Complete genomic sequence of the human *ABCA1* gene: Analysis of the human and mouse ATP-binding cassette A promoter

Silvia Santamarina-Fojo*†, Katherine Peterson*, Catherine Knapper*, Yang Qiu‡, Lita Freeman*, Jan-Fang Cheng‡, José Osorio§, Alan Remaley¶, Xiao-Ping Yang*, Changting Haudenschild*, Catherine Prades§, Giovanna Chimini‖, Eunice Blackmon*, Teena Francois*, Nicholas Duverger§, Edward M. Rubin‡, Marie Rosier§, Patrice Denèfle§, Donald S. Fredrickson*, and H. Bryan Brewer, Jr.*

*National Heart, Lung, and Blood Institute, and ¶Clinical Center, Clinical Pathology Department, National Institutes of Health, Bethesda, MD 20892; ‖Centre d'Immunologie de Marseille-Luminy, 13288 Marseille, France; §Evry Genomics Center, Aventis Pharma, 91047 Cedex, Evry, France; and ‡Lawrence Berkeley National Laboratories, Berkeley, CA 94720

Contributed by Donald S. Frederickson, May 11, 2000

The *ABCA1* gene, a member of the ATP-binding cassette A (ABCA1) transporter superfamily, encodes a membrane protein that facilitates the cellular efflux of cholesterol and phospholipids. Mutations in *ABCA1* lead to familial high density lipoprotein deficiency and Tangier disease. We report the complete human *ABCA1* gene sequence, including 1,453 bp of the promoter, 146,581 bp of introns and exons, and 1 kb of the 3′ flanking region. The *ABCA1* gene spans 149 kb and comprises 50 exons. Sixty-two repetitive Alu sequences were identified in introns 1–49. The transcription start site is 315 bp upstream of a newly identified initiation methionine codon and encodes an ORF of 6,783 bp. Thus, the ABCA1 protein is comprised of 2,261 aa. Analysis of the 1,453 bp 5′ upstream of the transcriptional start site reveals multiple binding sites for transcription factors with roles in lipid metabolism. Comparative analysis of the mouse and human ABCA1 promoter sequences identified specific regulatory elements, which are evolutionarily conserved. The human ABCA1 promoter fragment −200 to −80 bp that contains binding motifs for SP1, SP3, E-box, and AP1 modulates cellular cholesterol and cAMP regulation of *ABCA1* gene expression. These combined findings provide insights into ABCA1-mediated regulation of cellular cholesterol metabolism and will facilitate the identification of new pharmacologic agents for the treatment of atherosclerosis in humans.

high density lipoproteins | atherosclerosis | cholesterol efflux

Tangier disease is an autosomal codominant disease characterized by a low level of serum high density lipoprotein (HDL) and the accumulation of cholesteryl esters in peripheral tissues (1, 2). Recently, it has been shown that mutations in the ATP-binding cassette A (ABCA1) protein (3–7) lead to Tangier disease and familial hypoalphalipoproteinemia. The ABCA1 protein is a member of the ATP-binding cassette (ABC) family, which transports a wide variety of molecules, including proteins, lipids, ions, and sugars (8–11). Defects in the members of the ABC family are the cause of several genetic disorders that include adrenoleukodystrophy, cystic fibrosis, and macular degeneration (12–14). In patients with Tangier disease the cellular efflux of both phospholipid and cholesterol (15–18) is defective, leading to accumulation of excess cellular cholesterol. Thus, cellular cholesterol efflux, the initial step in reverse cholesterol transport (19), is impaired in Tangier disease. In the absence of ABCA1 function lipidation of the lipid-poor HDL is markedly reduced, resulting in the hypercatabolism of HDL in Tangier disease patients (2, 20). ABCA1 has been shown to be widely expressed, but is particularly abundant in macrophages (21), the cell type that is most affected in Tangier disease (2). Recent studies have demonstrated sterol-dependent regulation of *ABCA1* gene expression in macrophages (7, 21, 22).

To fully understand the role that ABCA1 plays in regulating cellular cholesterol metabolism and the process of reverse cholesterol transport, we have determined the complete gene sequence of the mouse and human *ABCA1* genes, including their promoter and regulatory elements. We show the human *ABCA1* gene is 149 kb long and contains 50 exons, one more than previously described (21). We also identified an initiation methionine, which extends the protein an additional 60 aa (21). In addition, we report that the fragment spanning −200 to −80 bp of the *ABCA1* gene promoter contain a cholesterol regulatory element that modulates *ABCA1* expression in macrophages, providing insights into the mechanisms that regulate the expression of this key receptor involved in cellular cholesterol efflux.

Materials and Methods

5′ Rapid Amplification of cDNA Ends (RACE). To determine the 5′ end of the ABCA1 mRNA, 5′ RACE was performed by using the SMART RACE cDNA amplification kit from CLONTECH. Human placental total RNA (CLONTECH) was used as a template to generate the 5′ cDNA end of ABCA1. The 5′ RACE fragment was generated by using CLONTECH's universal primer mix and a gene-specific primer 152R (5′-CGG AGA AGG GGA GAA AAC AGA ACC-3′). The amplified product was sequenced by using the Applied Biosystems Prism BigDye terminator cycle sequencing kit. Sequencing reactions were resolved on an Applied Biosystems 310 automated capillary DNA sequencer.

Identification of Bacterial Artificial Chromosome (BAC) Clones Containing Human ABCA1 Sequences and Generation of BAC Subclone Libraries. BAC clones containing the human *ABCA1* gene were identified by PCR screening of the human CIT D libraries, release I and II, and the GSI BAC human libraries, release I and II (Genome Systems, St. Louis). The screen identified BAC clones 22927, 22926, 23764, 23770, 23771, 23772, 23773, 23774. Purified DNA from BAC 22926 (http://genome.wustl.edu/gsc/Protocols/BAC.shtml) was kinetically sheared with a Hydroshear device (GeneMachines, San Carlos, CA). The resulting fragments were end-repaired with T4 DNA polymerase and Klenow fragment. *Bst*XI linkers were ligated to the end-repaired DNA fragments,

Abbreviations: ABC, ATP-binding cassette; ABCA1, ATP-binding cassette A; RACE, rapid amplification of cDNA ends; BAC, bacterial artificial chromosome; UTR, untranslated region; AP, activator protein; SREBP, sterol regulatory binding protein; HNF, hepatocyte nuclear factor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF275948 and AJ017356).

†To whom reprint requests should be addressed at: National Institutes of Health, National Heart, Lung, and Blood Institute, Molecular Disease Branch, 10/7N115, 10 Center Drive, Bethesda, MD 20892-1666. E-mail: silvia@mdb.nhlbi.nih.gov.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

size-selected to 1.5–3 kb on an agarose gel, and subcloned into the plasmid pOTWI3 (kind gift of Eric Lander, National Cancer Institute, Bethesda, MD). Subclone libraries of BACs 22927 and 23764 were prepared by digesting the BACs with restriction enzymes (*Bam*HI, *Eco*RI, *Hin*dIII, or *Xba*I) and cloning the fragments into pGEM7Zf(+) (Promega). The desired subclones were identified by PCR screening. Plasmid DNAs were prepared by using Qiagen plasmid kits QIAprep Spin Miniprep or QiaFilter Plasmid Midi (Valencia, CA). BAC DNA was prepared following the modified Qiagen protocol described by Kirschner and Stratakis (23).

Sequencing of the Human *ABCA1* Gene. The BAC clone 22926 was sequenced to high accuracy by using a shotgun strategy as described (24). Randomly selected subclones of BAC 22926 were sequenced from both ends to a final estimated redundancy of 10-fold. Fluorescent sequencing was performed with dye-terminator (BigDye, Perkin–Elmer/Applied Biosystems Division) chemistry using 377xl and 3700 automated DNA sequencing instruments (Perkin–Elmer/Applied Biosystems Division). Sequence gap closure and resequencing of low-quality regions were performed by using synthetic primers. Specific regions of BACs 22927, 23764, and 23774 and plasmid templates were sequenced by using BigDye Terminator Cycle Sequencing reagents and resolved on an Applied Biosystems Prism 310 Capillary Sequencer. Regions yielding poor sequencing data were resolved by using either Applied Biosystems Prism dRhodamine Terminator Cycle Sequencing reagents or Applied Biosystems Prism dGTP BigDye Terminator Cycle Sequencing reagents. Primers for sequencing and PCR were synthesized on an Applied Biosystems 394 DNA/RNA Synthesizer by using Applied Biosystems Masterpiece reagents. Subclones of the BAC 23764 were sequenced by using the EZ:TN <KAN-2> Insertion Kit (Epicenter Technologies, Madison, WI) to generate multiple transposon insertion plasmids.

Cloning and Sequencing of 5′ Portion of the Mouse *ABCA1* Gene. We used two primers (mABC1.5′fwd, 5′-AGTCACAGCTCTGTGCTCTGG-3′ and mABC1.5′rev, 5′-GTTTGTCTCCTTCGAAATGTCA-3′) derived from nucleotides 1–148 of the mouse AbcA1 mRNA sequence (X75926) to screen the mouse RPCI-11 BAC library by PCR. Clone 129K10 containing the exon 2 sequence but not exon 3 of the *AbcA1* gene was subjected to sequence analysis. This BAC was sheared, subcloned, and sequenced to 6-fold redundancy by a similar approach as was described in the human BAC sequencing except that a *Sma*I digested pUC18 vector (Amersham Pharmacia) was used to create the 3-kb subclone library. Sequence was collected from an Applied Biosystems 377 sequencer and assembled by using PHRAP/PHRED to an estimated error rate of 1 in 10^4 bp. For analysis of mouse intron exon junctions mouse genomic clones were isolated from a 129SV lambda phage library (Stratagene) by screening with probes A-1–1576, B-2635–3510, C-3510–4939, D-4939–5867, and E-5733–6940 (sequence deposited in GenBank, accession no. X75926).

Sequence Analysis. Individual sequences from BAC clone 22926 were assembled and edited by using the PHRED/PHRAP/CONSED suite of programs (25–27) to a final estimated error frequency of less than 1 in 10^4 bp. Assembly accuracy was confirmed by alignment with known cDNA sequences and by the concordance of read-pair sequences from individual shotgun subclones. Other sequences were assembled and analyzed by using SEQUENCHER from Gene Codes (Ann Arbor, MI). Oligonucleotides were selected by using OLIGO from Molecular Biology Insights (Plymouth, MN). Homology searches and alignments were generated by using BLAST and BLAST2 (28). Further analysis and annotation of the sequence was performed by using the VECTOR NTI suite from InforMax (North Bethesda, MD). Analysis of promoter regulatory elements was performed by using MATINSPECTOR (29) and MOTIF search service from the ICR, Kyoto, Japan (30, 31). The prediction of the transcription start site was performed with TSSW and TSSG.

Construction of Reporter Plasmids for Luciferase Assay. The human *ABCA1* promoter region spanning −990, −295, −200, and −80 bp to +120 bp was PCR-amplified by using ABC1-specific primers and BAC DNA as template. The PCR-amplified product was ligated into the PXP1 plasmid (32) and sequenced.

Cell Culture and Transfection. RAW cells (American Type Culture Collection) were grown in EMEM (minimum essential media with Earle's salts) with 10% FCS (BioWhittaker). Approximately 1.5×10^5 cells were plated in 12-well plates (2.5 cm), grown to 50–70% confluency and cotransfected with 1 μg of the ABC1-Lucif plasmid and 0.5 μg of the pBetagal vector (CLONTECH) by using the Superfectin Reagent Kit (Qiagen, Valencia, CA). Two hours after addition of DNA, the medium was removed and replaced with complete AMEM (minimal essential media Eagle's alpha modification). Twenty-four hours later, the cells were refed with fresh media with or without 50 mg/ml cholesterol and 0.3 mM cAMP (Sigma). The cells were harvested 16 h after refeeding by using Lysis Solution from the Tropix Luciferase Assay Kit (Tropix, Bedford, MA). Aliquots were used for protein quantitation by using the MicroBCA kit (Pierce) and for luciferase and β-galactosidase assays by using the Tropix Luciferase Assay Kit and Galacto-Light Plus Kits, respectively. Gel electrophoretic mobility shift and DNAse I protection assays were performed on fragment −171 to −71 bp as described (32).

Results

***ABCA1* Genomic Structure and Intron/Exon Boundaries.** The complete gene sequence of *ABCA1* was obtained by analysis of four BACs (22926, 22927, 23764, and 23774) isolated from the human CIT D and GSI BAC release I and II libraries (Genome Systems). Two contigs of approximately 88,990 bp and 93,360 bp, containing exons 3 through the 3′ *ABCA1* flanking region, were assembled from overlapping BAC subclones. These were combined with a 49,009-bp contig containing the 5′ end of the gene, generated by direct and subclone sequencing of BAC clones 23764 and 23774. The assembled *ABCA1* gene spanned 149 kb (Fig. 1), which included 1,453 bp upstream of the transcriptional start site, 1 kb downstream of the polyadenylation signal, and 146,581 bp of 49 introns and 50 exons.

The human *ABCA1* gene contains 50 exons (Fig. 1). The intron/exon junctions of 49 exons were previously reported (6). Exon 1, which encodes part of the 5′ untranslated region (UTR), was followed by a large, 24,156-bp intron I (Table 1). Exon 2 encodes the rest of the 5′ UTR and the initiation methionine codon and the first 21 amino terminal residues of ABCA1. With the exception of exon 50, which is 3,454 bp in length, exons were relatively small, ranging in size from 33 to 245 bp (Table 1). Introns ranged in size from 111 to 24,156 bp with the largest introns (nos. 1, 2, and 5) located primarily in the 5′ end of the gene. The mouse *ABCA1* gene (Table 1) also consists of 50 exons interrupted by 49 introns. The size of the introns and exons are very similar in the two species (Table 1). Sequence analysis using BLAST software identified 62 repetitive Alu elements within the 149 kb of the human *ABCA1* gene (Fig. 1). Sixty-two Alu elements were located in introns 1 and 2, and the remainder were dispersed throughout other introns. A HERV element was identified in intron 5. The genomic region of DNA sequenced for this study included two sets of STS markers (Fig. 1). The distal marker pair A009 × 28 is located in intron 3. The next proximal marker D9S53 is located in intron 36. The presence of these markers places the human *ABCA1* gene on the GeneBridge 4 radiation hybrid map between 109.8 cM and 112.0 cM on chromosome 9.

To determine the 5′ UTR of the human ABCA1 mRNA, 5′ RACE was performed by using human total placenta RNA as

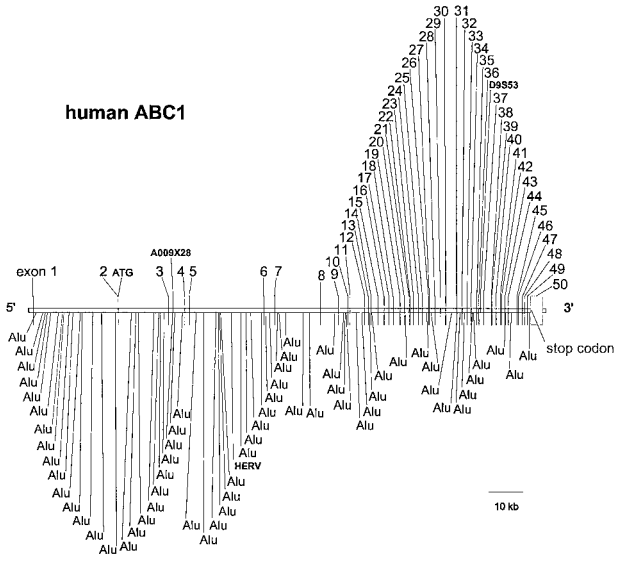

Fig. 1. The genomic organization of the human *ABCA1* gene is illustrated. The location of the 50 exons, two STS markers, and Alu repetitive elements present within the ABCA1 introns are shown.

template. A 213-bp amplification product was obtained when the *ABCA1* gene-specific primer 152R (see *Materials and Methods*) was used (data not shown). Sequence analysis of this product identified the transcription start site as G^{+1} (Fig. 2). These findings were confirmed in separate 5′ RACE studies performed by using exon 4 primers (data not shown). Promoter prediction algorithms such as TSSW and TSSG predict a transcription start site that is 3 bp upstream of the experimentally observed site, which would place the transcription start site 30 bp downstream of the TATA box. The 5′ UTR sequence of the RACE product was compared with the mouse *ABCA1* (33) as well as the human *ABCR* (34) cDNA sequences (Fig. 2). The ORF downstream of the identified transcription start site is 6,783 bp. It contains an in-frame methionine codon in exon 2 predicted to be the correct translation start site. This translation start site is 180 bp upstream of the previously reported initiation methionine codon located in exon 4 (21). The putative transcription start site identified by 5′ RACE is 315 bases upstream of the ATG (Fig. 2). The identification of this translation start site predicts an ABCA1 protein of 2,261 aa in length, 60 aa longer than originally described. We have altered our numbering of the amino acid sequence to reflect this change. Although the sequence similarity between the human exon 1 and the corresponding mouse sequence is only 73% and is disrupted by seven gaps in the alignment (Fig. 2), the conservation of the proposed transcription start site as well as the 5′ splice site of the intron 1 (Fig. 3) suggests the presence of a mouse exon 1 in this region. The predicted mouse exon 1 and intron 1 would be 219 bp and approximately 16 kb in size, respectively.

Analysis of the amino acid translation of both sequences using the SIGNALP web server (35) identified a potential signal peptide that included amino acids 1–45 in our sequence (Fig. 2). No signal peptide is predicted for the translation of AJ012376. The predicted signal peptide of the full-length cDNA conforms to the "positive-hydrophobic-polar" design consensus for signal peptides. The N-terminal portion of the signal peptide contains positively charged arginine and lysine residues. The middle portion of the signal peptide contains a 12-residue stretch of hydrophobic amino acids. The hydrophobic region is immediately adjacent to the polar C-terminal region that contains polar residues in positions −1 and −5 with respect to the cleavage site, and a positive charged residue at position −3 (36).

The first polyadenylation site, AATAAA, was located 3,284 bp 3′ of the stop codon. Comparison of our sequence to the human

Table 1. The human and mouse *ABCA1* gene organization

Exon number	Exon, bp		Intron, bp		Exon number	Exon, bp		Intron, bp	
	h	m	h	m		h	m	h	m
1	221	219	24,156	>16,000	26	49	49	195	230
2	159	149	14,359	8,000	27	114	114	1,395	800
3	94	94	4,539	4,300	28	149	149	1,649	1,335
4	142	142	1,269	700	29	125	125	1,236	1,000
5	119	119	21,187	15,000	30	99	99	3,031	3,400
6	122	122	2,980	1,300	31	190	190	1,520	700
7	177	177	12,952	10,500	32	95	95	1,309	1,300
8	93	93	4,957	1,900	33	33	33	1,122	1,100
9	241	241	2,711	2,250	34	105	106	1,474	6,500
10	140	140	331	330	35	75	75	521	430
11	117	117	4,208	3,540	36	170	170	1,230	2,000
12	198	198	746	660	37	178	178	1,997	3,000
13	206	206	520	500	38	116	116	111	230
14	177	177	1,786	530	39	145	145	1,040	1,350
15	223	223	1,746	1,605	40	124	124	1,086	1,400
16	222	222	1,060	780	41	130	130	264	290
17	205	205	1,104	884	42	121	121	787	800
18	114	114	1,797	1,385	43	63	63	907	830
19	172	172	989	730	44	107	107	2,354	2,010
20	132	132	1,305	1,100	45	142	142	371	200
21	143	143	203	200	46	135	135	943	1,500
22	138	138	702	505	47	104	104	482	750
23	221	221	1,257	1,100	48	93	93	658	700
24	73	73	986	700	49	245	244	940	1,040
25	203	203	1,667	1,600	50	3,454	>1,200		

h, human; m, mouse.

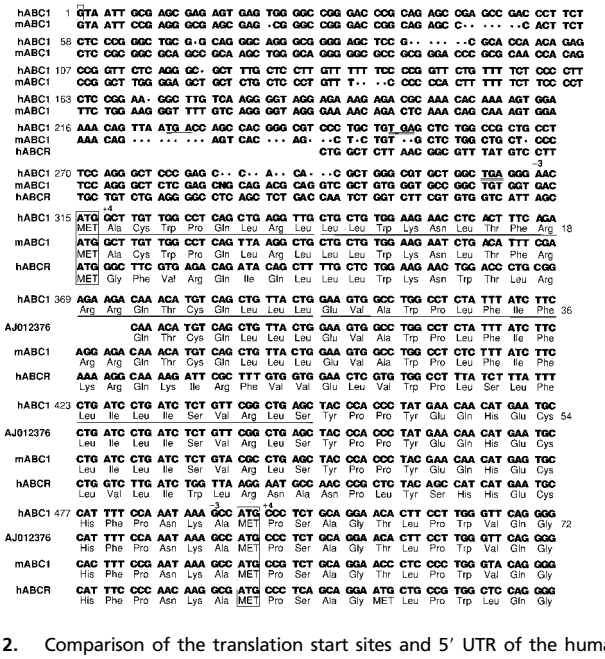

Fig. 2. Comparison of the translation start sites and 5′ UTR of the human *ABCA1*, mouse *ABCA1A,* and human *ABCR* genes. The transcription start site (G^{+1}) is boxed. The three stop codons upstream of the newly identified translation start site at position +315 bp (boxed) are underlined. The previously identified translation start site at position +495 bp (boxed) is also shown. The location of the highly conserved purine residues (−3) and G (+4) relative to both ATG start sites are indicated. The putative 45-aa signal peptide is underlined.

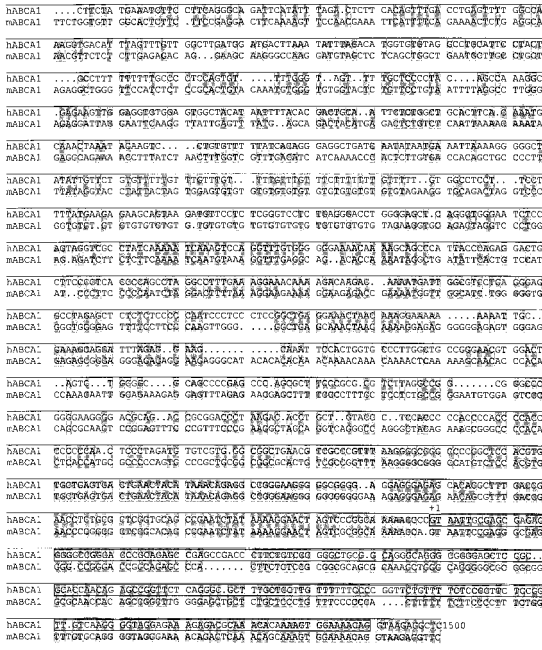

Fig. 3. Comparison of the mouse and human ABCA1 promoter sequence. BLAST2 software was used to search for homologous sequences (highlighted in gray) in the 5′ flanking regions of the mouse and human *ABCA1* genes. The proposed transcription start site is indicated (G^{+1}). Exon 1 is boxed.

expressed sequence tag database of GenBank places the last nucleotide of *ABCA1* 29 bases downstream of this poly(A) signal. Thus, the 3′ UTR of the human *ABCA1* gene spanned 3,313 bp in length. Comparison of our BAC exon sequence and recently reported *ABCA1* cDNA sequences (7) with the GenBank *ABCA1* sequence (accession no. AJ012376) (21) revealed several base differences resulting in amino acid changes. These included T1555I, P1648L, R1974K, and P2168L. Sequence analysis of the BAC clones revealed that the ABCA1 intron/exon splice junction sequences were as previously reported (6) with the following differences. The 3′ splice site for the newly identified exon 1 is 5′-GAAAACAGgtaagaggc-3′, and the 5′ splice site for exon 2 is 5′-tctttcagTTAATGAC-3′. In addition, five other sequence differences (see underlined bases) were identified in the intron/exon junctions of the *ABCA1* gene, including the 3′ splice site of exons 12 (5′-TCATGGAGgtgaatctg-3′), 19 (5′-ACCACCATgtaagaag-3′), 45 (5′-TTGGCAAGgtactgtg-3′), and 46 (5′-TGTTTCTGgtgagtat-3′) and the 5′ splice site of exon 46 (5′-tcactgtaGTTGGTGA-3′).

Analysis of the Mouse and Human ABCA1 Promoters. We have used BLAST2 software to search for homologous sequence in the 5′ flanking regions of the human and mouse *ABCA1* genes (Fig. 3). The analysis was performed by using the human ABCA1 sequence between −1453 bp and +300 bp to search for sequence matches in the mouse 129K10 sequence. A 189-bp sequence containing 90% similarity between the human sequence and the mouse sequence was identified from 1,081 bp to 1,089 bp of the human promoter (Fig. 3). The matching sequence starts immediately upstream of the human ABCA1 transcription initiation site. The sequence conservation strongly suggests the presence of a functional *ABCA1* gene promoter in this region. It is also likely that the transcription start site of the mouse *ABCA1* gene is located at the same position as the human ABCA1 gene.

The sequence of the human *ABCA1* promoter with predicted transcription binding sites is illustrated in Fig. 4. The first 250 bp of the human *ABCA1* promoter are GC rich with over 62% of the sequence containing either G or C. A TATA box (TCTATAAAG) was identified 33 bp upstream of the transcription start site mapped by 5′ RACE but a conventional CAT box motif was not present. Analysis of the −1453 bp of the human *ABCA1* promoter identified binding sites for the ubiquitously expressed transcription factors SP1, NF-κB, and activator protein (AP- 1, -2, and -4) as well as three E-box motifs (5′-CANNTG-3′). The *ABCA1* gene promoter also contains multiple binding motifs for the liver-enriched transcription factor hepatocyte nuclear factor (HNF)-3β (at positions −120, −532, −779, −865, −951, and −1050 bp relative to the transcription start site). Potential binding sites for transcription factors known to play a role in monocyte/macrophage differentiation, including STAT, c-myb, and GATA (37) also were identified. Interestingly, binding sites for SRY and SOX5, which like HNF-3β, are architectural factors that introduce strong bends in the chromatin were also present. Weak binding sites for sterol regulatory binding protein (SREBP) were located at −1415 bp, −1397 bp, −1212 bp, −1085 bp, −841 bp, −556 bp, −258 bp, and −150 bp of the human ABC1 promoter.

Comparison of the mouse and human *ABCA1* promoter sequences identified multiple predicted motifs that were conserved in both species (Fig. 4). In addition to the TATA box (−33 bp) several other potential binding motifs for transcription factors including SP1 (−100 and −166 bp), HNF-3β (−753 bp), AP-1 (−131 bp), AP-2 (−305 bp), SRY (−439, −530, and −599 bp) and NF-κB (−655 bp) were conserved in the mouse and human promoters. More weakly conserved motifs for the transcription factors HNF-3β (−121 bp and −952 bp), E-box (−147 bp), c-myb (−213 bp), STAT (−370 and −403 bp), SRY (−793 and −1072 bp), and AP1 (−1081 bp) also were present (Fig. 4). The high degree of conservation of these potential transcription binding sites between the two species indicate they may be biologically relevant.

To determine the effect of cholesterol on *ABCA1* promoter activity we transfected constructs containing the luciferase reporter cDNA under the control of the *ABCA1* promoter into RAW cells

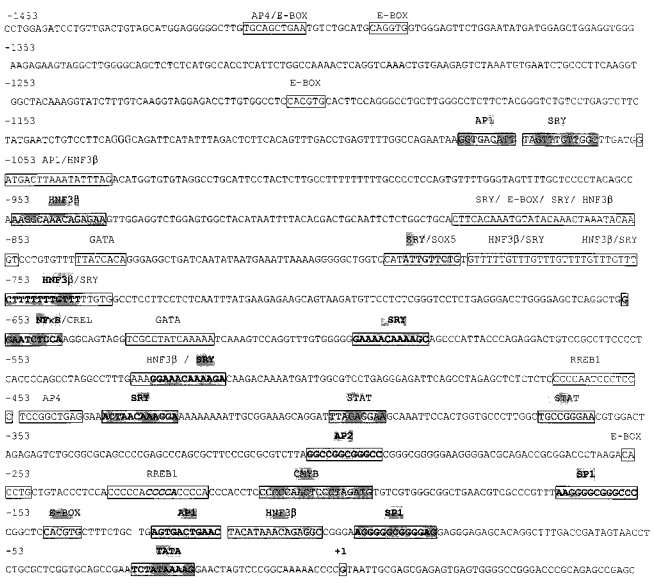

Fig. 4. The sequence of the human ABCA1 promoter is illustrated. The location of some potential binding sites for transcription factors identified by both MATINSPECTOR (core similarity >1 and matrix similarity > 0.87) and MOTIF programs is shown. Other motifs also identified by using the above criteria but not shown include: USF, myc-max, ARNT, myoD, E47, EVI1, MZF1, LMO2COM, and Nkx2.5. A potential TATA box is boxed. The transcription start site (G) identified by 5′ RACE is indicated as +1. Motifs that are conserved between the mouse and human sequences are highlighted in gray. Highly conserved motifs also are bolded.

in the presence or absence of cholesterol and cAMP (Fig. 5). Promoter fragment −220 to −80 bp contained sequences that regulated ABC1 gene expression by cholesterol and cAMP (Fig. 5). Gel shift assay (Fig. 5) demonstrated competition by DNA fragments containing the E-box sequence (F- and E-box). SREBP-2 antibodies inhibited the formation of DNA–protein complexes. RAW cell nuclear proteins were shown to bind to the E-box (−147 to −142 bp) motif by DNAse I footprint analyses (data not shown).

Summary

In the present study we report the entire gene sequence of the human *ABCA1* gene, including 1,153 bp of the promoter, 146,581 bp of the coding and intron sequence, and 1 kb of the 3′ flanking region. We identify potential binding sites for regulatory transcription factors common to both the mouse and human *ABCA1* gene promoters and provide insights into the mechanisms that regulate the expression of this key receptor involved in cellular cholesterol efflux. The genomic organization of the mouse and human *ABCA1* genes is similar to the *ABCR* gene (34), another member of the ABCA transporter family. All three transporter genes contain 50 exons interrupted by 49 introns. The mouse and human *ABCA1* genes span 100 kb and 149 kb, respectively, whereas the human ABCR gene is approximately 100 kb in size (34). Exons are relatively small but intron length varies greatly. Detailed analysis of all intron sequences in the human and mouse *ABCA1* genes reveal sizes ranging from 111 bp to 24,156 bp and 200 bp to 15,000 bp, respectively with the largest introns (nos. 1, 2, and 5) located primarily in the 5′ region of the gene. Multiple Alu repetitive elements were identified throughout the *ABCA1* gene introns. In addition a single human endogenous retrovirus element, a member of the LINE family of retrotransposon elements is present in intron 5. Identification of a polyadenylation site 3,284 bp downstream of the stop site predicts a large 3.3-kb 3′UTR for human ABCA1 and an ABCA1 transcript of approximately 10 kb in length. Analysis of GenBank expressed sequence tag data and human and mouse

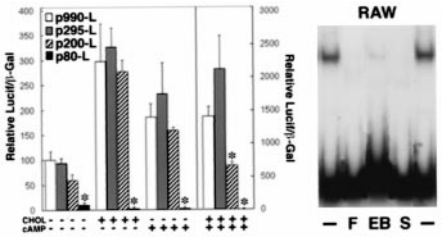

Fig. 5. Cholesterol and cAMP stimulation of the ABC1 promoter. Luciferase/β-galactosidase activity relative to the −990 bp to +120 bp fragment (F) is shown. Gel mobility shift assays demonstrates competition with 100× molar excess of fragment F, E-box (−147 to −142 bp), and SREBP-2 antibody (Santa Cruz Biotechnology). Data represent four independent studies. *, $P < 0.05$.

ABCA1 mRNA size (greater than 8 kb) (21, 33) support these findings.

A single ABCA1 transcription start site, identified by 5′RACE, was located 315 bp upstream of a newly identified translation start site that is identified as the first in-frame methionine codon after the transcription start site. It predicts an ORF of 6,783 bp instead of the originally reported 6,603 bp (21). The nucleotide sequence surrounding the initiation methionine conforms to the Kozak consensus sequence (RNNATGG), with a purine (R) at position −3 and a G at position + 4(38). The previously identified ATG start codon (position +60) in AJ012376 (21, 39) is a weak consensus sequence that contains a C instead of the conserved G at position +4. Terminator codons upstream of the ATG are another important indicator of translation start sites (38). Within 100 bases of the newly identified ATG in our cDNA sequence stop codons were identified in all three reading frames (Fig. 3), whereas the proposed ATG start site in AJ012376 has a terminator codon in only two of the three reading frames. Finally, the translation start site typically is preceded by a G/C-rich region (38). Sixty of the first 80 bases upstream (75%) of the translation start site were G or C. Whereas 35 of the first 80 bases (44%) of AJ012376 were G/C. The additional 60 aa of the human ABCA1 protein has 53% identical and 70% conserved amino acids with ABCR (34), a retinal ABC protein with overall high homology to ABCA1. Sequence analysis of 5′ RACE products and cDNA failed to provide evidence of alternative 5′ splicing of the human ABC1 gene. Based on this combined data, the human ABCA1 is comprised of 2,261 instead of the previously reported (21) 2,201 aa. The identification of a translation start site, 60 aa upstream of the originally described start site (21), has important implications for future structure-function analysis and expression studies of ABCA1.

Analysis of the amino acid translation of the *ABCA1* cDNA identified a potential signal peptide that includes amino acids 1–45 in our sequence. No leader sequence is predicted for the translation of AJ012376. The predicted signal peptide of the full-length *ABCA1* cDNA conforms to the "positive-hydrophobic-polar" design consensus and may be required for translocation of ABCA1 to the plasma membrane (36). The predicted cleavage site for the ABCA1 signal peptide resides 15 aa upstream of the previously reported ATG start site. Thus, a protein expressed by using this cDNA sequence would be truncated by at least 15 aa from its N-terminal domain.

Analysis of the mouse and human *ABCA1* promoters identified multiple motifs that were strongly conserved between the two species, suggesting important biological function. Some of these potential transcription factor binding sites are also present in the promoters of other receptors involved in lipid metabolism, including the low density lipoprotein (LDL) receptor, LDL receptor-related protein, CD36, SR-BI, and scavenger receptor A. These include binding motifs for SP1, AP-1, SRY/SOX5, NF-Y, and NF-κB (37, 40, 41). In addition, several E-box motifs, the consensus

recognition element for the basic helix–loop–helix leucine zipper containing proteins such as the SREBPs (42) were identified. Similar E-box motifs have been reported in the promoters for SR-BI (41, 43), fatty acid synthase (44), the human CD36 (40), and the LDL receptor (42). *ABCA1* gene expression has been reported in many tissues including placenta, lung, adrenal gland, and especially liver and macrophages (21, 33) where it facilitates the cellular efflux of cholesterol (3–5, 7). Analysis of the *ABCA1* gene promoter reveals several consensus binding sequences for transcription factors known to play a role in monocyte/macrophage differentiation (37) and for the liver-enriched transcription factor HNF-3β, which activates genes important for liver development and function as well as lipid metabolism (45–49). The HNF-3 motifs were closely linked to SRY/SOX5 binding sites. Particularly striking is the array of SRY/HNF-3B repeats from −780 bp to −737 bp in the ABCA1 promoter, preceded by SRY/SOX5 motifs at −793 bp. Like HNF-3, SRY and SOX5 are architectural factors that introduce a strong bend into the chromatin. Although their function is unknown, SRY/SOX5 binding motifs sites also are found in the SR-B1 promoter (41). The cluster of binding sites for these DNA bending factors at positions −533 bp, −780 bp, and −882 bp in the ABCA1 gene promoter suggest they may act to establish a promoter architecture or chromatin configuration permissive for transcription.

Recent studies have demonstrated sterol-dependent regulation of *ABCA1* gene expression in macrophages (7, 21). Several transcription factors that regulate sterol-modulated gene expression have been reported. These include the SREBPs and the orphan nuclear receptor LXR (liver receptor X). In cholesterol-loaded cells, SREBPs binds to the SRE element (5′-GCAGCCCAC-3′) in the promoter of various genes (42, 43, 50–52) leading to down-regulation of gene expression. SREBP-1 also has been shown to bind to a cholesterol response element (5′-ATGGTGNCAGATGGTG-3′) found in the cholesteryl ester transfer protein promoter (53). In contrast to SREBPs, oxysterol activation of the nuclear orphan receptor LXR (liver receptor X), which binds to the 5′ AG(G/T)TCA 3′ motif, has been shown to induce the expression of cholesterol 7 α-hydroxylase (54). Transfection analysis of 990 bp of the human *ABCA1* promoter revealed the presence of cholesterol and cAMP regulatory elements that enhanced the expression of the luciferase reporter gene in cholesterol-loaded macrophages in a region containing binding sites for SP1, SP3, E-box, and Ap1 transcription factors (−200 to −80 bp). The binding of RAW cell nuclear proteins could be competed by E-box fragments and SREBP-2 antibodies consistent with a role of SREBP in cholesterol-mediated *ABCA1* gene expression. The identification of potential promoter regulatory elements that modulate cholesterol-mediated *ABCA1* gene expression will facilitate the development of new pharmacologic agents for the treatment of low HDL and atherosclerosis in humans.

We thank Ms. Donna James for her excellent secretarial assistance.

1. Fredrickson, D. S., Attrocchi, P. H., Avioli, L. V., Goodman, D. S. & Goodman, H. C. (1961) *Ann. Intern. Med.* **55,** 1016–1031.
2. Assmann, G., von Eckardstein, A. & Brewer, H. B., Jr. (1995) in *The Metabolic and Molecular Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw–Hill, New York), pp. 2053–2072.
3. Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F., Brewer, H. B., Jr., Duverger, N., Denefle, P. & Assmann, G. (1999) *Nat. Genet.* **22,** 352–355.
4. Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L.-H., Roomp, K., van Dam, M., Yui, L., Brewer, C., Collins, J. A., Molhuizen, H. O. F., *et al.* (1999) *Nat. Genet.* **22,** 336–345.
5. Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., *et al.* (1999) *Nat. Genet.* **22,** 347–351.
6. Remaley, A. T., Rust, S., Rosier, M., Knapper, C., Naudin, L., Broccardo, C., Peterson, K. M., Koch, C., Arnould, I., Prades, C., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96,** 12685–12690.
7. Lawn, R. M., Wade, D. P., Garvin, M. R., Wang, X., Schwartz, K., Porter, J. G., Seilhamer, J. J., Vaughan, A. M. & Oram, J. F. (1999) *J. Clin. Invest.* **104,** R25–R31.
8. Dean, M. & Allikmets, R. (1995) *Curr. Opin. Genet. Dev.* **5,** 779–785.
9. Decottignies, A. & Goffeau, A. (1997) *Nat. Genet.* **15,** 137–145.
10. Allikmets, R., Gerrard, B., Hutchinson, A. & Dean, M. (1996) *Hum. Mol. Genet.* **5,** 1649–1655.
11. Schwiebert, E. M. (1999) *Am. J. Physiol.* **276,** C1–C8.
12. Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N. & Chou, J. L. (1989) *Science* **245,** 1066–1073.
13. Mosser, J., Douar, A. M., Sarde, C. O., Kioschis, P., Feil, R., Moser, H., Poustka, A. M., Mandel, J. L. & Aubourg, P. (1993) *Nature (London)* **361,** 726–730.
14. Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., Peiffer, A., Zabriskie, N. A., Li, Y., Hutchinson, A., *et al.* (1997) *Science* **277,** 1805–1807.
15. Remaley, A. T., Schumacher, U. K., Stonik, J. A., Farsi, B. D., Nazih, H. & Brewer, H. B., Jr. (1997) *Arterioscler. Thromb. Vasc. Biol.* **17,** 1813–1821.
16. Rogler, G., Truembach, B., Klima, B., Lackner, K. J. & Schmitz, G. (1995) *Arterioscler. Thromb.* **15,** 683–690.
17. Walter, M., Gerdes, U., Seedorf, U. & Assmann, G. (1994) *Biochem. Biophys. Res. Commun.* **205,** 850–856.
18. Francis, G. A., Knopp, R. H. & Oram, J. F. (1995) *J. Clin. Invest.* **96,** 78–87.
19. Fielding, C. J. & Fielding, P. E. (1995) *J. Lipid Res.* **36,** 211–228.
20. Schaefer, E. J., Kay, L. L., Zech, L. A. & Brewer, H. B., Jr. (1982) *J. Clin. Invest.* **70,** 934–945.
21. Langmann, T., Klucken, J., Reil, M., Liebisch, G., Luciani, M.-F., Chimini, G., Kaminski, W. E. & Schmitz, G. (1999) *Biochem. Biophys. Res. Commun.* **257,** 29–33.
22. Klucken, J., Buchler, C., Orso, E., Kaminiski, W. E., Porsch-Ozcurumez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97,** 817–822.
23. Kirschner, L. S. & Stratakis, C. A. (1999) *BioTechniques* **27,** 72–74.
24. Wilson, R. K. & Mardis, E. R. (1997) in *Analyzing DNA*, eds. Birren, B., Green, E. D., Klapholz, S., Myers, R. M. & Roskams, J. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 397–545.
25. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res.* **8,** 175–185.
26. Ewing, B. & Green, P. (1998) *Genome Res.* **8,** 186–194.
27. Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res.* **8,** 195–202.
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
29. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995) *Nucleic Acids Res.* **23,** 4878–4884.
30. Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A., *et al.* (1998) *Nucleic Acids Res.* **26,** 362–367.
31. Heinemeyer, T., Chen, X., Karas, H., Kel, A. E., Kel, O. V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F. & Wingender, E. (1999) *Nucleic Acids Res.* **27,** 318–322.
32. Previato, L., Parrott, C. L., Santamarina-Fojo, S. & Brewer, H. B., Jr. (1991) *J. Biol. Chem.* **266,** 18958–18963.
33. Luciani, M. F., Denizot, F., Savary, S., Mattei, M. G. & Chimini, G. (1994) *Genomics* **21,** 150–159.
34. Allikmets, R., Wasserman, W. W., Hutchinson, A., Smallwood, P., Nathans, J., Rogan, P. K., Schneider, T. D. & Dean, M. (1998) *Gene* **215,** 111–122.
35. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) *Int. J. Neural Syst.* **8,** 581–599.
36. von Heijne, G. (1990) *J. Membr. Biol.* **115,** 195–201.
37. Valledor, A. F., Borras, F. E., Cullell-Young, M. & Celada, A. (1998) *J. Leukocyte Biol.* **63,** 405–417.
38. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
39. Luciani, M.-F. & Chimini, G. (1996) *EMBO J.* **15,** 226–233.
40. Armesilla, A. L. & Vega, M. A. (1994) *J. Biol. Chem.* **269,** 18985–19991.
41. Cao, G., Garcia, C. K., Wyne, K. L., Schultz, R. A., Parker, K. L. & Hobbs, H. H. (1997) *J. Biol. Chem.* **272,** 33068–33076.
42. Brown, O. A. & Goldstein, J. L. (1997) *Cell* **89,** 331–340.
43. Lopez, D. & McLean, M. P. (1999) *Endocrinology* **140,** 5669–5681.
44. Magana, M. M., Koo, S. H., Towle, H. C. & Osborne, T. F. (2000) *J. Biol. Chem.* **275,** 4726–4733.
45. Allander, S. V., Durham, S. K., Scheimann, A. O., Wasserman, R. M., Suwanichkul, A. & Powell, D. R. (1997) *Endocrinology* **138,** 4291–4300.
46. Harnish, D. C., Malik, S., Kilbourne, E., Costa, R. & Karathanasis, S. K. (1996) *J. Biol. Chem.* **271,** 13621–13628.
47. Darlington, G. J. (1999) *Curr. Opin. Cell Biol.* **11,** 678–682.
48. Selden, C., Khalil, M. & Hodgson, H. J. (1999) *Gut* **44,** 443–446.
49. Cereghini, S. (1996) *FASEB J.* **10,** 267–282.
50. Kan, H. Y., Pissios, P., Chambaz, J. & Zannis, V. I. (1999) *Nucleic Acids Res.* **27,** 1104–1117.
51. Yang, W. S. & Deeb, S. S. (1998) *J. Lipid Res.* **39,** 2054–2064.
52. Sato, R., Miyamoto, W., Inoue, J., Terada, T., Imanaka, T. & Maeda, M. (1999) *J. Biol. Chem.* **274,** 24714–24720.
53. Gauthier, B., Robb, M., Gaudet, F., Ginsburg, G. S. & McPherson, R. (1999) *J. Lipid Res.* **40,** 1284–1293.
54. Lehmann, J. M., Kliewer, S. A., Moore, L. B., Smith-Oliver, T. A., Oliver, B. B., Su, J. L., Sundseth, S. S., Winegar, D. A., Blanchard, D. E., Spencer, T. A. & Willson, T. M. (1997) *J. Biol. Chem.* **272,** 3137–3140.