

ORIGINAL ARTICLE

Toshiaki Nakajima · Kikumi Iwaki · Takao Hamakubo
Tatsuhiko Kodama · Mitsuru Emi

Genomic structure of the gene encoding human 3-hydroxy-3-methyl-glutaryl coenzyme A reductase: comparison of exon/intron organization of sterol-sensing domains among four related genes

Received: April 25, 2000 / Accepted: June 12, 2000

Abstract We determined the genomic structure of the human gene encoding 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, which catalyzes the conversion of HMG-CoA to mevalonate and is the rate-limiting and major regulatory enzyme in sterol biosynthesis. The gene is more than 21 kb long, about five times the size of its corresponding cDNA. It consists of 20 exons, ranging in size from 68 to 1809 bp. An amino-terminal hydrophobic membrane-bound domain is encoded by exons 2–10, a flexible linker domain by exons 10 and 11, and the catalytic domain by exons 11–20. Exons 3–7 encode a sterol-sensing domain. We compared its genomic structure in this region with the sterol-sensing domains of three related genes, sterol regulatory element binding protein (SREBP) cleavage-activating protein (SCAP), Niemann-Pick type C1 protein (NPC1), and a morphogen receptor, Patched. Two of the five positions of introns in the sterol-sensing domain of the *HMG-CoA reductase* gene were identical to the exon/intron organization of this domain in the related human genes, but these positions of introns were not conserved in homologues from lower organisms, except in one instance. The data suggested that exon-shuffling may have occurred during relatively recent evolution; this would account for the structural similarity of this domain in four quite different human proteins.

Key words Exon shuffling · Sterol-sensing domain · 3-Hydroxy-3-methyl-glutaryl coenzyme A (HMG-CoA) reductase · Sterol regulatory element binding protein (SREBP) cleavage-activating protein (SCAP) · Niemann-Pick type C1 protein (NPC1) · Patched

T. Nakajima · K. Iwaki · M. Emi (✉)
Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, 1-396 Kosugi-cho, Nakahara-ku, Kawasaki 211-8533, Japan
Tel. +81-44-733-5230; Fax +81-44-733-5192
e-mail: memi@nms.ac.jp

T. Hamakubo · T. Kodama
Department of Molecular Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan

Introduction

HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A) reductase, which catalyzes the conversion of HMG-CoA to mevalonate, is the rate-limiting and major regulatory enzyme in sterol biosynthesis (Goldstein and Brown 1990). The complex system regulating levels of this enzyme in tissues involves both transcription of the gene and degradation of the protein (Gil et al. 1985; Luskey and Stevens 1985). HMG-CoA reductase is normally a relatively stable, slowly degraded enzyme, but in the presence of excess sterols it is degraded rapidly and selectively. Its membrane-bound domain plays a crucial role in regulating the degradation of this protein.

A sequence motif spanning five consecutive transmembrane regions is present in HMG-CoA reductase, and also in three distinctively different proteins: SREBP (sterol regulatory element binding protein) cleavage-activating protein (SCAP; Hua et al. 1996; Nohturfft et al. 1998); Niemann-Pick type C1 protein (NPC1), a defective protein in Niemann-Pick disease (Carstea et al. 1997; Loftus et al. 1997); and Patched, a morphogen receptor in *Drosophila*, and a defective protein in human basal cell nevus syndrome (Johnson et al. 1996; Hahn et al. 1996). This motif appears to function as a sterol-sensing domain (Lange and Steck 1998), because it has a critical role in regulating the degradation of HMG-CoA reductase (Gil et al. 1985; Luskey et al. 1985) and because a mutation in this motif blocks the responsiveness of SCAP to intracellular cholesterol levels (Hua et al. 1996; Nohturfft et al. 1998). We determined the genomic structure of all four genes in the regions encoding their sterol-sensing domains, to investigate the relationship between these otherwise distinct proteins.

Materials and methods

Isolation of the human HMG-CoA reductase, NPC1, and Patched genes

A bacteriophage P1-derived artificial chromosome (PAC) library containing human genomic DNA pooled in a three-

dimensional structure (Genome Systems, St. Louis, MO, USA) was used to isolate genomic clones of these three genes. We had determined the genomic structure of *SCAP* previously (Nakajima et al. 1999). We screened the PAC library by the polymerase chain reaction (PCR), according to methods described earlier for *SCAP*, using as primers the following pairs of oppositely oriented oligonucleotides: for *HMG-CoA reductase*, 5-AATGTTGTCAAGACTTTTTC GAATGC-3 and 5-TTCCAACCACAGATCTTATTGTT ACCA-3; for *NPCI*, 5-TACAAAGGAACAGAGCGCG AA-3 and 5-TTCACAGGCGCTACGTTCAA-3; and for *Patched*, 5-CATCACCAGTGTGACGGCTTC-3 and 5-GCTCAATGACTTCCACCTTCGA-3. *Escherichia coli* cells containing positive clones were cultured in the presence of kanamycin, and PAC DNA was isolated as described previously (Nakajima et al. 1999).

Structural analysis of entire human *HMG-CoA reductase* and the sterol-sensing domains of *NPCI* and *Patched* genes

The nucleotide sequences of exons and their boundaries were determined by directly sequencing the PAC clone, using primers designed from the corresponding cDNA sequences. Sequencing was performed by the BigDye Terminator cycle sequencing method, in a 377 ABI Prism automated DNA sequencer (Perkin Elmer, Norwalk, CT, USA). The size of each intron was either determined by direct sequencing of the PAC genomic clones, or estimated by amplification, using LA Taq (Takara, Tokyo, Japan) and oppositely oriented primers located in exons immediately flanking each intron.

Results

Genomic structure of human *HMG-CoA reductase*

More than 110kb of the genomic region containing the entire gene was cloned in a single PAC. This gene turned out to be more than 21 kb long, about five times the size of its corresponding cDNA, and consists of 20 exons interrupted by 19 introns (Table 1 and Fig. 1A) (Genbank accession numbers AF2737654, AF273755, AF273756, AF273757, AF273758, AF273759, AF273760, AF273761, AF273762, AF273763, AF273764, and AF273765). Exons ranged in size from 68 to 1809 bp. The positions of the introns were identical to those of the hamster *HMG-CoA reductase* gene (Genbank accession numbers L00166, L00169, and L00170). Sequences at all exon-intron boundaries were compatible with the consensus sequence for splicing junctions, including AG-GT (Mount 1982).

Comparison of the genomic sequence with the cDNA revealed that exon 1 and the 5' half of exon 2 encode the 5'-untranslated region (UTR); the protein-coding region extends from the 3' portion of exon 2 through exon 19; exon 20 encodes part of the carboxyl-terminal coding region and the entire 3'-UTR (Fig. 1B). Exon 20 contains a polyadenylation site preceded by a polyadenylation signal. The hydrophobic membrane-bound domain at the amino-terminal is encoded by exons 2–10, the flexible linker domain by exons 10 and 11, and the carboxyl-terminal catalytic domain by exons 11–20. The extensive region of homology with three other sterol-sensing proteins, thought to represent the sterol-sensing domain in each case, is encoded by exons 3–7.

Table 1. Exon/intron organization of human *HMG-CoA reductase* gene

Exon number	Exon size (bp)	Sequence at exon/intron junction ^a		Intron size (bp) ^b
		5' Splice donor	3' Splice acceptor	
1	Approx. 64	AGATCTGGAGgtgagggcggg...	... tctctccagGATCCAAGGA	~7000*
2	188	GTTTGAAGAGgttagtgaag...	... ctctattagGATGTTTTGA	1100*
3	112	TATATTTTGgtaataagt...	... tgcatttagGTATTGCTGG	280
4	88	CAGGCTTGAAgtaagtattt...	... tctttccagTGAAGCTTTG	1100*
5	85	CAACTCACAGgtaagtatta...	... tctgttttagGATGAAGTAA	1400*
6	106	ACCATGTCAggtttgtaagc...	... ctctaaccagGGGTACGTCA	2600*
7	107	GGTATTAGAGgtaagacaa...	... taattttagCTTTCTCGGG	109
8	117	GATGATTATGtaagtacat...	... tttattcagTCTCTAGGCT	500*
9	161	ATCTCTCTAAgtaagtaat...	... ttttatttagAATGATCAGC	118
10	248	GAAAGAAAAGgtaactgtt...	... tcttgaacagTTGAGGTTAT	108
11	179	GAATGCAGAGgtgaggatga...	... atttcttagAAAGGTGCAA	2600*
12	195	TTACTCCTTGgtatgtatt...	... taccttttagGTGATGGGAG	350*
13	159	AGCAATAGGTgtaagttggc...	... tactgttagCTTGGTGAG	150
14	158	GCACTAGCAGgtgtgtgagt...	... ttattatagATTTGCACGT	850*
15	106	GATTTCAAAGgtaagtgtag...	... ttgttaacagGGTACAGAGA	2000*
16	171	TGTCAGAGAAgtgagtgtact...	... tgttttagGTATTAAAGA	400*
17	141	CTGTGGACAGgtgagctctc...	... ttgttccagGATGCAGCAC	87
18	159	CTGTTTGCAGgtatgatgta...	... ctggctcagATGCTAGGTG	400*
19	155	TTCACAACAGgtaagactca...	... tctaccatagGTCGAAGATC	148
20	~1809			

^aThe sequences of exon/intron boundaries were determined as described in "Materials and methods". Lowercase letters refer to intronic sequences, and uppercase to exonic sequences

^bIntron sizes were determined by direct sequencing of the PAC genomic clone, or were estimated by polymerase chain reaction amplification, using LA Taq (asterisks) (TaKaRa, Tokyo, Japan)

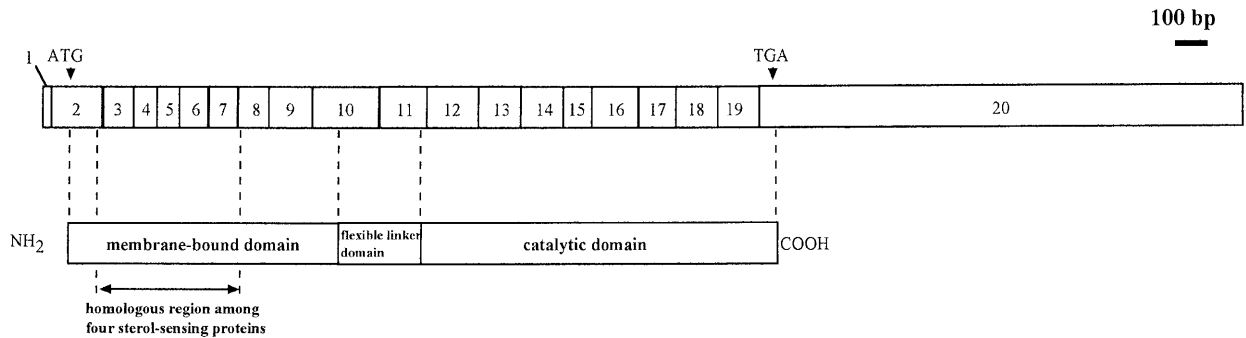
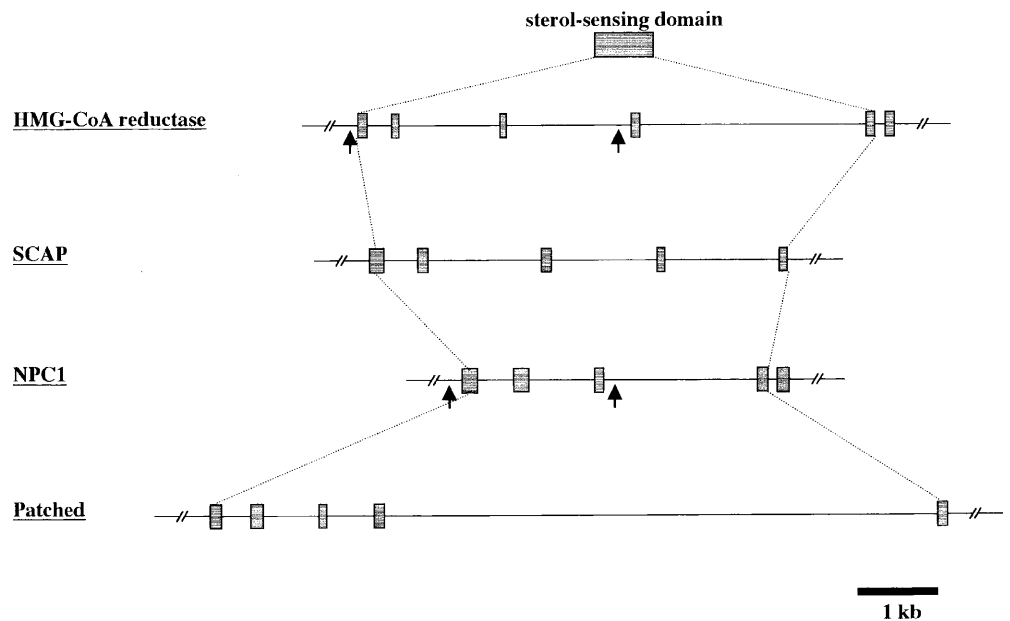
A**B**

Fig. 1. A Genomic structure of the human *HMG-CoA reductase* gene. **B** Relationship between exon organization and functional domains of *HMG-CoA reductase*. A translation initiation codon (*ATG*) is present in exon 2 and a translation stop codon (*TGA*) in exon 20

Fig. 2. Genomic structure of sterol-sensing domains of four distinct human genes. *Arrows* indicate the position of the *Alu* sequence



Genomic structure of sterol-sensing domains in four sterol-sensing proteins

We had previously shown that five exons encode the sterol-sensing domain of human SCAP (Nakajima et al. 1999). In the human *HMG-CoA reductase*, *NPC1*, and *Patched* genes, this domain is encoded by five, four, and five exons,

respectively (Fig. 2). The genomic structure we determined for this region of the human *NPC1* gene is identical to that reported by Morris et al. (1999). However, we identified an intron more than 6.5kb long, interrupting exon 12 of the human *Patched* gene; this intron had not been included in the genomic structure described by Hahn et al. (1996).

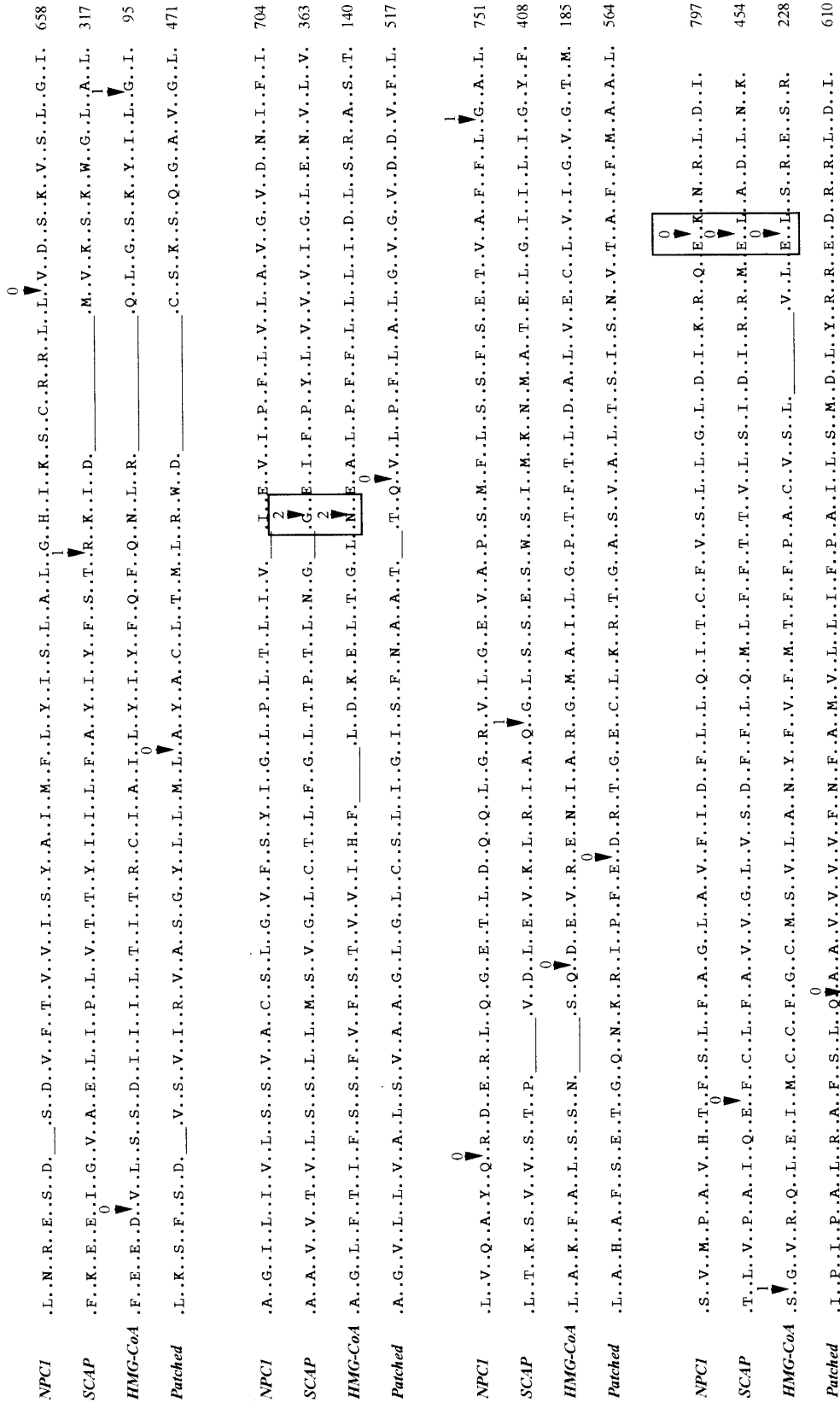


Fig. 3. Amino-acid sequence homologies among sterol-sensing domains of four genes, aligned by Basic Local Alignment Search Tool (BLAST) analysis. Arrowheads indicate positions of introns; numbers above arrowheads show intronic phases

Comparison of the genomic organization of sterol-sensing domains among the four genes

One intron, located at the 3' end of the region encoding the sterol-sensing domain, maintained an identical position among the *HMG-CoA reductase*, *SCAP*, and *NPC1* genes, and another position was identical between *HMG-CoA reductase* and *SCAP* (Figs. 3 and 4). No intron positions in human *Patched* were identical to those in the other three genes.

We also compared the genomic structure of human sterol-sensing domains with the *Caenorhabditis elegans* homologue of *SCAP* (D2013.8; GenBank accession number Z47808; ref. Hua et al. 1996), *NPC1* (F02E8.6; GenBank number U53340; ref. Carstea et al. 1997), and *Patched* (ZK675.1 and F21H12.4; GenBank numbers Z46812 and U23176, respectively) and also with the *Saccharomyces cerevisiae* homologues of *HMG-CoA reductase* (GenBank M22002) and *NPC1* (Lpa11p; GenBank number AAB68099; ref. Carstea et al. 1997). As shown in Fig. 4, only one intron position is identical between the human and invertebrate or yeast homologue, i.e., *NPC1* and its *C. elegans* homologue F02E8.6.

Discussion

The presence of sterol-sensing domains in four distinctly different proteins suggests that the genes encoding them may have evolved through exon-shuffling, a process by which exons can be transferred to other genes by recombination events involving introns or transposable elements (Gilbert et al. 1997; Long et al. 1995). Eukaryotic genes are often larger and more complex than their homologues from simple organisms. Gene elongation during evolution appears frequently to have involved repetition of existing amino acid sequences, often as a result of exon duplication. The structure of many eukaryotic polypeptides suggests that exchanges of structural or functional protein domains at the gene level have resulted in a complex mosaic capable of specifying a variety of different protein modules (Gilbert et al. 1997; Long et al. 1995). For example, the low-density lipoprotein (LDL) receptor, which is assembled from exonic sequences shared with the C9 component of complement and the epidermal growth factor (EGF) precursor, is a vivid example of such a mosaic protein (Sudhof et al. 1985a,b; Russell et al. 1984). As further evidence for exon-

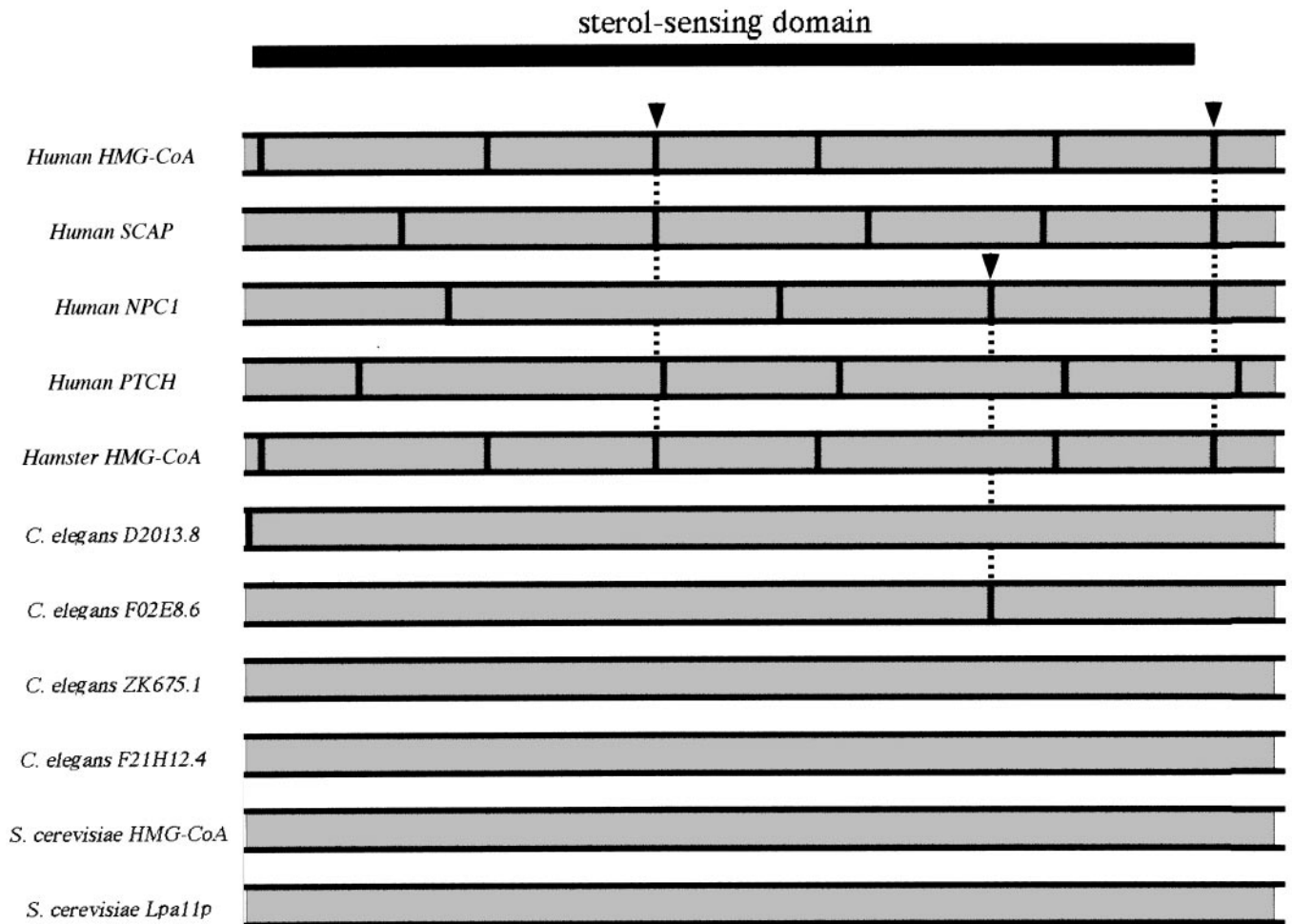


Fig. 4. Intron positions in sterol-sensing domains in human genes, hamster and *C. elegans HMG-CoA reductases*, and homologues of other human sterol-sensing proteins in *C. elegans* and *S. cerevisiae*

shuffling, homologous domains are sometimes interrupted by introns in the same position (Sudhof et al. 1985a,b). The four proteins investigated in our study (NPC1, HMG-CoA reductase, SCAP, and Patched) share a sequence motif that spans five consecutive transmembrane regions (Carstea et al. 1997; Loftus et al. 1997; Lange and Steck 1998). This motif is thought to function as a sterol-sensing domain, because it plays a critical role in regulating the degradation of HMG-CoA reductase (Gil et al. 1985; Luskey and Stevens 1985), and because a mutation within this motif blocks the responsiveness of SCAP to intracellular cholesterol levels (Hua et al. 1996; Nohturfft et al. 1998; Brown and Goldstein 1999). The presence of sterol-sensing domains in four such different proteins supports our view that the genes encoding them probably have evolved through exon-shuffling.

To obtain additional evidence for a common evolutionary origin of the sterol-sensing domains, we determined and compared the genomic structure of that portion of each gene and some of their homologues in lower organisms. The position of one intron, located at the 3' end of the region encoding this domain, was identical in three of the four human genes examined, and another position was identical in *HMG-CoA* reductase and *SCAP*. However, these positions were not conserved in *C. elegans* and *S. cerevisiae* homologues, except in one instance. These data suggest that exon-shuffling that produced a variety of human sterol-sensing domains may have been a recent event in the evolution of higher eukaryotes.

Exon-shuffling may occur through recombination between introns (Sudhof et al. 1985a,b; Lehrman et al. 1987) and/or transposable elements (Moran et al. 1999). Homologous recombination between repetitive DNA elements dispersed throughout ancestral genomes may have been responsible for generating gene duplications and other DNA rearrangements during evolution (Sudhof et al. 1985a,b). For example, *Alu* repeats, which are present near exons encoding the sterol-sensing domains of the *HMG-CoA* reductase and *NPC1* genes (Fig. 2), may have contributed to the evolution of different sterol-sensing proteins by serving as sites of recombination (Lehrman et al. 1987). We believe that the evidence presented here supports a hypothesis that sterol-sensing domains were incorporated into several proteins by recent exon-shuffling.

Acknowledgments This work was supported by a Grant-in-Aid for Scientific Research in the Priority Area "Genome Science" from Ministry of Education, Science, Sports and Culture (MESSC) of Japan, by research grants for osteoporosis from the Ministry of Health and Welfare of Japan, and by the Novartis Foundation for Gerontological Research.

References

Brown MS, Goldstein JL (1999) A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc Natl Acad Sci USA* 96:11041-11048

- Carstea ED, Morris JA, Coleman KG, Loftus SK, Zhang D, Cummings C, Gu J, Rosenfeld MA, Pavan WJ, Krizman DB, Nagle J, Polymeropoulos MH, Sturley SL, Ioannou YA, Higgins ME, Comly M, Cooney A, Brown A, Kaneski CR, Blanchette-Mackie EJ, Dwyer NK, Neufeld EB, Chang TY, Liscum L, Straus III JF, Ohno K, Marsha Z, Carmi R, Sokol J, Markie D, O'Neill RR, van Diggelen OP, Elleder M, Patterson MC, Brady RO, Vanier MT, Pentchev PG, Tagle DA (1997) Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* 277:228-231
- Gil G, Faust JR, Chin DJ, Goldstein JL, Brown MS (1985) Membrane-bound domain of HMG CoA reductase is required for sterol-enhanced degradation of the enzyme. *Cell* 41:249-258
- Gilbert W, de Souza SJ, Long M (1997) Origin of genes. *Proc Natl Acad Sci USA* 94:7698-7703
- Goldstein JL, Brown MS (1990) Regulation of the mevalonate pathway. *Nature* 343:425-430
- Hahn H, Wicking C, Zaphiropoulos PG, Gailani MR, Shanley S, Chidambaram A, Vorechovsky I, Holmberg E, Uden AB, Gillies S, Negus K, Smyth I, Pressman C, Leffell DJ, Gerrard B, Goldstein AM, Dean M, Toftgard R, Chenevix-Trench G, Wainwright B, Bale AE (1996) Mutations of the human homologue of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* 85:841-851
- Hua X, Nohturfft A, Goldstein JL, Brown MS (1996) Sterol resistance in CHO cells traced to point mutation in SREBP cleavage-activating protein. *Cell* 87:415-426
- Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM, Quinn AG, Myers RM, Cox DR, Epstein EH Jr, Scott MP (1996) Human homologue of patched, a candidate gene for the basal cell nevus syndrome. *Science* 272:1668-1671
- Lange Y, Steck TL (1998) Four cholesterol-sensing proteins. *Curr Opin Struct Biol* 8:435-439
- Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. *Cell* 48:827-835
- Loftus SK, Morris JA, Carstea ED, Gu JZ, Cummings C, Brown A, Ellison J, Ohno K, Rosenfeld MA, Tagle DA, Pentchev PG, Pavan WJ (1997) Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* 277:232-235
- Long M, de Souza SJ, Gilbert W (1995) Evolution of the intron-exon structure of eukaryotic genes. *Curr Opin Genet Dev* 5:774-778
- Luskey KL, Stevens B (1985) Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved domains responsible for catalytic activity and sterol-regulated degradation. *J Biol Chem* 260:10271-10277
- Moran JV, DeBerardinis RJ, Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. *Science* 283:1530-1534
- Morris JA, Zhang D, Coleman KG, Nagle J, Pentchev PG, Carstea ED (1999) The genomic organization and polymorphism analysis of the human Niemann-Pick C1 gene. *Biochem Biophys Res Commun* 261:493-498
- Mount SM (1982) A catalogue of splice junction sequences. *Nucleic Acids Res* 10:459-472
- Nakajima T, Hamakubo T, Kodama T, Inazawa J, Emi M (1999) Genomic structure and chromosomal mapping of the human SREBP cleavage-activating protein (SCAP) gene. *J Hum Genet* 44:402-407
- Nohturfft A, Brown MS, Goldstein JL (1998) Sterols regulate processing of carbohydrate chains of wild-type SREBP cleavage-activating protein (SCAP), but not sterol-resistant mutants Y298C or D443N. *Proc Natl Acad Sci USA* 95:12848-12853
- Russell DW, Schneider WJ, Yamamoto T, Luskey KL, Brown MS, Goldstein JL (1984) Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. *Cell* 37:577-585
- Sudhof TC, Russell DW, Goldstein JL, Brown MS, Sanchez-Pescador R, Bell GI (1985a) Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science* 228:893-895
- Sudhof TC, Goldstein JL, Brown MS, Russell DW (1985b) The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 228:815-822