munoassay (Fig. 2b), is shown by a titration curve practically identical with the standard curve for S antigen. Moreover, purified 48K protein induces the development of the characteristic pathology of experimental autoimmune uveitis with about the same efficiency as that observed with S antigen (Fig. 3a). The antibody levels in rats immunized with either protein are also similar (Fig. 3b).

These quantitative comparisons exclude the possibility that the pathological effects could be caused by a contaminant in one or both preparations. Because of the different purification procedures, there is little chance of a common contaminant, as confirmed in lanes 1 and 2 of Fig. 1. The 48K preparation contained as the major contaminant a few percent of a polypeptide of about 70 kD, most probably rhodopsin kinase [which copurifies with the 48K protein since it also binds reversibly to bleached disk membranes (*6*)]. In contrast, the S antigen preparation contained neither the 70-kD polypeptide, nor rhodopsin kinase activity (*14*). We therefore conclude that, in both preparations, the 50-kD polypeptide is responsible for the immunopathogenic properties of S antigen (Figs. 2 and 3). Zigler *et al.* (*15*) reported that extensively purified S antigen, containing only the 50-kD polypeptide, as shown by silver staining on gels, induces experimental autoimmune uveitis.

Immunofluorescent labeling of retinal sections has shown that S antigen is located exclusively in photoreceptor cells—not only in the outer segment but also in the inner segment and the synaptic terminal region (*2, 12*). This diffuse pattern may be related to intracytoplasmic diffusion of this highly soluble protein from its site of biosynthesis, the inner segment. Because the protein is absent from all other cells in the retina, in brain, and in other tissues (except the pinealocytes (*2*), which are phylogenetically related to photoreceptors (*16*), it is likely that its function is essentially restricted to phototransduction.

C. PFISTER
M. CHABRE
Biophysique Moléculaire et Cellulaire, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires de Grenoble, BP 85X, F-38041 Grenoble, France
J. PLOUET, V. V. TUYEN
Y. DE KOZAK, J. P. FAURE
Laboratoire d'Immunopathologie de l'Oeil, Hôtel Dieu, F-75181 Paris 04, France
H. KÜHN
Institut für Neurobiologie der KFA, D-5170 Jülich, Federal Republic of Germany

References and Notes

1. J. P. Faure, *Curr. Top. Eye Res.* **2**, 215 (1980).
2. W. B. Wacker, L. A. Donoso, C. M. Kalsow, J. A. Yankeelow, Jr., D. T. Organisciak, *J. Immunol.* **119**, 1949 (1977).
3. C. Dorey, J. Cozette, J. P. Faure, *Ophthalmic Res.* **14**, 249 (1982). Briefly, bovine retinas were suspended in 10 m*M* phosphate buffer, *p*H 7.6. The extracted soluble proteins were subjected to $(NH_4)_2SO_4$ precipitation (50 percent saturation) and then to chromatography, first on an Ultrogel AcA 34 molecular sieve column and then on an HA Ultrogel column. The fractions were tested at each step for their reaction with antibodies to S antigen. The active fractions were pooled and stored lyophilized. The yield was about 250 μg of S antigen per retina, or about 10 g S antigen molecules per 100 rhodopsins.
4. W. B. Wacker, N. A. Rao, G. E. Marak, Jr., *Ophthalmic Res.* **13**, 302 (1981).
5. Y. De Kozak, J. Sakai, B. Thillaye, J. P. Faure, *Curr. Eye Res.* **1**, 327 (1981).
6. H. Kühn, *Biochemistry* **17**, 4389 (1978).
7. H. Kühn, S. W. Hall, U. Wilden, *FEBS Lett.* **176**, 473 (1984).
8. U. Wilden and H. Kühn, in preparation.
9. For molecular mechanisms of photoreceptor transduction, see *Current Topics in Membranes and Transport*, F. Bonner and W. H. Miller, Eds. (Academic Press, New York, 1981), vol. 15.
10. C. Pfister, H. Kühn, M. Chabre, *Eur. J. Biochem.* **136**, 489 (1983); R. Zuckerman, B. Buzdygon, P. Liebman, *Biophys. J.* **45** (Abstr.), 292 (1984); N. J. Mangini *et al.*, paper presented at the ARVO Conference at Sarasota, Fla., May 1984 (abstract 2, p. 112, of conference proceedings).
11. H. Kühn, *Methods Enzymol.* **81**, 556 (1982). Briefly, extensively washed, dark-kept disk membranes were combined with an isotonic extract of dark-kept ROS in a medium consisting of 120 m*M* KCl, 20 m*M* Hepes, 1 m*M* $MgCl_2$, and 1 m*M* dithiothreitol, *p*H 7.0. The suspension was illuminated with orange light. The 48K protein, bound to R*, was sedimented with the membranes. Unbound proteins, contained in the supernatants upon repeated washings, were discarded. The 48K–containing membranes were incubated in buffer for 30 minutes at 25°C in the dark, reversing the binding of the 48K protein, which was then recovered in the supernatant.
12. J. P. Faure *et al.*, *Curr. Eye Res.* **3**, 867 (1984).
13. C. Pfister *et al.*, *C. R. Acad. Sci.* **299**, 261 (1984).
14. H. Shichi and R. L. Somers have proposed that rhodopsin kinase activity is associated with a ROS protein of 50 to 55 kD [*J. Biol. Chem.* **253**, 7040 (1978)]. The question arises if it is the S antigen that carries this kinase activity, as has been suggested by H. Shichi [*Jpn. J. Ophthalmol.* **25**, 306 (1981)], R. B. Nussenblatt *et al.* [*Br. J. Ophthalmol.* **65**, 778 (1981)], and R. L. Somers and D. C. Klein [*Science* **226**, 182 (1984)]. We find, in contrast, that purified S antigen is totally devoid of rhodopsin kinase activity but retains the properties of the 48K protein; that is, it binds to phosphorylated R* and quenches phosphodiesterase activity. In routine purification of S antigen (*3*), some rhodopsin kinase activity is found in the 50-kD peak eluted from the molecular sieve column; however, the hydroxyapatite column then completely separates S antigen (eluted at 100 m*M* phosphate) from rhodopsin kinase (eluted at 500 m*M* phosphate). Moreover, rhodopsin kinase activity is separated from S antigen activity (assayed immunologically) by gel filtration of soluble proteins extracted from ROS with Sephadex G-100; the kinase activity peak is eluted at about 65 kD, and the S antigen peak at 50-kD (data not shown). Finally, one of us (H.K.) has shown [*Neurochem. Int.* **1**, 269 (1980)] that ATP has opposite effects on the light-induced binding of the two proteins to disk membranes, strengthening the binding of 48K protein but weakening the binding of rhodopsin kinase. We therefore conclude that rhodopsin kinase and S antigen are two different proteins.
15. J. S. Zigler, Jr., *et al.*, *Invest. Ophthalmol. Visual Sci.* **22**, 977 (1984).
16. J. P. Collin and A. Oksche, in *The Pineal Gland*, R. J. Reiter, Ed. (CRC Press, Boca Raton, Fla., 1981), vol. 1, pp. 27–67.
17. The gel system of Laemmli [*Nature (London)* **227**, 680 (1970)] with 10 percent acrylamide was used. Gels were stained with Coomassie brilliant blue.
18. A 10 percent acrylamide-SDS Laemmli gel was run as in Fig. 1. The proteins were transferred to a nitrocellulose sheet by contact during 5 days at 4°C. Corresponding lanes on the same gel were subjected to Coomassie blue staining. The blot was saturated with bovine serum albumin, then incubated successively with the monoclonal antibody (overnight at 20°C) and finally with goat antibody to mouse immunoglobulin G coupled to peroxidase. The peroxidase activity was detected by aminobenzidine in the presence of Ni^{2+} and Co^{2+} salts [A. L. de Blas and H. M. Cherwincki, *Anal. Biochem.* **133**, 214 (1983)].
19. A very faint band appears at lower molecular weight, which could be due to a degradation product of the 48K protein.
20. M. M. Bradford, *Anal. Biochem.* **72**, 248 (1976).
21. Y. de Kozak *et al.*, *Jpn. J. Ophthalmol.* **27**, 598 (1983).
22. Fruitful discussions with E. Vadot and collaboration with J. Bigay, C. Boucheix, P. Deterre, C. Dorey, S. W. Hall, and M. Mirshahi are gratefully acknowledged. Supported by grant U. 86 from Institut National de la Sante et de la Recherche Médicale, grants U.A. 520 and 619 from Centre National de la Recherche Scientifique, by Universities of Paris VI and Grenoble I, and by grant SFB 160 from Deutsche Forschungsgemeinschaft.

25 January 1985; accepted 27 March 1985

Cassette of Eight Exons Shared by Genes for LDL Receptor and EGF Precursor

Abstract. *The amino acid sequences of the human low-density lipoprotein (LDL) receptor and the human precursor for epidermal growth factor (EGF) show 33 percent identity over a stretch of 400 residues. This region of homology is encoded by eight contiguous exons in each respective gene. Of the nine introns that separate these exons, five are located in identical positions in the two protein sequences. This finding suggests that the homologous region may have resulted from a duplication of an ancestral gene and that the two genes evolved further by recruitment of exons from other genes, which provided the specific functional domains of the LDL receptor and the EGF precursor.*

The low-density lipoprotein (LDL) receptor is a protein that carries a nutrient molecule, cholesterol, into cells by receptor-mediated endocytosis in clathrin-coated pits (*1*). Epidermal growth factor (EGF) is a 53-amino-acid peptide hormone that stimulates cells to divide (*2*); it is synthesized as a precursor of 1217 amino acids from which it is liberated by proteolysis (*3, 4*). Like the LDL receptor, the EGF precursor may exist as a membrane-bound molecule (*5–7*). The nucleotide sequences of cloned complementary DNA's (cDNA) for the EGF precursor and the LDL receptor revealed that the amino acid sequence of a

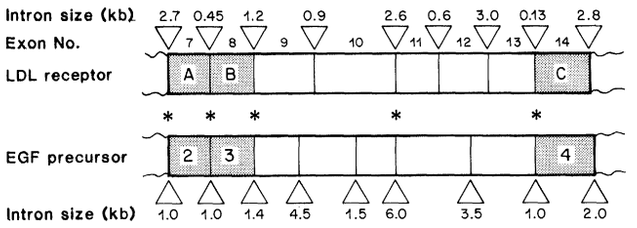

Fig. 1. Intron locations in homologous regions of the human LDL receptor and human EGF precursor genes. The amino acid sequences of the two proteins [residues 293 to 693 in the LDL receptor (*8*, *10*) and residues 403 to 797 in the EGF precursor (*11*)] are illustrated schematically. The three 40-amino-acid repeats are lettered A, B, and C in the LDL receptor protein (*10*) and numbered 2, 3, and 4 in the EGF precursor (*3*, *11*). Introns are indicated as large arrowheads. Asterisks indicate the five introns that interrupt identical residues in the amino acid alignment shown in Fig. 2. Intron sizes (in kilobases) are indicated above the arrowheads.

large segment of the mouse EGF precursor is homologous to a large region of the bovine and human LDL receptors (*7*, *8*). Over a span of 400 amino acids, 33 percent of the two sequences are identical.

The similarity in sequences might be caused by convergent evolution; that is, the homologous regions of the two proteins perform similar functions and therefore they evolved independently into similar structures. Alternatively, the resemblance might be caused by divergent evolution; that is, the homologous regions of the two proteins are descended from a common ancestral gene (*9*). The divergent evolution hypothesis is supported by the finding of nucleic acid homology between the messenger RNA's (mRNA's) of these two proteins (*7*). To obtain a further test of this hypothesis, we have compared the location of introns in the region of homology encoded by the human LDL receptor and the human EGF precursor genes. The results suggest that the homologous regions of these proteins are attributable to exons derived from a common ancestral gene.

The locations of introns in the human LDL receptor gene were determined through analysis of a series of genomic clones (*10*). The positions at which introns interrupt the human EGF precursor gene were determined by bacteriophage λ cloning of genomic DNA, comparison of restriction maps of genomic and cDNA sequences, and sequencing of cloned fragments containing the exon-intron junctions (*11*).

The contiguous region of homology between the two genes spans more than 15 kilobases (kb) in the LDL receptor gene and 23 kb in the EGF precursor gene (Fig. 1). It is interrupted by nine introns that vary in size from 0.13 kb to 6.0 kb. The eight homologous exons range in size from 105 to 228 base pairs. There does not appear to be a conservation of intron sizes. However, there is a striking conservation of the positions (asterisks in Fig. 1) at which the introns interrupt the genes of the LDL receptor and EGF precursor in the homologous region.

When these two protein sequences are aligned for maximum homology, five of the nine introns interrupt the coding sequence at precisely the same amino acid in the LDL receptor and EGF precursor (Fig. 2). These amino acids correspond to residues 293, 333, 375, 508, and 642 in the LDL receptor protein. Three additional pairs of introns are positioned in

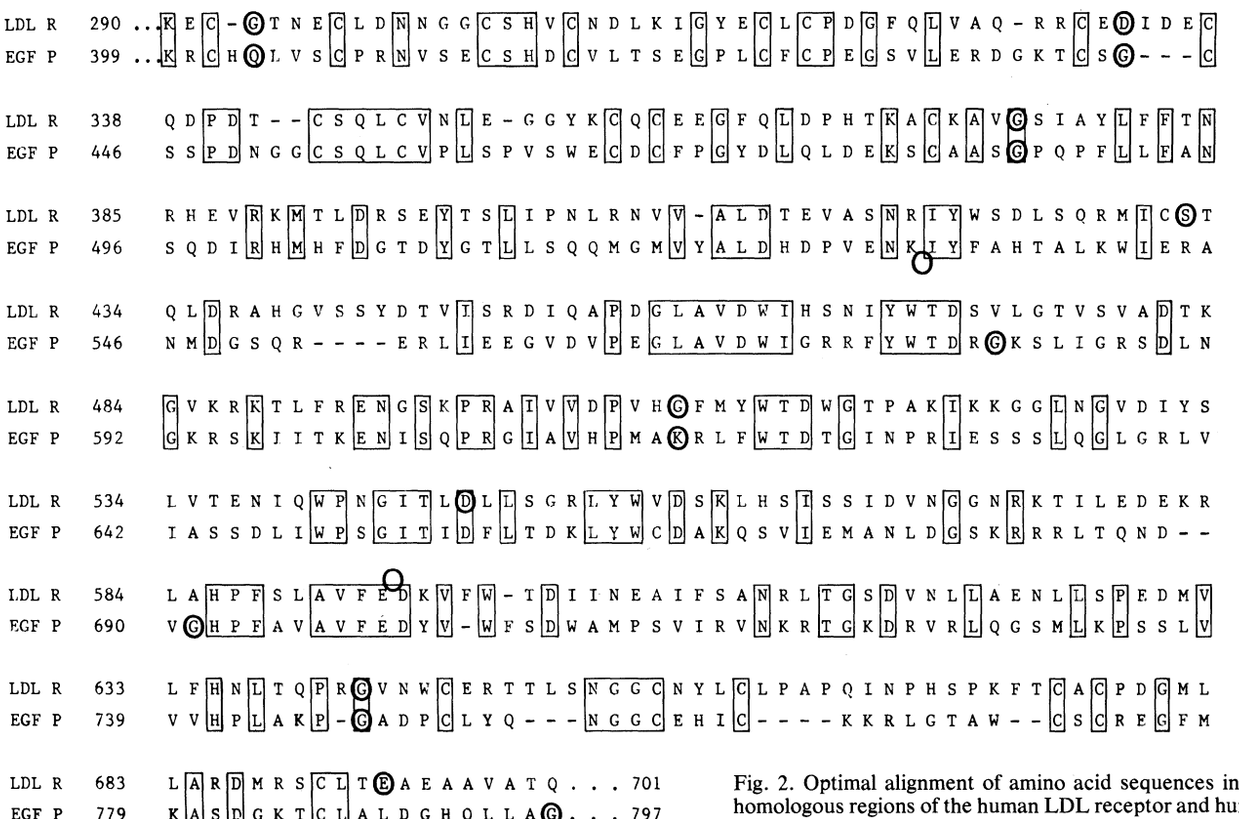

Fig. 2. Optimal alignment of amino acid sequences in the homologous regions of the human LDL receptor and human EGF precursor. The sequences were aligned with the use of the computer program ALIGN of the National Biomedical Research Foundation. Residues shared by the two proteins are boxed. The positions at which introns interrupt the sequences are denoted by the encircled amino acids. The single-letter amino acid code translates to the three-letter code as follows: A, Ala; C, Cys; D, Asp; E, Glu; F. Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

the homologous region within 12 amino acids of each other (residues 432, 595, and 693 in the LDL receptor protein). Finally, each gene appears to have one unshared intron in this region. In the LDL receptor gene, the unshared intron interrupts sequences that encode an aspartic acid residue at position 548; and in the EGF precursor gene the unshared glycine residue at position 581.

Within the 400-amino-acid region of homology are located three copies of a cysteine-rich repeat sequence consisting of about 40 amino acids (*10*). These repeats are labeled A, B, and C in the LDL receptor and 2, 3, and 4 in the EGF precursor as shown in Fig. 1. Of the five identically positioned introns, four serve to demarcate the boundaries of these three repeats (Fig. 1).

These data are consistent with the hypothesis that the extensive amino acid sequence homology between the LDL receptor and EGF precursor arose as a consequence of a gene duplication event in a common evolutionary ancestral gene. Many of the structural features of the progenitor gene have been conserved through time and are now found in the genes for these two present-day proteins (Fig. 1).

If the above argument is correct, how can we explain those instances in which the introns are located in different positions in the two genes? One of the differences arises from the presence of an unshared intron in each gene that has no counterpart in the other gene. A likely explanation is that the homologous counterpart of each of these introns has been lost through evolutionary time. A similar event has been postulated to explain the difference in intron number that occurs between the two insulin genes in the rat (*12*). In such instances, it is possible that intron loss resulted in the fusion of a region of the protein that was originally encoded by the adjacent exons. Such exon joining would then allow the region to evolve as a discrete exon unit, fixing it ultimately as an invariant functional domain.

It is more challenging to explain the three pairs of introns that interrupt the homologous region at similar but not identical positions. Intron sliding has been postulated to give rise to length differences between corresponding segments of homologous proteins (*13*). Intron sliding could explain the extra eight amino acids at the COOH-terminal end of the homologous region of the EGF precursor (amino acids 790 to 797), especially since there are no identities between the two sequences in these eight amino acids (Fig. 2). However, in the two other closely positioned pairs of introns (at residues 432 and 595 in the LDL receptor protein), there is conservation of amino acid sequence in the short stretch of protein that occurs between a given pair. This conservation is most clearly seen in the pair of introns that occur at residue 595 of the LDL receptor and residue 691 of the EGF precursor (Fig. 2). These two introns span a stretch of ten amino acids of which eight are identical. We cannot postulate an intron sliding mechanism that would result in amino acid homologies on both sides of a given intron pair. Intron loss is again a more likely explanation for this observation. Originally, there may have been introns at both ends of this homologous stretch, but one of these introns may have been eliminated by evolution in each gene. Thus, in the LDL receptor the intron that preceded this sequence may have been eliminated, whereas in the EGF precursor the intron that followed the region of identity may have been deleted.

As discussed (*10*), repeats A, B, and C of the LDL receptor and repeats 2, 3, and 4 of the EGF precursor are also homologous to sequences of ~40 amino acids each that occur in three proteins of the blood clotting system—factor IX, factor X, and protein C. Since these regions are contained within discrete exons in both the LDL receptor and the EGF precursor (Figs. 1 and 2), it seems reasonable to predict that they will be contained in discrete exons in these blood clotting proteins as well. Indeed, recent data for the human factor IX gene supports this hypothesis (*14*).

In addition to the exons that are homologous to the EGF precursor, the LDL receptor also has an exon that encodes an amino acid sequence that is homologous to a region in complement component C9 (*10, 15*). This sequence (and its exon) is repeated multiple times in the binding domain of the LDL receptor, but it appears only once in C9. On the basis of the current findings, we predict that this sequence will also be shown to be encoded by a discrete exon in the C9 gene.

The occurrence of shared sequences encoded by discrete exons in the LDL receptor gene provides strong evidence in support of Gilbert's original hypothesis that introns facilitate the evolution of apparently diverse genes by allowing recruitment of exons encoding functional domains (*16*). Our findings also lend support to the concept that the LDL receptor is a mosaic protein whose gene is built up of exons that have been borrowed from or loaned to other genes and is thus a member of several supergene families (*10*). Although the exon structure of other coated pit receptors is not yet known, it is clear from available primary amino acid sequences that some of these proteins are also mosaic structures. For example, the EGF receptor has a cytoplasmic domain that shows striking homology with the sequence of the *erb-B* oncogene, which is derived from a cellular gene (*17*). It may be that coated pit receptors for transport proteins and protein hormones are in general assembled as mosaics of exons borrowed from other genes.

THOMAS C. SÜDHOF
DAVID W. RUSSELL
JOSEPH L. GOLDSTEIN
MICHAEL S. BROWN
Department of Molecular Genetics, University of Texas Health Science Center, Southwestern Medical School, Dallas 75235
RAY SANCHEZ-PESCADOR
GRAEME I. BELL
Chiron Corporation, Emeryville, California 94608

References and Notes

1. J. L. Goldstein, R. G. W. Anderson, M. S. Brown, *Nature (London)* **279**, 679 (1979).
2. G. Carpenter and S. Cohen, *Annu. Rev. Biochem.* **48**, 193 (1979).
3. J. Scott *et al.*, *Science* **221**, 236 (1983).
4. A. Gray, T. J. Dull, A. Ullrich, *Nature (London)* **303**, 722 (1983).
5. R. F. Doolittle, D.-F. Feng, M. S. Johnson, *ibid.* **307**, 558 (1984); R. F. Doolittle, *Trends in Biochem. Sci.*, in press.
6. L. B. Rall *et al.*, *Nature (London)* **313**, 228 (1985).
7. D. W. Russell *et al.*, *Cell* **37**, 577 (1984).
8. T. Yamamoto *et al.*, *ibid.* **39**, 27 (1984).
9. R. F. Doolittle, in *The Proteins*, H. Neurath and R. L. Hill, Eds. (Academic Press, New York, ed. 3, 1979), vol. 4, pp. 1–118.
10. T. C. Südhof, J. L. Goldstein, M. S. Brown, D. W. Russell, *Science* **228**, 815 (1985).
11. Mouse EGF precursor cDNA clones, isolated as described (*3*), were used as probes to screen a human genomic library cloned in bacteriophage λ [R. M. Lawn *et al.*, *Cell* **15**, 1157 (1978)]. Overlapping positive clones were characterized by restriction endonuclease mapping and Southern blotting, and the DNA sequences at exon-intron junctions were determined by the enzymatic [F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977)] and chemical methods [A. M. Maxam and W. Gilbert, *Methods Enzmol.* **65**, 499 (1980)]. Exon DNA sequences were subsequently confirmed by analysis of cDNA clones isolated from a human kidney cDNA library (R. Sanchez-Pescador and G. I. Bell, in preparation). Since the complete amino acid sequence of the human EGF precursor has not yet been determined, we have arbitrarily used a numbering system in Fig. 2 that corresponds to that of the mouse EGF precursor (*3*). The number of residues between amino acids 399 and 797 in the mouse EGF precursor is identical to that for the human protein except for the addition of one leucine in the human protein at position 795 in the mouse protein.
12. F. Perler *et al.*, *Cell* **20**, 555 (1980).
13. C. S. Craik *et al.*, *Science* **220**, 1125 (1983).
14. D. S. Anson *et al.*, *EMBO J.* **3**, 1053 (1984).
15. K. K. Stanley *et al.*, *ibid.* **4**, 375 (1985).
16. W. Gilbert, *Nature (London)* **271**, 501 (1978).
17. J. Downward *et al.*, *ibid.* **307**, 521 (1984).
18. Supported by NIH research grants HL 20948 and HL 31346, a fellowship from the Deutsche Forschungsgemeinschaft (T.C.S.), and a Research Career Development Award from the NIH (HL 01287) (D.W.R.)

12 March 1985; accepted 1 April 1985