved

GERALD W. BOTH,[1]* LINDA J. SIEGMAN,[1] A. RICHARD BELLAMY,[2] AND PAUL H. ATKINSON[3]

Commonwealth Scientific and Industrial Research Organisation, Molecular and Cellular Biology Unit, North Ryde, New South Wales, 2113, Australia[1]; Department of Cell Biology, University of Auckland, Auckland, New Zealand[2]; and Department of Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York[3]

Received 23 May 1983/Accepted 21 July 1983

A cloned DNA copy of simian rotavirus SA11 genomic segment 10 was used to confirm the assignment of the nonstructural glycoprotein NCVP5 to this gene. Determination of the nucleotide sequence for gene 10 indicated that NCVP5 is 175 amino acids in length and has an N-terminal hydrophobic region with the characteristics of a signal sequence for membrane translocation. Unexpectedly, this region was also the location for the only two potential glycosylation sites within the molecule, asparagine residues 8 and 18. The carbohydrates carried by NCVP5 were of the high-mannose type, $Man_9GlcNAc$ and $Man_8GlcNAc$, with the mannose 9 species predominating; no complex oligosaccharides were present. If these asparagine residues are the sites for carbohydrate attachment, this implies that cleavage of the putative signal peptide does not occur during the maturation of this nonstructural glycoprotein.

Rotaviruses are a major cause of acute gastroenteritis in the young of a wide variety of animal species, including humans (9; M. K. Estes, E. L. Palmer, and J. F. Obijeski, Curr, Top. Microbiol. Immunol., in press). The rotavirus genome consists of 11 segments of double-stranded RNA (dsRNA) (18), which for simian rotavirus SA11 range in approximate size from 660 to 3,720 base pairs (4). The coding assignments for most segments of the SA11 genome have been completed by the use of a technique, first developed for reovirus (15), which involves direct translation of the denatured dsRNA segments in reticulocyte lysates. However, some difficulty has been encountered in the assignment of both the major outer structural protein VP7 (which constitutes the neutralizing antigen) and the nonstructural protein NCVP5 because these proteins are glycosylated and the carbohydrate side chains are the high-mannose type (1, 4b, 7).

Recently, the structural glycoprotein VP7 was assigned to genomic segment 9 by using a cloned DNA copy of gene 9 to select the specific mRNA for translation in vitro in the presence of dog pancreatic microsomes (4a). However, the origin of the nonstructural glycoprotein NCVP5 has been the subject of conflicting reports. Both Dyall-Smith and Holmes (6) and Arias et al. (1) assigned a protein called NS3 or pNCVP5, respectively, to gene 10 of SA11, but they disagreed as to whether this polypeptide was the precursor of the glycoprotein which they referred to as O_4 or NCVP5, respectively.

Here we used independent methods to show that gene 10 encodes pNCVP5, confirming the results of Arias et al. (1) that this protein is the precursor of the nonstructural viral glycoprotein NCVP5. We also determined the nucleotide sequence of gene 10 and deduced the amino acid sequence of NCVP5. Surprisingly, the amino acid sequence reveals that the only potential glycosylation sites are both located within the domain of an N-terminal region that could serve as a signal peptide for membrane translocation (2). The unusual location on the molecule of the potential glycosylation sites implies that NCVP5 contains a signal sequence which remains uncleaved during membrane translocation.

MATERIALS AND METHODS

Growth and purification of virus: preparation of cell lysates. Methods for SA11 growth and purification and for preparation of [^{35}S]methionine-labeled infected cell lysates and [^{3}H]mannose-labeled virus have been described elsewhere (4a, 4b).

Genomic clones and viral mRNA. Cloned DNA copies of SA11 dsRNA segments were obtained as described previously (4). Viral mRNA was synthesized in vitro with endogenous transcriptase in viral cores (12).

Hybrid selection and translation of viral mRNAs. Hybrid selection of gene-specific viral mRNAs was carried out with cloned gene copies immobilized on nitrocellulose filters (4a, 5). mRNAs were translated in a rabbit reticulocyte lysate system in the presence of [^{35}S]methionine (Amersham Corp.). Canine pancreas membranes for in vitro glycosylation were prepared as described previously (11). Translation products were analyzed on a sodium dodecyl sulfate-urea-polyacrylamide gel system (13).

Nucleic acid sequencing. Sequencing was performed by a combination of methods. Parts of the DNA clone were subcloned into bacteriophage M13 mp 7.1 (16) and sequenced by the Sanger method (19); other regions were sequenced by the method of Maxam and Gilbert (14). Some of the sequence was determined indirectly by the Sanger method with SA11 mRNA as template and DNA fragments from the clone as primers for cDNA synthesis by reverse transcriptase (4, 4a). This method provided a convenient means for identifying the coding (plus) strand of the cloned DNA and permitted overlaps between the blocks of sequence data to be obtained.

RESULTS

Coding assignments for SA11 gene segments 10 and 11. Recently, cloned DNA copies of SA11 dsRNA segments were used to complete the coding assignments for segments 7, 8, and 9 (4, 4b); gene-specific mRNAs were purified by hybrid selection and translated in vitro (4a). Here we used a similar approach to clarify the coding assignments for SA11 segments 10 and 11. mRNAs specific for the two genes were translated in a reticulocyte cell-free system, and the proteins synthesized were analyzed by polyacrylamide gel electrophoresis (Fig. 1). It is clear that genes 10 and 11 encode 20,000-dalton (20K) (Fig. 1c) and 28K (Fig. 1d) proteins, respectively, and that the 20K protein comigrates with the authentic pNCVP5 produced in infected cells treated with tunicamycin (Fig. 1e). This coding assignment is consistent with earlier data (1, 6). To confirm the glycosylated nature of the gene 10 product, we added canine pancreatic microsomes to the translation system (Fig. 2). The 20K translation product (Fig. 2c) was modified and now migrated as a 28K protein (Fig. 2d). This transition was similar to the conversion of

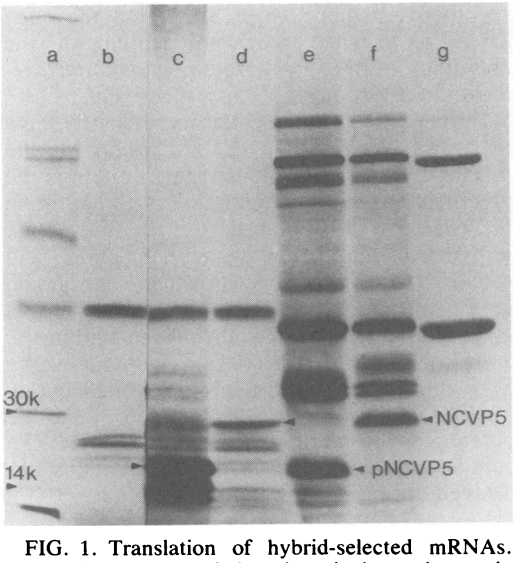

FIG. 1. Translation of hybrid-selected mRNAs. Translations were carried out in reticulocyte lysates in the absence of mRNA (b) or in the presence of gene 10-specific (c) or gene 11-specific (d) mRNAs. Translation products were analyzed by polyacrylamide gel electrophoresis and compared with [^{35}S]methionine-labeled proteins from SA11-infected cells prepared in the presence (e) and absence (f) of tunicamycin. (g) [^{35}S]methionine-labeled proteins from purified virus. (a) Standard molecular-weight markers. The protein nomenclature is that of Arias et al. (1).

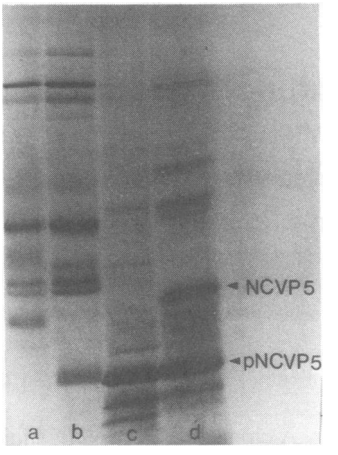

FIG. 2. Translation of hybrid-selected gene 10 mRNA in the presence of pancreatic microsomal membranes. mRNA was translated in the absence (c) and in the presence (d) of microsomal membranes, and the synthesized proteins were compared by polyacrylamide gel electrophoresis. (a) and (b) are identical to (f) and (e) in Fig. 1, respectively.

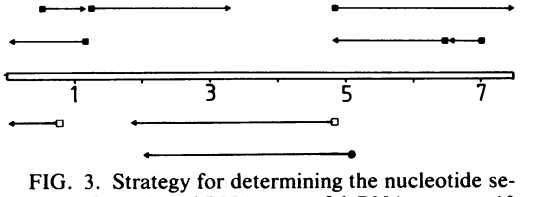

FIG. 3. Strategy for determining the nucleotide sequence of the cloned DNA copy of dsRNA segment 10 of SA11. The numbers refer to the distance in nucleotides times 10^{-2}. The sequence was determined either by copying cDNA from SA11 mRNA (□) (4a), by the method of Maxam and Gilbert (■) (14), or by the dideoxy method (19) after subcloning into phage M13 mp 7.1 (●) (16).

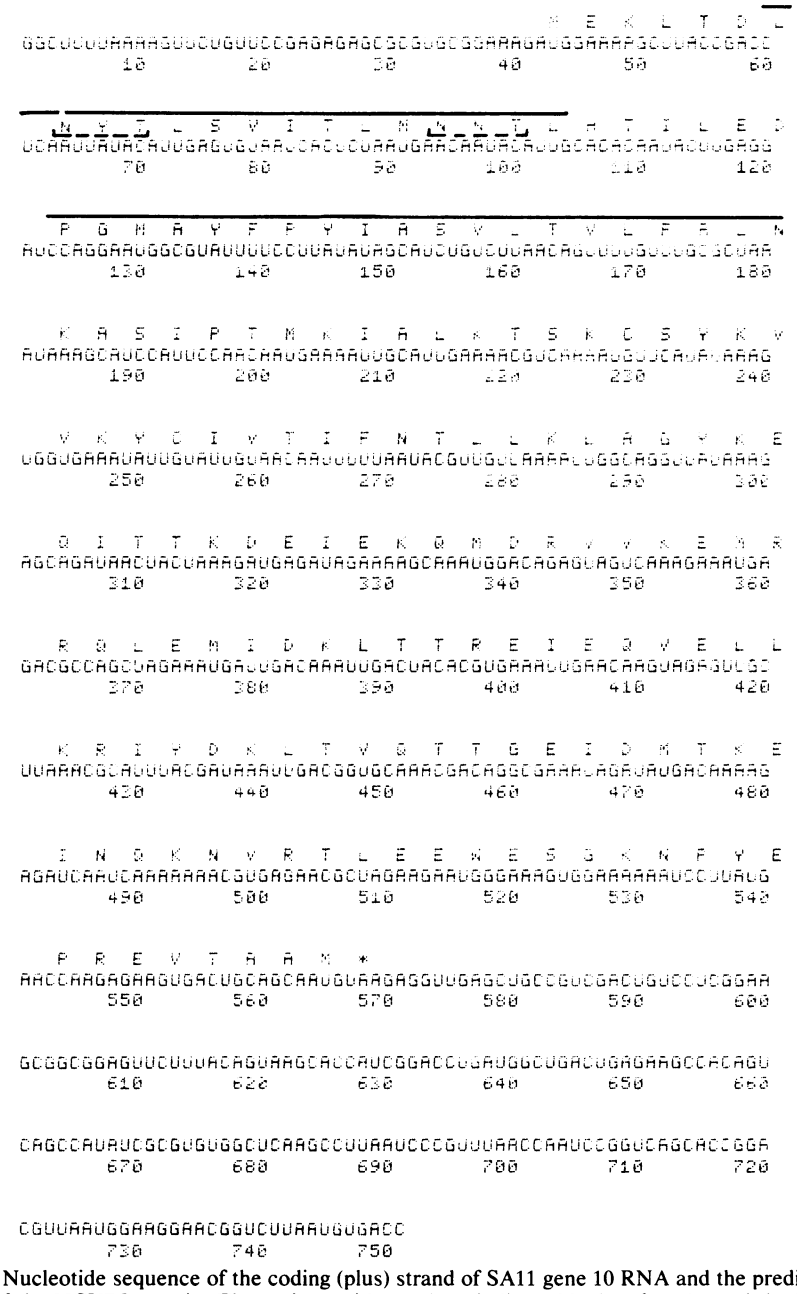

FIG. 4. Nucleotide sequence of the coding (plus) strand of SA11 gene 10 RNA and the predicted amino acid sequence of the NCVP5 protein. The regions with overbars indicate blocks of uncharged, hydrophobic amino acids. The broken lines indicate the glycosylation sites.

pNCVP5 to NCVP5 observed in cells infected with SA11 in the presence (Fig. 2b) and absence (Fig. 2a) of tunicamycin. These results confirm those of Arias et al. (1) and unequivocally assign NCVP5 as the product of SA11 gene 10 and confirm pNCVP5 as its precursor.

Structure and sequence of SA11 gene segment 10. Our cloned DNA copy of dsRNA segment 10 is known to be complete except for seven base pairs missing from the 5′ end of the plus strand (4). The sequence of gene 10, therefore, was determined by the strategy outlined in Fig. 3. The sequence of the missing nucleotides was determined from the mRNA (4).

The nucleotide sequence of the plus strand of gene 10 and the protein sequence which it pre-

dicts are shown in Fig. 4. The RNA segment was 751 bases in length and possessed 5'- and 3'-terminal noncoding regions of 41 and 185 bases, respectively. The 3'-terminal noncoding region was considerably longer, compared with the total length of the segment, than that observed for other rotavirus genes (4, 4a, 4b; unpublished observations). There was one open reading frame which coded for a protein of 175 amino acids. This predicts a size of 19.25K for pNCVP5, which is in good agreement with the observed value of 20K (7). The other two reading frames were open for no more than 46 codons. There were two potential sites for glycosylation, both close to the N-terminus, at asparagine residues 8 and 18.

Analysis of carbohydrates attached to NCVP5. [^{3}H]mannose-labeled SA11 proteins pulsed for 1 h were separated by polyacrylamide gel electrophoresis, and NCVP5 was digested with pronase and endoglycosidase H (4b). The resulting oligosaccharides were analyzed by cochromatography on Bio-Gel P4 with [^{14}C]mannose-labeled Sindbis virus marker oligosaccharides (Fig. 5). The SA11 oligosaccharides that were released were predominantly $Man_9GlcNAc$ with a small amount of $Man_8GlcNAc$. The compositions of the carbohydrates were consistent with those found in earlier studies in which the SA11 proteins became radiolabeled only when [^{3}H]mannose and [^{3}H]glucosamine were used (1, 7). There did not appear to be any sialic acid or galactose residues attached to NCVP5 (Fig. 5), in contrast to the Sindbis virus glycoproteins.

DISCUSSION

Examination of the deduced protein sequence revealed that the 21 N-terminal amino acids of NCVP5 include 15 consecutive uncharged residues (from 7 to 21), which potentially could fulfill the characteristics of a signal sequence of the type known to be present on other viral and cellular glycoproteins (2, 4a, 8, 17). There is also a second block of uncharged amino acids from residues 28 to 47 (Fig. 4). The hydrophobic profile of the first 50 residues of NCVP5 is, therefore, very similar to that observed in the equivalent region of the structural glycoprotein VP7, for which the feature of two regions of hydrophobic residues near the N-terminus was first described (4a, 4b).

Another feature of the sequence is the location of the only potential glycosylation sites of the Asn-X-Ser or Asn-X-Thr type, at asparagine residues 8 and 18, i.e., within the domain of the proposed signal sequence. Therefore, for the glycosylation sites to be retained, few, if any, amino acids could be cleaved from pNCVP5 during membrane translocation. The high-mannose composition of the NCVP5 carbohydrates (Fig. 5) is consistent with their sensitivity to tunicamycin (7) and endoglycosidase H (1) and implies a complex of ~1,864 daltons of carbohydrate attached to each of the two proposed glycosylation sites. However, the addition of a total of 3,728 daltons of carbohydrate is insufficient to account for the apparent size difference between pNCVP5 (19.3K) and NCVP5 (28K).

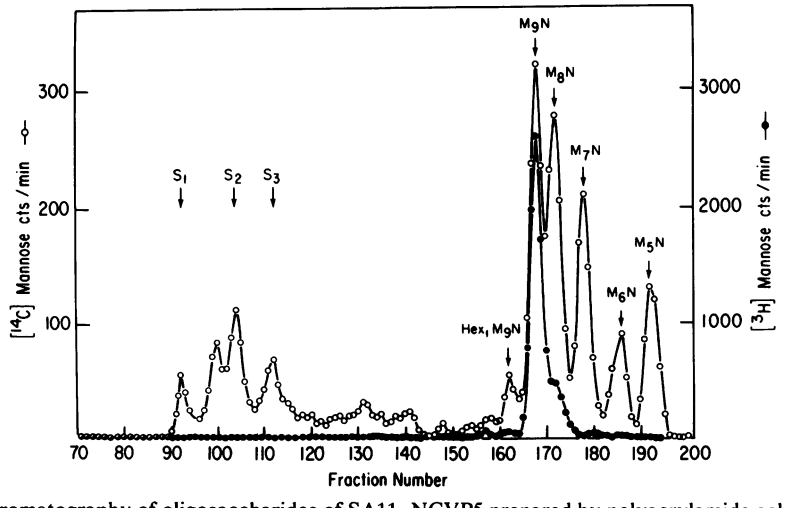

FIG. 5. Chromatography of oligosaccharides of SA11. NCVP5 prepared by polyacrylamide gel electrophoresis of purified SA11 labeled for 1 h with [^{3}H]mannose was exhaustively digested with pronase, and the glycopeptides were prepared and digested with endoglycosidase H (4b). [^{3}H]mannose-labeled NCVP5 (●) and [^{14}C]mannose-labeled Sindbis virus marker (○) oligosaccharides were separated by Bio-Gel P4 chromatography. cts, Counts. S_1, S_2, and S_3 indicate the positions of complex sialic acid-containing oligosaccharides.

This anomalous gel behavior may reflect the proximity of the glycosylation sites to the N-terminus, increasing the apparent length of a fairly small protein. The observation of only one (23K) glycosylated molecule intermediate in molecular weight between pNCVP5 and NCVP5 in partial digests with endoglycosidase H (1) and in biosynthesis studies (7) would support that there are only two glycosylation sites. Because of the rough endoplasmic reticulum location of NCVP5, the appearance of $Man_8GlcNAc$ (Fig. 5) suggests an α-1,2 mannosidase activity in this compartment.

The carbohydrate analysis of NCVP5 indicates that this protein, like VP7 (4b), is an example of a glycoprotein which is modified in the rough endoplasmic reticulum by an α-mannosidase but which does not reach the golgi apparatus of the cell (4b), where some of the oligosaccharide-trimming enzymes and glycosyltransferases are known to be located (10). If the prediction that NCVP5 is not cleaved during membrane translocation is correct, this unusual nonstructural viral glycoprotein would fall into a class similar to bovine opsin and the p62 precursor glycoproteins of Sindbis and Semliki Forest viruses. These proteins also remain uncleaved but become glycosylated at asparagine residues located within their putative signal polypeptides (3, 8, 17, 20).

During the preparation of this manuscript, we became aware of the work of Ericson et al. (B. L. Ericson, D. Y. Graham, B. B. Mason, H. H. Hanssen, and M. K. Estes, Virology, in press), who demonstrated that NCVP5 is not cleaved during membrane translocation.

ACKNOWLEDGMENTS

We thank Mary K. Estes for making a copy of her manuscript available before its publication. We also thank J. Mattick for comments on the manuscript.

This work was supported by grants from the New Zealand Medical Research Council, the New Zealand National Child Health Research Foundation, and the World Health Organization.

LITERATURE CITED

1. **Arias, C. F., S. López, and R. T. Espejo.** 1982. Gene protein products of SA11 simian rotavirus genome. J. Virol. **41**:42–50.
2. **Blobel, G., P. Walter, C. N. Chang, B. M. Goldman, A. H. Erikson, and V. R. Lingappa.** 1979. Translocation of proteins across membranes: the signal hypothesis and beyond. Proc. Soc. Exp. Biol. Med. **33**:9–36.
3. **Bonatti, S., and G. Blobel.** 1979. Absence of a cleavable signal sequence in Sindbis virus glycoprotein pE2. J. Biol. Chem. **254**:12261–12264.
4. **Both, G. W., A. R. Bellamy, J. E. Street, and L. J. Siegman.** 1982. A general strategy for cloning double-stranded RNA: nucleotide sequence of the Simian-11 rotavirus gene 8. Nucleic Acids Res. **10**:7075–7088.

4a. **Both, G. W., J. S. Mattick, and A. R. Bellamy.** 1983. The serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence. Proc. Natl. Acad. Sci. U.S.A. **80**:3091–3095.

4b. **Both, G. W., J. S. Mattick, L. J. Siegman, P. H. Atkinson, S. Weiss, A. R. Bellamy, J. E. Street, and P. Metcalf.** 1983. Cloning of SA11 rotavirus genes: gene structure and polypeptide assignment for the type-specific glycoprotein, p. 73–82. *In* D. H. L. Bishop and R. W. Compans (ed.), The double-stranded RNA viruses. Elsevier Biomedical, New York.

5. **Cleveland, D. W., M. A. Lopata, R. J. MacDonald, N. J. Cowan, W. J. Rutter, and M. W. Kirschner.** 1980. Number and evolutionary conservation of α-tubulin and β-tubulin and cytoplasmic β-actin and γ-actin genes using specific cloned complementary DNA probes. Cell **20**:95–106.
6. **Dyall-Smith, M. L., and I. H. Holmes.** 1981. Gene-coding assignments of rotavirus double-stranded RNA segments 10 and 11. J. Virol. **38**:1099–1103.
7. **Ericson, B. L., D. Y. Graham, B. B. Mason, and M. K. Estes.** 1982. Identification, synthesis, and modifications of simian rotavirus SA11 polypeptides in infected cells. J. Virol. **42**:825–839.
8. **Garoff, H., A.-M. Frischauf, K. Simons, H. Lehrach, and H. Delius.** 1980. Nucleotide sequence of the cDNA coding for Semliki Forest virus membrane glycoproteins. Nature (London) **288**:236–241.
9. **Holmes, I. H.** 1979. Viral gastroenteritis. Prog. Med. Virol. **25**:1–36.
10. **Hubbard, S. C., and R. J. Raymond.** 1981. Synthesis and processing of asparagine-linked oligosaccharides. Annu. Rev. Biochem. **50**:555–583.
11. **Katz, F. N., J. E. Rothman, V. R. Lingappa, G. Blobel, and H. F. Lodish.** 1977. Membrane assembly *in vitro*: synthesis, glycosylation and asymmetric insertion of a transmembrane protein. Proc. Natl. Acad. Sci. U.S.A. **74**:3278–3282.
12. **Mason, B. B., D. Y. Graham, and M. K. Estes.** 1980. In vitro transcription and translation of simian rotavirus SA11 gene products. J. Virol. **33**:1111–1121.
13. **Mattick, J. S., Z. E. Zehner, M. A. Calabro, and S. J. Wakil.** 1981. The isolation and characterization of fatty-acid-synthetase mRNA from rat mammary gland. Eur. J. Biochem. **114**:643–651.
14. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
15. **McCrae, M. A., and W. K. Joklik.** 1978. The nature of the polypeptide encoded by each of the 10 double-stranded RNA segments of reovirus type 3. Virology **89**:578–593.
16. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
17. **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.
18. **Rodger, S. M., and I. H. Holmes.** 1979. Comparison of the genomes of simian, bovine, and human rotaviruses by gel electrophoresis and detection of genomic variation among bovine isolates. J. Virol. **30**:839–846.
19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
20. **Schechter, I., Y. Burgstein, R. Zamell, E. Ziv, F. Kantor, and D. S. Papermaster.** 1979. Messenger RNA of opsin from bovine retina: isolation and partial sequence of the *in vitro* translation product. Proc. Natl. Acad. Sci. U.S.A. **76**:2654–2658.