DATABASES

The UMD-LDLR Database: Additions to the Software and 490 New Entries to the Database

Ludovic Villéger,[1] Marianne Abifadel,[1,2] Delphine Allard,[1] Jean-Pierre Rabès,[1,3] Rochelle Thiart,[4†] Maritha J. Kotze,[4†] Christophe Béroud,[1] Claudine Junien,[1,3] Catherine Boileau,[1,3]* and Mathilde Varret[1]

[1]INSERM UR383, Hôpital Necker-Enfants Malades, Université René Descartes, Paris, France; [2]Faculté de Pharmacie, Université Saint-Joseph, Beirut, Lebanon; [3]Laboratoire Central de Biochimie et de Génétique Moléculaire, CHU Ambroise Paré AP-HP, Boulogne, France; [4]MRC Cape Heart Group, Division of Human Genetics, Faculty of Medicine, University of Stellenbosch, Tygerberg, South Africa

Communicated by Jean-Claude Kaplan

Mutations in the LDL receptor gene (*LDLR*) cause familial hypercholesterolemia (FH), one of the most frequent hereditary dominant disorders. The protein defect was identified in 1973, the gene was localized by in situ hybridization in 1985, and since, a growing number of mutations have been reported. The UMD-LDLR database is customized software that has been developed to list all mutations, and also to provide means to analyze them at the nucleotide and protein levels. The database has been recently modified to fulfill the recommendations of the Nomenclature Working Group for human gene mutations. However, in the current version, both the nomenclature and usual *LDLR* gene mutation names are reported since the latter are more commonly used. The software has also been modified to accommodate the splicing mutations and alleles that carry two nucleotide variations. The current version of UMD-LDLR contains 840 entries, of which 490 are new entries. Point mutations account for 90% of all mutations in the *LDLR* gene; the remaining are mostly major rearrangements, due to the presence of *Alu* sequences. Three new routines have been implemented in the software, thus giving users access to 13 sorting tools. In addition to the database, a Web site containing information about polymorphisms, major rearrangements, and promoter mutations is available. Both are accessible to the scientific community (www.umd.necker.fr) and should help groups working on *LDLR* to check their mutations and identify new ones, and greatly facilitate the understanding of functional classes/genotype relationships and of genotype/phenotype correlations. Hum Mutat 20:81–87, 2002. © **2002 Wiley-Liss, Inc.**

KEY WORDS: familial hypercholesterolemia; LDLR; mutations; database

DATABASES:
LDLR-OMIM: 143890; Genbank: NM_000527
http://www.umd.necker.fr (UMD-LDLR Database)
http://www.umd.necker.fr/LDLR/Home_Page.html (LDLR/FH Web Site)
http://www.ucl.ac.uk/fh (LDLR Gene in FH – Mutation Database)
http://archive.uwcm.ac.uk/uwcm/mg/search/119362.html (HGMD- LDLR Mutation Data)

INTRODUCTION

The LDL receptor (MIM# 143890) is produced in the ER as a pro-receptor, whose 21 amino acid signal peptide is cleaved, mostly glycosylated to give rise to a mature receptor. The 160 kDa transmembrane receptor is ubiquitously distributed, playing a major role in cholesterol homeostasis [Goldstein et al.,

Received 20 December 2001; accepted revised manuscript 22 April 2002.

*Correspondence to: Catherine Boileau, INSERM U383, Clinique Maurice Lamy, Hôpital Necker-Enfants Malades, 149-161 rue de Sèvres, 75743 Paris Cedex 15, France. E-mail: boileau@necker.fr

The Supplementary Material referred to in this article can be found http://www.wiley.com/humanmutation/suppmat/2002/v20.html

Contract grant sponsors: Pfizer; Université René Descartes Paris V; Faculté de Médecine Necker-Enfants Malades; la Fondation de France; PROGRES-INSERM; ARCOL; Ministère de l'Education Nationale, de la Recherche, et de la Technologie; ARCOL; SFA.

†Current address for Rochelle Thiart and Maritha J. Kotze: Genecare, Cape Town 8000, PO Box 15734, Vlaabora 8018, South Africa.

DOI: 10.1002/humu.10102
Published online in Wiley InterScience (www.interscience.wiley.com).

1985]. Impairment of LDL receptor activity results in the accumulation of LDL cholesterol in the circulation, leading to familial hypercholesterolemia (FH). Affected individuals display arcus corneae, tendon xanthomas, and premature symptomatic coronary heart disease [Goldstein and Brown, 1989]. FH is an autosomal dominant disorder, homozygotes being more severely affected than heterozygotes, and one of the most common inherited disorders with frequencies of heterozygotes and homozygotes estimated to be 1/500 and $1/10^6$, respectively. In certain communities (French Canadians [Moorjani et al., 1989], Finns [Koivisto et al., 1992], Afrikaners [Kotze et al., 1989; Leitersdorf et al., 1989], Druze [Landsberger et al., 1992], and Lebanese [Lehrman et al., 1987]), FH frequency can be as high as 1/67 because of founder effects. The LDL receptor gene (*LDLR*) lies at 19p13.1-p13.3 [Lindgren et al., 1985; Yamamoto et al., 1984]. It contains 18 exons encoding the six functional domains of the protein: signal peptide, ligand-binding domain, epidermal growth factor (EGF) precursor like, O-linked sugar, transmembrane, and cytoplasmic [Sudhof et al., 1985].

THE LDLR DATABASE

The database structure contains five tables. These are schematically represented in Figure 1. For each mutation record, the amount of information that can be included in the clinical table is unlimited. Currently, we include clinical symptoms, lipid values, and the age of the proband at which these data were gathered. The current version of the LDLR database contains 840 records. Table 1 shows 80 mutations among the 490 new entries of the database. They have been selected either because of their interesting characteristics or because their proper name with respect to the present nomenclature is noteworthy. However, the complete Table S1 reporting the data of the 490 new entries is available online as supplementary material at the website: http://www.wiley.com/humanmutation/suppmat/2002/v20.html.

As in previous reports [Varret et al., 1997b; Varret et al., 1998], usual mutation names are given and are often followed by the name of the city or country from which the proband's family originated. This last version includes also the mutation name recommended by S. Antonarakis and the Nomenclature Working Group for Human Gene Mutations [1998; Den Dunnen and Antonarakis, 2000]. The software has been modified to include splicing mutations that have been named as recommended. For each mutation, information is provided at several levels: gene (exon and codon number, wild-type and mutant codon, mutational event, mutation name), protein (wild-type and mutant amino acid, affected domain, activity, mutation class), personal (ethnic background, age, sex, body mass index, familial history of coronary heart disease), clinical (values of plasma total cholesterol, LDL-cholesterol, HDL-cholesterol and triglycerides, presence or absence of xanthomas, arcus corneae, and symptomatic coronary heart disease), and impact (private, recurrent, founder). Identical mutations reported by different teams with no haplotype analysis were considered potentially recurrent when they were identified in probands with distant geographic or ethnic background.

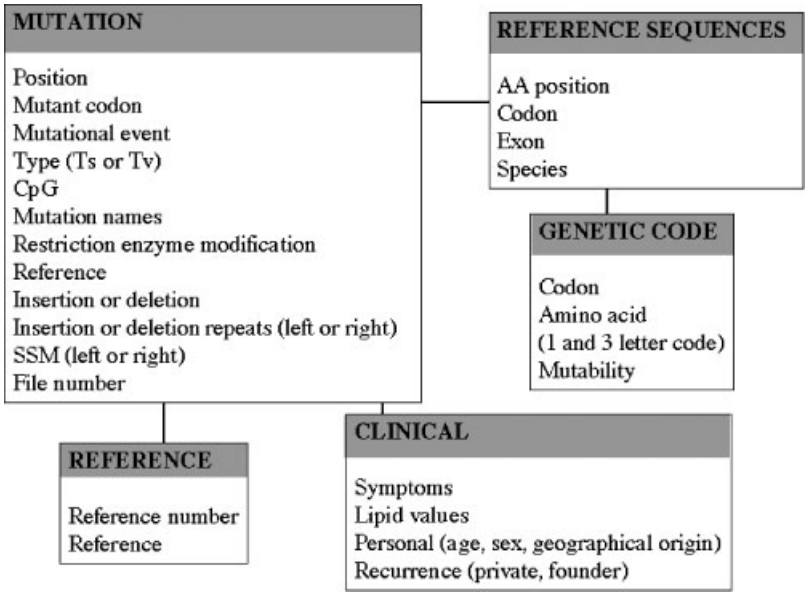

FIGURE 1. **Structure of the UMD-LDLR database. Each box represents one of the five tables of the software. AA: amino acid; SSM: slipped strand mispairing; Ts or Tv: transition or transversion.**

TABLE 1. The 80 Selected Mutations Among the 490 New Mutation Reports of the LDLR Database*

A	B	C	D	E	F	G	H	I	J	K	L
477	1	1A>C	Unknown	M-21L	SP	Htz	Wa	–	?	Spanish	164
670	1	1A>G	Unknown	M-21V	SP	Htz	Wa	–	?	German	161
478	1	11G>A	W4X	W-18X	SP	Htz	Wa	–	?	Spanish	164
510	1	12G>A	W4X	W-18X	SP	Htz	W(a1/a2)	182	?	Spanish	79
655	2	81T>G	C27W	C6W	LB #1	Htz	Wa	–	?	Greek	155
726	2	81T>G	C27W	C6W	LB #1	Htz	Wa	–	?	German	163
741	2	81T>G	C27W	C6W	LB #1	Htz	Wa	–	?	Anglo-Saxon	163
759	2	95T>G	F32C	F11C	LB #1	Htz	Wa	–	P	Dutch	167
377	2	97C>T	Q33X	Q12X – MILAN 4	LB #1	Hmz	aa	–	?	Italian	143
441	2	103C>T	Q35X	Q14X – MILAN 3	LB #1	Hmz	ab	100	P	Italian	143
760	2	108C>A	D36E	D15E	LB #1	Htz	Wa	–	P	Dutch	167
742	2	114dupA	C39fsX51	112insA	LB #1	Htz	Wa	–	P	Spanish	164
542	2	118delA	I40fX205	118delA	LB #1	Htz	Wa	–	P	English	120
761	2	148delG	50fsX205	148delG	LB #1	Htz	Wa	–	P	Dutch	167
722	2	171.173delTGA	D57.E58delinsE	171del3	LB #1	Htz	Wa	–	P	Anglo-Saxon	163
727	2	187T>C	C63R	C42R	LB #1	Htz	Wa	–	P	Italian	163
599	2–3	IVS2+1G>A	Probable fs	190+1(G>A)	LB #2	Htz	Wa	–	P	SAfr	136
762	2–3	IVS2-2G>A	Probable fs	191-2 (G>A)	LB #2	Htz	Wa	–	P	Dutch	167
498	3	194.195insAT	V66fsX206	195ins2	LB #2	Hmz	aa	–	F 4/218	German	99
626	3	211delG	G71fsX205	211delG	LB #2	Htz	Wa	–	?	Spanish	132
497	3	224G>A	C75Y	C54Y	LB #2	Hmz	aa	–	P	German	99
763	3	241C>A	R81S	R60S	LB #2	Htz	Wa	–	?	Dutch	167
638	3	242dupG	R81fsX129	242insG - BOMBAY 1	LB #2	Htz	Wa	-	P	Indian	151
696	3	246C>A	C82X	C61X	LB #2	Htz	Wa	–	P	Czech	162
764	3	250C>T	P84S	P63S	LB #2	Htz	Wa	–	P	Dutch	167
499	3	256.264del	F86.R88del	256del9	LB #2	Hmz	aa	–	P	German	99
694	3	257.265del	F86.C89delinsC	257del9	LB #2	Htz	Wa	–	P	German	161
697	4	318dupC	K107fsX129	313insC	LB #3	Htz	Wa	–	P	Czech	162
372	4	320dupA	K107fsX129	319insA	LB #3	Htz	Wa	–	P	North Irish	98
643	4	323C>T	T108M	T87M	LB #3	Htz	Wa	–	P	German	153
443	4	352G>T	D118Y	D97Y-NAPLES 3	LB #3	Hmz	ab	394	P	Italian	143
545	4	353delA	D118fsX205	353delA	LB #3	Htz	Wa	–	P	English	120
584	4	355.361del	G119fsX203	355del7	LB #3	Htz	Wa	–	P	Japanese	127
598	4	370C>T	R124W	R103W	LB #3	Htz	Wa	–	P	USA - White	135
547	4	530C>T	S177L	S156L	LB #4	Htz	Wa	–	?	English	120
704	4	651.653delTGG	D217.G218delinsD	651del3	LB #5	Htz	Wa	–	P	Czech	162
548	4	652.654delGGT	G218del	652del3	LB #5	Htz	Wa	–	?	English	120
774	4	652.654delGGT	G218del	652delGGT	LB #5	Htz	Wa	–	?	Dutch	167
746	6	884delT	V295fsX369	884delT	LB #7	Htz	Wa	–	P	Spanish	164
711	6	896delC	A299fsX369	896delC	LB #7	Htz	Wa	–	P	Czech	162
562	7	1008.1031del	Y336.F344delinsY	1008del24C	EGF p A	Htz	Wa	–	P	English	121
714	7	1053.1060dup8	D354fsX357	1053ins8	EGF p A	Htz	Wa	–	P	Czech	162
422	7	1056C>G	C352W	C331W - AVELLINO 1	EGF p A	Htz	Wa	–	R	Italian	165
471	7	1056C>A	C352X	C331X	EGF p A	Htz	Wa	–	P	Polish	118
794	7	1057G>A	E353K	E332K	EGF p A	Htz	Wa	–	P	Dutch	167
799	8	1118G>A	G373D	G352D	EGF p B	Htz	Wa	–	?	Dutch	167
551	8	1118.1121dupGTGG	G374fsX380	1118ins4	EGF p B	Htz	Wa	–	P	English	120
593	8	1119.1122dupTGGC	Y375fsX380	1122ins4 - PISA	EGF p B	Htz	Wa	–	P	Italian	129
690	9–10	IVS9+2T>A	Probable fs	1358+2 (T>A)	EGF p like	Htz	Wa	–	P	German	161
752	9–10	IVS9-1G>A	Probable fs	1359-1 (G>A)	EGF p like	Htz	Wa	–	?	Spanish	164
521	9–10	IVS9-1G>A	Probable fs	1359-1(G>A)	EGF p like	Htz	Wa	–	?	Dutch	48
564	10	1415.1418dupACAT	I473fsX535	1415ins4	EGF p like	Htz	Wa	–	P	English	121
432	10	1418T>A	I473N	I452N	EGF p like	Htz	Wa	–	P	Danish	166
454	10	1418.1419ins4	I473fs	1419ins4 - SAVONA	EGF p like	Htz	Wa	–	F 18/100	Italian	165
820	10	1420C>T	Q474X	Q453X	EGF p like	Htz	Wa	–	P	Dutch	167
651	10	1428dupC	D477fsX535	1428insC	EGF p like	Htz	Wa	–	P	German	153
620	11	1591.1626del	M531.I542del	1591del36	EGF p like	Htz	Wa	–	P	Korean	123
583	11	1599G>A	W533X	W512X	EGF p like	Htz	Wa	–	P	Japanese	127
621	11	1600.1608del	T534.W536del	1600del9	EGF p like	Htz	Wa	–	P	Korean	123
565	11	1618G>A	A540T	A519T	EGF p like	Htz	Wa	–	?	English	121
829	11	1637G>A	G546D	G525D	EGF p like	Htz	Wa	–	?	Dutch	167
496	11	1644-1645insC	G549fsX558	1645insC	EGF p like	Htz	Wa	–	? - F	Dutch	92
832	11–12	IVS11-10G>A	Unknown	1706-10 (G>A)	EGF p like	Htz	Wa	–	?	Dutch	167
532	11–12	IVS11-10G>A	Unknown	1706-10(G>A)	EGF p like	Htz	W(a1/a2)	505	P	Spanish	79
842	13	1867.1868delAT	I623fsX643	1867delAT	EGF p like	Htz	Wa	–	P	Dutch	167
755	13	1871.1873delTCA	I624.N625delinsN	1871del3	EGF p like	Htz	Wa	–	P	Spanish	164
682	13	1874A>C	N625T	N604T	EGF p like	Htz	Wa	–	P	German	161
586	14	2035.2036insT	Y679fsX679	2035ins1	EGF p C	Htz	Wa	–	P	Japanese	127
573	14	2041T>A	C681S	C660S	EGF p C	Htz	Wa	–	P	English	121
623	14	2043C>A	C681X	C660X	EGF p C	Hmz	aa	–	?	Cypriot	148
472	14	2046.2049delCCCT	L682fsX707	2046del4	EGF p C	Htz	Wa	–	P	Polish	118
844	14	2050.2063del14	A683fsX711	2050del14	EGF p C	Htz	Wa	–	P	Dutch	167
845	14	2053C>A	P685T	P664T	EGF p C	Htz	Wa	–	P	Dutch	167
509	14	2054C>T	P685L	P664L	EGF p C	Htz	Wa	–	?	Czech	102

TABLE 1. Continued.

574	14	2054C>T	P685L	P664L	EGF p C	Hmz	aa	–	?	English	122
575	14	2054C>T	P685L	P664L	EGF p C	Htz	Wa	–	F 6/100	Korean	147
577	14	2054C>T	P685L	P664L	EGF p C	Htz	Wa	–	?	Chinese	131
473	14	2056C>T	Q686X	Q665X	EGF p C	Htz	Wa	–	?	Polish	118
846	14	2056.2068del13	Q686fsX704	2056del13	EGF p C	Htz	Wa	–	P	Dutch	167
572	14	2060delT	I687fsX708	2060delT	EGF p C	Htz	Wa	–	P	English	121

*The full list of 490 new mutation reports is available as Supplementary Material, Table S1, online at www.interscience.wiley.com/humanmutation/suppmat/2002/v20.html.

Each line represents a single *LDLR* gene mutation report. The columns contain the following information and abbreviations: **A**: Report number. **B**: Exon number in which the mutation occurred. Exons are numbered according to Südhof et al. [1985]. **C:** Nomenclature name at the DNA level, according to the international nomenclature and the Nomenclature Working Group [Antonarakis 1998; Den Dunnen and Antonarakis, 2000]. Nucleotide position and modification are shown. Substitutions are indicated with the wild-type and mutated nucleotide, deletion (del) with the first and last number of the deleted nucleotides and their nature, duplications (dup) with the first and last number of the duplicated nucleotides, and insertion (ins) with its precise location (flanking nucleotides) and the nature of the new nucleotides. Intron mutations are named IVS followed by the position and substitution of the modified nucleotide. "+1" represents nucleotide G of the donor site, and "−1" represents nucleotide G of the acceptor site. **D**: Nomenclature name at the protein level, according to the international nomenclature [Antonarakis, 1998; Dunnen and Antonarakis, 2000]. The first amino acid of the LDL receptor is the initiator methionin of the signal peptide domain. Missense mutations are designated by the codon number flanked by the single letter code of the normal amino acid prior and of the mutant amino acid after (e.g. Cys to Tyr at codon 75 is designated C75Y). Nonsense mutations are designated similarly except that X is used to indicate any termination codon (e.g. Cys to stop at codon 352 is designated C352X). Deletions are designated by "del" after the amino acid interval (e.g. F86_R88del designates a deletion from phenylalanine 86 to arginine 88); duplications are designated by "dup" following the duplicated amino acids (e.g. C803_L808dup designates a duplication from cysteine 803 to leucine 808). Deletions, that do not alter the reading frame are designated by the first and last affected amino acid, followed by "del" (deletion) and the inserted amino acid (e.g. F86_C89delinsC designates a deletion from phenylalanine 86 to cysteine 89 and the insertion of a cysteine). Finally, frameshift mutations are designated by "fs" preceded by the first affected amino acid and followed by the location of the premature Stop codon (e.g D118fsX205 designates a frameshift affecting the aspartic acid 118 and the creation of a new Stop codon at 205). "Unknown" is used when no information is available at the protein level. "Probable fs" is used for splicing mutations affecting the consensus nucleotides of the acceptor and the donor splice sites (+1, +2, −1 and −2) when no protein data are available, but an aberrant splicing mechanism is highly probable. **E**: Former mutation names according to the amino acid numbering from Yamamoto et al. [1984]. Therefore, the C75Y and C352X mutations are named in this column C54Y and C331X, respectively. Splicing mutations are named with the number of the codon at the splice junction followed by a "+" for donor sites or a "−" for acceptor sites, the number of the base in the intron sequence and the nature of the substitution (e.g. 190+1G>A designates a G>A substitution at the first base of intron 2). The original names also appear in this column. These names were given according to the population or the city in which the mutation was first reported (e.g. TUNIS). **F**: Protein domain in which the mutation occurs. SP, Signal Peptide; LB, Ligand-Binding domain; EGF, Epidermal Growth Factor precursor-like domain. **G**: Clinical status according to Goldstein and Brown [1989]: Hmz, homozygotes; Htz heterozygotes. **H**: Genotype: aa, homozygotes; ab, compound heterozygotes; Wa, heterozygotes; W(a1/a2), patients with two mutations on the same allele. **I**: Number of the report in which the second mutation identified in a compound heterozygote or a two-mutation allele carrier is described. When the second mutation is one of those omitted in the database, this mutation is briefly described with respect to the coding sequence. Finally, "?" indicates that the second mutation has not been identified; **J**: Recurrence of the mutation. F, a founder effect; F 18/100, mutation was found in 18 unrelated probands in a sample of 100 FH patients; R, recurrent mutations; ?, mutations that have been identified in at least two unrelated probands of different ethnic backgrounds but for which *LDLR* gene haplotypes are not described or not comparable; ?-F, mutations for which *LDLR* gene haplotypes are not described (or incomplete) and that either are associated with a founder effect in the proband's ethnic or geographic origin, or have been identified in at least two unrelated probands of the same ethnic or geographic background; and P, mutations identified, to date, in a single proband. **K:** Ethnic or geographic background of the proband. **L:** Reference number indicating the publication in which the mutation is described. Full citations (authors, year, title, journal, volume, pages) are provided on the UMD-LDLR Web site: http://www.umd.necker.fr. If the same mutation has been reported for the same patient in different papers, only one entry is given.

DEVELOPED SOFTWARE ROUTINES

The software package now contains 13 routines available on the Web site, for the analysis of the UMD-LDLR database that were developed with the 4th Dimension® (4D, San Jose, CA) package from ACI. The purpose of the software is to facilitate the mutational analysis of the *LDLR* gene at the molecular level and to provide the tools to promote the analysis of relationships between phenotype and genotype:

1. "Position" studies the distribution of mutations at the nucleotide level to identify preferential mutation sites.
2. "Statistical evaluation of mutational events" is comparable to 1, but also indicates the type of mutational event. The result can either be displayed as a table or in a graphic representation.
3. "Frequency of mutation" allows one to study the relative distribution of mutations at all sites and to sort them according to their frequency. A graphic representation is also available and displays a cumulative chart of mutation distribution.
4. "Stat exons" studies the distribution of mutations in the different exons and enables detection of a statistically significant difference between observed and expected mutations.

5. "Protein" studies the distribution of mutational events in various protein domains (ligand-binding and EGF-precursor-like motifs) after alignment of the amino acids of repeat motifs for each domain type.
6. "Insertions and deletions analysis" searches for repeated sequences surrounding the mutation and possibly involved in the mutational mechanism.
7. "Restriction enzyme" appears on the first page of the mutation record. If the mutation modifies a restriction site, the program shows a restriction map displaying the new or abolished site and the enzymes of interest.
8. "Amino acid type search" studies the mutations with respect to phylogenic conservation. In effect, the LDLR gene has been identified, sequenced, and converted to protein sequence in four mammalian species (complete coding sequence of the Chinese hamster (SWISS-PROT accession number: P35950), the rabbit (P20063), the rat (P35952), and the mouse (P35951) LDL receptor and in *Xenopus laevis* [Mehta et al., 1991]). The identity at the amino acid level between the human and Chinese hamster, rabbit, rat, mouse, and *Xenopus laevis* sequences are 81%, 79%, 77%, 76%, and 70%, respectively. The routine lists the mutations affecting conserved or non-conserved amino acids in the four mammals, in the *Xenopus laevis*, or in all these sequences.
9. "Phylogeny" studies the distribution of mutations (missense, stop, and frameshift) in conserved amino acids between humans and mammals or vertebrates and in amino acids specifically found in the human protein.
10. "Binary comparison" compares two mutation groups, each group being defined by distinct research criteria chosen from the database records (molecular, clinical, personal, etc.). The result can be displayed as either of several graphic representations (by amino acids, by exon, or by protein domain) of the distribution of the sorted mutations. Furthermore, the sorted mutations can also appear in a cumulated or detailed format (insertion, deletion, missense, nonsense).
11. "CpG" displays the consequences of the various mutations along the LDL receptor with special emphasis on CpG dinucleotides.
12. "AA modifications" displays the consequences of the mutations at the amino acid level. This function compares for each amino acid the expected and observed number of mutations.
13. "Base modifications" displays the consequences of mutations at the nucleotide level, according to their position in the codon.

DESCRIPTION OF MUTATIONS

To date, 920 mutations have been identified in the LDLR gene. At the cDNA level, most mutations (840) are point mutations and distributed as follows: 610 substitutions, 121 small deletions, 53 small insertions or duplications. At the genomic level, there are also 57 intronic substitutions, of which 37 (65%) affect consensus splice sequences (+1, +2, –1, and –2). At the protein level, mutations are distributed as follows: 507 missense mutations, 103 nonsense mutations, 139 premature Stop codons, and 35 in frame amino acid insertions or deletions. Major rearrangements (79) account for 10% of all mutations and are in majority the result of unequal recombination between the 30 *Alu* sequences identified throughout the gene [Hobbs et al., 1990]. Only seven mutations (1%) have been found in the promoter. Presently, major rearrangements, promoter mutations, and polymorphisms are not reported in the database since the software cannot accommodate this type of variation. However, they are recorded in the LDLR/Familial Hypercholesterolemia Web site with a complete description.

MUTATION NAMES (RECOMMENDED AND USUAL ONES)

The major modification of this new version of the UMD-LDR database is the addition of the name of the mutation in accordance with the internationally accepted nomenclature (Table 1, columns C and D). Indeed, the first mutations identified were numbered according to the first amino acid of the mature receptor (alanine), which was thought to be the first amino acid of the protein. It was subsequently shown that the protein is synthesized as a precursor with a signal peptide of 21 amino acids. Therefore, mutations occuring in exon 1 were named with negative amino acids, whereas the first amino acid of the protein is a methionin, and mutations should be numbered accordingly [Antonarakis and the Nomenclature Working Group, 1998; Den Dunnen and Antonarakis, 2000]. The recommended mutation name (the methionine being the number 1 amino acid) was added in this new version but the old name, still the most commonly used in the field, was also kept (Table 1, columns C, D, and E).

DESCRIPTION OF FILES

For many of the 840 mutations included in UMD-LDLR, data are incomplete for functional, personal, or clinical information. Indeed, lipid values are provided in 289 cases for total cholesterol, 406 cases for LDL-cholesterol, 251 cases for triglycerides, and 226 cases for HDL-cholesterol. Presence or absence of symptomatic CHD is given in 174 cases and presence or absence of xanthomas is reported for 216 cases. To

overcome this shortage, we are currently developing direct submission through the Web site that should facilitate the input of high quality clinical information for each mutation. To date, a mutation report sheet can be loaded on the LDLR/FH Web site in PDF or Word format.

Recurrent mutations have been recorded if carriers of the same mutation were from distant ethnic or geographic backgrounds (when no comparable haplotypes of the *LDLR* gene were available). This is the case for the P685L (previous name: P664L) mutation for which 12 reports appear in the database (Indian, English, Japanese, Italian, Norwegian, Chinese, German, Korean, Dutch, British, Belgian/Flemish-Wallon, and Czech patients). The probands originated from 12 areas, indicating that this particular substitution is one of the most frequent *LDLR* gene mutations. Founder effects were reported in five countries: England, Japan, Korea, Italy, Norway, and Holland, with a frequency of 2%, 3.3%, 6.2%, 8.9%, 4%, and 0.8% respectively. The same mutation identified in Belgium, Czech Republic, China, and Germany was found in single probands with no haplotype data and each report was annotated as "?". It is noteworthy that this mutation found in two different probands in England is associated with different haplotypes, while it is carried by the same haplotype and consistent with an old founder mutation in Indian, Norwegian, and Dutch probands. Finally in this current version of UMD-LDLR, independent mutations that are found on the same allele (or "two-mutation alleles") are reported in two separate files. Such mutational events represent 16 files in the database. The carrier of a two-mutation allele is indicated with a W(a1/a2) genotype and the name of the file reporting the second mutational event is given (Table 1, columns H and I). These two-mutation alleles should not be confused with compound heterozygosity (two different alleles each carrying a single mutation) that are indicated as "ab" (column H).

CONCLUSION

The UMD-LDLR database not only lists reported mutations in the *LDLR* gene, but also provides powerful tools to analyze precisely such a large number of mutations. The analysis of the proportion of mutations shows that point mutations account for 90% and major rearrangements for 10% of all *LDLR* gene mutations. Previous reports estimated a higher proportion of major rearrangements in the *LDLR* gene: 1/3 [Hobbs et al., 1990], and 25% [Varret et al., 1997a]. This difference is due to the fact that in the last 10 years the majority of *LDLR* mutation screening programs only searched for point mutations. In fact, techniques such as SSCP, DGGE, DHPLC, and sequencing are continually improving in efficiency and speed. Thus, they are more commonly used than the classical Southern method that remains the only way to search for major rearrangements. Therefore, it is probable that the present proportion of major rearrangements in the *LDLR* gene is underestimated.

Mutations in the *LDLR* gene are numerous and frequently recurrent but, conversely, rarely sporadic. These observations reveal not only the high mutability at one time of this gene, but also that these mutations were probably selected through time. It can be postulated that an hypercholesterolemic mutation could have given a selective advantage to carriers and may be a member of the pool of alleles that constitute the "thrifty genotype." Another hypothesis is that the mutation conferred resistance to rhinovirus HRV2 binding, since the LDL receptor was shown to directly interact with this class of virus [Marlovits et al., 1998; Rong et al., 1998; Schober et al., 1998].

Many mutations have been reported several times, and for most of them, no haplotype studies are available, leading probably to inclusion of more mutational events. This raises the problem of recurrent mutations that could be resolved by haplotyping. However, there is no consensus on the *LDLR* gene polymorphisms that should be tested. Moreover, many reports do not include clinical data thus limiting genotype/phenotype studies. Recently, through a constructive collaboration, we were able to include clinical data for 48 mutations previously reported by the South Africa team of Maritha Kotze [Kotze et al., 1989] for the C143Y and E317X mutations (C122Y and E296X) previously reported by Genschel et al. (Germany) [2001], and for two Indian mutations (242insG and 395insG) reported by Ashavaid et al. [2000]. Therefore, data collection is possible and should be enhanced through the availibilty on the LDLR/FH Web site of the mutation report sheet that can be directly downloaded. Furthermore, more information concerning previously reported mutations could also be provided by authors. Altogether, this should facilitate curatorship and provide regular input of good quality molecular and clinical data into this reference database.

DATABASE ON THE WEB

The UMD-LDLR database is available through the World Wide Web at www.umd.necker.fr. The LDLR/Familial Hypercholesterolemia Web site is available at www.umd.necker.fr/LDLR/Home_Page.html, where a link to the software can be found. Users of the database must cite this article. Finally, notification of omissions and errors in the current version would be gratefully received by the corresponding authors.

ACKNOWLEDGMENTS

This work is supported by grants from Ministère de l'Education Nationale, de la Recherche, et de la Technologie (to L.V.), and ARCOL and SFA (to M.A.). We gratefully acknowledge the help of collaborators, and especially from Dr. Genschel (Germany) and Dr. Ashavaid (India).

REFERENCES

Antonarakis SE, Nomenclature Working Group. 1998. Recommendations for a nomenclature system for human gene mutations. Hum Mutat 11:1–3.

Ashavaid TF, Kondkar AA, Nair KG. 2000. Identification of two LDL-receptor mutations causing familial hypercholesterolemia in Indian subjects by a simplified rapid PCR-heteroduplex method. Clin Chem 46:1183–1185.

Den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

Genschel J, Thomas HP, Kassner U, Lochs H, Steinhagen-Thiessen E, Schmidt HH. 2001. Two novel LDL receptor mutations in familial hypercholesterolemia: C122Y and E296X. Hum Mutat 17:354.

Goldstein JL, Brown MS, Anderson RG, Russell DW, Schneider WJ. 1985. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. Annu Rev Cell Biol 1:1–39.

Goldstein JL, Brown MS. 1989. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, editors. The metabolic basis of inherited diseases, 6th ed. New-York: Mc Graw-Hill. p 1215–1250.

Hobbs HH, Russell DW, Brown MS, Goldstein JL. 1990. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Annu Rev Genet 24:133–170.

Koivisto UM, Turtola H, Aalto-Setala K, Top B, Frants RR, Kovanen PT, Syvanen AC, Kontula K. 1992. The familial hypercholesterolemia (FH)-North Karelia mutation of the low density lipoprotein receptor gene deletes seven nucleotides of exon 6 and is a common cause of FH in Finland. J Clin Invest 90:219–228.

Kotze MJ, Langenhoven E, Warnich L, du Plessis L, Marx MP, Oosthuizen CJ, Retief AE. 1989. The identification of two low-density lipoprotein receptor gene mutations in South African familial hypercholesterolaemia. S Afr Med J 76: 399–401.

Landsberger D, Meiner V, Reshef A, Levy Y, van der Westhuyzen DR, Coetzee GA, Leitersdorf E. 1992. A nonsense mutation in the LDL receptor gene leads to familial hypercholesterolemia in the Druze sect. Am J Hum Genet 50:427–433.

Lehrman MA, Schneider WJ, Brown MS, Davis CG, Elhammer A, Russell DW, Goldstein JL. 1987. The Lebanese allele at the low density lipoprotein receptor locus. Nonsense mutation produces truncated receptor that is retained in endoplasmic reticulum. J Biol Chem 262:401–410.

Leitersdorf E, Van der Westhuyzen DR, Coetzee GA, Hobbs HH. 1989. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. J Clin Invest 84:954–961.

Lindgren V, Luskey KL, Russell DW, Francke U. 1985. Human genes involved in cholesterol metabolism: chromosomal mapping of the loci for the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl-coenzyme A reductase with cDNA probes. Proc Natl Acad Sci USA 82:8567–8571.

Marlovits TC, Zechmeister T, Gruenberger M, Ronacher B, Schwihla H, Blaas D. 1998. Recombinant soluble low density lipoprotein receptor fragment inhibits minor group rhinovirus infection in vitro. Faseb J 12:695–703.

Mehta KD, Brown MS, Bilheimer DW, Goldstein JL. 1991. The low density lipoprotein receptor in Xenopus laevis. II. Feedback repression mediated by conserved sterol regulatory element. J Biol Chem 266:10415–10419.

Moorjani S, Roy M, Gagne C, Davignon J, Brun D, Toussaint M, Lambert M, Campeau L, Blaichman S, Lupien P. 1989. Homozygous familial hypercholesterolemia among French Canadians in Quebec Province. Arteriosclerosis 9:211–216.

Rong L, Gendron K, Bates P. 1998. Conversion of a human low-density lipoprotein receptor ligandbinding repeat to a virus receptor: identification of residues important for ligand specificity. Proc Natl Acad Sci USA 95:8467–8472.

Schober D, Kronenberger P, Prchla E, Blaas D, Fuchs R. 1998. Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. J Virol 72:1354–1364.

Sudhof TC, Russell DW, Goldstein JL, Brown MS, Sanchez-Pescador R, Bell GI. 1985. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. Science 228: 893–895.

Varret M, Rabès J-P, Boileau C. 1997a. L'hypercholestérolémie familiale 25 ans après Brown et Goldstein. I: Le point sur la maladie du gène du récepteur LDL, ses modèles animaux et sa thérapie génique. m/s 12:1399–1408.

Varret M, Rabes JP, Collod-Beroud G, Junien C, Boileau C, Beroud C. 1997b. Software and database for the analysis of mutations in the human LDL receptor gene. Nucleic Acids Res 25:172–180.

Varret M, Rabes JP, Thiart R, Kotze MJ, Baron H, Cenarro A, Descamps O, Ebhardt M, Hondelijn JC, Kostner GM, Miyake Y, Pocovi M, Schmidt H, Schuster H, Stuhrmann M, Yamamura T, Junien C, Beroud C, Boileau C. 1998. LDLR Database (second edition): new additions to the database and the software, and results of the first molecular analysis. Nucleic Acids Res 26:248–252.

Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, Russell DW. 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39:27–38.