Biochemical and Biophysical Research Communications **280,** 121–131 (2001)
doi:10.1006/bbrc.2000.4089, available online at http://www.idealibrary.com on IDEAL®

Genomic Sequence and Structure of the Human ABCG1 (ABC8) Gene

Stefan Lorkowski,*,†,[1] Stephan Rust,*,[1] Thomas Engel,* Elisabeth Jung,* Karin Tegelkamp,* Erwin Arno Galinski,† Gerd Assmann,*,‡ and Paul Cullen*,‡,[2]

**Institut für Arteriosklerosеforschung, †Institut für Biochemie, and ‡Institut für Klinische Chemie und Laboratoriumsmedizin, Westfälische Wilhelms-Universität Münster, Münster, Germany*

Received November 27, 2000

The human ATP-binding cassette half transporter G1 (hABCG1) may play a role in cholesterol transport in macrophages. Using RACE assays we determined the structure of this gene. The hABCG1 gene spans more than 97 kb comprising 20 exons, 20 kb and 5 exons more than hitherto described. Four of the novel exons are upstream and one is downstream of previous exon 1, and they are predicted to encode at least five novel transcripts. We also detected two separate promoters, upstream of exons 1 and 5, respectively. The region 650 bp upstream of exon 1 was predicted to contain putative binding sites for SP1 and nuclear factor κB (NF-κB), but no sterol response elements (SREs) or retinoid X receptor (RXR) binding sites. The region 650 bp upstream of exon 5 contained 19 possible SP1 binding sites, one possible SRE, two possible NF-κB, and two putative RXR binding sites. Nevertheless, both promoters responded in macrophages to stimulation by hydroxycholesterol and retinoic acid.

Key Words: **5′-RACE; ATP-binding cassette transporter G1; ABC8; ABCG1; *Drosophila* white gene; promoter region; genomic structure; macrophages.**

Abbreviations used: 22(*R*)-HC, 22(*R*)-hydroxycholesterol; 9-cRA, 9-*cis* retinoic acid; aa, amino acid(s); ABC, ATP-binding cassette; ABCA1, ATP-binding cassette transporter A1; AP, activator protein; ATP, adenosinetriphosphate; bp, base pair(s); BRE, B recognition element; cAMP, cyclic-adenosinmonophosphate; C/EBP, CCAAT-enhancer-binding proteins; CREB, cAMP responsive element binding; CREBP, cAMP-responsive element binding protein; GCF, GC-binding factor; GM, granulocyte/macrophage; GR, glucocorticoid receptor; hABCG1, human ATP-binding cassette half transporter G1; HMG-CoA, hydroxymethylglutaryl coenzyme A; HNF, hepatic nuclear factor; IRF, interferon regulatory factor; ISGF, interferon-stimulated gene factor; ISRE, interferon-stimulated response element; LXR, liver-specific X receptor; mABCG1, murine ATP-binding cassette half transporter G1 (*Mus musculus*); MAZ, myc-associated zinc finger protein; NF-GMa, nuclear factor GMa; NF-κB, nuclear factor kappa B; NFY, nuclear factor Y (Y-box binding factor); N-Oct, nervous-system specific POU domain transcription factor binding to the octamer DNA motif; Oct, octamer transcription factor; RACE, rapid amplification of cDNA ends; RAR, retinoic acid receptor; RLM-RACE, RNA ligase-mediated rapid amplification of 5′ cDNA ends; ROR RAR-related orphan receptor; RXR retinoid X receptor; SP, stimulating protein; SRE, sterol response element; SREBP, sterol response element binding protein; T3R, thyroid hormone receptor; TCF, T-cell transcription factor; TFIID, transcription factor IID; TESS, Transcription Element Search Software; VDR, vitamin D receptor.

[1] Stefan Lorkowski and Stephan Rust contributed equally to this paper.

[2] To whom correspondence and reprint requests should be addressed at Ogham Diagnostics GmbH, Mendelstrasse 11, 48149 Münster, Germany. Fax: +49-251-9801389. E-mail: cullen@uni-muenster.de.

ATP-binding cassette (ABC) transporters form a highly conserved superfamily of proteins, at least 50 members of which have been identified in humans and other eukaryotes (1–4). Functional ABC full transporters comprise two symmetrical halves, each of which consists of the well-defined ATP-binding fold first characterized by Walker in 1982 (5), and a hydrophobic part, containing six membrane spanning domains, that provides substrate specificity. Half transporters contain six transmembrane domains and a single ATP-binding cassette (1, 6) and are thought to form homo- or heterodimers with a second half transporter to build up a functional transporter. The ATP-binding fold is composed of the conserved Walker A and B domains and a "signature" motif upstream of the Walker B domain (1, 7). These three domains are required for the binding and hydrolysis of ATP, which in turn provides energy for the transport of a large variety of substrates across cellular and organelle membranes (1, 3, 8).

The half transporter ABCG1, a member of the ABC transporter subfamily G, encoded by the ABCG1 gene, was first described by Chen *et al.* (9) and independently by Croop *et al.* and Savary *et al.* as a homologue of the *Drosophila* white protein (3, 10). Kuwana *et al.* described an ABC transporter expressed in trachea of *Drosophila melanogaster*, which shows higher homology to the human ABCG1 gene product than the *Drosophila* white protein (11). Chen *et al.* isolated the coding region of the hABCG1 gene and reported evidence of alternative splicing in the coding region (9).

0006-291X/01 $35.00

TABLE 1

Primers Used for ABCG1 5′-RACE PCRs, ABCG1 3′-RACE PCRs, and Real-Time RT-PCR

Primer	Sequence	Hybridization region	Orientation
hABCG1 5′-RACE 1 reverse	5′-gag gag aag cgc tgg cct tcc gtg agg t-3′	Exon 7 (former exon 2)	Antisense
hABCG1 5′-RACE 2 reverse	5′-cag atg tcc att cag cag gtc cgt ctc ag-3′	Exon 7 (former exon 2)	Antisense
hABCG1 5′-RACE 3 reverse	5′-gtt gctg gac acc acc tca tcc acc gag-3′	Exon 7 (former exon 2)	Antisense
hABCG1 3′-RACE 1 forward	5′-tgc tgt tcc tca tgt tcg-3′	Exon 16 (former exon 11)	Sense
hABCG1 3′-RACE 2 forward	5′-tgg atg tcc tac atc tcc-3′	Exon 19 (former exon 14)	Sense
hABCG1 promoter A forward	5′-atg ttt gtg gcg cta cac caa-3′	Exons 1 and 2	Sense
hABCG1 promoter A reverse	5′-tgc gtt ccc agc atc aag a-3′	Exon 2	Antisense
hABCG1 promoter B forward	5′-tct gat ggc cgc ttt ctc g-3′	Exon 5 (former exon 1)	Sense
hABCG1 promoter B reverse	5′-cca tgt tgc tgg aca cca cc-3′	Exon 7 (former exon 2)	Antisense
mABCG1 5′-RACE 1 reverse	5′-cct tga att cga tgt tca cgg ccg ccc tcc-3′	—	Antisense
h18S rRNA forward	5′-cgg cta cca cat cca gga a-3′	—	Sense
h18S rRNA reverse	5′-gct gga tta ccg cgg ct-3′	—	Antisense

hABCG1 cDNAs with different 5′-ends encoding alternative methionine start codons, and which are thus predicted to encode different proteins have also been described (3, 10).

The human ABCG1 gene has been mapped to chromosome 21q22.3 (3, 9, 10). A DNA polymorphism with 62% heterozygosity due to variation of a poly(T) region in the 3′-untranslated region has also been reported (9). Furthermore, Nakamura *et al.* described an A → G transversion in the 3′-untranslated region of the hABCG1 gene and found significant associations between the (A/G)-polymorphism and the poly(T)-polymorphism first described by Chen *et al.* with mood and panic disorders in males (9, 12).

In recent reports, hABCG1, as well as many other human ABC genes, was shown to be regulated by cholesterol uptake or HDL_3-mediated lipid release in human macrophages (13). Venkateswaran *et al.* showed that hABCG1 mRNA levels are regulated by the liver-specific X receptor/retinoid X receptor (LXR/RXR) pathway (14), which recently has been shown to play an important role in lipid metabolism (15, 16). These findings suggest a potential function for ABCG1 in macrophage cholesterol homeostasis.

It is now becoming clear that not only the ATP-binding cassette transporter A1 (ABCA1, ABC1) gene which was recently linked to Tangier disease (17–20), but also the ABCG1 gene is involved in cholesterol efflux and atherosclerotic plaque formation. Therefore, ABCG1 may be a potential target for antiatherosclerotic therapies. Knowledge of regulatory elements of the ABCG1 gene will be of relevance in this regard.

We have determined the complete structure of the hABCG1 gene and carried out examination of the 5′-end to determine putative regulatory elements. Our studies indicate an unexpected degree of complexity in the expression and regulation of this gene product.

MATERIALS AND METHODS

Cell culture. Monocytes were isolated from the peripheral blood of healthy volunteers by cell separation and countercurrent elutriation as previously described (21). All procedures were approved by the Hospital Ethics Committee. The purity of monocyte fractions was tested by flow cytometry (FACScan, Becton–Dickinson, Heidelberg, FRG). The cells were cultured in RPMI 1640 (Gibco BRL Life Technologies, Eggenstein, FRG) plus 20% heterologous human serum (PAA Laboratories, Linz, Austria) for 14 days to allow differentiation into macrophages. Differentiated macrophages were stimulated for 24 h using 10 μM 22(*R*)-hydroxycholesterol and 10 μM 9-*cis* retinoic acid to induce transcription of ABCG1 (14).

A THP-1 monocytic cell line was purchased from the American Tissue Culture Collection (Rockville, MD). THP-1 cells were maintained in RPMI 1640 containing L-glutamine, 10% FCS, 100 U/ml penicillin, and 100 μg/ml streptomycin supplemented with 50 nM β-mercaptoethanol. Subconfluent THP-1 cells were differentiated into macrophages using 0.2 mM phorbol 12-myristate 13-acetate during 48 h. Differentiated THP-1 cells were stimulated for 24 h using 10 μM each of 22(*R*)-hydroxycholesterol and 9-*cis* retinoic acid, respectively, to induce transcription of ABCG1 (14).

RAW 264.7 cells were maintained in DMEM, 10% FBS, and 1% antibiotic, antimycotic solution (Sigma). Cells were passaged by scraping. For extraction of RNA a confluent layer of RAW macrophages was incubated in DMEM, 0.1% BSA, and 1% antibiotic, antimycotic solution with or without addition of 10 μM 9-*cis* retinoic acid and 22(*R*)-hydroxycholesterol, each for 16 h.

5′-RACE assays. The 5′-untranslated region of human ABCG1 mRNA was determined using rapid amplification of cDNA ends (RACE) using the SMART RACE cDNA Amplification Kit (Clontech Laboratories, Heidelberg, FRG). For this purpose total cellular RNA was isolated from macrophages using the RNeasy Mini kit (Qiagen) and oligonucleotides *hABCG1 5′-RACE 1 reverse*, *hABCG1 5′-RACE 2 reverse*, and *hABCG1 5′-RACE 3 reverse* (Table 1) which were designed from published mRNA sequences (GenBank Accession Nos. NM_016818 and NM_004915) (3, 9, 10, 13). Poly$(T)_{12–18}$ oligonucleotide (Gibco BRL Life Technologies) or oligonucleotide hABCG1 5′-RACE 1 reverse were used to synthesize first strand cDNA from 1 μg total RNA isolated from human monocyte-derived macrophages or stimulated THP-1 cells using SuperScript II reverse transcriptase (Gibco BRL Life Technologies) and SMART II oligonucleotide (Clontech, Heidelberg, FRG). 5′-RACE PCRs were performed according to the manufacturer's protocol (Clontech) using Hotstar *Taq* polymerase (Qiagen). The oligonucleotide *hABCG1 5′-RACE 2 reverse* together with the Universal Primer Mix (UPM) (Clontech) was used for the PCR of the RACE assay, which was carried out using an Applied

Biosystems 9600 thermocycler (Applied Biosystems, Weiterstadt, FRG). Conditions for the RACE reaction were 94°C for 15 min followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 45 s, followed by a final extension phase of 72°C for 7 min. PCR products were gel purified and a diffuse band of about 400 bp was reamplified in a nested PCR using the oligonucleotide *hABCG1 5′-RACE 3 reverse* together with the nested universal primer (Clontech) and Hotstar *Taq* polymerase (Qiagen). The nested PCR conditions were 94°C for 15 min followed by 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 45 s, followed by a final extension of 72°C for 7 min.

The 5′-untranslated region of *Mus musculus* ABCG1 mRNA was determined using RNA ligase-mediated rapid amplification of 5′ cDNA ends (RLM-RACE) using the GeneRacer Kit (Invitrogen, Groningen, The Netherlands). Total cellular RNA was isolated from RAW 264.7 cells as described above and primers were designed from published mRNA sequences (GenBank Accession No. Z48745; Table 1) (10). After ligation of the GeneRacer RNA oligo (Invitrogen) to the RNA, the GeneRacer oligo dT primer (Invitrogen) was used to synthesize first strand cDNA from 1 μg total RNA isolated from stimulated RAW 264.7 cells using AMV reverse transcriptase (Invitrogen). 5′-RACE PCRs were performed according to the manufacturer's protocol (Invitrogen) using Hotstar *Taq* polymerase (Qiagen). The oligonucleotide *mABCG1 5′-RACE 1 reverse* (Table 1) together with the GeneRacer 5′-primer (Invitrogen) was used for the 5′-RACE reaction, which was carried out using an Applied Biosystems 9600 thermocycler (Applied Biosystems, Weiterstadt, FRG). Conditions for the 5′-RACE reaction were 94°C for 15 min followed by 5 cycles of 94°C for 30 s, and 72°C for 60 s, 5 cycles of 94°C for 30 s, 70°C for 30 s and 72°C for 60 s, and 30 cycles of 94°C for 30 s, 68°C for 30 s and 72°C for 60 s, followed by a final extension phase of 72°C for 10 min. PCR products were gel purified.

3′-RACE assays. Five micrograms of total RNA from human macrophages were reverse transcribed using SuperScript II (Gibco BRL Life Technologies) as recommended by the manufacturer using the adapter primer (Gibco BRL Life Technologies). RNase H treatment of the cDNA was performed as described by the manufacturer (Gibco BRL Life Technologies). 3′-RACE PCRs were performed according to the manufacturer's protocol using Hotstar *Taq* polymerase and the oligonucleotide *hABCG1 3′-RACE 1 forward* (Table 1) together with the universal amplification primer (Gibco BRL Life Technologies) of the 3′-RACE system for rapid amplification of cDNA ends (Gibco BRL Life Technologies). The PCR conditions were 94°C for 15 min followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 45 s followed by a final extension phase of 72°C for 7 min.

DNA sequencing. PCR products were gel purified and cloned into pCRII-TOPO vector (Invitrogen, Leek, NL) using the TOPO TA Cloning Kit (Invitrogen). DNA sequencing was performed directly on PCR products or cloned cDNA using the Big Dye Terminator Cycle Sequencing kit (Applied Biosystems) with primers *hABCG1 5′-RACE 2 reverse, hABCG1 5′-RACE 3 reverse, mABCG1 5′-RACE 2 reverse, hABCG1 3′-RACE 2 forward* (Table 1) or *M13/pUC reverse* and *M13/pUC forward* (Gibco BRL Life Technologies) on an ABI Prism 3700 DNA sequencer or an ABI Prism 377 DNA sequencer (Applied Biosystems, Weiterstadt, FRG).

Analysis of predicted protein sequences. Predicted amino terminal protein sequences were analyzed for the presence of signal peptides, the position of cleavage sites and cellular localization using SignalP (22) (URL: http://130.225.67.199/services/SignalP/index.html), the NCBI Conserved Domain Database with reverse position-specific BLAST Search Service (23–26) (URL: http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and PSORT II (27–29) (URL: http://psort.ims.u-tokyo.ac.jp/).

Analysis of 5′-flanking regions. The *Homo sapiens* genomic DNA derived from chromosome 21q section 90/105 (GenBank Accession No. AP001746) was used to identify the 5′-flanking regions of the hABCG1 gene. To locate regulatory elements which potentially control expression of the hABCG1 gene a comparison of scores was generated on the world-wide-web from the Transcription Factor database (TRANSFAC) using Transcription Element Search Software (TESS) (30) (URL: http://www.cbil.upenn.edu/tess/index.html), and MatInsepctor V2.2 (31) (URL: http://www.gsf.de/biodv/matinspector.html) to analyse genomic 5′-flanking sequences of up to 2000 bp in length derived from the first exons of the different hABCG1 transcripts. The 5′-flanking regions and parts of the first exons of the hABCG1 transcripts were analyzed for the presence of CpG islands as defined by Gardiner-Garden and Frommer (32) using Webgene software (URL: http://www.itba.mi.cnr.it/webgene/).

Real-time RT-PCR assays. Five micrograms of total RNA were transcribed to cDNA using SuperScript II reverse transcriptase (Gibco BRL Life Technologies) and an oligo(dT)$_{12–18}$ primer (Gibco BRL Life Technologies). Real-time PCR analysis was performed using the SYBR Green PCR Reagents Kit (PE Applied Biosystems) with an ABI Prism 7700 Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions. hABCG1 cDNAs were generated by reverse transcription from 5 μg total RNA. Each PCR amplification of cDNA was then performed with the cDNA derived from 50 ng total RNA. The primers used for PCR amplification of the cDNAs were *hABCG1 promoter A forward* and *hABCG1 promoter A reverse* for promoter A transcripts, and *hABCG1 promoter B forward* and *hABCG1 promoter B reverse* for promoter B transcripts; amplification of 18S ribosomal RNA cDNA, the internal standard, was performed using the oligonucleotide *h18S rRNA forward* and *h18S rRNA reverse* (Table 1). Each sample was measured in triplicate and a blank containing no template cDNA was used as a negative control.

RESULTS

Human ABCG1 mRNA Sequences Generated by 5′-RACE

5′-RACE using total RNA from THP-1 cells revealed several hABCG1 transcripts with different 5′-ends (Fig. 1) and various length (2844 bp to 3121 bp calculated length of the respective hABCG1 transcripts). When these transcripts were compared to the genomic DNA sequence on chromosome 21, we found that some mapped upstream of the exons which were previously described as exons 1 and 2. Based on the position of the transcripts, we predict the presence of five novel exons in the ABCG1 gene, indicated as exons 1, 2, 3, 4, and 6 in Fig. 1. The exon which was previously described as exon 1 should now be designated exon 5 and the exon which was previously described as exon 2 should now be termed exon 7. Accordingly, those exons that were previously given the numbers 3 to 15 in the ABCG1 gene should now be assigned the numbers 8 to 20.

Based on our analyses, at least five different splice variants with at least 10 different 5′ ends are encoded by the hABCG1 gene. Moreover, we found four different start regions of RNA transcription located in exon 1 (transcripts A, B, C, D, and E), exon 4 (transcript F), exon 5 (transcripts G, H, and I), and exon 6 (transcript J), respectively (Fig. 1). Figure 2 shows the predicted amino acid sequences of the proteins predicted from the mRNA sequences detected in our studies. Our data indicate the presence of a new putative TATA box-containing promoter nearly 20 kb upstream of the previously described putative TATA box-free promoter

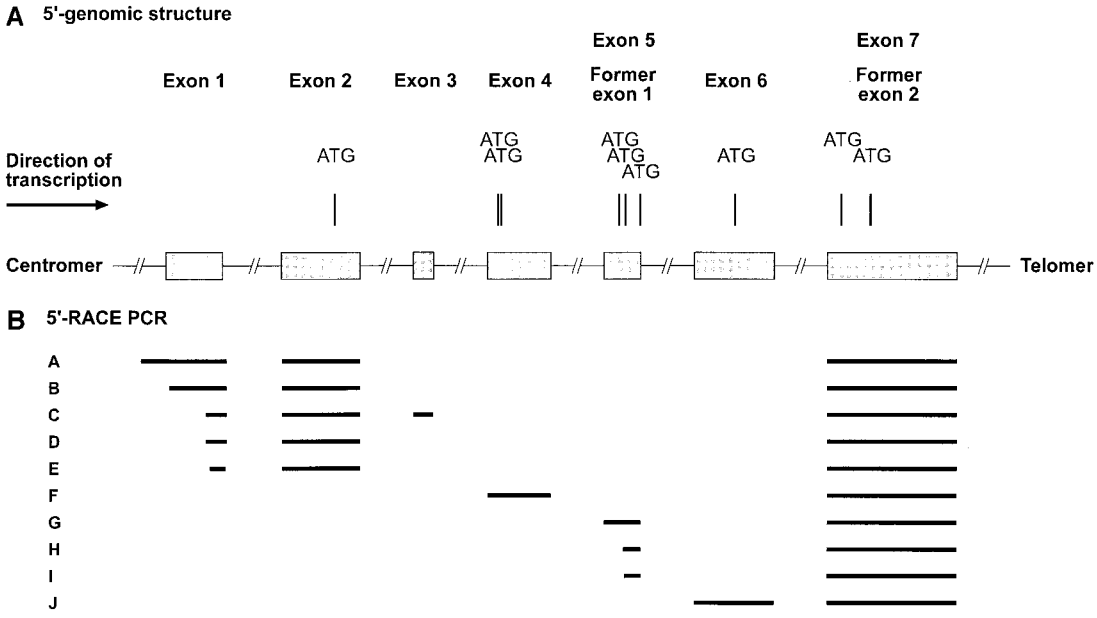

FIG. 1. Analysis of human ABCG1 5′-sequence: (A) partial gene structure and (B) structure of 5′-ends of cDNAs. (A) The structure of the human ABCG1 gene upstream of the former exon 2 was determined by sequence analysis. The genomic sequence of the 5′-end of human ABCG1 was identical to the respective sequence in genomic DNA database sequence of human chromosome 21q section 90/105 (Accession No. AP001746). The positions of the five new exons (exons 1, 2, 3, 4, and 6) and two previously known exons (exons 5 and 7) in the human sequence are indicated by gray boxes. Possible ATG start codons in frame with the translational frame of published ABCG1 protein sequences are present in exons 2, 4, 5, 6, and 7 (indicated by vertical lines in A). The relative positions of the exons are given in Table 2. (B) 5′-RACE PCRs were performed on RNA isolated from THP-1 macrophages stimulated with 10 μM 22(*R*)-hydroxycholesterol and 10 μM 9-*cis* retinoic acid. These cells express exons 1 to 7 as splice variants (transcripts A to J; sorted by 5′ end point). The 5′-ends of single RNA transcripts showed different starting points in their first exons. The relative genomic positions of the exons in the ABCG1 gene on chromosome 21 are shown in Fig. 3.

(Accession No. AJ289317) and immediately upstream of exon 1 (our numbering). Furthermore, we identified two additional putative promoters, as indicated by two hABCG1 transcripts with start sites in exon 4 and 6 (transcripts F and J) both of which lack TATA boxes (data not shown). Further, 5′-ends of transcripts F and J were identified only once, suggesting that these promoters are of minor physiological importance.

Murine ABCG1 mRNA Sequences Generated by 5′-RACE

5′-RACE using total RNA from the RAW 264.7 murine macrophage cell line revealed two murine ABCG1 (mABCG1) transcripts (calculated length of the mABCG1 transcripts 3572 and 3601 bp, respectively). One of these transcripts shared 100% homology with

FIG. 2. Alignment of conceptually translated amino terminal human ABCG1 peptide sequences encoded by the identified human ABCG1 transcripts with different N-termini. (A) The 10 different 5′-ends of the human ABCG1 transcripts identified in these studies revealed six variant amino termini of the ABCG1 protein. In all cases the first ATG in the human ABCG1 open reading frame was used as the initiation site for the prediction of the amino acid sequences. (B) The two murine (*Mus musculus*) ABCG1 transcripts encode for the same ABCG1 protein which shares 96% (643 of 666) identity and more than 97% (654 of 666) amino acid homology with the human ABCG1 protein sequence (Accession Nos. CAA88636 and NP_058198). Homologous amino acids are marked with a +.

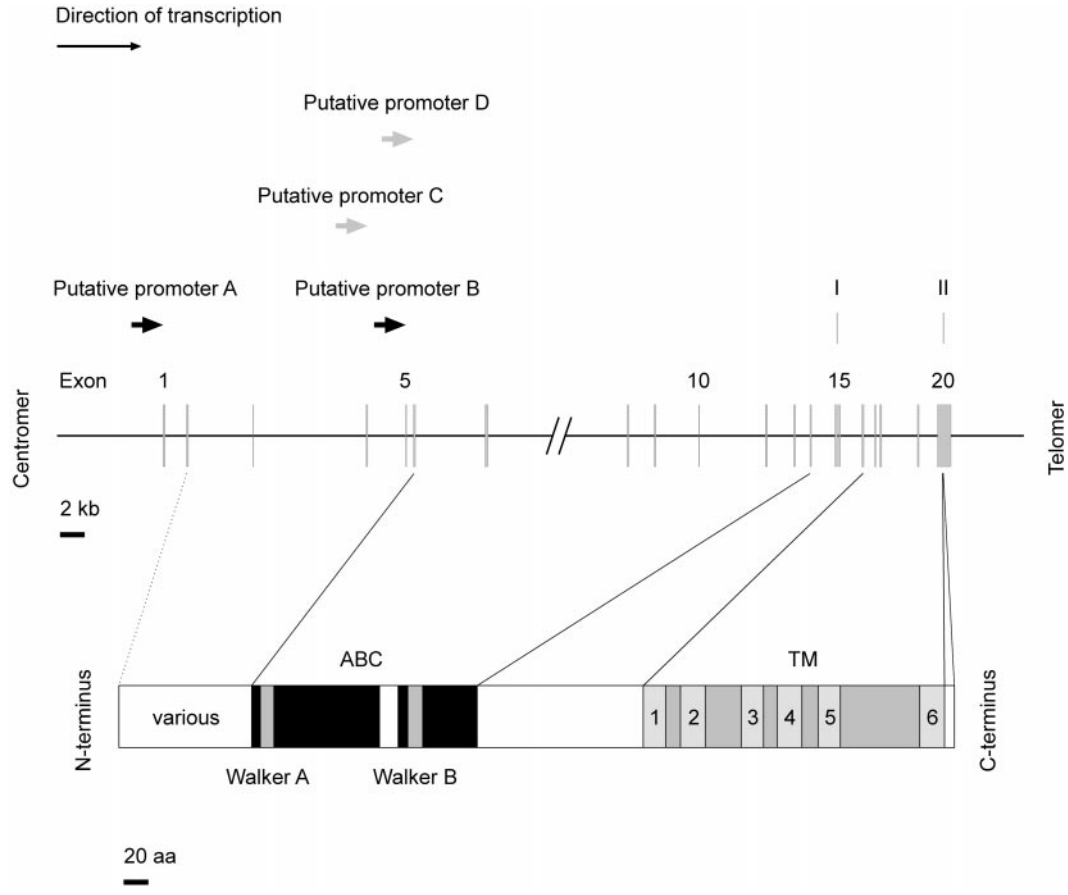

FIG. 3. Correspondence between genomic structure of the human ABCG1 gene and protein domains of the human ABCG1 protein. (Top) Exon–intron structure of the human ABCG1 gene on human chromosome 21. Exons are shown as gray boxes and vertical lines. I marks a 36-bp insertion reported by Chen *et al.* (9). II marks a poly(T) polymorphism reported by Chen *et al.* and Nakamura *et al.* (9, 12). The two putative promoter regions A and B of the two main start regions of hABCG1 transcripts are shown as solid black arrows. The position of two further putative, but weak promoters (C and D; not further analyzed) in the 5′-region upstream of exons 4 and 6 (transcripts C and J in Fig. 1) are indicated by solid gray arrows. (Bottom) Structure of the human ABCG1 protein. Exon-encoded protein domains are highlighted by graphic motifs in the predicted protein using transcript G. Various: Region of different peptides of the amino terminus of the human ABCG1 protein as encoded by human ABCG1 transcripts with different 5′-ends (see Fig. 2). ABC (black box): ATP-binding cassette including Walker A and B motifs (gray) and the "signature" motif (white). TM (gray box): transmembrane spanning domain with six transmembrane helices (1 to 6). White boxes: area between the nucleotide binding fold and the transmembrane domain with no known function and cytosolic C-terminus of the human ABCG1 protein.

the published sequence of the 5′-region of the mABCG1 mRNA (Accession No. Z48745) (10), with the exception of nine additional bases (GGAACCCGC) at the 5′-end of our sequences (data not shown). The new bases at the 5′-end do not result in a new ATG start codon, so that the published protein sequence (Accession No. CAA88636) would appear to represent the full-length mABCG1 protein. Moreover, the first ATG codon in the mABCG1 sequence agrees with Kozak's rules (33) and is therefore likely to be the authentic site at which translation of mABCG1 mRNA is initiated in the mouse. The predicted sequence of the first 49 amino acids of murine ABCG1 shows 96% identity and 97% homology to the proteins predicted from one of the transcripts of human ABCG1 (transcript G in Fig. 1) with its start site in exon 5 (former exon 1) (Fig. 2). The second mABCG1 transcript identified in our studies shares 100% homology with the published sequence of the 5′-region of the mABCG1 mRNA (Accession No. Z48745) (10), apart from the absence of 20 bases at the 5′ end of our sequences (data not shown). Our sequence contains the same ATG start codon as the published mRNA sequence (Accession No. Z48745) (10) so that the translation initiation site of the mABCG1 protein is the same for both our and the previously published sequence.

Human ABCG1 mRNA Sequences Generated by 3′-RACE

A single hABCG1 transcript representing the 3′-end of the hABCG1 mRNA was identified using 3′-RACE of

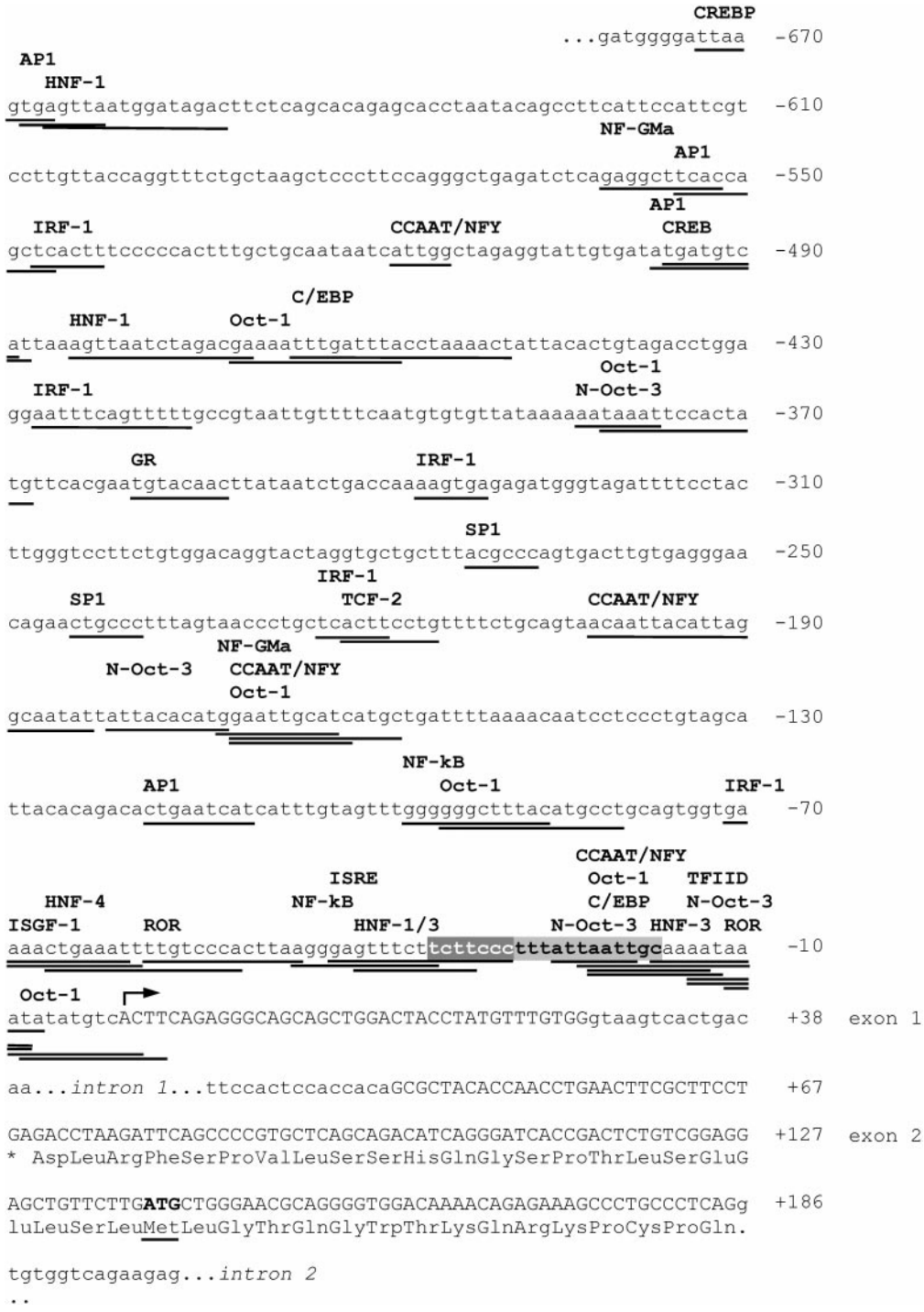

FIG. 4. Chromosomal nucleotide sequence of the 5′-flanking region of the human ABCG1 gene upstream of exon 1 (promoter A). The translation into amino acids of the first two exons of transcript D (the most common variant at promoter site A) in frame to the main part of the protein is given below the appropriate codons, beginning the translation after a stop codon (indicated as *) in exon 2. The first in frame ATG is highlighted in bold letters and the respective potential methionine start amino acid of protein sequences is underlined. Numbers to the right are relative to the transcription start site, which is indicated by an arrow. Negative values refer to the 5′-genomic sequence, positive values count only the bases of the cDNA (uppercase letters in the sequence). The genomic intron sequences are in lowercase letters, while the coding regions are shown in upper case. Sequences similar to consensus sequences for transcription factors are underlined. A TATA box consensus sequence is colored in light gray (41). The consensus sequence of a putative B recognition element (BRE) is colored in dark gray (42). Further explanations are given in the text.

TABLE 2

Position of the 20 Exons Encoding the Human ABCG1 Gene

Exon	Exon size (bp)	Position[a]	Intron size (bp)	Splice acceptor site intron′exon	Donor acceptor site exon′intron
1[b]	161	33,547–33,707	1,727	′AGACACTGAA	ATGTTTGTGG′gtaagtcact
2	147	35,435–35,581	5,222	tccaccacag′CGCTACACCA	CTGCCCTCAG′gtgtggtcag
3	41	40,804–40,844	9,090	ttgttttcag′TGAGAAATAT	TCCCCTATAG′gtaaggtatt
4	112	49,935–50,046	3,047	′GCTTTATAAA	ACATGGAACT′gtgagtccgt
5[c]	69	53,093–53,162	6,367	′GCCGCCGCCG	CACCGCCATG′gtgagtgagc
6	149	53,756–53,904	5,627	′GGCGAGGGGC	CACAGTTACT′gtaagtgctg
7	244	59,529–59,772	45,155	gtttttctag′AATGCCAGCA	AGGAAGAAAG′gtagggaggg
8	118	104,928–105,045	2,115	ttttttgcag′GATACAAGAC	CTGGATACAG′gtgagcagcc
9	133	107,161–107,293	3,459	ggttctgcag′GGAGACGGGC	GGCCATGATG′gtgagctccg
10	51	110,753–110,803	5,328	tcccctgcag′GTGTCGGCAC	AAGGGAAATG′gtaagtgggt
11	146	116,132–116,277	2,140	ttgccccccag′GTCAAGGAGA	AGCCCACCAG′gtaagtcagg
12	124	118,418–118,541	1,196	tgacttgcag′CGGCCTGGAC	GTTCGACCAG′gtacgcgggc
13	115	119,738–119,852	1,894	ttttttccag′CTTTACGTCC	GCAGATTTTG′gtaagcggag
14[d]	185	121,747–121,931	137	tctcctgcag′TCATGGAGGT	GTTGAGAAAG′gtaatgcaaa
15	102	122,069–122,170	1,737	gtgtcctcag′GACTCCTCGT	GAGGGACTCG′gtaaggctgc
16	169	123,908–124,076	873	ctgtccccag′GTCCTGACAC	GTTCTGACAT′gtgagtgaca
17	101	124,950–125,050	305	tatctcctag′TTCCCCTGGA	GCCCTTTCAG′gtgtgttagc
18	159	125,356–125,514	2,885	ttgtttccag′ATCATGTTCC	GTCCCTGCAG′gtgccagccc
19	119	128,400–128,518	1,503	ccaactccag′GTGGCCACTT	CCTATGTCAG′gtagcgggcg
20	1081	130,022–131,102	—	ttcctcctag′GTATGGGTTC	TTCTTCCTAA′

Note. The human ABCG1 gene spans 97,556 bp comprising 20 exons. The genomic intron sequences are in lowercase letters, while the coding regions are in uppercase letters. Sequences of the exons of the ABCG1 gene have been published in the GenBank database under Accession Nos. AF323639 to AF323658. Sequences of the predicted full-length ABCG1 transcripts encoded by the ABCG1 gene have been published in the GenBank database under Accession Nos. AF323659 to AF323664.

[a] Derived from the *Homo sapiens* genomic DNA sequence of chromosome 21q section 90/105 (Accession No. AP001746).

[b] Alternative start points of transcription results in various starts of exon 1 (transcript A, shown above; transcript B, 107 bp, position 33,601–33,707; transcripts C and D, 38 bp, position 33,670–33,707 (most frequently found); transcript E, 29 bp, position 33,679–33,707).

[c] Alternative start points of transcription results in various starts of exon 5 (transcript G, shown above; transcript H, 33 bp, position 53,130–53,162; transcript I, 31 bp, position 53,132–53,162).

[d] Alternative splicing produces a variant of exon 14 with 149 bp (position 121,747 to 121,895).

total RNA from human monocyte-derived macrophages treated with 22(*R*)-hydroxycholesterol and 9-*cis* retinoic acid. This transcript showed 100% identity with the human ABCG1 mRNA database sequence (Accession Nos. NM_004915 and NM_016818) and confirmed the published 3′-end of hABCG1 mRNA (data not shown) (3, 9, 10, 13).

Signal Peptide Sequences

Several of the hABCG1 mRNA transcripts encode proteins with possible cleavage sites near the amino terminus. None of the predicted proteins contains a signal peptide. In addition there are no target signals in the amino acid sequence predictive for the subcellular localization of ABCG1. Using the PSORT II algorithm, all putative ABC proteins were predicted to have the highest probability of localization in the endoplasmic reticulum or the plasma membrane.

Genomic Structure of the Human ABCG1 Gene

We calculated that the human ABCG1 gene spans 97 kb comprising 20 exons, 20 kb comprising 5 exons more than currently published in public databases (Fig. 3). The exon size was between 41 (exon 3) and 1081 bp (exon 20) (our numbering, Table 2). The introns ranged in size from 137 bp (intron 14) to more than 45,155 bp (intron 7) (our numbering, Table 2). All exon–intron boundaries displayed the canonical GT/AG sequence (34). The 36 bp insert reported by Chen *et al.* was present in the human ABCG1 mRNAs detected by us (9).

Location of Putative Regulatory Elements in the 5′-Flanking Region

Figures 4 and 5 show the results of an analysis of potentially regulatory elements (promoters A and B) immediately upstream of exons 1 and 5 (our numbering), respectively, using online programs as well as the transcripts C, D, and E to identify the most common start site of the hABCG1 transcript with its start sites in exon 1. The region from the transcription start site of exon 1 to −670 bp upstream has a G/C content of 39.6% and contains two possible SP1 site (Fig. 4). In contrast to the promoter region of ABCA1, the region upstream of exon 1 of ABCG1 contains no putative sterol response elements. Of note is the presence of two putative binding sites for nuclear factor κB (NF-κB) at positions −40 and −90. No

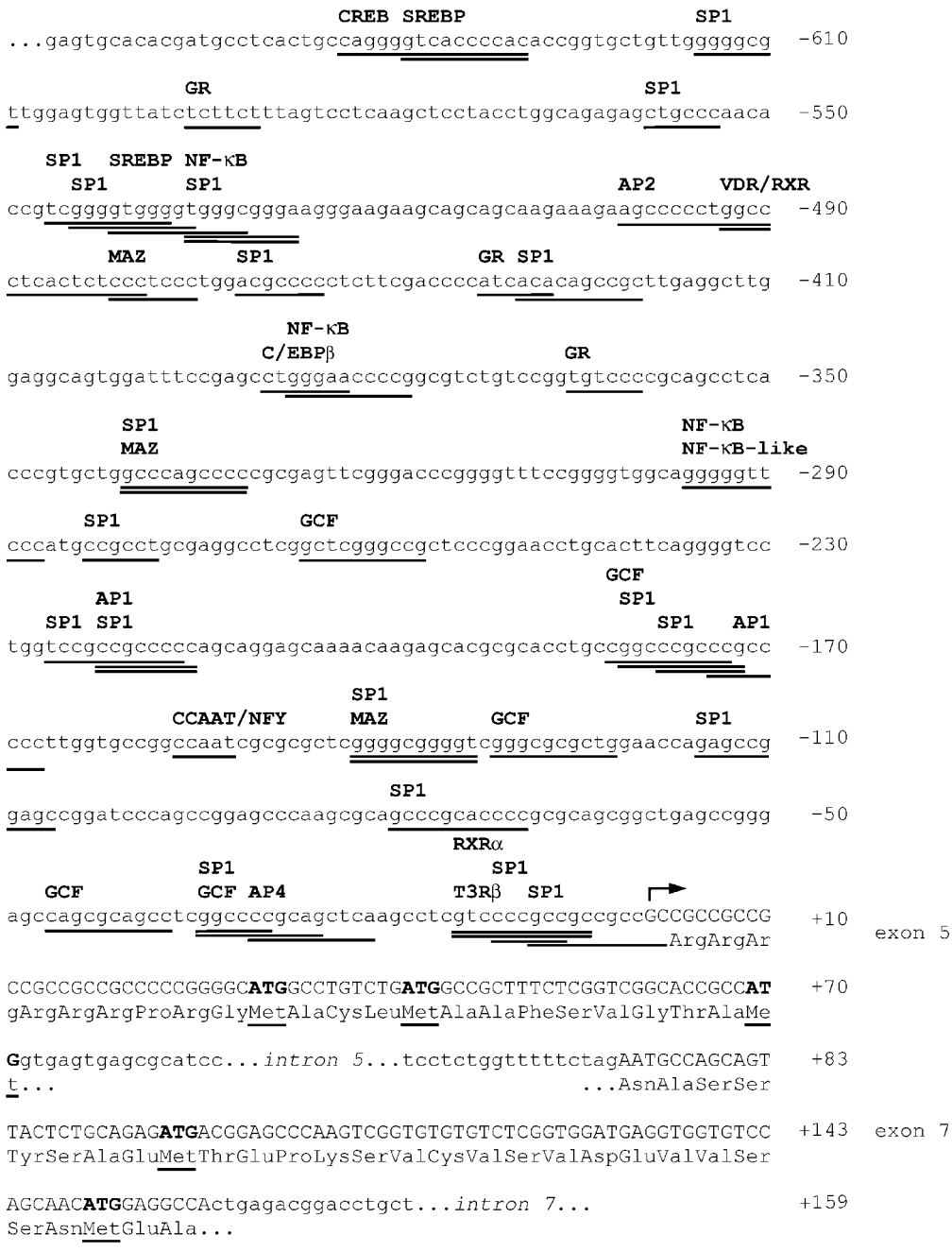

FIG. 5. Chromosomal nucleotide sequence of the 5′-flanking region of the human ABCG1 gene upstream of exon 5 (promoter B). The amino acid residues encoded by the first two exons of transcripts G to I are indicated below the appropriate codons with the start of the encoded protein sequences underlined. Numbers to the right are relative to the transcription start site, which is indicated by an arrow. The genomic intron sequences are in lower case letters, while the coding regions are shown in upper case. Sequences similar to consensus sequences for transcription factors are underlined. Further explanations are given in the text.

binding sites for the RXR were predicted within the 650 bp upstream of the transcription start. Nevertheless, a PPAR/RXR heterodimer binding site was predicted in the next 1000 bp immediately upstream of the region shown in Fig. 4 (data not shown). The region from the transcription start site of exon 5 to −650 bp upstream has a G/C content of 71.1% and contains no fewer than 19 possible SP1 binding sites (Fig. 5), which are often found in core promoters of TATA box-less gene (35). One possible sterol response element is present at position −540 and two possible NF-κB binding sites at positions −290 and −535. In contrast to promoter A, this promoter contains two putative RXR binding sites, one at position −20 and one at position −494.

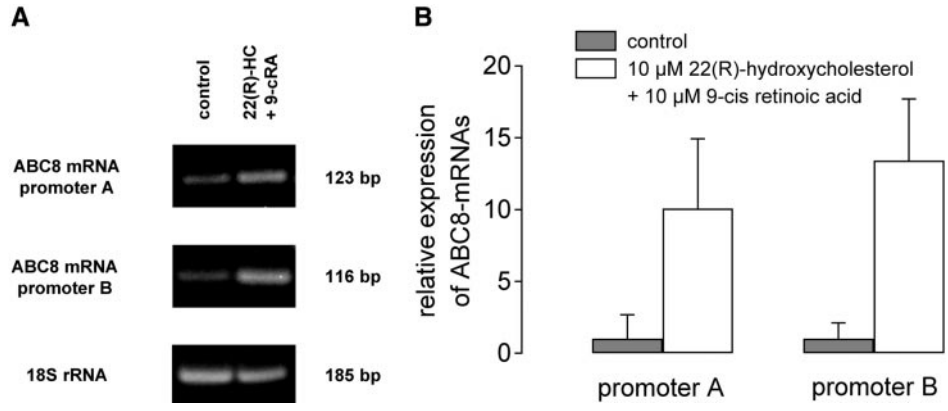

FIG. 6. Expression of hABCG1 transcripts in THP-1 macrophages. THP-1 cells were incubated for 48 h with 0.2 mM phorbol 12-myristate 13-acetate to induce their differentiation into macrophages. On day 3 a portion of the cells was treated for 24 h with 10 μM 22(*R*)-hydroxycholesterol (22(*R*)-HC) and 10 μM 9-*cis* retinoic acid (9-cRA). Real-time RT-PCR specific for human ABCG1 transcripts starting from promoter A and B and for human 18S rRNA, which was used for normalization, was performed with 50 ng reverse transcribed total RNA from each sample. (A) Promoter-specific RT-PCRs of human ABCG1 transcripts. Agarose gel analysis of real-time RT-PCR products after 50 PCR cycles: promoter A-specific mRNA (top), promoter B-specific mRNA (middle), and 18S rRNA (bottom). Expression of both human ABCG1 transcripts is increased by stimulation with 22(*R*)-hydroxycholesterol and 9-*cis* retinoic acid. No change in expression was observed for 18S rRNA during stimulation. (B) Relative expression of specific hABCG1 transcripts measured by real-time RT-PCR. The expression of both promoter A-specific and promoter B-specific transcripts of human ABCG1 is significantly increased (10.1 $\pm$ 4.9)-fold and (13.4 $\pm$ 4.3)-fold, respectively, by incubation with 22(*R*)-hydroxycholesterol and 9-*cis* retinoic acid.

Differential Expression of hABCG1 mRNAs

To investigate the activity of the two putative promoter regions, mRNA levels of the relevant transcripts were quantified by reverse transcription PCR. For the investigation of promoter A, primers (Table 1) were chosen which hybridize within the first and the second exon (our numbering) as shown in Fig. 1. Promoter B was investigated using primers which anneal to exon 5 and exon 7 (Fig. 1). In both cases, promoter activity was stimulated by addition of 10 μM 9-*cis* retinoic acid and 10 μM 22(*R*)-hydroxycholesterol to the culture medium of the cells. As shown in Fig. 6, activity of both promoters was stimulated approximately 10-fold by these substances.

DISCUSSION

The major novel finding of the present report is the discovery of several novel transcripts of the human ABCG1 gene in human macrophages. These were encoded by 5 novel exons, 4 upstream and one immediately downstream of the exon which was previously termed exon 1. These exons are predicted to encode at least five novel transcripts from the hABCG1 gene, which are due to different transcription start points and variant splicing of mRNA. In addition, we found evidence for two separate promoters for this gene, both of which appear to respond to stimulation by hydroxycholesterol and retinoic acid. The latter finding is of interest in view of the fact that only the putative promoter B contains putative binding sites for the RXR (Fig. 5). The finding of two promoters initiating separate transcripts starting with different exons from the same gene is also somewhat unusual; in most genes that have alternative transcripts these are produced by differential splicing within the gene (36).

Our data indicate the presence of a $CC(CCG)_{10}$-CCCCCGGGGC sequence extending upstream of the 5′-end of exon 5 (our terminology). GC-rich sequences and in particular CCG repeats are known to form stable secondary structures (37). Moreover, the mouse ABCG1 mRNA extends 91 bp further 5′ than the human ABCG1 mRNA and these 91 bp show more than 30% homology with the human genomic (promoter) ABCG1 sequence upstream of exon 5 (our numbering). It is therefore possible that the RACE-PCR or reverse transcription may not have been able to pass through this region and that hABCG1 transcripts containing exon 5 may be longer than we report here.

Very recent studies have highlighted the surprising and central role of members of the ABC gene family in cellular cholesterol regulation (13). Most of these efforts have focussed on the ABCA1 gene, a defect of which is known to cause the recessively inherited condition known as Tangier disease (17–20) which in the homozygous state is characterized by cholesterol accumulation in the cells of the reticuloendothelial system, virtually complete absence of circulating high density lipoprotein and, possibly, premature atherosclerosis (38). In addition, a recent paper showed that ABCA1 is involved in absorption of cholesterol from the gut, and thus in regulation of whole-body cholesterol balance in humans (15, 16).

It is now becoming clear, however, that other members of the ABC family are also intimately involved in cellular lipid metabolism. Among these the hABCG1 gene is very important. In studies of differential gene expression, the ABCG1 gene showed the highest level of upregulation in response to cholesterol loading of human macrophages (Cullen *et al.*, unpublished observations). Transcription of both the ABCA1 and ABCG1 genes has been shown to be increased by stimulation of the LXR/RXR pathway (14) and was confirmed for transcripts regulated by both putative promoters A and B of the ABCG1 gene in the present report. Both promoters, but in particular the promoter upstream of exon 5, contain potential SP1 sites, and both contained putative NF-κB binding sites, while the promoter upstream of exon 5, but not that upstream of exon 1, contained potential RXR and sterol response element binding protein (SREBP) binding sites. Both SP1 and NF-κ B have been shown to be involved in the coactivation of transcription of genes involved in cholesterol metabolism such as the low density lipoprotein receptor, hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase, HMG-CoA synthase and farnesyl pyrophosphate synthase (39), while SREBP speaks for itself. The role of the RXR receptor in the regulation of many lipid-related genes including ABCA1 has already been alluded to (15, 16). Thus, both promoters of hABCG1 contain binding sites which support a possible involvement of this protein in lipid metabolism (40).

Our findings suggest a hitherto unsuspected degree of complexity in the regulation of the hABCG1 gene. Further studies will be required to elucidate the physiological function of the gene at each of these sites and to clarify the role of the two putative promoters in regulating gene activity. Because of the central role of cholesterol metabolism in atherosclerosis, the ABCG1 gene may also prove to be a useful target for antiatherogenic therapies. A detailed understanding of the regulation of this gene is an important step in the design of such therapies.

ACKNOWLEDGMENTS

We thank Anja Wittelmann, Anja Mischke, and Hildegard Hünting for their excellent technical assistance and Renate Berfeld for monocyte isolation. This study was supported by Grant Cu 31/2 from the Deutsche Forschungsgemeinschaft to Dr. Paul Cullen and Grant B7 from the Interdisziplinäres Zentrum für Klinische Forschung to Prof. Dr. Gerd Assmann and Dr. Stephan Rust. This work forms part of the doctoral thesis of Stefan Lorkowski.

REFERENCES

1. Higgins, C. F. (1992) ABC transporters: From microorganism to man. *Annu. Rev. Cell Biol.* **8,** 67–113.
2. Luciani, M. F., and Chimini, G. (1996) The ATP binding cassette transporter ABC1 is required for the engulfment of corpses generated by apoptotic cell death. *EMBO J.* **15,** 226–235.
3. Croop, J. M., Tiller, G. E., Fletcher, J. A., Lux, M. L., Raab, E., Goldenson, D., Son, D., Arciniegas, S., and Wu, R. L. (1997) Isolation and characterization of a mammalian homolog of the Drosophila white gene. *Gene* **185,** 77–85.
4. Luciani, M. F., Denizot, F., Savary, S., Mattei, M. G., and Chimini, G. (1994) Cloning of two novel ABC transporters mapping on human chromosome 9. *Genomics* **21,** 150–159.
5. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) Distantly related sequences in the a and b subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951.
6. Dean, M., and Allikmets, R. (1995) Evolution of ATP-binding cassette transporter genes. *Curr. Opin. Genet. Dev.* **5,** 779–785.
7. Allikmets, R., Gerrard, B., Hutchinson, A., and Dean, M. (1996) Characterization of the human ABC superfamily: Isolation and mapping of 21 new genes using the expressed sequence tags database. *Hum. Mol. Genet.* **5,** 1649–1655.
8. Ames, G. F. L., Mimura, C. S., and Shyamala, V. (1990) Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli* to human: Traffic ATPases. *FEMS Immunol. Med. Microbiol.* **75,** 429–446.
9. Chen, H. M., Rossier, C., Lalioti, M. D., Lynn, A., Chakravarti, A., Perrin, G., and Antonarakis, S. E. (1996) Cloning of the cDNA for a human homologue of the Drosophila white gene and mapping to chromosome 21q22.3. *Am. J. Hum. Genet.* **59,** 66–75.
10. Savary, S., Denizot, F., Luciani, M. F., Mattei, M. G., and Chimini, G. (1996) Molecular cloning of a mammalian ABC transporter homologous to Drosophila white gene. *Mamm. Genome* **7,** 673–676.
11. Kuwana, H., Shimizu-Nishikawa, K., Iwahana, H., and Yamamoto, D. (1996) Molecular cloning and characterization of the ABC transporter expressed in Trachea (ATET) gene from *Drosophila melanogaster. Biochim. Biophys. Acta* **1309,** 47–52.
12. Nakamura, M., Ueno, S., Sano, A., and Tanabe, H. (1999) Polymorphisms of the human homologue of the Drosophila white gene are associated with mood and panic disorders. *Mol. Psychiatry* **4,** 155–162.
13. Klucken, J., Büchler, C., Orsó, E., Kaminksi, W. E., Porsch-Özcürümez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M., Allikmets, R., and Schmitz, G. (2000) ABCG1 (ABC8), the human homolog of the *Drosophila white* gene, is a regulator of macrophage cholesterol and phospholipid transport. *Proc. Natl. Acad. Sci. USA* **97,** 817–822.
14. Venkateswaran, A., Repa, J. J., Lobaccaro, J. M., Bronson, A., Mangelsdorf, D. J., and Edwards, P. A. (2000) Human white/murine ABC8 mRNA levels are highly induced in lipid-loaded macrophages. A transcriptional role for specific oxysterols. *J. Biol. Chem.* **275,** 14700–14707.
15. 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) Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. *Science* **289,** 1524–1529.
16. Ferber, D. (2000) Lipid research. Possible new way to lower cholesterol. *Science* **289,** 1446–1447.
17. Bodzioch, M., Orso, E., Klucken, T., Langmann, T., Bottcher, L., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Özcürümez, M., Kaminski, W. E., Hahmann, H. W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K. J., and Schmitz, G. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* **22,** 347–351.
18. Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J. A., Molhuizen, H. O. F., Loubser, O., Ouelette, B. F. F., Fichter, K., Ashbourne-Excoffon, K. J. D., Sensen, C. W., Scherer, S., Mott, S., Denis, M., Martindale, D., Frohlich, J., Morgan, K., Koop, B., Pimstone, S., Kastelein, J. J. P., Genest, J., and Hayden, M. R. (1999) Muta-

tions in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat. Genet.* **22,** 336–345.

19. Marcil, M., Brooks-Wilson, A., Clee, S. M., Roomp, K., Zhang, L. H., Yu, L., Collins, J. A., van Dam, M., Molhuizen, H. O., Loubster, O., Ouellette, B. F., Sensen, C. W., Fichter, K., Mott, S., Denis, M., Boucher, B., Pimstone, S., Genest, J., Jr., Kastelein, J. J., and Hayden, M. R. (1999) Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux. *Lancet* **354,** 1341–1346.
20. Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F., Brewer, H. B., Duverger, N., Denefle, P., and Assmann, G. (1999) Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat. Genet.* **22,** 352–355.
21. Cullen, P., Mohr, S., Brennhausen, B., Cignarella, A., and Assmann, G. (1997) Downregulation of the selectin ligand-producing fucosyltransferases Fuc-TIV and Fuc-TVII during foam cell formation in monocyte-derived macrophages. *Arterioscler. Thromb. Vasc. Biol.* **17,** 1591–1598.
22. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6.
23. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.
24. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) The Pfam protein families database. *Nucleic Acids Res.* **28,** 263–266.
25. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95,** 5857–5864.
26. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28,** 231–234.
27. Horton, P., and Nakai, K. (1996) A probabilistic classification system for predicting the cellular localization sites of proteins. *ISMB* **4,** 109–115.
28. Horton, P., and Nakai, K. (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *ISMB* **5,** 147–152.
29. Nakai, K., and Horton, P. (1999) PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24,** 34–36.
30. Schug, J., and Overton, C. G. (1997) Technical report (CBIL-TR-1997-1001) of the Computational Biology and Informatics Laboratory, School of Medicine, Univ. Pennsylvania.
31. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23,** 4878–4884.
32. Gardiner-Garden, M., and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.* **196,** 261–282.
33. Kozak, M. (1996) Interpreting cDNA sequences: Some insights from studies on translation. *Mamm. Genome* **7,** 563–574.
34. Breathnach, R., and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50,** 349–383.
35. McKnight, S., and Tjian, R. (1986) Transcriptional activity of viral genes in mammalian cells. *Cell* **46,** 795–805.
36. Smith, C. W., and Valcarcel, J. (2000) Alternative pre-mRNA splicing: The logic of combinatorial control. *Trends Biochem. Sci.* **25,** 381–388.
37. Darlow, J. M., and Leach, D. R. (1998) Secondary structures in d(CGG) and d(CCG) repeats. *J. Mol. Biol.* **275,** 3–16.
38. Assmann, G., von Eckardstein, A., and Brewer, H. B. (1995) *in* The Metabolic and Molecular Basis of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., Eds.), pp. 1981–2030, McGraw-Hill, New York.
39. Osborne, T. F., and LaMorte, V. J. (1998) Molecular aspects in feedback regulation of gene expression by cholesterol in mammalian cells. *Methods* **16,** 42–48.
40. Schroepfer, G. J., Jr. (2000) Oxysterols: Modulators of cholesterol metabolism and other processes. *Physiol. Rev.* **80,** 361–554.
41. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212,** 563–578.
42. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998) New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12,** 34–44.