Characterization of the Human ABCG1 Gene

LIVER X RECEPTOR ACTIVATES AN INTERNAL PROMOTER THAT PRODUCES A NOVEL TRANSCRIPT ENCODING AN ALTERNATIVE FORM OF THE PROTEIN*

Received for publication, June 24, 2001, and in revised form, July 31, 2001
Published, JBC Papers in Press, August 10, 2001, DOI 10.1074/jbc.M105863200

Matthew A. Kennedy‡‡, Asha Venkateswaran§, Paul T. Tarr§, Ioannis Xenarios¶, Jun Kudoh**, Nobuyoshi Shimizu**, and Peter A. Edwards†† ‡‡ From the Departments of Biological Medicine and Medicine and the Molecular Biology Institute, University of California, Los Angeles, California 90095 and the ‡Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan.

The human ABCG1 gene encodes a member of the ATP-binding cassette (ABC) superfamily of transporter proteins and is highly induced when macrophages are incubated with oxysterols. Using mRNA from oxysterol-treated human THP-1 cells together with 5′-rapid amplification of cDNA ends and polymerase chain reaction, we identified a novel ABCG1 transcript that encodes a putative protein of 786 residues containing a new amino terminus of 203 amino acids. Characterization of the genomic organization and structure of the human ABCG1 gene demonstrates that: (i) the gene consists of 23 exons spanning 98 kilobase pairs (kb) on chromosome 21q22.3, (ii) the 203 amino acids are encoded on three previously unidentified exons, 8–10, and (iii) a promoter, containing a TATA box and two liver X receptor (LXR) α response elements (LXREs), is located upstream of exon 8. Northern analysis using exon-specific probes confirms that oxysterol treatment results in >10-fold induction of ABCG1 transcripts that are derived from either exons 8–23 or exons 5, 7, and 11–23. Electromobility shift assays demonstrate that LXRs and retinoid X receptor α bind to the two LXREs in intron 7. Cells were transiently transfected with reporter luciferase constructs under the control of either (i) 9 kb of genomic DNA corresponding to intron 7 and part of exon 8 and containing either wild-type or mutant LXREs or (ii) two copies of the wild-type or mutant LXRE. In all cases, the wild-type construct was regulated in an LXR- and oxysterol-dependent manner, and this regulation was attenuated when the LXREs were mutated. In conclusion, the human ABCG1 gene contains multiple promoters, spans more than 98 kb and comprises 23 exons that give rise to alternative transcripts encoding proteins with different amino-terminal sequences. Elucidation of the various roles of different ABCG1 isoforms will be important for our understanding of mammalian cholesterol homeostasis.

* This work was supported in part by National Institutes of Health Grant HL 30568 and a grant from the Laubisch Fund (both to P. A. E.), by Grant-in-aid for Scientific Research on Priority Areas (C) “Medical Genome Science” from the Ministry of Education, Science, Sports and Culture of Japan (to J. K.), and by a grant from the “Research for the Future” program of the Japan Society for the Promotion of Science (to N. S.). The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY048757.

† Supported by a postdoctoral fellowship from the National Institutes of Health.
‡ Fellow of the Swiss National Fund.
§ To whom correspondence should be addressed: Dept. of Biological Chemistry, CHS 33-257, UCLA, Los Angeles, CA 90095-1769. Tel.: 310-206-3717; Fax: 310-794-7345; E-mail: pedwards@mednet.ucla.edu.

In the presence of oxidized, aggregated, or acetylated LDL, macrophages take on a “foamy” appearance as a result of the cytoplasmic accumulation of cholesteryl ester lipid droplets (1–3). Such macrophage foam cells are found in both fatty streaks and more advanced lesions in the artery wall and appear to be important in the development of atherosclerosis (2). Recent studies have begun to explore the changes in macrophage gene expression that occur during lipid loading, on the assumption that such changes effect the development of fatty streaks and/or the stability of advanced plaques. These studies led to the identification of three mRNAs, ABCG1, ABCA1, and apoE, that are highly induced when macrophages are incubated in the presence of either modified LDL or specific oxysterols (4–7). Induction of each mRNA requires the nuclear hormone receptor LXR (4–6, 8, 9).

ABCG1 (also referred to as human White or murine ABC8), and ABCA1 are two members of the ATP-binding cassette (ABC) transporter superfamily of proteins (10–12). The ABC superfamily consists of membrane-bound proteins that mediate the ATP-dependent translocation of a variety of amphiphilic and lipophilic molecules (12). The ABC transporter family is large and currently is comprised of over 359 members, including 129 in Arabidopsis thaliana (13), 55 in Drosophila melanogaster, 56 in Caenorhabditis elegans, 32 in Saccharomyces cerevisiae, 36 in Mus musculus, and 51 in Homo sapiens (www.med.rug.nl/mdl/humanabc.htm). ABC transporters fall into one of two broad groups: full ABC transporters, which contain two symmetrical halves, each half containing one ABC and six transmembrane domains; half transporters, which contain one ABC and six transmembrane domains, and which dimerize to form a functionally active transporter (10, 12). To date, 27 full and 25 half transporters have been identified in humans (14).

Each ABC is characterized by a conserved nucleotide binding domain comprising three motifs: a Walker A and Walker B motif (each approximately 9 amino acids in length) spaced by 90–120 amino acids, and a “signature” motif of 12 amino acids located 8 amino acids upstream of the Walker B motif (10, 12, 14). The ATP-binding fold is required for the binding and hydrolysis of ATP, providing the energy for the transport ac-

The abbreviations used are: LDL, low density lipoprotein; HDL, high density lipoprotein; LXR, liver X receptor; RXR, retinoid X receptor; ABC, ATP-binding cassette; ABCA1, ATP-binding cassette transporter A1; ABCG1, ATP-binding cassette transporter G1; apoE, apolipoprotein E; apoA-1, apolipoprotein A-1; CPT2A1, cholesterol 7α-hydroxylase; RACE, rapid amplification of cDNA ends; LXRE, LXR response element; LPDS, lipoprotein-deficient serum; PCR, polymerase chain reaction; RT-PCR, reverse transcribed polymerase chain reaction; TK, thymidine kinase; BAC, bacterial artificial chromosome; bp, base pair(s); kb, kilobase pair(s); nt, nucleotide(s); GAPDH, glyceraldehyde-3-phosphate dehydrogenase.
tivity, while the transmembrane domains determine substrate specificity (12).

The human ABCG1 cDNA was first cloned when Chen et al. (15) used exon trapping experiments and cosmid, containing DNA from human chromosome 21, to identify cDNAs expressed in human retina cells. Independently, Croop et al. (16) isolated a cDNA homologous to Drosophila White from a human fetal brain cDNA library. The cDNAs isolated by these two groups have ~34% sequence identity with the Drosophila White gene (15, 16) and are identical to each other except that the cDNA isolated by Chen et al. contains an additional 108 nucleotides at the 5’ end (15). Murine ABCG1 cDNAs, isolated from a pre-B cDNA library (16) or a macrophage cell line, P38D1 (17), have 97% identity with the human cDNA (16, 17). Interestingly, neither the murine nor human cDNAs have an in-frame stop codon 5’ of the putative initiating ATG (15–17).

There are conflicting data on the genomic organization of the human ABCG1 gene. For simplicity, we refer here to exons 1–23, as illustrated in Fig. 1A. Langmann et al. (18) used data available from an early sequencing draft of the Human Genome Project, that included sequences derived from two BAC clones (AC015555 and AP001623), and reported that the entire cDNA sequence was contained on 15 exons (see Fig. 1A-IA). Langmann et al. (18) also performed 5’-RACE, using a cDNA library derived from human fetal liver and a primer corresponding to exon 7 (exon 2 in their nomenclature; see Fig. 1A) and identified a transcript that differed at the 5’ end from that reported by Chen et al. (15). In contrast, Lorkowski et al., performed 5’-RACE with a primer corresponding to exon 7 and RNA from THP1 cells treated with LXRα and RXRα ligands and identified five new exons (19) (exons 1–4 and 6 in Fig. 1A-III). The latter 5’-RACE studies identified ten different transcripts that were assumed to result from alternative splicing and/or the use of four putative promoters that were postulated to be located upstream of either exon 1, 4, 5, or 6 (19). Based on these results, and the observation that there were multiple in-frame methionines in exons 1–6, Lorkowski et al. (19) proposed that there might be 4–6 different ABCG1 protein isoforms that differed only in the —25 amino-terminal amino acids.

Previous studies have led to the proposal that the nuclear receptor LXR is required for the induction of ABCG1 mRNA levels in lipid-loaded macrophages; this proposal is based on the observations that induction of ABCG1 mRNA levels (i) is enhanced when cells are incubated with LXR ligands, (ii) is stimulated when cells overexpress LXR, (iii) is attenuated when cells overexpress dominant negative forms of LXRα or LXRβ, and (iv) is attenuated in peritoneal macrophages derived from LXRα−/−, LXRβ−/− double knockout mice (4, 5). Despite these observations, neither LXXRs nor other cis-elements that are involved in the oxysterol-dependent induction of ABCG1 mRNA levels have been identified. In contrast, functional LXXRs have been identified in either the proximal promoter or distal enhancer elements of the five other genes that have been shown to be activated in an oxysterol- and LXR-dependent manner (6, 9, 20–25); these genes encode cholesterol 7α-hydroxylase (20), apoE (6), cholesteryl ester transfer protein (21), sterol regulatory element-binding protein-1c (22, 23), and ABCA1 (9, 24, 25).

In the current study, we identify three new exons (designated exons 8–10; see Fig. 1) in the human ABCG1 gene that are part of a transcript derived from exons 8–23. In addition, we show that this transcript encodes a novel ABCG1 isoform containing a unique amino-terminal sequence of 203 amino acids. We also identify and characterize a novel promoter that lies upstream of exon 8 and contains a TATA box and two functional LXXRs. These findings identify the molecular mechanism by which oxysterols activate the human ABCG1 gene. The results suggest that ABCG1 expression, regulation, and function is complex and that identification of different ABCG1 isoforms will be important for our understanding of mammalian cholesterol homeostasis.

**EXPERIMENTAL PROCEDURES**

Reagents—pCMX expression plasmids for LXRα and RXRα were a gift from Ron Evans (Salk Institute). Oxysterols (Sigma), phorbol 12-myristate 13-acetate (Sigma), and the RXR-specific agonist LG 100153 (gift from Dr. Richard Heyman, Ligand Pharmaceuticals, La Jolla, CA) were dissolved in ethanol prior to the addition to cells (<1 μg/ml medium). DNA modification and restriction enzymes were obtained from New England Biolabs and Life Technologies, Inc. (α23-Pi dCTP and γ-32P[i]ATP were supplied by ICN Biomedicals. Lipoprotein-deficient fetal bovine serum (LPDS) and lipopolysaccharide were purchased from PerImmune and Lissatz, respectively. RXRα peptide polyclonal antibody (D20, sc-553X) was purchased from Santa Cruz Biotechnologies, Inc. All other reagents have been described previously (26).

Oligonucleotides and DNA Sequencing—Oligonucleotides used in this study were synthesized by Life Technologies, Inc. and are listed in Table I. 5’-RACE and PCR products were gel-purified and cloned into the pcR2.1-TOPO vector (Invitrogen) using the TOPO TA Cloning Kit (Invitrogen). DNA sequencing was performed using Big Dye Termina
tive Cycle sequencing (Applied Biosystems) with the M13 forward and reverse primers (Invitrogen) and DNA Sequencing was performed using Big Dye Termina
tive Cycle sequencing (Applied Biosystems) with the M13 forward and reverse primers (Invitrogen) and gene-specific primers (see Table I) by Laragen, Inc. (Los Angeles, CA).

Cell Culture—THP-1 human monocytic cells were cultured in RPMI medium containing 10% fetal bovine serum, 100 units/ml penicillin, and 100 units/ml streptomycin. Eight × 105 cells/ml were incubated with 0.2 μM phorbol 12-myristate 13-acetate for 48 h to promote differentiation into macrophages. The cells were then incubated in RPMI containing 10% LPDS and mevalonic acid (100 μM) in the presence of 5 μM 20(S)-hydroxycholesterol for an additional 24 h. HepG2 cells were cultured and maintained in modified Eagle’s medium containing 10% fetal bovine serum as described (27).

RNA Isolation and Northern Blot Analysis—Total RNA was isolated using Trizol reagent (Life Technologies, Inc.). Northern analyses (10 μg of RNA/lane) were performed as described (26) using the specific probes indicated in the text. Poly(A)+ mRNA was isolated with the mRNA purification kit (Amersham Pharmacia Biotech) from oxysterol-treated THP-1 macrophages. The relative abundance of transcripts was deter
dized by PhosphorImager analysis (ImageQuant software, Molecular Dynamics) and normalized against glyceraldehyde-3-phosphate dehy
denase (GAPDH) (4).

5’-RACE Analysis and PCR Confirmation—5’-RACE was performed with the SMART RACE cDNA amplification kit (CLONTECH) using 1 μg of poly(A)+ mRNA from oxysterol-treated THP-1 cells. First strand cDNA synthesis was carried out according to the manufacturer’s protocol (CLONTECH) using the SuperScript II reverse transcriptase (Life Technologies, Inc.). SMART II oligonucleotide, and 5’-RACE cDNA synthesis primer (CLONTECH). 5’-RACE PCR was performed using the UPM primer (CLONTECH) and the gene-specific primer GSP1 (Table I) with Advantage 2 polymerase (CLONTECH) according to the manufacturer’s recommendation in a PerkinElmer GeneAmp 2400 thermocycler. The sample was then diluted 50-fold, and 2.5 μl used in nested 5’-RACE with a second gene-specific primer, NGSIP1 (Table I), and the UPM primer (CLONTECH), according to the manufacturer’s protocol.

PCR amplifications were performed using various gene-specific primers (see Table I) and Advantage 2 polymerase (CLONTECH). The PCR conditions were: 94°C for 2 min followed by 30 cycles of 94°C for 15 s, 64°C for 30 s, 70°C for 2.5 min.

Electromobility Shift Assays—In vitro translated RXRα and LXXRs were generated from the pCMX-RXXα and pCMX-LXXα plasmids with the TNT quick-coupled transcription/translation system (Promega). Radiolabeled double-stranded oligonucleotides (see Table I) were prepared and used in electromobility shift assays (EMSA) with the in vitro translated proteins and, where indicated, anti-RXXα as described (28).

Construction of Luciferase Reporter Genes and Transfections—The wild-type and mutant two copy LXRκ TK-Luc reporter constructs (2xLXRκ-1, 2xLXRκ-2, 2xLXRκ-1m, and 2xLXRκ-2m) were generated by annealing complementary oligonucleotides (see Table I) containing BamHI overhangs and ligated into an empty TK-Luc reporter vector (29) at the BamHI site. Constructs were verified by se

quencing, as described above, using a TK-Luc-specific primer.

The ABCG1 promoter contained within intron 7 was amplified from
the BAC clone KB1430A10 by PCR using primers EV-1 and EV-4 (see Table I for primer sequences). First, a PCR product obtained with primers EV-5 (EV-5) was cloned into the pGL3-luciferase vector (Promega) to create pG-9.3. LXRE

- **Table I**

**Primers used in this study**

| Primer          | Sequence | Hybridization region | Orientation | Use in this study |
|-----------------|----------|----------------------|-------------|------------------|
| hABCG1 GSP1     | 5'-GGAGACGAGGGCGGCTGCAGG-3' | Exon 15 | Antisense | 5'-RACE |
| hABCG1 NSP1     | 5'-GGAGACGAGGGCGGCTGCAGG-3' | Exon 12 | Antisense | 5'-RACE |
| hABCG1 NSP5     | 5'-GGAGACGAGGGCGGCTGCAGG-3' | Exon 14 | Antisense | 5'-RACE |
| hABCG1 10-REV   | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 10 | Antisense | 5'-RACE |
| hABCG1 CONF1    | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 9 | Antisense | 5'-RACE |
| hABCG1 8-F      | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 5 | Sense | RT-PCR |
| hABCG1 12-F     | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 8 | Sense | RT-PCR |
| hABCG1 11-F     | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 10 | Sense | Northern analysis |
| hABCG1 8-R      | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 8 | Sense | Northern analysis |
| hABCG1 9-F      | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 8 | Sense | Northern analysis |
| hABCG1 10-F     | 5'-GGGGTGGCTACAGACGAGGGCGG-3' | Exon 10 | Sense | Northern analysis |

**Characterization of the Human ABCG1 Gene**—Homology searches were generated with the BLAST suite of algorithms at the NCBI (31). Sequence assemblies were performed with GeneTool (Doubletwist, San Francisco, CA). Simple and multiple alignments were generated with the BESTFIT and PILEUP algorithms at the Genetic Computing Group (GCG) Wisconsin package (32). exon and promoter predictions were performed with the GENSCAN and GRAIL algorithms (33, 34). Specific cis-element searches were performed with the FINDPATTERNS algorithm from the Genetic Computing Group (GCG) Wisconsin package (32). Analyses of potential promoter regulatory elements for hABCG1 were performed using the MATINSPECTOR (35) and TESS Transcription Element Search Software (36) algorithms. The prediction of putative transcription start sites was performed with the TSSW and TSSS algorithms (37, 38). Analysis for potential signal sequences was performed with the SignalP algorithms in the Genetics Computer Group (GCG) Wisconsin package (32). Potential CpG islands were predicted with the Gardiner-Garden algorithm (39).
and A-III). However, in the current study, we identify three novel exons and conclude that the human gene has 23 exons and a novel internal promoter (see Fig. 1A-I). To minimize confusion, we refer in the text to exons 1–23, as shown in Fig. 1A, unless otherwise noted.

Identification of New Exons and Transcripts in the Human ABCG1 Gene Using 5'-RACE from Exon 15—Human and murine ABCG1 cDNAs were reported to have 87% sequence identity (15–17). However, the initiating methionine was not identified definitively since both putative 5'-untranslated regions lacked a stop codon (15–17). Based on these results and our inability to identify a functional promoter upstream of exon 5 (originally defined as exon 1; see Fig. 1A-II), we hypothesized that the ABCG1 cDNA sequences might be incomplete and that additional nucleotides/exons might remain to be identified.

To test this hypothesis, we treated human monocytic leukemia THP-1 cells with a phorbol ester to promote their differentiation into macrophages. The cells were then incubated for 24 h in the absence or presence of 20(S)-hydroxycholesterol, a ligand for LXR (40–42). As expected (4), such treatment led to a marked induction of the ABCG1 mRNA (data not shown). Poly(A)+ mRNA was isolated from the oxysterol-treated cells and used as the template for first strand cDNA synthesis and 5'-RACE, as detailed under "Experimental Procedures," using the universal primer and a primer corresponding to exon 15 (see Fig. 1A-I).

This approach identified two distinct DNA products that were 509 and 482 bp (Fig. 2, B-I and B-II). One product, illustrated in Fig. 2B-I, had 100% sequence identity with the indicated exons. The exon-specific primers are indicated by arrows. C. RT-PCR identified three ABCG1 DNA products that had 100% sequence identity with the indicated exons. The gene-specific primers are indicated by arrows.

Characterization of the Human ABCG1 Gene

A. Genomic Organization of human ABCG1

B. 5'-RACE Products

B-I (509 bp)

B-II (482 bp)

B-III (874 bp)

B-IV (724 bp)

C. PCR Products

C-I (451 bp)

C-II (874 bp)

C-III (285 bp)

Fig. 1. The ABCG1 gene has 23 exons and multiple transcripts. A-I, the 23 exons of the human ABCG1 gene, as reported in this study, are illustrated. The locations of the domains encoding the Walker A (WA), Walker B (WB), and signature motif (SM) are shown. Exons 8–10 (11), reported in this study, and exons 5 and 7 (12), are identified. A-II and A-III show the organization of the human ABCG1 gene as reported in earlier studies by Langmann et al. (18) and Lorkowski et al. (19), respectively. B, the human ABCG1 transcripts either identified in this study (B-I and B-II) or reported by Chen et al. (15) and Coop et al. (16) (B-II) or by Lorkowski et al. (19) (B-III), are shown.

Fig. 2. 5'-RACE and RT-PCR identify a novel ABCG1 transcript. A, schematic representation of the 23 exons of the human ABCG1 gene. B, 5'-RACE generated four DNA products that had 100% sequence identity with the indicated exons. The exon-specific primers are indicated by arrows. C, RT-PCR identified three ABCG1 DNA products that had 100% sequence identity with the indicated exons. The gene-specific primers are indicated by arrows.
Characterization of the Human ABCG1 Gene

Fig. 3. The nucleotide and predicted amino acid sequence of the ABCG1 transcript derived from exons 8–23. New sequence, corresponding to exons 8–10 (nt 1–874), is shaded. The Walker A (Ⅲ), signature motif (Ⅲ), and Walker B (Ⅲ) motifs are boxed. The alternate ATG (nt +149) is underlined. The polyadenylation signal is double underlined.

Fig. 4. Exon-specific probes identify three oxysterol-regulated ABCG1 transcripts. THP-1 cells were incubated for 24 h in medium supplemented with 10% LPDS in the absence (−) or presence (+) of 5 μg/ml 20(S)-hydroxycholesterol (20(S)-OHC). Total RNA was isolated and Northern blot assays performed, as described under “Experimental Procedures.” Using the indicated radiolabeled ABCG1 probes, the membranes were exposed to x-ray film for 16 h before they were re-probed with radiolabeled DNA corresponding to GAPDH. The three ABCG1 and GAPDH transcripts are indicated. Minor, unregulated bands observed in some panels correspond to nonspecific hybridization of the probe to 18 and/or 28 S.

| ABCG1 Probe | Exons 11–23 | Exon 7 | Exon 8 | Exon 9 | Exon 10 | Exons 11–14 |
|-------------|-------------|--------|--------|--------|---------|-------------|
| Panel A     | B           | C      | D      | E      | F       |             |
| ABCG1       |             |        |        |        |         |             |
| ACBG1       |             |        |        |        |         |             |
| GAPDH       |             |        |        |        |         |             |

RT-PCR.—In order to determine whether the exons predicted by Genscan and Grail algorithms were present in ABCG1 transcripts, we performed RT-PCR using RNA from oxysterol-treated THP-1 cells and human ABCG1 gene-specific primers, as described under “Experimental Procedures.” A single PCR product (874 bp) was produced when primers corresponding to the putative 5′ end of exon 8 and 3′ end of exon 10 were used (Fig. 2C-I). Sequencing of the product revealed 100% identity with the putative exons predicted by the Genscan and Grail algorithms (exons 8–10) (Fig. 2A). Based on the results of the 5′-RACE and RT-PCR, we conclude that one ABCG1 transcript in THP-1 cells contains sequences corresponding to exons 8–14.

Another single DNA product (451 bp) was produced when PCR was performed with primers corresponding to exons 5 and 12 (Fig. 2C-I). These 451 bp had 100% sequence identity with nucleotides contained in exons 5, 7, 11, and 12 (Fig. 2C-I) and to the published cDNA sequence (15, 16). No DNA product was obtained with this primer pair that contained sequences corresponding to exons 6, 8, 9, or 10. These data suggest that exons 6 and 8–10 are not part of any major macrophage ABCG1 transcript that contains exons 5, 7, 11, and 12.

The Genscan algorithm predicted that an additional exon might be located between exons 13 and 14. However, PCR analysis, using gene-specific primers to exons 12 and 14 gave a single DNA product of 285 bp that had 100% identity to exons 12–14 and lacked any additional nucleotides (Fig. 2C-III). Thus, we find no evidence for ABCG1 transcripts containing additional sequences between exons 13 and 14. Based on additional PCR studies and Northern blot assays, we also found no evidence that ABCG1 transcripts contain sequences corresponding to exons that were predicted to lie downstream of exon 23 (data not shown).

5′-RACE from the Novel Exons 9 and 10.—The studies described above suggest that macrophages express one or more ABCG1 transcripts that contain sequences corresponding to the novel exons 8–10. In order to determine the 5′ end of the transcript, 5′-RACE was performed with gene-specific primers corresponding to exon 10 or exon 9 (Fig. 2, B-III and B-IV). Each exon-specific primer produced one distinct DNA product (874 or 724 bp) that showed 100% identity with the sequences of the predicted exons 8–10 or 8 and 9, respectively (Fig. 2, B-III and B-IV). Based on these data, the three new exons were designated as exons 9, 8, and 10 (Fig. 1A and 2A). Since both 5′-RACE products terminated at the same nucleotide, at the start of exon 8, we hypothesize that this corresponds to a transcriptional start site. In addition, we used the TSSW and TSSG (Transcriptional Element Search System) algorithms to analyze the genomic sequence that encompasses exons 1–23 of the ABCG1 gene. These algorithms predicted the presence of a putative TATA box-containing promoter directly upstream of exon 8 (see Fig. 5A).

The Novel ABCG1 Transcript Encodes a Protein with a Unique Amino Terminus.—Fig. 3 shows the nucleotide sequence of the novel ABCG1 transcript encoded by exons 8–23 and the corresponding amino acid sequence. Within the nucleotides corresponding to exon 8, there are two in-frame methionines located at +141 and +266 and four in-frame stop codons upstream of nucleotide +141 (Fig. 3). The second in-frame ATG...
Characterization of the Human ABCG1 Gene

Table II
Position of the 23 exons encoding the human ABCG1 gene

| Exon | Exon size (bp) | Position * | Intron size (bp) | Splice acceptor site (intron-exon) | Donor acceptor site (exon-intron) |
|------|----------------|------------|------------------|----------------------------------|----------------------------------|
| 1    | 161            | 33,547–33,707 | 1,727           | ag-ACACTGAATC                 | AGTTTTGGTG-gtaagtact              |
| 2    | 147            | 35,435–35,581 | 5,222           | tccacacag-CCCTACACCA          | CTGCTCCCTG-gtagtgcag              |
| 3    | 41             | 40,804–40,844 | 4,990           | ttgttcag-TGAGAAATAT          | TCCCTCTATAG-gtaaagtatt           |
| 4    | 112            | 49,955–50,046 | 3,020           | agatcctag-GCTTTTAATAA         | ACATGACACT-gtagtcgct              |
| 5    | 107            | 53,056–53,163 | 593             | gcagctcaag-CCTCGTCCCC         | CACGCCGAT-Gtagtcgctg              |
| 6    | 152            | 53,753–53,904 | 5,624           | gcttctcag-ATGGACACGA          | AGGAAGAAGAG-gtaagggagg            |
| 7    | 244            | 59,529–59,772 | 30,434          | ttctcttca-ATGGACACGA          | AGGAAGAAGAG-gtaagggagg            |
| 8    | 534            | 89,941–90,474 | 2,696           | TTTTCTCCAA                          |
| 9    | 191            | 92,173–92,363 | 1,190           | tcttcttcatg-CCACAGCCGT        | CGCTGCCACTC-gtagtcgac             |
| 10   | 118            | 94,102–95,045 | 1,045           | ttttttgtac-GTATACAGAG          | CTTGATAGAC-gtagcgacc              |
| 11   | 10             | 107,161–107,293 | 3,459          | ggtttccag-GGAGACGGGC          | GGCCTAGATG-gtagcagc              |
| 12   | 51             | 110,753–110,803 | 5,328        | tctcttcag-GTGGCGAGCA          | AGGAAGGAAT-gtagtgggt              |
| 13   | 146            | 116,132–116,277 | 2,140          | tttgcctctag-GTCAAGAGA          | AGCCACACAG-gtagtcagg              |
| 14   | 124            | 118,418–118,541 | 1,196          | tttttttttag-CTTTACGTTTC      | GCCATTTTG-gtagcgag                |
| 15   | 115            | 113,799–113,852 | 1,894          | tttttttttag-CTTTACGTTTC      | GCCATTTTG-gtagcgag                |
| 16   | 143            | 121,747–121,931 | 137           | tctcttcag-TGATGGAGGT          | GTTGAGAAAGAG-gtaatgcac           |
| 17   | 131            | 121,895–122,089 | 173           | tctcttcag-TGATGGAGGT          | GTTGAGAAAGAG-gtaatgcac           |
| 18   | 102            | 122,069–122,170 | 1,170          | tctcttcag-GGCTGCTGTG          | GAGGGAGCTG-gtagtgctg              |
| 19   | 169            | 125,988–126,078 | 873           | tctttccag-GTCCTTCGCA          | GTTGATGACAT-gtagtgact             |
| 20   | 101            | 124,950–125,050 | 2,050          | tctcttcag-GTCTTCGCA          | GTTGATGACAT-gtagtgact             |
| 21   | 159            | 125,356–125,514 | 2,885          | tctcttcag-GTCTTCGCA          | GTTGATGACAT-gtagtgact             |
| 22   | 119            | 128,400–128,518 | 1,503          | tctttccag-GTGGCCACTT          | CCTATGTCAG-gtagcgccg             |
| 23   | 1085          | 130,022–131,106 | 1,084          | tctttccag-GTATGGGGCA          | GTTGATGACAT-gtagcgccg             |

* Position number corresponds to the \( S.\) \( sapa\) s genomic DNA sequence of chromosome 21q22.3 section 90/105 (accession no. \( A\) \( P\)001746).

b Alternative splicing produces a variant of exon 17 (14).

(Fig. 3, nt +266) lies within a nucleotide sequence that conforms to the Kozak rules (43) and thus likely encodes the initiating methionine. Assuming that this second ATG functions as the initiating methionine, the transcript encodes a protein of 786 amino acids that contains a novel 203-amino-terminal amino acids, encoded by exons 8–10, followed by 583 amino acids that are encoded by exons 11–23. Analysis of these 203 amino acids, using PROSITE and SignalP algorithms, failed to identify any conserved domains/motifs or membrane signaling sequences. Translational initiation at the first methionine (Fig. 3, nt +149) would increase the number of novel amino acids from 203 to 242 and size of the protein derived from exons 8–23 to 825 amino acids. Analysis of these additional 39 amino acids with the PROSITE and SignalP algorithms failed to identify any conserved domains/motifs or membrane signaling sequences. The carboxyl-terminal 583 amino acids are known to be part of an ABCG1 isoform of 674 amino acids that is derived from exons 5, 7, and 11–23 (see Fig. 1B-II) (18).

Oxysterol-regulated ABCG1 Transcripts Contain the Novel Exons 8–10—In order to confirm that sequences corresponding to exons 8–10 were part of an oxysterol- and LXR-regulated ABCG1 transcript, we probed Northern blots with exon-specific radiolabeled probes. As shown in Fig. 4, radiolabeled probes corresponding to exons 11–23 (panel A) or exons 11–14 (panel F) identified three distinct transcripts that are induced when THP-1 cells are treated for 24 h in the presence of 20(S)-hydroxycholesterol. The three transcripts are thought to result from use of alternative polyadenylation sites (data not shown). An identical pattern of mRNA regulation was observed when the radiolabeled probes corresponded to either exon 7, 8, 9, or 10, respectively (Fig. 4, panels B-E). In contrast, no transcript was identified when the membrane was probed with radiolabeled DNA corresponding to exons 1, 2, 4, or 6, despite the fact that the specific activities of all probes were similar (data not shown).

The data in Figs. 1, 2, and 4 suggest that activation of the human ABCG1 gene by LXR ligands results in a novel macrophage transcript containing exons 8–23. A second regulated transcript contains sequences derived from exons 5, 7, and 11–23. Both transcripts contain the Walker A motif (exon 11), signature and Walker B motifs (exon 14), and six transmembrane domains (exons 18, 20–23). These motifs/domains are present in all members of the ABC half transporter subfamily (12, 14).

Genomic Organization and Exon/Intron Boundaries of the Human ABCG1 Gene—The ABCG1 gene is encoded by 23 exons on 98 kb of genomic DNA that is located on chromosome 21 between nucleotides 29,167,125 and 29,264,680 (44). The genomic organization of ABCG1 is shown in Fig. 4, panels A–F. The introns range in size from 41 bp (exon 3) to the largest is 1081 bp (exon 23). The introns range in size from 137 bp (intron 17) to 30,434 (intron 7). All exon/intron boundaries displayed the canonical GT/AG sequence (45). Interestingly, exon 1, identified by Lorkowski et al. (see Fig. 1A-III) (19), is also preceded by an AG (see Table II).

EMSAs Identify Two Putative LXREs Upstream of Exon 8—We used the FINDPATTERNS algorithm in the GCG suite of analyses software to identify potential LXRE cis-elements within the gene, as described under “Experimental Procedures.” Two putative LXREs, termed LXRE-1 and LXRE-2, were identified within intron 7 of the ABCG1 gene, −5 and 9 kb 5’ of exon 8, respectively (Fig. 5A). We used EMSAs and in vitro transcribed/translated LXRs and RXRα to determine if the LXRE/RXRα heterodimer bound to these putative LXREs. The sequence of the wild type (LXRE-1, LXRE-2) and mutant LXRE-1 (LXRE-1m, LXRE-2m) is shown in Fig. 5B. The data of Fig. 5C demonstrate that incubation of radiolabeled oligonucleotides containing either LXRE-1 (Fig. 5C, lane 3) or LXRE-2 (Fig. 5C, lane 1) with LXRs and RXRα results in the formation of a shifted complex that is supershifted in the presence of antibody to RXRα (Fig. 5C, lanes 4 and 12). In contrast, no complex was observed when LXRE-1 or LXRE-2 were incubated with either LXRα or RXRα alone (Fig. 5C, lanes 1, 2, 9, 39443
Competition experiments were also performed to determine the specificity of the interaction. Fig. 5D shows that the formation of the LXRα/RXRα-radiolabeled LXRE-1 complex was re-
duced in the presence of increasing concentrations of unlabeled oligonucleotide containing wild-type LXRE-1 (lanes 2–5) but not mutant LXRE-1m (lanes 6–9). An unlabeled oligonucleotide containing the LXRE derived from the CYP7A1 promoter served as a positive control and also inhibited the formation of the radiolabeled shifted complex (lanes 10–13). After 24 h the cell lysate was assayed for reporter gene activity and normalized for minor variations in transfection efficiency as described under "Experimental Procedures." The relative luciferase units are given. The results shown are representative of three separate experiments, each performed in triplicate. Note that the left and right panels represent a single experiment, but the scale differs because of the greater activity of LXRE-2.

Identification of a Functional Promoter in Intron 7 of the Human ABCG1 Gene—The data from Fig. 5 suggest that the LXREs identified in intron 7 may be important for the oxysterol-dependent regulation of ABCG1 in response to LXRa. To determine if this hypothesis is correct, transient transfection assays were performed with luciferase reporter constructs, each containing two copies of either the wild type or mutant LXRE-1 and LXRE-2 linked to a minimal heterologous thymidine kinase (TK) promoter (Fig. 6). Each reporter construct was transfected into HepG2 cells in the presence or absence of LXRa (50 ng) and RXRa (5 ng) and either Me2SO (DMSO), LG100153 (100 nM), and/or 22(R)-hydroxycholesterol (22(R)-OHC) (5 µg/ml), as indicated. After 24 h the cell lysate was assayed for reporter gene activity and normalized for minor variations in transfection efficiency as described under "Experimental Procedures." The relative luciferase units are given. The results shown are representative of three separate experiments, each performed in triplicate. Note that the left and right panels represent a single experiment, but the scale differs because of the greater activity of LXRE-2.

FIG. 7. The ABCG1 promoter contained within intron 7 is activated by LXR and ligands for LXR and RXR. Triplicate dishes of HepG2 cells were transiently transfected with 100 ng of the indicated ABCG1 promoter-reporter gene construct in the presence or absence of LXRa (50 ng) and RXRa (5 ng) and either Me2SO (DMSO), LG100153 (100 nM), and/or 22(R)-hydroxycholesterol (22(R)-OHC) (5 µg/ml), as indicated. After 24 h the cell lysate was assayed for reporter gene activity and normalized for minor variations in transfection efficiency, as described under "Experimental Procedures." Normalized activities varied less than 10%, and relative luciferase units are given.
is present within intron 7 and immediately upstream of exon 8 of the human ABCG1 gene, we generated a series of reporter genes, as shown in Fig. 7. The parent reporter (pG-9.3) contained genomic sequence corresponding to +167 to −9318, where +1 corresponds to the initial nucleotide of exon 8. Other reporter genes contained the same genomic DNA except that mutations were incorporated into either LXRE-1, LXRE-2, or both LXRE-1 and LXRE-2 to generate plasmids pG-9.3-m1, pG-9.3-m2, or pG-9.3-m1,2, respectively (Fig. 7). The mutations within the LXREs were identical to those shown in Fig. 5A. Each reporter gene was transiently transfected into HepG2 cells in the presence or absence of LXR and RXR and the indicated ligands for LXR or RXR (see Fig. 7). As shown in Fig. 7, the results of these studies demonstrate that the reporter containing the wild-type promoter (pG-9.3) was most highly induced in the presence of LXR, RXR, and ligands for both nuclear receptors. However, luciferase activity was also significantly induced in response to ligands for either LXR (22(R)-hydroxycholesterol) or RXR (LG100153) (Fig. 7). Under the same conditions, induction of the reporter gene was partially attenuated when the promoter contained mutations in either LXRE-1 (pG-9.3-m1) or LXRE-2 (pG-9.3-m2) (Fig. 7). In contrast, the activity of pG-9.3-m1,2 (containing mutations in both LXRE-1 and LXRE-2) was unaffected by the co-transfected nuclear receptors or ligands, and thus it functioned in a manner that was similar to that of the promoterless pGL-3 (Fig. 7). These data demonstrate that intron 7 of the human ABCG1 gene contains a promoter with two functional LXREs.

**DISCUSSION**

Previous studies have shown that both the human and murine ABCG1 genes are activated in response to oxysterols, by a mechanism that involves LXR and RXR (5). The current studies identify two functional LXREs that are likely to mediate the transcriptional activation of the human ABCG1 gene in response to oxysterols or other LXR ligands. Surprisingly, the promoter containing these LXREs was identified in intron 7 and shown to lie upstream of three new exons (exons 8–10). We utilized 5′-RACE, PCR, and Northern blot assays using exon-specific probes to demonstrate the activation of a novel transcript derived from exons 8–23 (Figs. 2 and 4). Translation of this transcript produces a new form of ABCG1 that contains 786 amino acids with a unique 203 amino acids at the amino terminus (Fig. 3). The current studies also demonstrate that a second transcript, derived from exons 5, 7, and 11–23, is also highly induced in response to oxysterol treatment (Fig. 4). We hypothesize that a second promoter, yet to be identified and presumably containing one or more additional LXREs, will be involved in the regulation of this latter transcript. Additional studies are currently under way to determine whether these distinct ABCG1 transcripts are expressed in a tissue-specific manner.

The two oxysterol-regulated ABCG1 transcripts encode proteins with common amino acids derived from exons 11–23 but distinct amino termini derived from either exons 8–10 or 5 and 7 (Fig. 1, B-I and B-II). The finding that these ABCG1 isoforms share sequences derived from exons 11–23 is not surprising since all functional half-transporters contain Walker A and Walker B domains, a signature motif, and six transmembrane domains (12, 14). These important domains are encoded on exons 11, 14, 18, and 20–23 of the ABCG1 gene.

The function of ABCG1 protein(s) is not known definitively. Incubation of cells with antisense oligonucleotides resulted in decreased ABCG1 mRNA levels and a decrease in the rate of efflux of cellular cholesterol to HDL_3 (7). These data suggest that ABCG1 may have a role in controlling the efflux of lipids to extracellular lipid acceptors. Two other gene products, ABCA1 and apoE, that are known to be involved in lipid efflux/metabolism are also highly induced when macrophages are treated with oxysterols by a process that requires LXR (5, 6, 9). Mutations in the ABCA1 gene are associated with Tangier disease or familial HDL deficiency, disorders in which patients have little or no HDL (46–52). Studies with cells derived from Tangier patients or from ABCA1 null mice demonstrate that ABCA1 is normally involved in the export of cellular cholesterol and/or phospholipids to apoA-1 (53). Interestingly, a number of patients with Tangier disease have mutations in the amino terminus of the ABCA1 protein (25). Based on these studies of ABCA1, we hypothesize that the amino termini of the different isoforms of ABCG1 may also contain domains that are important for function. Studies to test this hypothesis are currently under way. apoE, the third gene that is induced in oxysterol-treated macrophages (6), has three major polymorphisms that have been linked to both hypercholesterolemia and Alzheimer's disease (54).

Recently, Lorkowski et al. (19) reported that human THP-1 cells express novel ABCG1 transcripts that contain variable combinations of five new exons (1–4 and 6) together with exons 5, 7, and 11–23 (19) (see Fig. 1B-III). In addition, these authors used RT-PCR and primers to exons 1 and 2 or 5 and 7 and reported that exons 1–2 or exons 5 and 7 (but not exon 6) were part of two distinct transcripts that were induced in response to oxysterola (19). In contrast, we used specific radiolabeled probes corresponding to exons 1, 2, 4, or 6, in Northern blot assays and were unable to identify any transcript of any size (data not shown). However, the regulated ABCG1 transcripts were easily identified when the same membranes were hybridized with radiolabeled probes corresponding to either exon 5, 7, 8, 9, 10, 11–14, or 11–23 (Fig. 4 and data not shown). Studies are currently under way to understand these discrepancies.

Many ABC half transporters have been shown to function as heterodimers (55–57), and at least one has been localized to the plasma membrane (58). At this time, the functional partner for human or murine ABCG1 has not been identified. The studies described above indicate that at least two major transcripts, that encode ABCG1 proteins with alternative amino termini, are induced from one gene when macrophages are incubated in the presence of oxysterols. It is possible that these ABCG1 isoforms form functional dimers and that the transport function is modulated by the amino terminus.

**Acknowledgments**—We thank Drs. R. Heyman for LG100153, and R. Evans for pCMX-LXRs and pCMX-RXRs. We thank Dr. Peter Tontonoz and the Edwards laboratory for critically reading the manuscript, and to Drs. Kazunori Shibuya and David Eisenberg and the members of the Edwards and Tontonoz laboratory for helpful discussions.

**REFERENCES**

1. Brown, M. S., Goldstein, J. L., Krieger, M., Ho, Y. K., and Anderson, R. G. (1979) J. Cell Biol. 82, 597–613

2. Gerrity, R. G. (1981) Am. J. Pathol. 103, 181–190

3. Ross, R. (1985) Annu. Rev. Physiol. 47, 791–804

4. Venkatesswaran, A., Repa, J. J., Lobaccaro, J. M., Bronson, A., Mangelbdf, D. J., and Edwards, P. A. (2000) J. Biol. Chem. 275, 14700–14707

5. Venkatesswaran, A., Laffitte, B. A., Joseph, S. B., Mak, P. A., Wiputz, D. C., Edwards, P. A., and Tontonoz, P. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 12097–12102

6. Laffitte, B. A., Repa, J. J., Joseph, S. B., Wiputz, D. C., Kast, H. R., Mangelbdf, D. J., and Tontonoz, P. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 507–512

7. Klucnen, J., Buchler, C., Orso, E., Kaminski, W. E., Porsch-Ozcurumez, M., Liebich, G., Kapinsky, M., Diderich, W., Drobnik, W., Dean, M., Allikmets, R., and Schmitz, G. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 817–822

8. Repa, J. J., Turley, S. D., Lobaccaro, J. A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R. A., Dietschy, J. M., and Mangelsdorf, D. J. (2000) Science 289, 1524–1529

9. Costet, P., Lun, Y., Wang, N., and Tall, A. R. (2000) J. Biol. Chem. 275, 29240–29245

10. Dean, M., and Allikmets, R. (1995) Curr. Opin. Genet. Dev. 5, 779–785

11. Ambudkar, S. V., and Gottesman, M. M. (eds) (1998) Methods Enzymol. 292, 504–514

12. Higgins, C. F. (1992) Annu. Rev. Cell Biol. 8, 67–113
