

Structure of the hamster low density lipoprotein receptor gene

Richard W. Bishop¹

Department of Molecular Genetics, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75235

Abstract The metabolism of low density lipoprotein (LDL) in the hamster is substantially similar to that of the human. To extend the usefulness of the hamster as an experimental model, the hamster LDL receptor gene was isolated and characterized. The gene is composed of 18 exons and 17 introns which span 26 kilobases. The introns occur at precisely the same positions as those previously determined for the human LDL receptor gene. The 18 exons of the hamster gene predict an LDL receptor protein of 854 amino acids that is similar in organization and sequence to those predicted from the cDNAs of rat, rabbit, cow, and human. Within the 5'-flanking region of the hamster LDL receptor gene are three highly conserved imperfect direct repeat sequences of 16 nucleotides each that in the human gene have been demonstrated to regulate transcription. In addition, a similar arrangement of direct repeat sequences was also isolated from the 5'-flanking region of the rat LDL receptor gene using the polymerase chain reaction. ■ These results indicate a strong sequence and structural conservation of the LDL receptor among several species and further support the hamster as an experimental model for the study of human LDL-cholesterol metabolism.—**Bishop, R. W.** Structure of the hamster low density lipoprotein receptor gene. *J. Lipid Res.* 1992.33: 549–557.

Supplementary key words LDL receptor • promoter • transcriptional regulation • rat • cholesterol homeostasis • sterol regulatory element • nucleotide sequence • cloning • RNA

The cellular uptake of low density lipoprotein (LDL) through receptor-mediated endocytosis has been characterized by thorough genetic and biochemical analysis of the LDL receptor (1). Numerous mutants have been described that affect the synthesis, processing, ligand binding, and recycling of the LDL receptor protein (2). These mutations occur naturally and are found in humans, rabbits, and monkeys with the genetic disorder familial hypercholesterolemia (2–4). The physiological consequence of such defects is to impair the efficient removal of LDL particles from the blood which in turn can lead to increased serum LDL concentrations and atherosclerosis (5, 6).

While the characterization of these mutants has been productive, the inherent physiological limitations illustrate the need for an animal model that ap-

proximates human cholesterol metabolism. The various species currently in use include the monkey, pigeon, pig, guinea pig, rat, rabbit, and hamster (4, 7–9). Among these, the metabolism of cholesterol in the hamster most closely resembles that in humans (9–13).

Support for the hamster model comes from kinetic analysis of LDL production and degradation (9–13). These studies demonstrate similarities between the male hamster and humans with regard to rates of hepatic cholesterol synthesis (9, 10), the relative significance of receptor-dependent and receptor-independent LDL transport (11), and the response of plasma LDL cholesterol to dietary additions (12) or the administration of cholesterol-lowering agents (10, 13). The usefulness of the hamster model is further enhanced by the availability of established cell lines such as the Chinese Hamster Ovary (CHO) line which contain defects in the pathways of cholesterol synthesis (14), regulation (15), esterification (16), and uptake by the LDL receptor protein (17).

The isolation of cDNA clones for the bovine, human, rabbit, and rat LDL receptors has been reported (3, 18–20). In addition, the gene for the human LDL receptor has been isolated and characterized (21). To complement these studies and to extend the genetic analysis of the LDL receptor in the hamster, the cloning and structural analysis of the hamster LDL receptor gene is presented, including a highly conserved transcriptional regulatory region in the immediate 5'-flanking DNA.

Abbreviations: LDL, low density lipoprotein; CHO, Chinese hamster ovary; EGF, epidermal growth factor; SRE, sterol regulatory element; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; cDNA, complementary DNA.

¹Present address: Department of Metabolic Diseases, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, NJ 08543-4000.

METHODS

General methods

Unless specified, all methods were performed according to standard procedures (22).

Genomic library construction and screening

Probes derived from the human LDL receptor cDNA, pLDLR-2 (19), were used to obtain clone λ U4 from approximately 1×10^6 clones of a Chinese hamster-derived UT-1 cell genomic library (23). Likewise, plasmid subclone pRB2 was isolated from a XbaI partial UT-1 cell genomic library constructed in λ Charon 35. The screening of both libraries was performed under low stringency hybridization and wash conditions (19). To isolate more 5'-flanking regions of the hamster gene, the clones pRB3 and pEX1 were obtained by chromosome walking (24). UT-1 cell genomic DNA was digested with the appropriate restriction endonuclease and size-fractionated by density gradient centrifugation on 10–40% (w/v) sucrose gradients (22). Hybridization-positive fractions were used to construct size-fractionated libraries in either pUC18 or the bacteriophage vector λ EMBL-4 (Promega, Madison, WI) and screened as described (19) except that the hybridization buffers contained 50% (v/v) formamide. Plasmid clone pRB3 was isolated from 5×10^5 colonies of a pUC18 StuI library and plasmid pEX1 was subcloned from a phage clone isolated from 1×10^6 recombinants of an EMBL-4 EcoRI library.

Gene mapping and nucleotide sequence analysis

A crude map of the cloned DNA was established using ^{32}P -labeled human cDNA fragments (19, 22). Restriction fragments were then subcloned into bacteriophage M13mp18 or mp19 vectors (Pharmacia) for automated or manual nucleotide sequence analysis of both DNA strands by the dideoxy chain termination method (25). Gradient gels, ^{35}S -radiolabeled nucleotides, and Sequenase (US Biochemical Corp., Cleveland, OH) were used according to the manufacturer's recommendations. Automated sequencing was done using fluorescence-labeled primers (26) and a DNA sequencer (Applied Biosystems, Inc., model 370A).

cDNA cloning

A primer-specific, size-fractionated cDNA library was constructed from CHO cell poly(A)⁺ RNA by standard methodologies (22) using an antisense oligonucleotide primer, complementary to nucleotide positions 106 to 140 of Fig. 2, in place of oligo(dT). After second strand synthesis and EcoRI linker addition, the cDNA

products were digested with EcoRI, fractionated on a 5% (w/v) neutral polyacrylamide gel, and the appropriately sized products were inserted into the bacteriophage λ gt10 vector (Stratagene, La Jolla, CA). Positive clones were identified from approximately 2×10^5 recombinants using a ^{32}P -labeled antisense oligonucleotide complementary to nucleotide positions 68 to 104 of Fig. 2 and subcloned into bacteriophage M13 vectors for DNA sequencing as described above.

Nuclease S1 and primer extension analysis

Total and poly(A)⁺ RNA were prepared by standard methods from CHO cells (22). For nuclease S1 analysis, a uniformly labeled single-stranded probe of 227 nucleotides was prepared as described (27) using an oligonucleotide primer complementary to nucleotides +14 to +48 (Fig. 4A) and a hamster LDL receptor M13 template containing nucleotides -194 to +61 (Fig. 4A). After EcoRI cleavage and gel purification, 20,000 cpm of the ^{32}P -labeled probe was annealed to 25 μg of either tRNA or total CHO RNA overnight at 60°C. Nuclease S1 analysis was performed as described (3) except that 100 units of S1 nuclease (BRL, Gaithersburg, MA) was used in the digestion reaction. Primer extension analysis with reverse transcriptase (Life Sciences, St. Petersburg, FL) was performed as described by Sühof et al. (28). The annealing reaction was carried out at 65°C using a ^{32}P -labeled oligonucleotide complementary to nucleotides 68 to 104 of Fig. 2 and either 5 μg of poly(A)⁺ RNA or 10 μg of total RNA isolated from CHO cells grown in the presence or absence of sterols (29).

RESULTS AND DISCUSSION

Isolation of the hamster LDL receptor gene

Several techniques were used to isolate the hamster LDL receptor gene. A total of four genomic clones spanned all 18 exons and 17 introns of the gene and included 9.6 kb of the 5'-flanking region; the 3' untranslated region was not represented in the isolated clones (Fig. 1). Initially, a gene fragment spanning exons 4 through 18 was isolated from a UT-1 CHO library (23) by screening at reduced stringency with a human LDL receptor cDNA probe (λ U4, Fig. 1). Subsequently, a cDNA probe corresponding to human exons 2 and 3 was used to screen a lambda genomic library generated by partial digestion of UT-1 CHO DNA with the enzyme XbaI. A 5.5 kilobase fragment of the gene was isolated from one positive clone (pRB2, Fig. 1). The two remaining upstream clones were isolated by chromosome walking. DNA fragments from

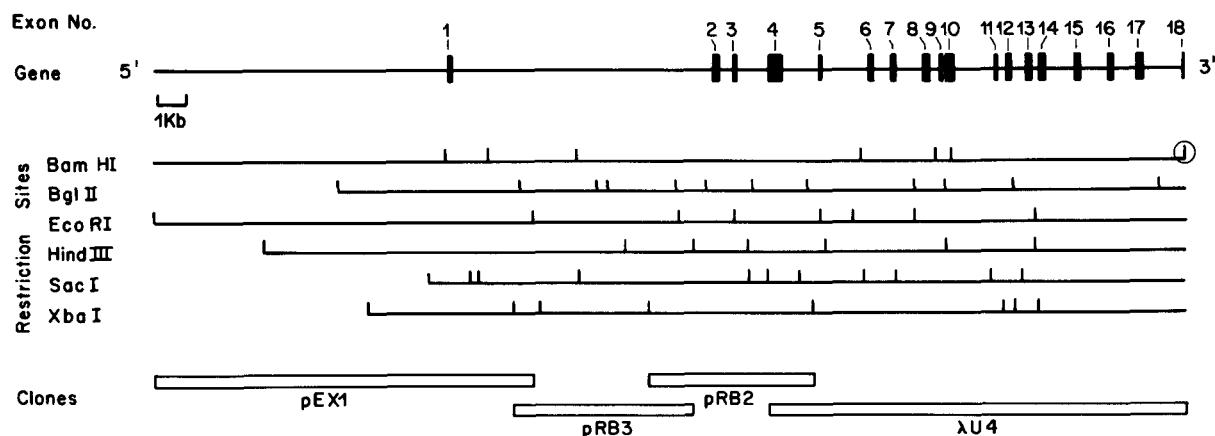


Fig. 1. Map of the hamster LDL receptor gene. The gene structure is shown to scale in the upper portion of the diagram. Exons are denoted by numbers above the black boxes. The regions included in the various genomic bacteriophage and plasmid clones are indicated at the bottom. Cleavage sites for six selected restriction endonucleases are shown. The encircled Bam HI site in the 3' untranslated region, representing a cloning junction for λ U4, may not be present in the gene.

the 5' region of pRB2 were used as probes in genomic Southern blots to identify upstream restriction fragments for the construction of size-fractionated genomic libraries (see Methods). Screening of one such library yielded a clone containing the 5.8 kilobase insert, pRB3 (Fig. 1). Similar techniques yielded the 12.3 kilobase insert in pEX1 (Fig. 1).

Exon-intron structure of hamster LDL receptor gene

Exons within the cloned genomic fragments were localized by several methods. Human LDL receptor cDNA probes were used in restriction mapping and Southern blotting of plasmid subclones to yield tentative exon placements. A precise location was subsequently determined by DNA sequence analysis and by comparison to the structure of the human gene. Due to the poor identity between the two species in exons 15 and 16, approximately 2.8 kilobases of DNA encompassing these exons were sequenced and analyzed for amino acid sequence similarity with analogous regions of the human protein.

The position of exon 1 was established by primer-specific cDNA cloning. Sequence information derived from exon 2 in clone pRB2 (Fig. 1) was used to synthesize two oligonucleotides complementary to adjacent positions in the mRNA. Primer extension analysis of CHO RNA with the more 5' oligonucleotide indicated that there were an additional 188 nucleotides to the transcriptional start site (see below). Using the more 3' oligonucleotide as a primer for the first strand cDNA synthesis reaction, UT-1 mRNA was converted into cDNA, size-fractionated, and cloned into λ gt10. Screening of the library was performed using the more 5' oligonucleotide as a probe. The resulting cDNA extended to nucleotide position -142

and was later used to position exon 1 in genomic clone pEX1 (Fig. 1).

The conservation of structure between the hamster and human LDL receptor genes is remarkable. The 17 introns of the hamster gene interrupt the same amino acids and codons as those of the human gene (21). In addition, all of the intron-exon boundaries agree with the consensus GT...AG rule established for these junctions in eukaryotic genes (data not shown) (30). The sizes of the 17 introns ranged from less than 100 base pairs (intron 9) to 8.5 kilobases (intron 1).

Nucleotide and predicted amino acid sequence of the hamster LDL receptor mRNA and protein

The merged nucleotide sequence derived from the 18 exons of the hamster LDL receptor gene and its predicted translation product are shown in Fig. 2. The sequence of exon 12 is identical to that obtained previously by Sege, Kozarsky, and Krieger (17) from an LDL receptor-deficient CHO mutant. With the A of the initiator methionine residue designated as +1, the 18 exons are spliced to generate an open reading frame of 2562 base pairs. Analysis of the mRNA from CHO cells by RNA blotting and hybridization indicated a single major transcription product of 4.9 kilobases (data not shown). This result predicts a 3' untranslated region of approximately 2.3 kilobases and suggests that the total size of the hamster gene is about 26 kilobases. This size is nearly one half that of the 45 kb human gene (21).

Translation of the predicted mRNA sequence results in a protein of 854 amino acids. The position and composition of a 21 amino acid segment at the amino terminus of the hamster protein closely resembles that of a signal sequence (31). Assuming that this sequence

ATGAGCACCCGGGATCTCAGGCTGCGCTGGCCATCGCCCTGCTCTGGCCGCTGCGGGGGTGCAGCAGAAGACACATGCGACAGGAATGAGTTCGGGTGCAGGGACGGGAAATGTATC 120
 MetSerThrAlaAspLeuArgLeuArgTrpAlaIleAlaLeuLeuLeuAlaAlaAlaGlyAlaAlaAlaGluAspThrCysAspArgAsnGluPheArgCysArgAspGlyLysCysIle
 -20 -10 1 10
 GCCAGCAAGTGGGTCTGTGACCGCAGCCCTGAATGCCCGGATGGCTCTGATGAGTCTCTGAGACATGATGTCTGTCACTGTCAATCCAAGGAATTGAGTGTGGGGCCGTGTCAGC 240
 AlaSerLysTrpValCysAspGlySerProGluCysProAspGlySerAspGluSerSerLeuThrCysMetSerValThrCysGlnSerLysGluPheSerCysGlyGlyArgValSer
 20 30 40 50
 CGTGCACTTCTAACTCTCGGCTGTGTGATGGGCAACAGACTGTGAGAAACGGCTCAGATGAACAGGCTGTGCCCCAAGCGTGTCCAGGATGAGTTCGGCTGCCAGGATGGCAAG 360
 ArgCysIleProAsnSerTrpArgCysAspGlyGlnThrAspCysGluAsnGlySerAspGluGlnGlyCysAlaProLysThrCysSerGlnAspGluPheArgCysGlnAspGlyLys
 60 70 80 90
 TGCATCTCCAGAAAGTTGTGTGACAGGACCAAGATTGCGTGGATGGCTCTGACGAGGCCACTGCCAGGCCGCTTGTGGGCCCGGCAATTCGGTTCGAATCTGCCCTGTC 480
 CysIleSerGlnLysPheValCysAspGlnAspGlnAspCysValAspGlySerAspGluAlaHisCysGlnAlaAlaThrCysGlyProAlaHisPheArgCysAsnSerTrpProCys
 100 110 120 130
 ATCCCGACCTGTGGCCCTGTGACGGGATGACGACTGTGAGGACGGCTCCGATGAGTGGCCACAGAATGTGGGGGGCGGACACCGCCGCTGCCTGGAGTAGCAGCCCTGTCTCTCC 600
 IleProSerLeuTrpAlaCysAspGlyAspAspCysGluAspGlySerAspGluTrpProGlnAsnCysGlyGlyArgAspThrAlaAlaAlaTrpSerSerSerProCysSerSer
 140 150 160 170
 CTGGAGTTCCTACTGTGGCAGCAGTGAATCATTTCAGCTCGCAGTGGCTCTGTGATGGCAGCGCTGACTGCAAGGCAAGTCAGATGAGGACCACTGCTGCACGCCACTGCCCGCTGAC 720
 GluGluPheHisCysGlySerSerGluCysIleHisArgSerTrpValCysAspGlySerAlaAspCysLysAspLysSerAspGluGluHisCysValThrAlaThrCysArgProAsp
 180 190 200 210
 GAATTCAGTGTGCAGATGGCACCTGCATCCATGGTAGCCCGCACTGTGACCGTGAATATGACTGCAAGGACATGAGTGCAGGCTGGCTCCATCAATGTGACACAGTGTGACGGGGCC 840
 GluPheGlnCysAlaAspGlyThrCysIleHisGlySerArgGlnCysAspArgGluTrpAspCysLysAspMetSerAspGluLeuGlyCysIleAsnValThrGlnCysAspGlyPro
 220 230 240 250
 AACCAAGTTCAGTGCACAGTGGCAATGTATCGCCCTGGACAAAGTGTGCGACTCCATCGGGACTGCCGGACTGGTCCAGATGAGCCCATCAAGGACTGCAGGACCAATGAGTGTCTG 960
 AsnLysPheLysCysHisSerGlyGluCysIleAlaLeuAspLysValCysAspSerMetArgAspCysArgAspTrpSerAspGluProIleLysAspCysArgThrAsnGluCysLeu
 260 270 280 290
 GACAACATGGTGGCTGTCCCAAGTGTGCAAGGACCTCAAGATTGGCTACGAGTCTGTGTGCCAATGGCTTCCAGCTGGTGGACACAGGTTGTAAGACATTGACGAGTGTGAG 1080
 AspAsnAsnGlyGlyCysSerHisValCysLysAspLeuLysIleGlyTyrGluCysLysCysProAsnGlyPheGlnLeuValAspGlnHisArgCysGluAspIleAspGluCysGln
 300 310 320 330
 GAGCCGGACACCTGTGACCACTCTGCGTGAACCTTGAAGGACGCTACAGTGGAGTGGCGGGCTGGATTCCACATGGACCCCTCACACCAGGCTGCAAGGCTGTAGCTCCGTAGCC 1200
 GluProAspThrCysAspGlnLeuCysValAsnLeuGluGlySerTyrLysCysGluCysArgAlaGlyPheHisMetAspProHisThrArgValCysLysAlaValGlySerValAla
 340 350 360 370
 TATCTGCTCTCACTAACCGCCACGAGGTACGGAAGATGACCCCTGGACCGAAGCAGTACACAGCCCTGATCCCCAACCTGAAGAAGCTGGTGGCCCTGGACACTGAAGTGGCCAACT 1320
 TyrLeuLeuPheThrAsnArgHisGluValArgLysMetThrLeuAspArgSerGluTyrThrSerLeuIleProAsnLeuLysAsnValValAlaLeuAspThrGluValAlaAsnAsn
 380 390 400 410
 AGAATCTACTGCTCCGACCTGTCCAGGAAAGATCTATAGTCCCTGATGGACCAAGCACCTACTCTGTCTATGACACCATCATCAGTGGGGATCTGACAGGCCCCAGATGGGTGGCC 1440
 ArgIleTyrTrpSerAspLeuSerGlnGlyLysIleTyrSerAlaLeuMetAspGlnAlaProThrLeuSerTyrAspThrIleIleSerGlyAspLeuGlnAlaProAspGlyLeuAla
 420 430 440 450
 GTAGACTGGATCCATGCCAACATCTACTGGACAGATTCACTCCAGGCACTGTTCTGTGGCTGACACCAAGGGTATAAGGAGAGGACACTGTTCCAGCAGAGGGGTCCAGACCCAG 1560
 ValAspTrpIleHisGlyAsnIleTyrTrpThrAspSerValProGlySerValSerValAlaAspThrLysGlyIleArgArgArgThrLeuPheGlnGluLysGlySerArgProArg
 460 470 480 490
 GACATTTGGTGGACCCCTGTGATGCTCTCATGTACTGGACAGATTGGGGACACTGCCAAGATCAAGAAAGGGGGCCCTGAATGGTGTAGATATCTACTCGCTGGTACTGAGACATC 1680
 AspIleValValAspProValHisGlyPheMetTyrTrpThrAspTrpGlyThrProAlaLysIleLysLysGlyGlyLeuAsnGlyValAspIleTyrSerLeuValThrGluAspIle
 500 510 520 530
 CAGTGGCCAAATGGCATCACACTAGATATCCCACTGGCCGCTCTATGGGTTGATTCCAACTCCACTCTACTCCAGCATGTATGTCATGGGGGCAATCGGAAGCCATTTTGGAA 1800
 GlnTrpProAsnGlyIleThrLeuAspIleProSerGlyArgLeuTyrTrpValAspSerLysLeuHisSerIleSerSerIleAspValAsnGlyGlyAsnArgLysThrIleLeuGlu
 540 550 560 570
 GATGAGAAGCAGCTGGCCCAACCCCTCTCTCTGGCCATCTATGAGGACAAAGATTTTGGACGGATGTCATAAATGAAGCCMTTTCAGTGGCAATCGCCTCACAGGCTCAGATGTTAAT 1920
 AspGluLysGlnLeuAlaHisProPheSerLeuAlaIleTyrGluAspLysValPheTrpThrAspValIleAsnGluAlaIlePheSerAlaAsnArgLeuThrGlySerAspValAsn
 580 590 600 610
 TTGGTGGCTGAAAACCTCTGTCCCGAGGACATTTCTCTGTCCCAACATCACACAGCTAGAGGGGTGAACCTGGTGTGAGAGAACAGCCCTCCCAACGGTGGCTGCCAATACCTG 2040
 LeuValAlaGluAsnLeuLeuSerProGluAspIleValLeuPheHisAsnIleThrGlnProArgGlyValAsnTrpCysGluArgThrAlaLeuProAsnGlyGlyCysGlnTyrLeu
 620 630 640 650
 TGCCTGCCGCCCCACAGATCAATCTCACTCCCAAAAATTCACCTGTGCTGCTGATGGCATGCTACTGGCCAAGGATATGCGAACGTGCCTCACAGAAGTGGCACTGTACTGACC 2160
 CysLeuProAlaProGlnIleAsnProHisSerProLysPheThrCysAlaCysProAspGlyMetLeuLeuAlaLysAspMetArgThrCysLeuThrGluValAlaProValLeuThr
 660 670 680 690
 ACCCAGGGAACATCCACCATCAGGCTGAGATCACTGCAGGACCGAAGGCTGCCGAAGCACAAAGGAGATCAGTCGGCTTCCAGTACTTCCAGGCGCCTGCACCTTAGCACAGTGGAG 2280
 ThrGlnGlyThrSerThrIleArgProGluIleThrAlaGlyAlaGluGlyCysProLysHisLysGluAspGlnSerAlaSerSerThrSerArgGlnProAlaLeuSerThrValGlu
 700 710 720 730
 TCGGTGACAAATGCCCAAGTCCAGGCTGACAGAAGGAAACGAGGACCGCCACAGGCTGTGGGGTCTTGTCCATCACCCCTCCCATCCCACTGCTCATCTCTTGTGGGGCC 2400
 SerValThrMetSerHisGlnValGlnGlyAspArgArgAsnGluGluArgProGlnGlyValGlyValLeuSerIleThrLeuProIleAlaLeuValIleLeuLeuValPheGlyAla
 740 750 760 770
 ATCTCTGTGGAGAACTGGCCGCTGAGGAACTCAACAGCAATAAATTTGACACCCAGTGTACCAAGAGCCACAGAGATGAACCTCCATTTTGGGCAAGCCAGGACGGCTACTCC 2520
 IleLeuLeuTrpArgAsnTrpArgLeuArgAsnIleAsnSerIleAsnPheAspAsnProValTyrGlnLysThrThrGluAspGluLeuHisIleCysArgSerGlnAspGlyTyrSer
 780 790 800 810
 TACCCCTCCGACAGATGCTCAGCCTGGAGATGATGTGGCATGA 2565
 TyrProSerArgGlnMetValSerLeuGluAspAspValAla***
 820 830 833

Downloaded from www.jlr.org by guest, on June 18, 2012

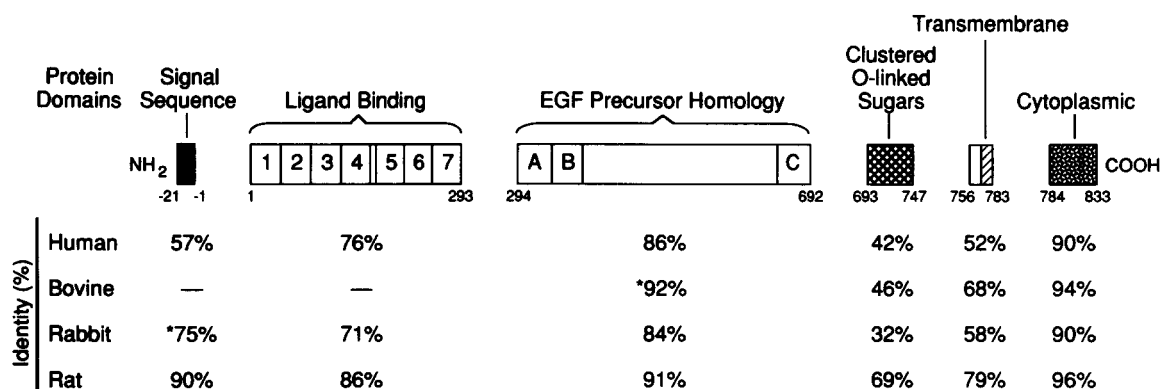


Fig. 3. Interspecies comparison of LDL receptor protein sequences. A schematic of the LDL receptor protein consisting of six domains is shown at the top (21). The amino acid residues corresponding to the hamster LDL receptor protein domain boundaries are numbered below the schematic. The percent identity as deduced by computer-assisted comparison between the hamster protein domains and those of four other species is shown below the model. Asterisks denote percentages that are tabulated from incomplete sequence information; dashes represent missing sequence information. The sequences of the human, bovine, rabbit and rat LDL receptors were taken from references 27, 26, 5, and 28, respectively.

is processed in a manner analogous to that of the bovine LDL receptor protein (19), the mature hamster protein would consist of 833 amino acids with a theoretical mass of 94,524 daltons. In addition, the hamster protein is 6 amino acids shorter than the human protein in the transmembrane domain.

Fig. 3 illustrates the structural conservation in the LDL receptor protein domains observed between those of the hamster and those of four other species. Numerous studies with the human LDL receptor protein and gene have assigned discreet functions to each of these domains (6). They include an amino terminal signal sequence, a ligand binding domain, a large segment encompassing 400 amino acids with homology to the epidermal growth factor precursor, a clustered O-linked sugar domain, a membrane-anchoring domain, and a carboxyl-terminal cytoplasmic domain. The high degree of sequence conservation at both the gene and protein level among the LDL receptors of different species (**Fig. 3**) strongly suggests that these domains perform the same functions in all species. Interestingly, domains of the LDL receptor that are known to interact with components of the endocytic pathway, such as coated pits (with the cytoplasmic domain (32)) and the recycling apparatus of the endosome compartment (with the EGF precursor homology domain (33)) appear to be the most highly conserved (**Fig. 3**). Studies in which the human LDL receptor has been shown to function normally in

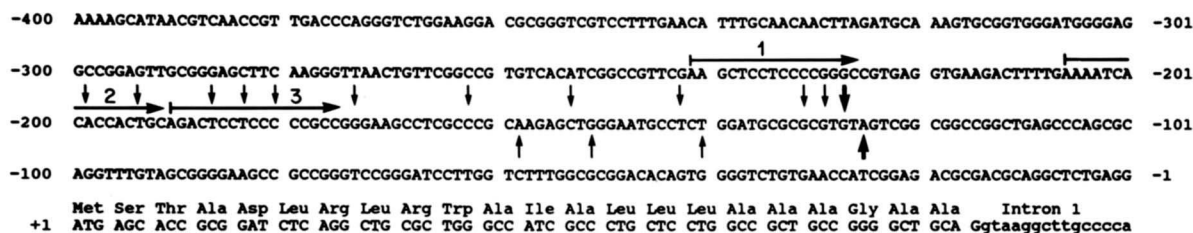
Xenopus laevis oocytes support the notion of a highly conserved endocytic apparatus (34). These observations suggest that regions of high sequence identity detected in interspecies comparisons of other coated pit receptors may be useful predictors of endocytic domains. Overall, the hamster LDL receptor protein shares the most sequence identity with the rat (88%) and cow (partial comparison, 86%), followed by the human (78%) and rabbit receptors (75%).

Characterization of the 5'-flanking sequence of the hamster LDL receptor gene

Fig. 4 displays the nucleotide sequence of the hamster gene 5'-flanking region and the transcriptional initiation sites determined by primer extension (**Fig. 5**) and nuclease S1 (**Fig. 6**) analyses. Both methods detect one major and several minor transcript initiation sites which appear to be down-regulated when sterols are present in the growth media (**Fig. 5**, right). While the absence of a discernible TATA sequence may underlie the multiple transcription initiation sites detected, the responsiveness to sterols is most likely a consequence of *cis*-acting elements present as three imperfect direct repeats of 16 nucleotides (see below). A similarly oriented and very homologous series of repeats present in the human LDL receptor gene (**Table 1**) have been shown to confer transcriptional activation and regulation by sterols (28).

Fig. 2. Predicted nucleotide sequence of the hamster LDL receptor mRNA and protein. Nucleotides are numbered in the right-hand column beginning with the first nucleotide of the initiator methionine codon. Dots are placed above the sequence every tenth nucleotide. Amino acids are numbered below the sequence; residues in the predicted signal sequence are assigned negative numbers while those in the mature protein are assigned positive numbers. Introns are denoted by numbered triangles appearing above the nucleotide sequence. Asterisks denote the translational stop codon.

A. Hamster 5' Flanking Sequence



B. Rat 5' Flanking Sequence

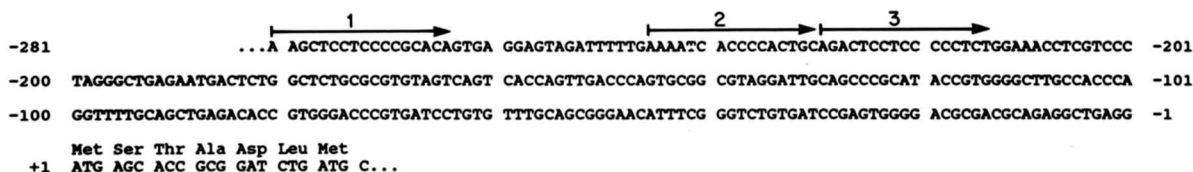


Fig. 4. Nucleotide sequence of the hamster and rat gene 5'-flanking regions. (A) The nucleotide sequence of the hamster 5'-flanking region was derived from genomic and cDNA clones as described in Methods. Nucleotides are numbered on the right and left with position +1 assigned to the A of the ATG codon specifying the initiator methionine; negative numbers refer to 5'-flanking sequences. The predicted amino acids of exon 1 and the position of intron 1 appear on the bottom line. Vertical arrows denote major (thick arrows) and minor (thin arrows) sites of transcription initiation as determined by primer extension (beneath the sequence, Fig. 5) and nuclease S1 analysis (above the sequence, Fig. 6). Solid horizontal lines above the sequence indicate three imperfect direct repeats of 16 nucleotides. (B) The nucleotide sequence of the rat LDL receptor gene 5'-flanking region as determined by PCR and chemical sequencing (22) using primers corresponding to nucleotide positions -269 to -243 and positions 23 to 47 of the hamster DNA sequence in (A). The numbering of nucleotides and landmarks are as in (A).

The 5'-flanking sequence conservation suggested that other rodent species, such as the rat, might contain similar sequences that could be easily examined by PCR using oligonucleotides that flanked this conserved region. Except for the addition of a guanine at position -117, the nucleotide sequence derived by this procedure (Fig. 4B) agrees precisely with a segment (-153 to +22) of the rat LDL receptor cDNA (20). The rat 5'-flanking sequence also contains three imperfect direct repeats with a high degree of sequence identity

to those of the hamster and human LDL receptor genes (Fig. 4B, Table 1).

The sequence analysis of the 5'-flanking regions from the hamster, rat, and human (21) LDL receptor genes suggests that the structural conservation observed in the protein and in the gene may also apply to the promoter regions of these genes (Fig. 4). Three homologous imperfect direct repeats of 16 nucleotides are almost identically positioned with respect to one another in the 5'-flanking regions of the three genes

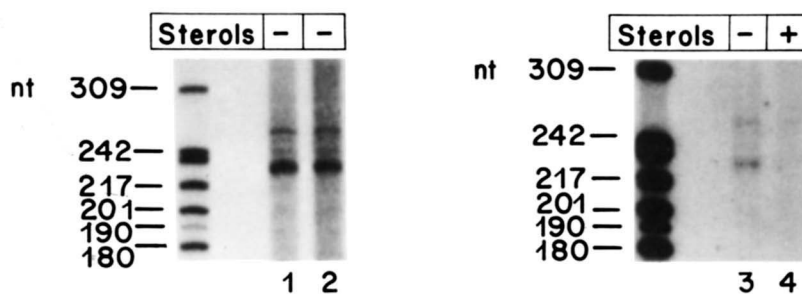


Fig. 5. Primer extension analysis of the hamster LDL receptor mRNA. Poly(A)⁺ (left) or total RNA (right) was subjected to primer extension analysis as described in Methods. The presence or absence of sterols in the growth media is indicated above each lane. Size standards are indicated on the left. Lanes 1 and 2, 5 μ g of CHO poly(A)⁺ RNA from two different preparations was used as template; lanes 3 and 4, 10 μ g of total CHO RNA grown in the absence (lane 3) or presence (lane 4) of sterols (29) was used as template. An assignment of the primer extension products to the nucleotide sequence of the hamster gene was carried out by comparison to the standards and is shown in Fig. 4A.

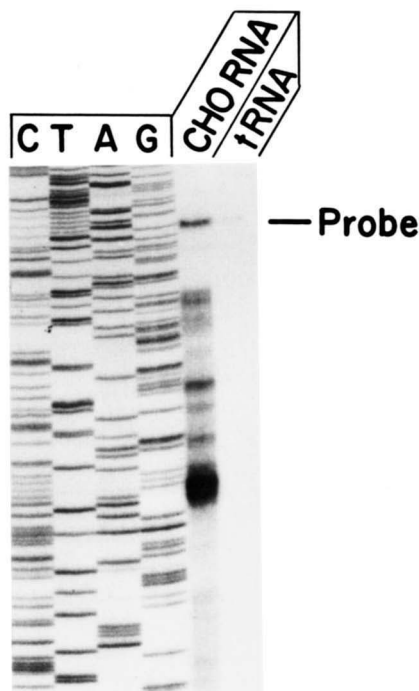


Fig. 6. Nuclease S1 analysis of transcription initiation sites in the hamster LDL receptor gene. Twenty-five μg of total RNA from CHO cells or 25 μg of tRNA was subjected to nuclease S1 analysis as described under Methods. The DNA sequencing ladder derived by the dideoxy terminator method used the identical primer and M13 template as that used to generate the nuclease S1 probe. The position to which the undigested probe migrated is indicated. The major and minor protected bands correspond to nucleotides in the hamster gene as indicated in Fig. 4A.

(Fig. 4). Although we have not carried out functional studies on these sequences in the rodent LDL receptor genes described here, it seems likely that they perform the same regulatory roles as their extensively studied (28, 29, 35, 36) counterparts in the human LDL receptor gene. In these studies, direct repeats 1 and 3 have been shown to interact with the transcription factor Sp1 (35), whereas direct repeat 2 acts as a sterol-responsive element (SRE) (28, 29, 35, 36). The consensus DNA sequence recognized by Sp1 is somewhat weak (35), and consistent with this finding there is sequence variation in direct repeats 1 and 3 of the rodent and human genes (Table 1). In contrast, the sequences of the SRE differ by only a single nucleotide (A for C at nucleotide -198 in the hamster gene, Table 1). This particular change has been recreated in the human SRE and shown not to affect the sterol-responsiveness of the human LDL receptor promoter (36). This is supported by the fact that LDL receptor mRNA expression appears to be down-regulated when CHO cells were grown in the presence of sterols (Fig. 5, right). The strong conservation in the sequence of the SRE and its near identical location relative to the

TABLE 1. Interspecies comparison of direct repeat sequences in LDL receptor 5'-flanking region and similarity to binding site for transcription factor Sp1

Repeat 1	
Human	A A A C T C C T C C T C T T G C
Hamster	A A G C T C C <u>T</u> C C C C G G G C
Rat	A A G C T C C <u>T</u> C C C C G C A C
Repeat 2	
Human	A A A <u>A</u> T C A C C C C A C T G C
Hamster	A A A <u>A</u> T C <u>A</u> C C A C C A C T G C
Rat	A A A <u>A</u> T C <u>A</u> C C C C A C T G C
Repeat 3	
Human	A A A C T C C T C C C C C T G C
Hamster	A G A C T C C <u>T</u> C C C C G G C C
Rat	A G A C T C C <u>T</u> C C C C C T C T
Sp1 Consensus	G C C C C G C C C C A T T A C G C C T A

The nucleotide sequences of repeats 1-3 (overlined in Fig. 4) from the human (28), hamster, and rat genes are indicated. The consensus sequence for nuclear factor Sp1 (39) is shown at the bottom. Nucleotides in repeats 1, 2, and 3 that differ from the Sp1 consensus sequence are underlined.

direct repeats that interact with Sp1 suggests that the protein or proteins that mediate the conditional-positive response to sterols via the SRE (35, 36) have a low tolerance for sequence and positional variation. If this hypothesis is true, then the SRE elements in other genes involved in cholesterol metabolism which are quite different at the DNA sequence level (37, 38) may interact with different transcription factors.

The general features of the hamster LDL receptor gene presented here support the use of this species as an experimental model for the study of human LDL-cholesterol metabolism. Furthermore, the high degree of sequence and structural conservation observed in both the LDL receptor gene and protein of humans and hamsters suggests a molecular basis for the previously described metabolic similarities (11-13). This analogy, however, must extend beyond those features described here since the rat also displays significant sequence homology in both the 5'-flanking region (Fig. 4) and protein sequence (20) of the LDL receptor, yet its cholesterol metabolism is different from that of humans (10). This could be explained by as yet undiscovered structurally important regions of heterogeneity in the LDL receptor protein and promoter structure. Alternatively, other components of LDL-cholesterol metabolism may be involved. Nevertheless, this amount of genetic identity between two vastly divergent species serves to illustrate the evolutionary importance of the LDL receptor gene and its role in cellular cholesterol homeostasis of humans and the hamster.

I thank Daphne Davis, Daphne Norsworthy, and Jim Cali for excellent technical assistance. Special thanks to M. S. Brown, J. L. Goldstein, T. C. Südhof, and D. W. Russell for advice or

critical reading of the manuscript. This research was supported by National Institutes of Health Grant HL 20948 awarded to D. W. Russell. ■■

Manuscript received 19 September 1991 and in revised form 27 December 1991.

REFERENCES

1. Brown, M. S., and J. L. Goldstein. 1986. A receptor-mediated pathway for cholesterol homeostasis. *Science*. **232**: 34–47.
2. Hobbs, H. H., D. W. Russell, M. S. Brown, and J. L. Goldstein. 1990. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu. Rev. Genet.* **24**: 133–170.
3. Yamamoto, T., R. W. Bishop, M. S. Brown, J. L. Goldstein, and D. W. Russell. 1986. Deletion in cysteine-rich region of LDL receptor impedes transport to cell surface in WHHL rabbit. *Science*. **232**: 1230–1237.
4. Scanu, A. M., A. Khalil, L. Neven, M. Tidore, G. Dawson, D. Pfaffinger, E. Jackson, K. D. Carey, H. C. McGill, and G. M. Fless. 1988. Genetically determined hypercholesterolemia in a rhesus monkey family due to a deficiency of the LDL receptor. *J. Lipid Res.* **29**: 1671–1681.
5. Bilheimer, D. W., N. J. Stone, and S. M. Grundy. 1979. Metabolic studies in familial hypercholesterolemia: evidence for a gene-dosage effect in vivo. *J. Clin. Invest.* **64**: 524–533.
6. Goldstein, J. L., and M. S. Brown. 1989. Familial hypercholesterolemia. In *The Metabolic Basis of Inherited Disease*. C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, editors. McGraw-Hill Inc., New York, NY. 1215–1250.
7. Reagan, J. W., Jr., L. R. Miller, and R. W. St. Clair. 1990. In vivo clearance of low density lipoprotein in pigeons occurs by a receptor-like mechanism that is not down-regulated by cholesterol feeding. *J. Biol. Chem.* **265**: 9381–9391.
8. Nordestgaard, B. G., E. Hjelm, S. Stender, and K. Kjeldsen. 1990. Different efflux pathways for high and low density lipoproteins from porcine aortic intima. *Arteriosclerosis*. **10**: 477–485.
9. Spady, D. K., and J. M. Dietschy. 1983. Sterol synthesis in vivo in 18 tissues of the squirrel monkey, guinea pig, rabbit, hamster, and rat. *J. Lipid Res.* **24**: 303–315.
10. Spady, D. K., S. D. Turley, and J. M. Dietschy. 1985. Rates of low density lipoprotein uptake and cholesterol synthesis are regulated independently in the liver. *J. Lipid Res.* **26**: 465–472.
11. Spady, D. K., and J. M. Dietschy. 1985. Dietary saturated triacylglycerols suppress hepatic low density lipoprotein receptor activity in the hamster. *Proc. Natl. Acad. Sci. USA*. **82**: 4526–4530.
12. Spady, D. K., and J. M. Dietschy. 1988. Interaction of dietary cholesterol and triglycerides in the regulation of hepatic low density lipoprotein transport in the hamster. *J. Clin. Invest.* **81**: 300–309.
13. Plancke, M. O., P. Olivier, V. Clavey, D. Marzin, and J. C. Fruchart. 1988. Aspects of cholesterol metabolism in normal and hypercholesterolemic Syrian hamsters. Influence of fenofibrate. *Methods Find. Exp. Clin. Pharmacol.* **10**: 575–579.
14. Chin, D. J., K. L. Luskey, R. G. W. Anderson, J. R. Faust, J. L. Goldstein, and M. S. Brown. 1982. Appearance of crystalloid endoplasmic reticulum in compactin-resistant Chinese hamster cells with a 500-fold increase in 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Proc. Natl. Acad. Sci. USA*. **79**: 1185–1189.
15. Metherall, J. E., J. L. Goldstein, K. L. Luskey, and M. S. Brown. 1989. Loss of transcriptional repression of three sterol-regulated genes in mutant hamster cells. *J. Biol. Chem.* **264**: 15634–15641.
16. Cadigan, K. M., J. G. Heider, and T. Y. Chang. 1988. Isolation and characterization of Chinese hamster ovary cell mutants deficient in acyl-coenzyme A:cholesterol acyltransferase activity. *J. Biol. Chem.* **263**: 274–282.
17. Sege, R. D., K. F. Kozarsky, and M. Krieger. 1986. Characterization of a family of gamma-ray-induced CHO mutants demonstrates that the *ldlA* locus is diploid and encodes the low-density lipoprotein receptor. *Mol. Cell. Biol.* **6**: 3268–3277.
18. Russell, D. W., T. Yamamoto, W. J. Schneider, C. J. Slaughter, M. S. Brown, and J. L. Goldstein. 1983. cDNA cloning of the bovine low density lipoprotein receptor: feedback regulation of a receptor mRNA. *Proc. Natl. Acad. Sci. USA*. **80**: 7501–7505.
19. Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple *alu* sequences in its mRNA. *Cell*. **39**: 27–38.
20. Lee, L. Y., W. A. Mohler, B. L. Schafer, J. S. Freudenberg, N. Byrne-Connolly, K. B. Eager, S. T. Mosley, J. K. Leighton, R. N. Thrift, R. A. Davis, and R. D. Tanaka. 1989. Nucleotide sequence of the rat low density lipoprotein receptor cDNA. *Nucleic Acids Res.* **17**: 1259–1260.
21. Südhof, T. C., J. L. Goldstein, M. S. Brown, and D. W. Russell. 1985. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science*. **228**: 815–822.
22. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1.1–1.47.
23. Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey. 1984. HMG-CoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. *Cell*. **38**: 275–285.
24. Spoerel, N. A., and F. C. Kafatos. 1987. Isolation of full-length genes: walking the chromosome. *Methods Enzymol.* **152**: 598–610.
25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. **74**: 5463–5467.
26. Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*. **321**: 674–679.
27. Church, G. M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA*. **81**: 1991–1995.
28. Südhof, T. C., D. W. Russell, M. S. Brown, and J. L. Goldstein. 1987. 42 bp element from LDL receptor gene confers end-product repression by sterols when inserted into viral TK promoter. *Cell*. **48**: 1061–1069.
29. Südhof, T. C., D. R. Van Der Westhuyzen, J. L. Goldstein, M. S. Brown, and D. W. Russell. 1987. Three direct repeats and a TATA-like sequence are required for regulated expression of the human low density lipoprotein receptor gene. *J. Biol. Chem.* **262**: 10773–10779.

30. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**: 459–472.
31. Boyd, D., and J. Beckwith. 1990. The role of charged amino acids in the localization of secreted and membrane proteins. *Cell.* **62**: 1031–1033.
32. Davis, C. G., I. R. Van Driel, D. W. Russell, M. S. Brown, and J. L. Goldstein. 1987. The LDL receptor: identification of amino acids in cytoplasmic domain required for rapid endocytosis. *J. Biol. Chem.* **262**: 4075–4082.
33. Davis, C. G., J. L. Goldstein, T. C. Südhof, R. G. W. Anderson, D. W. Russell, and M. S. Brown. 1987. Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. *Nature.* **326**: 760–765.
34. Peacock, S. L., M. P. Bates, D. W. Russell, M. S. Brown, and J. L. Goldstein. 1988. Human low density lipoprotein receptor expressed in *Xenopus* oocytes. Conserved signals for O-linked glycosylation and receptor-mediated endocytosis. *J. Biol. Chem.* **263**: 7838–7845.
35. Dawson, P. A., S. L. Hofmann, D. R. van der Westhuyzen, T. C. Südhof, M. S. Brown, and J. L. Goldstein. 1988. Sterol-dependent repression of low density lipoprotein receptor promoter mediated by 16-base pair sequence adjacent to binding site for transcription factor Sp1. *J. Biol. Chem.* **263**: 3372–3379.
36. Smith, J. R., T. F. Osborne, J. L. Goldstein, and M. S. Brown. 1990. Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. *J. Biol. Chem.* **265**: 2306–2310.
37. Smith, J. R., T. F. Osborne, M. S. Brown, J. L. Goldstein, and G. Gil. 1988. Multiple sterol regulatory elements in promoter for hamster 3-hydroxy-3-methylglutaryl-coenzyme A synthase. *J. Biol. Chem.* **263**: 18480–18487.
38. Osborne, T. F., G. Gil, J. L. Goldstein, and M. S. Brown. 1988. Operator constitutive mutation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase promoter abolishes protein binding to sterol regulatory element. *J. Biol. Chem.* **263**: 3380–3387.
39. Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science.* **234**: 47–52.