Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6975–6979, October 1987
Biochemistry

Primary and secondary structure of bovine retinal S antigen (48-kDa protein)

(amino acid sequence/cDNA/vision/retina)

T. SHINOHARA*, B. DIETZSCHOLD†, C. M. CRAFT*, G. WISTOW*, J. J. EARLY†, L. A. DONOSO‡, J. HORWITZ§, AND R. TAO¶

*Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, MD 20892; †The Wistar Institute of Anatomy and Biology, Philadelphia, PA 19104; ‡Wills Eye Hospital, Philadelphia, PA 19107; §Jules Stein Institute, University of California–Los Angeles Medical School, Los Angeles, CA 20024; and ¶Lions of Illinois Eye Research Institute, College of Medicine, University of Illinois, Chicago, IL 60612

Communicated by Igor B. Dawid, May 20, 1987

ABSTRACT **The complete amino acid sequence of bovine S antigen (48-kDa protein) has been determined by cDNA and partial amino acid sequencing. A 1623-base-pair (bp) cDNA contains an open reading frame coding for a protein of 404 amino acids (45,275 Da). Tryptic peptides and cyanogen bromide peptides of native bovine S antigen were purified and partially sequenced. All of these peptides were accounted for in the long open reading frame. Searching of the National Biomedical Research Foundation data bank revealed no extensive sequence homology between S antigen and other proteins. However, there are local regions of sequence similarity with α transducin, including the sites subject to ADP-ribosylation by Bordetella pertussis and cholera toxins and the phosphoryl binding-sites. Secondary structure prediction and circular dichroic spectroscopy show that S antigen is composed predominantly of β-sheet conformation. Acid-catalyzed methanolysis suggests the presence of low levels of carbohydrate in the molecule.**

The visual process begins with the absorption of a photon by rhodopsin in the retinal rod outer segments (ROS). This event triggers the activation of an amplifying cascade of ROS proteins, including a guanine nucleotide-binding protein (G protein; transducin) and a phosphodiesterase, eventually generating a nerve impulse (1, 2). This system has strong similarities with the G protein/adenylate cyclase system for hormone receptor signal transduction (3–5). S antigen is another major protein of the ROS (6–8). It has recently been identified as the 48-kDa ("48K") protein that binds to photoactivated–phosphorylated rhodopsin, thereby apparently preventing the transducin-mediated activation of phosphodiesterase (9, 10). These findings suggest that S antigen has an important role in phototransduction and may represent a class of proteins involved in the modulation of signal transduction.

S antigen is also a highly pathogenic protein and is responsible for the induction of experimental autoimmune uveitis (6, 7, 11, 12), a T-cell-mediated disease (13) that results in the destruction of photoreceptor cells of the retina and the pinealocytes of the pineal gland (14).

Here we report the complete primary sequence of S antigen deduced from its cDNA and confirmed by partial amino acid sequencing.‖ The results of secondary structure prediction and carbohydrate analysis are also presented.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Preparation of S Antigen and Antisera. S antigen was purified from bovine retinas as described (15, 16). Polyclonal antisera were provided by I. Gery (National Eye Institute, Bethesda, MD). Two monoclonal antibodies, MAbA9-C6 and MAbA1-G5, were obtained as described (17, 18).

Preparation of Tryptic and CNBr Peptides. One milligram of carboxymethylated and iodoacetamine-treated S antigen was cleaved with L-1-*P*-tosylamino-2-phenylethyl chloromethyl ketone-treated trypsin (1 mg/ml) for 24 hr at 37°C. The resulting tryptic peptides were purified by two cycles of reverse-phase HPLC. The peak fractions were collected (19). S antigen (500 μg) was cleaved by 100 mg of CNBr in 1 ml of 70% formic acid for 24 hr. After lyophilization, the CNBr-cleavage peptides were purified by PAGE (5% stacking gel; 25% separation gel) followed by electroelution (20, 21).

Peptide Sequence Determination. The tryptic or CNBr-cleavage peptides were sequenced by Edman degradation on a gas-phase sequencer (22) (model 470A, Applied Biosystems, Foster City, CA). Analysis of the N- and C-terminal amino acid residues was performed by the Harvard University Biological Laboratories (Cambridge, MA).

cDNA Libraries and Isolation of cDNA Clones. Two bovine phage λgt11 retinal libraries and one λgt10 retinal library were used. One bovine λgt11 retinal library was provided by D. Oprian (Massachusetts Institute of Technology, Cambridge, MA), and the other was purchased from Clontech Laboratories (Palo Alto, CA). An additional λgt10 library was provided by J. Nathans (Stanford, CA). The methods for isolation of cDNA clones from these libraries have been described (23, 24).

cDNA Sequence Determination. The *Eco*RI fragment cDNA inserts obtained from screening the cDNA libraries were subcloned into phage M13 mp19 or mp18 vector by using reagents from Bethesda Research Laboratories. DNA sequences were obtained by the dideoxy chain-termination method (25) modified for adenosine [(^{35}S)thio]triphosphate (Amersham) (24). Specific oligonucleotide primer probes (15- or 20-mer) were synthesized chemically on an automated DNA synthesizer (model 380B, Applied Biosystems).

CD Spectral Analysis. CD spectra were recorded on a modified Beckman CD spectrophotometer as described (26). The CD intensity was calibrated with an aqueous solution of D-10 camphorsulfonic acid ($\Delta\varepsilon$ = 2.20 $M^{-1}\cdot cm^{-1}$) (27).

Abbreviations: ROS, rod outer segment; G protein, guanine nucleotide-binding protein.

‖This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02955).

TP 1	L M H P Q P E D P D T A K (355-367)
TP 2	A G E Y K E E K (387-395)
TP 3	D Y I D H V E R (30-37)
TP 4	L A V S L S K (205-211)
TP 5	A F A T H S T D V E E D K I P (151-165)
TP 6	S C G V D F E I K (142-150)
TP 7	S V T I Y L G K (21-28)
TP 8	I K H E D T N L A S (299-308)
TP 9	E I Y Y H G E P I P V T - A V (212-226)
TP 10	V Y V *D L T C A - R (57-67)
TP 11	D L Y F *D Q V Q V - P - V G A (82-97)
TP 12	V E P V D - V V L V *S P E L V (38-52)
TP 13	V M G L (74-77)
TP 14	A E A S W Q F F M S D K (190-201)
TP 15	K V Q H A P R (176-182)
TP 16	D M G P Q P R (183-189)
TP 17	G I *Q V (322-325)
TP 18	D A G E Y (387-391)
TP 19	M S D K P L (198-203)
TP 20	S S D Y Y I K (251-257)
TP 21	L Q D S L I K (103-109)
TP 22	L G A N T Y P (111-117)

FIG. 1. N-terminal partial amino acid sequences of tryptic fragments. Numbers at the end of the sequences indicate the position of these peptides. Asterisks indicate residues not agreeing with the cDNA sequence.

Protein concentration was determined by the method of Whitaker and Granum (28).

Carbohydrate Analysis. S antigen (0.5 mg) was desalted by dialysis and subjected to acid-catalyzed methanolysis (29). The recovered methyl glycosides were estimated by capillary GLC as *O*-trimethylsilyl derivatives (30).

RESULTS

Amino Acid Sequences. Thirty-nine tryptic peptides were resolved by HPLC, with an average yield of ≈75%. Unique and unambiguous sequence data were obtained from 22 peptides (Fig. 1).

Five major and five minor CNBr-cleavage peptides were isolated. Amino acid sequence data were obtained from seven CNBr-cleavage fragments (Fig. 2). Although the N terminus of native S antigen is blocked (31, 32), a 17-amino acid sequence was obtained from S antigen purified by HPLC (20% yield) as reported (23). This sequence was identical to that of CNBr-cleavage peptide C6 (Fig. 2), which lacks the N-terminal tripeptide Met-Lys-Ala. The blocked N-terminal

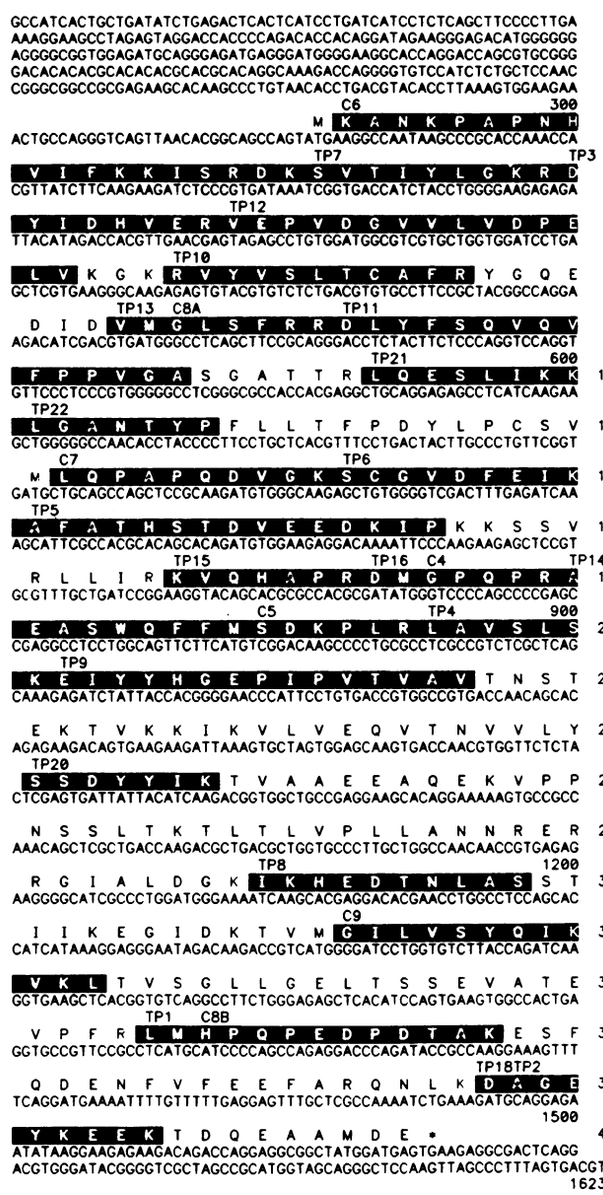

FIG. 3. Nucleotide sequence of the cDNA and predicted amino acid sequence for bovine retinal S antigen. The regions corresponding to the sequences of the tryptic and CNBr-treated peptides are shown boxed. Subscript numbers indicate nucleotides from the 5′ end. The amino acid sequence is numbered at the ends of the lines. Numbers over boxes indicate tryptic fragments or CNBr-cleavage fragments. An asterisk marks the termination codon.

C4	G P Q P R A E A S W Q F F (185-197)
C5	S D K P L R L A V S L S K E I Y Y H G E P I P V T V A V (199-226)
C6	K A N K P A P N H V I F K K I S R D K S V T I Y (2-25)
C7	L Q P *E P Q D V G K S C G V D F E I K A F A T H S T D V E E D K I P P (132-165)
C8A	G L S F R R D L Y F S Q V Q V (76-90)
C8B	H P Q P E D P D T A K (357-367)
C9	G I L V S Y Q I K V K L (322-333)

FIG. 2. N-terminal partial amino acid sequences of CNBr-cleavage fragments. Their position in the sequence is indicated. The asterisk indicates a residue not agreeing with the cDNA sequence.

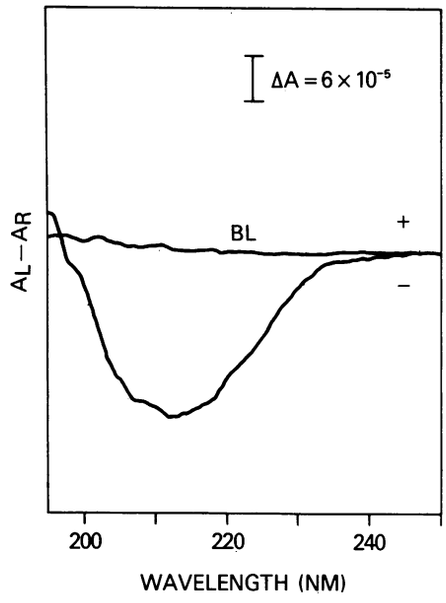

FIG. 4. Far-ultraviolet CD spectrum of retinal S antigen. The spectrum represents the average of 16 scans. The protein concentration was 0.6 mg/ml; the pathlength was 0.2 mm. A_L, left-handed polarized light; A_R, right-handed polarized light; BL, base line.

amino acid was inferred to be methionine, while the C-terminal amino acid was determined to be glutamic acid.

cDNA Analysis. Initially, several incomplete cDNAs were obtained by immunoscreening of the λgt11 libraries. Larger cDNAs were then obtained by hybridization, and a partial sequence for S antigen was deduced (23, 24). Finally, a cDNA insert, BSC70 [1623 base pairs (bp)], was obtained and sequenced.

The nucleotide sequence of the BSC70 cDNA insert is shown in Fig. 3. The insert contains a long open reading frame that begins at base 332 and encodes 404 amino acids, corresponding to an unmodified molecular mass of 45,275 Da.

CD Spectral Analysis. The far-ultraviolet CD spectrum of purified S antigen is shown in Fig. 4. The CD profile is typical of all-β-sheet proteins, exhibiting a broad, relatively weak minimum in the 210 to 220 nm range (33, 34). The α-helical content was calculated as no more than 10% (Table 1). These results are in good agreement with the secondary structure predicted from the deduced amino acid sequence by the method of Garnier *et al.* (36).

Carbohydrate Analysis. GLC of S antigen subjected to acid-catalyzed methanolysis showed the presence of mannose and other carbohydrates. Total carbohydrate bound to S antigen was estimated as 4 mol/mol of S antigen, quantified by GLC using mannitol as an internal standard. Native S antigen was also subjected to direct trimethylsilylation and analyzed by GLC, revealing no traces of contaminating carbohydrates.

Table 1. Estimation of the α-helical fraction in retinal S antigen

Wavelength, nm	Extinction coefficient, $M^{-1} \cdot cm^{-1}$	% helix
222	524	13.0*
208	769	9.2†

*The percent helix was calculated by the method of Chen *et al.* (33) and Chang *et al.* (34), assuming a protein with a helix having an average length of six residues.

†The percent helix was calculated by the method of Greenfield and Fasman (35).

DISCUSSION

The open reading frame between bases 332 and 1545 appears to contain the entire coding sequence for S antigen, a protein consisting of 404 amino acids with a calculated molecular mass of 45,275 Da. The sequence accounts for all of the tryptic and CNBr-treated peptides obtained for authentic S antigen and includes glutamic acid as the C-terminal amino acid. The N-terminal sequence of the HPLC-purified S antigen (23) matches CNBr-cleavage peptide C6, except for cleavage between Ala-3 and Asn-4. The codon at base 332 is in a sequence that conforms well to the initiator consensus (CAGTATG, compared with CANCATG) (37). However, the cDNA insert contains a fairly long (331 base) 5′ sequence that has some unusual features. It contains three ATG

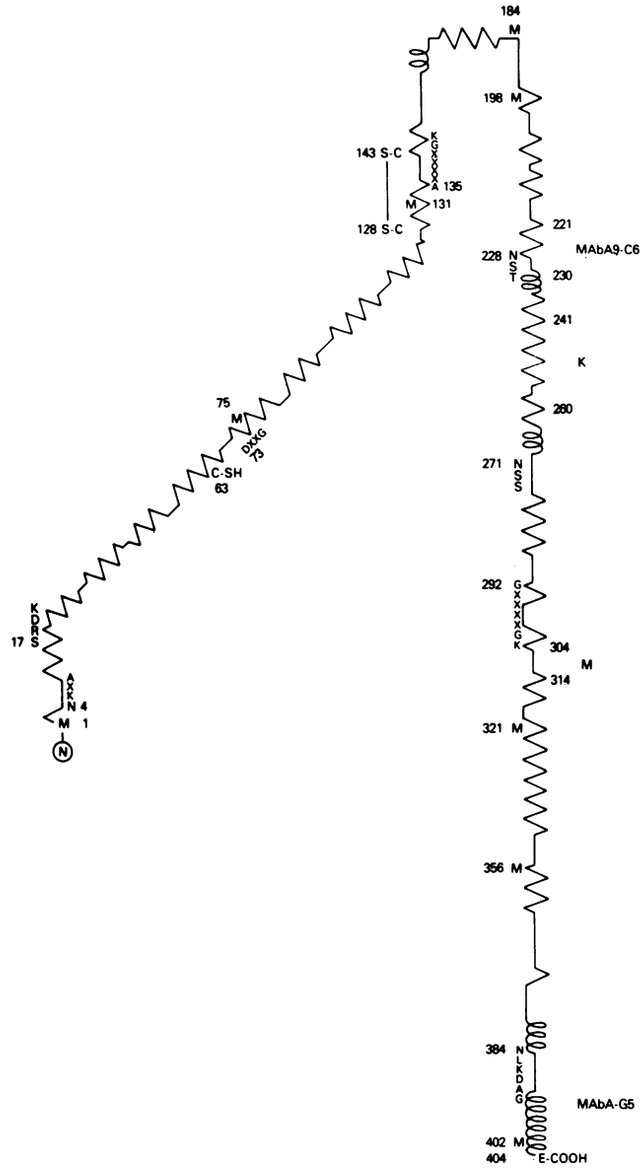

FIG. 5. Schematic diagram of the predicted secondary structure of S antigen. The secondary structure predictions were generated by the method of Garnier *et al.* (36). Residues are represented as α-helical (loop pattern), β-sheet (saw-tooth pattern), and random coil (straight line). Predicted turns are represented according to the number of residues confidently predicted at 30° per residue. Methionine and cysteine residues and the other interesting sequences are numbered. MAb6A9-C6 and MAbA-G5 indicate the monoclonal antibody binding sites (24, 44). K (near 241) and M (near 304) indicate the uveitopathogenic sites (44).

codons, none of which has a close match to the initiator consensus (37). Two of them are in-frame with the putative initiation codon and are part of an open reading frame (bases 113–171) that could encode 53 amino acids. This sequence has no similarity with any of the tryptic or CNBr-cleavage peptides sequenced for S antigen or with any sequences in the National Biomedical Research Foundation data base. The significance of this unusual 5′ sequence remains to be determined. The cDNA seems to be incomplete at the 3′ end because no polyadenylylation signal is apparent. A full size of 1700 nucleotides has been estimated for bovine S antigen mRNA (23).

Although the function of S antigen in the visual system is still not well understood, a possible relationship with purine nucleotide-binding proteins exists. Both S antigen and the G-protein α transducin interact with photoactivated rhodopsin, and there are local sequence similarities between them. As described previously (24), a sequence near the C terminus of S antigen (371–391) is similar to the C-terminal sequence of α transducin (326–350), including the site of ADP-ribosylation by pertussis toxin (38), although the modifiable residue, Cys-347 is alanine in S antigen. Another similarity is found between residues 17 and 20 of S antigen (Ser-Arg-Asp-Lys) and between residues 173 and 176 of transducin (Ser-Arg-Val-Lys), the site for ADP ribosylation by cholera toxin (39). Since the modification of α transducin by these toxins presumably occurs at functionally important sites, the similarity with S antigen suggests possible related functions.

G proteins in general contain highly conserved sequences for GTP/GDP-binding. The most characteristic of these are Gly/Ala-Xaa-Xaa-Xaa-Xaa-Gly-Lys and Asp-Xaa-Xaa-Gly at the phosphoryl-binding site and Asn-Lys-Xaa-Asp at the guanine-binding site (40, 41). Residues 135–142 and 292–298 of S antigen have the sequences Ala-Pro-Gln-Asp-Val-Gly-Lys and Gly-Ile-Ala-Leu-Asp-Gly-Lys, respectively, and residues 73–76 of S antigen has the sequence Asp-Val-Met-Gly, which fits this pattern. However, S antigen has no potential guanine-binding site, the closest sequence being Asn-Lys-Pro-Ala at positions 4–7.

The predicted sequence contains three cysteine residues at positions 63, 128, and 143, located in three separate CNBr-cleavage peptides C6, C7, and C8A. Peptides C7 and C8A are found cross-linked after CNBr cleavage of unreduced S antigen, suggesting the existence of a disulfide bridge between Cys-128 and Cys-143.

The sequence also reveals two Asn-Xaa-$\begin{matrix}\text{Thr}\\\text{Ser}\end{matrix}$ sequences (at residues 228 and 271), which are characteristic of N-linked carbohydrate attachment sites (42). Indeed, carbohydrate analysis suggested the presence of at least 4 mol of sugar per mol of S antigen. Low, variable levels of glycosylation in S antigen may explain the differing estimates of molecular size (43).

The results of CD analysis (Fig. 4) were used to determine the initial values of the decision constants, and the method of Garnier *et al.* (36) was used to predict the secondary structure of S antigen. A total of 54% extended (β-sheet) conformation was predicted, with only 8% α-helix—mostly in a region at the protein C terminus. A schematic representation of the predicted secondary structure and other features is shown in Fig. 5.

Some regions of internal similarity (rather than identity) were detected by using the program DIAGON (by R. Staden) (Fig. 6). In particular, residues 76–143 and 206–273 (region A) have strikingly similar hydrophobicity profiles (Fig. 6) and can be aligned to show a general similarity in amino acid sequence, perhaps reflecting a repeated structure in S antigen.

The partial amino acid sequence of S antigen (24) has already proved useful in identifying sites in the molecule responsible for the pathogenesis of experimental autoimmune uveitis (15).

As this work was being completed, a cDNA sequence for bovine S antigen from the same cDNA library was published by Yamaki *et al.* (44). The predicted amino acid sequences are in agreement, except for four residues in their sequence that appear to be misidentified due to typographical errors (K. Yamaki, personal communication).

We thank D. Oprian and J. Nathans for providing retinal cDNA libraries, I. Gery for providing S-antigen antisera, A. Katial for DNA

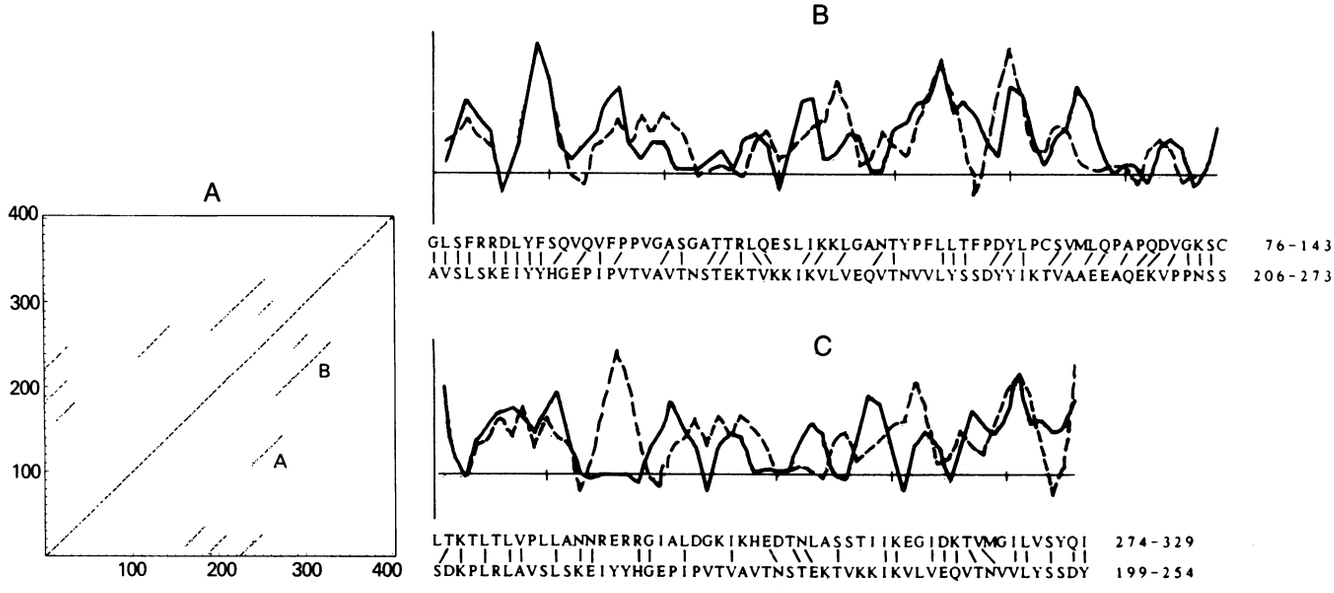

FIG. 6. Search for regions of internal similarity in S antigen. (*A*) A self-comparison matrix produced by DIAGON searching for regions of similarity. Span length was 99; similarity score was 1005; and the significance level was 0.65×10^{-3}. The two major regions of similarity are labeled A and B. (*B*) Region A: residues 76–143 and 206–273 of S antigen. (*B Upper*) Qualitative comparison of hydrophobicity profiles by the method of Nozaki and Tanford (45) as implemented in HCOMP in the IDEAS package. (*B Lower*) Comparison of the amino acid sequences. Lines connect identical or structurally conservative residues that occupy similar positions in the two sequences. (*C*) Part of region B: residues 199–254 and 274–329 of S antigen. (*C Upper*) Hydrophobicity comparison as in *B Upper*. (*C Lower*) Amino acid sequence comparison as in *B Lower*.

sequencing, and Mrs. D. Chicchirichi for expert typing. This work was supported in part by National Institutes of Health Grant EY5095 (to L.A.D.).

1. Kuhn, H. (1984) in *Progress in Retinal Research*, eds. Osborne, N. & Chader, G. (Pergamon, New York), Vol. 3, pp. 123–156.
2. Shichi, H. (1983) *Biochemistry of Vision* (Academic, New York).
3. Gilman, A. G. (1984) *Cell* **36,** 577–579.
4. Yatsunami, K. & Khorana, H. G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4316–4320.
5. Bray, P., Carter, A., Simon, C., Guo, V., Puckett, C., Kamholz, J., Spiegel, A. & Nirenberg, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8893–8897.
6. Wacker, W. B., Donoso, L. A., Kalsow, D. M., Yankeelov, J. A., Jr., & Organisciak, D. T. (1977) *J. Immunol.* **119,** 1949–1958.
7. Faure, J. P. (1980) *Curr. Top. Eye Res.* **2,** 215–302.
8. Kuhn, H. (1978) *Biochemistry* **17,** 4389–4395.
9. Pfister, C., Chabre, M., Plouet, J., Tuyen, V. V., DeKozak, Y., Faure, J. P. & Kuhn, H. (1985) *Science* **228,** 891–893.
10. Wilden, U., Hall, S. W. & Kuhn, H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1174–1178.
11. Nussenblatt, R. B., Kuwabara, R., deMonasterio, F. M. & Wacker, W. B. (1981) *Arch. Ophthalmol.* **99,** 1090–1092.
12. Nussenblatt, R. B., Gery, I., Ballintine, E. J. & Wacker, W. B. (1980) *Am. J. Ophthalmol.* **89,** 173–179.
13. Mochizuki, M., Kuwabara, T., McAllister, C., Nussenblatt, R. B. & Gery, I. (1985) *Invest. Ophthalmol. Visual Sci.* **26,** 1–14.
14. Kalsow, C. M. & Wacker, W. B. (1978) *Invest. Ophthalmol. Visual Sci.* **17,** 774–783.
15. Donoso, L. A., Merryman, C. F., Shinohara, T., Dietzschold, B., Wistow, G., Craft, C., Morley, W. & Henry, R. T. (1986) *Curr. Eye Res.* **5,** 995–1004.
16. Zigler, J. S., Mochizuki, M., Kuwabara, T. & Gery, I. (1984) *Invest. Ophthalmol. Visual Sci.* **25,** 977–980.
17. Donoso, L. A., Merryman, C. F., Edelberg, K. E., Naids, R. & Kalsow, C. (1985) *Invest. Ophthalmol. Visual Sci.* **26,** 561–567.
18. Witt, P. L., Hamm, H. E. & Bownds, M. D. (1984) *J. Gen. Physiol.* **84,** 251–263.
19. Mahoney, W. C. (1982) *Biochim. Biophys. Acta* **704,** 284–289.
20. Gross, E. (1967) *Methods Enzymol.* **11,** 238–255.
21. Higgins, R. C. & Dahmus, M. E. (1979) *Anal. Biochem.* **93,** 257–260.
22. Hewick, R. M., Hunkapilar, M. W., Hood, L. E. & Dreyer, W. J. (1981) *J. Biol. Chem.* **256,** 7990–7997.
23. Shinohara, T., Craft, C., Stein, P., Zigler, S., Wistow, G., Katial, A., Gery, I. & Klein, D. (1986) in *Pineal and Retinal Relationships*, eds. O'Brien, P. & Klein, D., pp. 331–342.
24. Wistow, G., Katial, A., Craft, C. & Shinohara, T. (1986) *FEBS Lett.* **196,** 23–28.
25. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
26. Horwitz, J., Bullard, B. & Mercola, D. (1979) *J. Biol. Chem.* **254,** 350–355.
27. Cassim, J. Y. & Yang, J. T. (1969) *Biochemistry* **8,** 1947–1950.
28. Whitaker, R. J. & Granum, E. P. (1980) *Anal. Biochem.* **109,** 156–159.
29. Sweeley, C. C., Wall, W. W. & Bentley, A. (1966) *Methods Enzymol.* **8,** 95–108.
30. Reinhold, V. N. (1972) *Methods Enzymol.* **25,** 244–249.
31. Gregerson, D. S. & Putterman, G. J. (1984) *J. Immunol.* **133,** 843–848.
32. Gregerson, D. S., Fling, S. P. & Wohlhueter, R. M. (1986) *Exp. Eye Res.* **43,** 803–818.
33. Chen, H. Y., Yang, T. J. & Chau, H. K. (1974) *Biochemistry* **13,** 3350–3359.
34. Chang, T. C., Wu, C. S. C. & Yang, T. J. (1978) *Anal. Biochem.* **91,** 13–31.
35. Greenfield, N. & Fasman, D. G. (1969) *Biochemistry* **8,** 4108–4116.
36. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
37. Kozak, M. (1984) *Nucleic Acids. Res.* **12,** 857–873.
38. Manning, D. R., Fraser, B. A., Kahn, R. A. & Gilman, A. G. (1984) *J. Biol. Chem.* **259,** 749–756.
39. VanDop, C., Tsubokawa, M., Bourne, H. R. & Ramachandran, J. (1984) *J. Biol. Chem.* **259,** 696–698.
40. Dever, T. E., Glynias, M. J. & Merrick, W. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1814–1818.
41. McCormic, F., Clark, B. F. C., LaCour, T. F. M., Kjelgaard, M., Norskov-Lauristen, L. & Nyborg, J. (1985) *Science* **230,** 78–82.
42. Sharon, N. (1975) *Complex Carbohydrates: Their Chemistry, Biosynthesis, and Function* (Addison-Wesley, Reading, MA), p. 108.
43. Broekhuyse, R. M. & Bessens, H. J. H. (1985) *Exp. Eye Res.* **40,** 763–766.
44. Yamaki, K., Takahashi, Y., Sakuragi, S. & Matsubara, K. (1987) *Biochem. Biophys. Res. Commun.* **142,** 904–910.
45. Nozaki, Y. & Tanford, C. (1971) *J. Biol. Chem.* **246,** 2211–2217.