

ARTICLE

Molecular characterization and structural implications of 25 new ABCB4 mutations in progressive familial intrahepatic cholestasis type 3 (PFIC3)

Dario Degiorgio¹, Carla Colombo², Manuela Seia¹, Luigi Porcaro¹, Lucy Costantino¹, Laura Zazzeron², Domenico Bordo³ and Domenico A Coviello^{*1}

¹Laboratorio di Genetica Medica, Fondazione IRCCS, Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milano, Italy; ²Centro Fibrosi Cistica, Università degli Studi di Milano, Fondazione IRCCS, Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milano, Italy; ³Bioinformatica e Proteomica Strutturale, Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy

Progressive familial intrahepatic cholestasis type 3 (PFIC3) is an autosomal-recessive disorder due to mutations in the ATP-binding cassette, subfamily B, member 4 gene (*ABCB4*). *ABCB4* is the liver-specific membrane transporter of phosphatidylcholine, a major and exclusive component of mammalian bile. The disease is characterized by early onset of cholestasis with high serum γ -glutamyltranspeptidase activity, which progresses into cirrhosis and liver failure before adulthood. Presently, about 20 distinct *ABCB4* mutations associated to PFIC3 have been described. We report the molecular characterization of 68 PFIC3 index cases enrolled in a multicenter study, which represents the largest cohort of PFIC3 patients screened for *ABCB4* mutations to date. We observed 31 mutated *ABCB4* alleles in 18 index cases with 29 distinct mutations, 25 of which are novel. Despite the lack of structural information on the *ABCB4* protein, the elucidation of the three-dimensional structure of bacterial homolog allows the three-dimensional model of *ABCB4* to be built by homology modeling and the position of the mutated amino-acids in the protein tertiary structure to be located. In a significant fraction of the cases reported in this study, the mutation should result in substantial impairment of *ABCB4* floppase activity. The results of this study provide evidence of the broad allelic heterogeneity of the disease, with causative mutations spread along 14 of the 27 coding exons, but with higher prevalence on exon 17 that, as recently shown for the closely related paralogous *ABCB1* gene, could contain an evolutionary marker for mammalian *ABCB4* genes in the seventh transmembrane segment.

European Journal of Human Genetics (2007) 15, 1230–1238; doi:10.1038/sj.ejhg.5201908; published online 29 August 2007

Keywords: progressive familial intrahepatic cholestasis; PFIC3; *ABCB4*; MDR3; multicenter study; allelic heterogeneity; evolutionary marker

*Correspondence: Dr DA Coviello, Laboratorio di Genetica Medica, Fondazione IRCCS, Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Via Commenda 12, Milano, MI 20122, Italy.
Tel: +39 02 5503 2173; Fax: +39 02 5503 2019;
E-mail: coviello@unige.it
Received 14 April 2007; revised 2 July 2007; accepted 17 July 2007;
published online 29 August 2007

Introduction

Progressive familial intrahepatic cholestasis type 3 (PFIC3) is a very rare autosomal-recessive human disease (Online Mendelian Inherited in Man (OMIM) no. 602347), which causes a chronic and progressive cholestatic liver disease in patients with high γ -glutamyltranspeptidase activity (γ -GT) levels and

histologic evidence of bile duct proliferation.¹ Age of onset may range from 1 month to 20.5 years, symptoms may include jaundice, pruritus, or there may be only biochemical cholestasis. Most patients progress to cirrhosis, portal hypertension and liver failure and often require liver transplantation² even during the first years of life. Response to treatment with ursodeoxycholate is variable.

PFIC3 is caused by mutations of ATP-binding cassette subfamily B (MDR/TAP) member 4 gene (*ABCB4*; OMIM*171060), also known as multidrug resistance 3 gene (*MDR3*), a member of the superfamily of ABC transporters. This ubiquitous molecular family uses energy derived from ATP hydrolysis to efflux a wide range of substrates across the cell membrane.³ The name *MDR3* was originally assigned because of the high homology (78% identical residues) with the ABC transporter P-glycoprotein coded by the multidrug resistance 1 (*MDR1*) gene, also called *ABCB1* gene, which shares with *MDR3* the same domain organization.⁴ Both ABC transporters contain four distinct protein domains and their ancestor appears to have arisen by gene duplication.³ *ABCB4* polypeptide chain is organized in two repeats, each about 610-amino-acid long, joined by a 60-residue-long linker region. Each repeat contains two structural modules (Figure 1a), a transmembrane (TM) domain (TMD) comprised of six TM α -helices and a

cytoplasmatic nucleotide-binding domain (NBD). On the cytoplasmatic side of the protein four small linker peptides, also referred to as intracellular domains (ICDs), connect the TM helices 2–3, 4–5, 8–9 and 10–11; on the extracellular side six short loops (ECs) connect TM segments^{5–7} (Figure 1a). The two TMDs contain specific sites for substrate binding and translocation, whereas the two NBDs, which display a high degree of sequence similarity with the equivalent domain of ABC transporters, couple the energy obtained from ATP hydrolysis to substrate transport.⁸ The ICDs are deemed to be involved in mediating the coupling between NBD conformational changes and the reorientation of TM helices concomitant with substrate extrusion.⁹

The *ABCB1* gene, one of the most extensively studied ABC transporters, is responsible for the human multidrug resistance phenotype that is a rapidly growing obstacle to the treatment of numerous infectious diseases, including human immunodeficiency¹⁰ and malaria.¹¹ The properties of this transporter are also exploited in cancer pharmacological therapy where *ABCB1* translocates the chemotherapeutic drugs and other molecules with a broad but defined specificity.¹²

A gene duplication of *ABCB1* and additional mutations selected as advantageous have created in mammals the

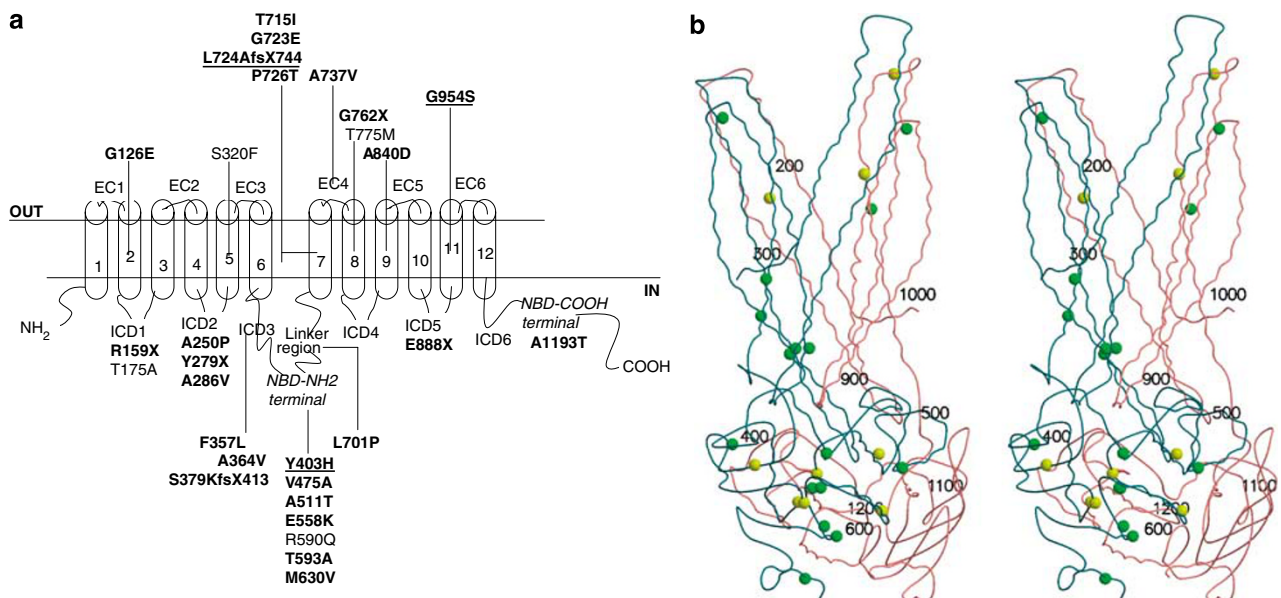


Figure 1 (a) Localization of the 29 mutations identified in this study in the *ABCB4* protein, schematically represented in its domains. All P-glycoproteins are composed of 12 TM segments (TM1–TM12) organized in two TMDs, connected by six intracellular segments (ICD1–ICD6) and by six short extracellular loops (EC1–EC6). A 60-amino-acid-long linker peptide connects the N terminal to the C terminal TMD–NBD domain. Of the 29 mutations identified in this study, 28% are located in the one of the two NBDs, 34% in TMDs, and 31% in ICDs. The 25 mutations newly identified in this study are marked in bold. Homozygous mutations are underlined; amino-acid numbering derived from *ABCB4* isoform B is used throughout this report. (b) Stereo representation of the three-dimensional model of *ABCB4*, build by homology with the multidrug ABC transporter Sav1866. It is assumed that the N- and C-terminal domains of *ABCB4* maintain relative to each other the same position of the two molecules of the (functional) crystallographic Sav1866 dimer. The N- and C-terminal domains are represented in blue and red, respectively. The mutated residues described in this work, which could be mapped in the homology model, are depicted as green spheres. Previously described mutations are represented as yellow spheres.

paralog *ABCB4* gene¹³ that, instead of drugs, translocates phosphatidylcholine (PC) from the inner to the outer leaflet of the canalicular membrane of the hepatocyte. The mammalian *ABCB4*-mediated PC translocation, also known as floppase activity, makes the phospholipids available for extraction into the canalicular lumen by bile salts.¹⁴ In the mouse, *ABCB4* floppase activity has been discovered by inactivation of the orthologous *Mdr2* gene, which resulted in progressive liver damage and cirrhosis that highly resemble the pathology of PFIC3 patients.¹⁵ The pathophysiology resides in lack of phospholipid protection in the bile against the detergent effect of bile salts, resulting in damage to the biliary epithelium, bile ductular proliferation, and progressive portal fibrosis.

The human *ABCB4* gene is located on chromosome 7q21.1 and spans approximately 74 kb (GenBank accession no. CH236949, region 10391396...10464818). It contains 27 coding exons and the B isoform of the transcript encodes a 1286-amino-acid-long polypeptide.

To date, about 20 distinct *ABCB4* mutations causing PFIC3 have been reported.^{1,2,16} In this study, we describe 29 distinct mutations that have been identified in a cohort of 68 unrelated patients with PFIC3 phenotype. The elucidation of the three-dimensional structure of *ABCB4* bacterial homolog makes it possible to map the position of the mutated residues in the protein structure, under the assumption of conserved three-dimensional fold. Although not conclusive, this study contributes to complete the picture of the molecular variability associated to PFIC3 that

will eventually provide a clear insight on the molecular mechanisms leading to the onset of the disease.

Materials and methods

Patients and controls

This study includes 68 proband index cases affected by PFIC3 enrolled in a multicenter study coordinated by one of us (CC) and carried out within the Italian Society for Pediatric Gastroenterology, Hepatology and Nutrition.

Inclusion criteria were increased values of γ -GT at presentation, associated with at least one of the following: (a) symptomatic cholestasis with pruritus; (b) jaundice; (c) abnormal alanine aminotransferase (ALT) at presentation. The exclusion of other known causes of cholestatic liver disease in our patient population reduced to 68 the number of patients enrolled for the molecular analysis. Each of the 13 participating centers collected the same clinical information using a common database and sent the blood samples to our laboratory for *ABCB4* analysis. Molecular analysis of *ABCB4* was also carried out on a control set of 100 healthy subjects.

DNA amplification

The DNA was extracted from peripheral blood by Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. Mutation analysis of all coding regions of the human *ABCB4* gene was performed using intronic primers (Table 1). Exon

Table 1 Oligonucleotides and specific annealing temperature used for DNA amplification of the *ABCB4* gene

Product including	Sequence forward primer 5' → 3'	Sequence reverse primer 5' → 3'	Annealing Temperature
Exon 2	CCGGAGAGGGTGTACTTGGTTCTGAG	GATTGGCGTGTAACGGAAAAGCCAG	60
Exon 3	TATCTTTGTAGTACCTTCGACAGTTAC	GTCAGCTCGCCTAGCTAGACAATC	56
Exon 4	GAGGAGAAATTCATTCCACAGC	CTTCTATGATATCTGGAGTCAACCAG	61
Exon 5	CAATAGCATCTTCTAGCTTTTCATACA	TGAATCTGGGTAAAGAGTACACG	55
Exon 6	GGTATTTAATAGAGCCTTTCTC	CCTTGACATATTTTCACACAG	53
Exon 7	GGCTTGCAGTCAGTGAACAA	CCTGAACAGGTACAAGTACGAGA	55
Exon 8	GGTTAAAGGGTTGACCAGAGTG	GCCATCAGTAAAGGGTGCTT	55
Exon 9	GGTCTCACCATGGGTTTCATT	TCATCTTTCAAAAAGGAGCGA	53
Exon 10	TAATGAATGCCAGAATGTGAC	TCAAAAATATGCAAACTAAAGCCA	59
Exon 11	CTTGTTTGTGCTATGATGGAATAGTC	GATTCAACAATCAACCTCAGTTAGG	56
Exon 12	TATGTGCCTTTTCCTTTGGG	AATTTGTATCCAGCTTCCCAC	56
Exon 13	GGATGTTTTTCATGAATGGTCC	TGGACAATCTTGCATCTCAAA	55
Exon 14	CAAAGCTCCATGTTGTCTTTATG	TGTTTCTCAGCCCAGACTCC	55
Exon 15	TTTGCCATAATCACGCAGAG	TGCTCAGTATAGCATTCACTGGA	55
Exon 16	CTTTAATGTCTTGATATCTTTTCAG	CCTGAAAAATATTTGCAAGGCT	55
Exon 17	CGAACAACCCATACTCAGCTTATG	GAGGTTGGGAGAAGCAGCAGC	64
Exon 18	AACATGTGACACTCAAGCCA	TACCCTCCAGCAGAGCCTTA	55
Exon 19	TCCGTGGCAACTGTAAAACA	GGGAAAACATGCATATCGAC	55
Exon 20	AGAGATGCCCTCCCTGCTAC	CAGTGGGTCAATCAACCTTGA	60
Exon 21	AGCCAATGTAGCTCAGGCAT	AGTTGTAGTGGGCACAAACATT	61
Exon 22	TGGCAGGCTGGGAATAGAA	AGAATTGTTGAAAAAGGCCACC	58
Exon 23	AAGCCGTGCTCTTTCCACTA	TTCTACTTAAACTTTTGCATAAAC	55
Exon 24	GTATGTCGGGGAGAAAGGG	TGAAACACATGTTGGTGAATCA	55
Exon 25	CAGTCTTTGGTAAAGTTTCTTGAA	TTCCATTATGACAATATTGGTTGG	53
Exon 26	AACGATCCTCCAAATGGACA	GTTTGGGGGATAAAAAGTAGTCTC	58
Exon 27	AAACACTCTGTTAAGTTGAAACAACG	TTTTTCATGGTTGACAGCAAA	55
Exon 28	CACATGCATAAAGTTCAAAGATT	TGCCGTAATAAACCCCAAT	53

numbering was used according to Lincke *et al.*¹⁷ Polymerase chain reaction (PCR) was carried out in a 30 μ l reaction volume containing 50–100 ng DNA, 1.5 Unit Taq DNA Polymerase (Roche Applied Science, Mannheim, Germany; URL: <http://www.roche.com/home/science.htm>), 6–30 pmol of each primer, 3.75 nmol dNTP (Finnzymes Company, Espoo, Finland; URL: <http://www.finnzymes.fi>) and 1.0–2.0 mM MgCl₂ in the buffer supplied by the manufacturer. The amplification was performed using a denaturation step at 94°C for 30 s, an annealing step (temperature primer-specific) for 30 s, extension step at 72°C for 30 s, repeated for 35 cycles. PCR products were cleaned up in Microcon[®] Centrifugal Filter (Millipore Corporation, Billerica, MA, USA; URL: <http://www.millipore.com/offices.nsf/home>).

DNA analysis

Sequence reactions were performed using 5–15 ng of PCR product, 5 pmol of the same primers used for PCR reactions, 1 μ l BigDye[®] Terminator v1.1. Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA; URL: <http://www.appliedbiosystems.com>), and 2.5 μ l of a dilution buffer (400 mM Tris–HCl, 10 mM, pH 9.0) in a total volume of 10 μ l. Thermal cycling was done at 96°C for 60 s followed by 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The sequencing reactions were purified by Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore Corporation, Billerica, MA, USA; URL: <http://www.millipore.com/offices.nsf/home>) and separated on a 16-capillary ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems). Sequence data were analyzed using the software Sequencer 4.2 (Gene Codes, Ann Arbor, MI, USA; URL: <http://www.genecodes.com>). Each individual mutation was confirmed in two independent PCR reactions and sequencings and all mutations were verified in at least two different members of the family. To exclude the possibility that the identified missense mutations could be polymorphisms, we analyzed 200 chromosomes from the general population by nucleotide. The fragments with abnormal chromatograms were sequenced to identify the exact denaturing high-performance liquid chromatography (DHPLC) change.

Sequence and structural analysis

Amino-acid multiple sequence alignment of human ABCB4 protein (NP_061337) was produced, after the inclusion of the paralogous human ABCB1 enzyme (NP_000918), four orthologous Eutheria mammalians ABCB4 proteins (*Pan troglodytes* (XP_001160982), *Mus musculus* (NP_032856), *Rattus norvegicus* (NP_036822) and *Bos taurus* (XP_602101)), the ABCB4 orthologous Metatheria mammalian *Monodelphis domestica* or Opossum (from Ensembl release 44: ENSMODP00000032941), the orthologous MsbA proteins from *Escherichia coli* (AP_001544), *Salmonella typhimurium* (NP_459959), *Vibrio cholera* (Q9KQW9) and the ABC transporter Sav1866 from *Staphylococcus aureus*

(2ONJA). The alignment was generated using the ClustalW web server (<http://www.ebi.ac.uk/clustalw/>). The three-dimensional model of ABCB4 was generated by homology model building with the SwissModel server¹⁸ using the three-dimensional structure of Sav1866 determined crystallographically⁹ (PDB code, 2HYD) and visually analyzed with the O program suite.¹⁹ For this purpose, the sequence alignment Sav1866 with either the N- and C-terminal domain of ABCB4 was determined with BL2SEQ program suite.²⁰ Significant sequence homology of the N- and C-terminal ABCB4 domain with Sav1866 spans amino-acid positions 56–600 (41% identical residues) and positions 807–1280 (32% identity), respectively. On the basis of the obtained sequence alignment, the models of the N- and C-terminal ABCB4 domains were built and energy minimized with SwissModel, using the molecule A and B of the crystallographic asymmetric unit, respectively.

Results

The mutation analysis of the human *ABCB4* gene carried out on 68 patients from different families led to the identification of 29 distinct mutations, of which 25 are not yet described. The observed mutations include two short insertions, 23 missense mutations and four nonsense mutations. The new sequences were submitted to Genbank (accession numbers are shown in Table 2). Overall, in this study 31 mutated alleles were identified in 18 families, for a total number of 35 mutations. Clinical inclusion criteria of each patient with identified mutations are described in Table 3. In these patients, in addition to increased values of γ -GT at presentation (18/18), other associated signs and/or symptoms were symptomatic cholestasis with pruritus (11/18), abnormal ALT at presentation (11/18) and jaundice (3/18).

In 13 patients mutations were found on both alleles, whereas in five cases only one mutated allele was identified. Four patients carried a double mutant allele showing a genotype with three mutations, one on the maternal allele and two on the paternal allele (Table 3). In three index cases (IC2, IC7 and IC10) with an affected brother, mutations were concordantly segregating.

In summary, more than half (51.7%) of the observed mutations were located in exons 6, 9, 15 and 17 (Table 2), with highest mutation frequency observed on exon 17. From a structural viewpoint, the mutations described here were located in one of the NBDs (eight cases, accounting for the 27.6% of the total; Figure 1a), in one of the TMDs (10 cases, 34.5%) and in one of the ICDs (nine cases, 31%). Only one mutation was found in the EC4 and one was found in the linker region.

Insertions

Two families showed two distinct short insertions. In one patient, the homozygous insertion c.2169_2170insG (exon

Table 2 Mutations identified in ABCB4

Exons	cDNA locus ^a	Type of mutation ^b		ABCB4-predicted domain	GenBank accession number ^c
		Missense	Frameshift or nonsense		
Exon 6	c.377G>A	G126E		TM2	DQ861346
Exon 6	c.523A>G	T175A		ICD1	
Exon 6	c.475C>T		R159X	ICD1	DQ861347
Exon 8	c.748G>C	A250P		ICD2	DQ861349
Exon 9	c.837T>A		Y279X	ICD2	DQ861348
Exon 9	c.857C>T	A286V		ICD2	DQ861350
Exon 9	c.959C>T	S320F		TM5	
Exon 10	c.1069T>C	F357L		ICD3	DQ861351
Exon 10	c.1091C>T	A364V		ICD3	DQ861352
Exon 11	c.1135_1136insAA		S379KfsX413	ICD3	DQ861353
Exon 11	c.1207T>C	Y403H		NBD-NH ₂ A-loop	EF035007
Exon 13	c.1424T>C	V475A		NBD-NH ₂ ter	DQ861354
Exon 13	c.1531G>A	A511T		NBD-NH ₂ ter	DQ861355
Exon 14	c.1672G>A	E558K		NBD-NH ₂ ter	DQ861356
Exon 15	c.1769G>A	R590Q		NBD-NH ₂ ter	
Exon 15	c.1777A>G	T593A		NBD-NH ₂ ter	DQ861357
Exon 15	c.1888A>G	M630V		NBD-NH ₂ ter	DQ861358
Exon 17	c.2102T>C	L701P		Linker region	DQ861359
Exon 17	c.2144C>T	T715I		TM7	DQ861360
Exon 17	c.2168G>A	G723E		TM7	DQ861361
Exon 17	c.2169_2170insG		L724AfsX744	TM7	DQ861362
Exon 17	c.2176C>A	P726T		TM7	DQ861363
Exon 17	c.2210C>T	A737V		EC4	DQ861364
Exon 18	c.2284G>T		G762X	TM8	DQ861365
Exon 19	c.2324C>T	T775M		TM8	
Exon 21	c.2519C>A	A840D		TM9	DQ861366
Exon 21	c.2662G>T		E888X	ICD5	DQ861367
Exon 23	c.2860G>A	G954S		TM11	DQ861368
Exon 27	c.3577G>A	A1193T		NBD-COOH ter	DQ861369

^acDNA sequence is based on reference sequence GenBank NM_018849. The A of the ATG of the initiator Met codon is reported as nucleotide +1.

^bThe protein GenBank reference number is NP_061337.

^cAccession number of the new mutations.

17) was observed, leading to the formation of a stop codon at position 744 (p.L724AfsX744; Table 3); the resulting truncated protein is therefore missing the C-terminal TMD and NBD. In the second case, the insertion c.1135_1136insAA (exon 11) was identified. This mutation leads to the formation of a stop codon at position 413 (p.S379KfsX413) resulting into a truncated protein comprised of the first TMD, with likely impairment of floppase activity; the patient was compound heterozygous with the missense mutation c.2102T>C (p.L701P) (Table 3).

Missense mutations

This is the largest group of ABCB4 mutations, with 23 distinct nucleotide substitutions identified, 19 of which are novel (Table 2). Two patients were homozygous, one for the mutation p.G954S and the second for p.Y403H (Table 3). Notably, two mutations already described were found in members of different families that are not part of a genetic isolate: p.T175A and p.S320F were identified in unrelated Caucasian patients (three and two cases, respectively). The comparison among six orthologous ABCB4 proteins and paralogous human ABCB1 shows that most (80%) of the mutated amino-acids identified in this study are conserved in eukaryotic P-glycoproteins (red and blue residues in

Figure 2). Moreover, 200 alleles from the general population were analyzed by DHPLC, and none of the missense mutations identified in the PFIC3 families were observed.

Nonsense mutations

The four novel nonsense mutations identified in this study are p.R159X, p.Y279X, p.G762X and p.E888X (Table 2). These mutations result in the synthesis of truncated proteins likely lacking floppase activity. The patient carrying the p.R159X mutation is heterozygous and the second mutated allele is still unknown; the other three cases show one nonsense mutation associated with three distinct missense mutations (Table 3).

Discussion

In this study, we analyzed the genomic DNA from 68 patients with PFIC3 diagnosis enrolled in the context of a pediatric multicenter study.

There are no PFIC3 epidemiologic data available to date; however, knowing that the number of newborns in Italy has been on average 500000/year in the last 14 years (<http://demo.istat.it/>), since we observed 18 patients with ABCB4-mutated alleles born within a 14-year period (with

Table 3 Clinical inclusion criteria and ABCB4 genotype in our PFIC3 patients with identified mutations

Index cases	Clinical inclusion criteria	Base change (maternal allele)	Effect on protein	Mutation category	Base change (paternal allele)	Effect on protein	Mutation category
IC1	ϵ_i, γ	c.475C>T	R159X	X	Unknown	L724AfsX744	X
IC2	δ, γ	c.2169_2170insG	L724AfsX744	X	c.2169_2170insG	L724AfsX744	X
IC3	$\alpha_i, \epsilon_i, \gamma$	c.523A>G	T175A	B ^a	c.1069T>C+c.2324C>T	F357L+ T775M	B+B ^a
IC4	ϵ_i, γ	c.2662G>T	E888X	X	c.748G>C+c.1888A>G	A250P+M630V	B+B
IC5	$\alpha_i, \epsilon_i, \gamma$	c.1135_1136insAA	S379KfsX413	X	c.2102T>C	L701P	C
IC6	ϵ_i, γ	c.2144C>T	T715I	X	Unknown		
IC7	α_i, γ	c.1672G>A	E558K	A	c.2168G>A+c.3577G>A	G723E+A1193T	C+A
IC8	α_i, δ, γ	c.2176C>A	P726T	B	c.1769G>A+c.2284G>T	R590Q+G762X	A ^a +X
IC9	α_i, γ	c.2860G>A	G954S	B	c.2860G>A	G954S	B
IC10	$\alpha_i, \epsilon_i, \gamma$	c.377G>A	G126E	B	c.1531G>A	A511T	A
IC11	δ, γ	c.523A>G	T175A	B ^a	Unknown		
IC12	ϵ_i, γ	c.959C>T	S320F	C ^a	c.857C>T	A286V	B
IC13	α_i, γ	c.2519C>A	A840D	B	c.1424T>C	V475A	A
IC14	ϵ_i, γ	c.1091C>T	A364V	B	c.2210C>T	A737V	C
IC15	$\alpha_i, \epsilon_i, \gamma$	c.1777A>G	T593A	A	Unknown		
IC16	α_i, γ	c.523A>G	T175A	B ^a	Unknown		
IC17	$\alpha_i, \epsilon_i, \gamma$	c.959C>T	S320F	C ^a	c.837T>A	Y279X	X
IC18	$\alpha_i, \epsilon_i, \gamma$	c.1207T>C	Y403H	A	c.1207T>C	Y403H	A

X = truncated protein; A = fully conserved residues; B = residues conserved in mammals ABCB4 and human ABCB1; C = residues conserved only in mammals ABCB4; α = symptomatic cholestasis with pruritus; δ = jaundice; ϵ = abnormal ALT at presentation; γ = elevated values; γ GT at presentation.
^aAlready described mutation in other ABCB4-deficient phenotype.

an average of 1.28 cases per year), we can assume that the frequency of the disease is not less than 1/400 000 live births in Italy. However, considering that the clinical centers participating in this study encompass no more than half of Italian population, and that the positive subjects for the molecular analysis are about one-fourth of the patients with PFIC3 clinical phenotype, we can estimate a more realistic frequency of 1/50 000.

To evaluate the effect of the mutations described in this study on protein function, the positions of the mutated residues were located on the predicted tertiary structure of ABCB4; the latter was obtained by homology building as previously illustrated starting from the known three-dimensional structure of homologous bacterial ABC transporter Sav1866 determined with crystallographic methods.⁹

The bacterial ABC transporter Sav1866 from *S. aureus* displays significant sequence homology with either the N- and C-terminal half of ABCB4 (41 and 32% identical residues spanning 545 and 474 amino-acid positions, respectively). Therefore, they are likely to share in common with Sav1866 the same tertiary structure, consisting of a TMD composed of six membrane-spanning helical segments followed by a NBD. The structure of the multidrug ABC transporter Sav1866, determined with crystallography at 3.0 Å resolution⁹ (PDB entry code, 2HYD), was used as a template structure to locate the position of mutated amino-acids in the protein tertiary structure, under the hypothesis of conserved overall topology between ABCB4 and the two molecules of the asymmetric unit of Sav1866 crystals (Figure 1b). In fact, the functional Sav1866 dimer is observed in the crystallographic asymmetric unit. It is assumed here that the relative position of the N- and C-terminal domain of ABCB4 is the same as the dimeric arrangement of the molecules A and B in the crystallographic Sav1866 structure, respectively. Although this has to be considered a working hypothesis, the deduced considerations may provide a useful framework to address validating site-directed mutagenesis experiments, aimed also at determining the functional role of the residues involved.

The amino-acid variability as observed in 10 distinct ABC transporters was also considered. For this purpose, the multiple sequence alignment was built with six mammalian proteins and four bacterial proteins (Figure 2). The multiple alignment provides a clear view of the amino-acids conserved among mammalian species, as well as those conserved across the whole molecular family encompassing eukaryota and prokaryota.

The missense mutations described in this study can be classified into three distinct categories, according to the degree of amino-acid conservation observed in the multiple alignment:

- (a) Mutations affecting fully conserved residues (represented in red in Figure 2). This category (named A in

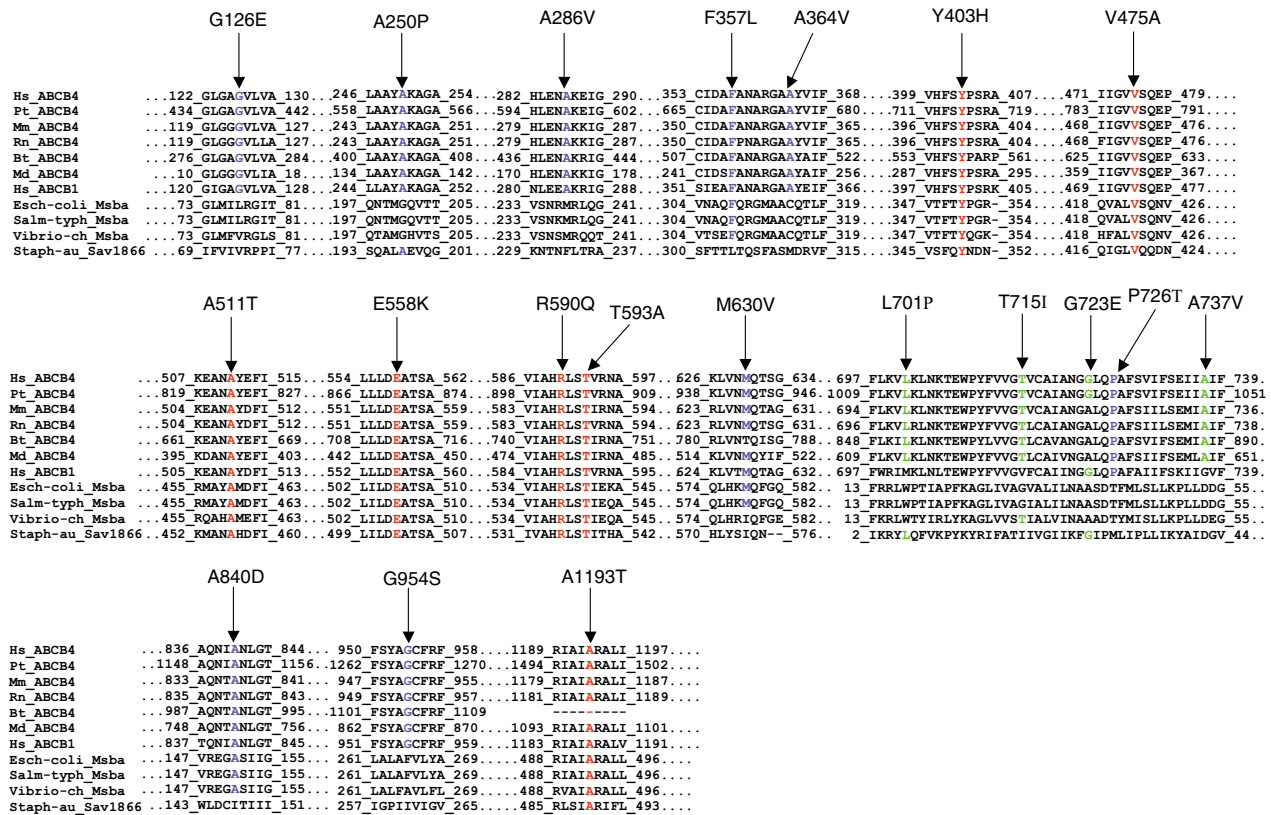


Figure 2 Multiple sequence alignment of ABCB4 protein sequences. The alignment includes the amino-acid sequences of the human ABCB4 protein (Hs_ABCB4), *Pan troglodytes* (Pt_ABCB4), *Mus musculus* (Mm_ABCB4), *Rattus norvegicus* (Rn_ABCB4), *Bos taurus* (Bt_ABCB4) and *Monodelphis domestica* or Opossum (Md_ABCB4). The sequence of the human paralogous ABCB1 (Hs_ABCB1) and that of four ABC bacterial proteins, *Escherichia coli* (Esch-coli_Msba), *Salmonella typhimurium* (Salm-typh_Msba), *Vibrio cholera* (Vibrio-ch_Msba) and *Staphylococcus aureus* (Staph-au_Sav1866) have also been included. The multiple alignment was obtained with ClustalW. Only sequence stretches containing amino-acid positions involved in new changes in PFIC3 are shown. Residues fully conserved are shown in red, amino-acid residues highly conserved are depicted in blue, residues less conserved are represented in green.

Table 3) includes seven cases: p.Y403H, p.V475A, p.A511T, p.E558K, p.R590Q, p.T593A and p.A1193T. Six of them (see Figure 1a) are located in the N-terminal NBD (spanning amino-acid positions 394–630), whereas only one is found in the C-terminal NBD (residues 1034–1272). Notably, among the 11 missense mutations previously described in the literature for PFIC3 phenotype,² six involve the N-terminal NBD, and none involve the C-terminal NBD. This bias is indicative of a distinct functional role of the two NBDs in the floppase activity. The strict conservation of the residues involved suggests either a functional or a structural role maintained in both mammalian and bacterial ABC transporters. In particular, the two missense mutations p.Y403H and p.E558K are located in protein sites already extensively investigated in other ABC transporters. Amino-acid Y403 forms the A-loop and represents the aromatic residue interacting with the adenine ring of ATP. Biochemical and site-directed mutagenesis studies carried out on human

ABCB1 (Y401) and on the HisP subunit of the bacterial ABC transporter histidine permease (Y16) have shown that this residue is critical for ATP binding and hydrolysis.^{21,22} The amino-acid E558 is part of the ‘extended’ Walker B motif; this carboxylate residue is completely conserved in the sequence of 266 ABC transporters;²³ site-directed mutagenesis studies involving the homologous position (E552) in mouse ABCB1 and the equivalent position (E179) in the HisP subunit of the bacterial histidine permease indicate that this residue is involved in ATP hydrolysis but not in ATP binding.^{21,23}

(b) Mutations involving amino-acid positions conserved in mammals ABCB4 and human ABCB1 P-glycoprotein (represented in blue in Figure 2); nine residues, located in functional domains, have been identified in this study (category named B in Table 3). Apart M630, located in the NBD, the remaining mutations involve either the helical loops (A250, A286, F357, A364) or TM segments (G126, P726, A840, G954; Figure 1b); the

changes p.G126E, p.P726T, p.A840D and p.G954S that involve substitution of neutral residue with a polar residue, could cause structural perturbations affecting the correct interaction of the respective TM segment with the cell membrane phospholipid bilayer.

Previous mutagenesis and biochemical studies carried out on human ABCB1 have assessed that Ala841, part of TM9, structurally equivalent to Ala840 in ABCB4, is involved in drug binding^{24,25} and interferes with the ability of the two structural modules (TMDs and NBDs) to form correct interactions.²⁵ Furthermore, the position Gly954 was already described as ABCB1 mutant G955V that shows altered drug-stimulated ATPase activities.²⁶

- (c) The remaining four novel missense mutations described here (green bars, Figure 2) are affected residues conserved only in mammals ABCB4 (category named C in Table 3). Among them, p.L701P involves the linker region (Figure 1a). Although no information is available concerning the structural role of this amino-acid, its substitution with the sterically constrained Pro residue may result in conformational stress. In this regard, transfected mammalian cell lines with ABCB1 mutants showed that insertion of a rigid structure (α -helical peptide) in this flexible connecting segment affects the cell surface localization of ABCB1.²⁷ The mutation p.G723E brings a charged residue within the hydrophobic environment of the membrane, with very unfavorable energetic effects. In fact, in the multiple alignment only ambivalent or hydrophobic residues are observed at this position; it has been previously shown that the Gly–Ala change at the adjacent position Gly722 in human ABCB1 disrupts interactions between the NBDs.²⁵

By cysteine-scanning mutagenesis, it has been recently shown that seventh transmembrane (TM7) of human ABCB1 forms part of the drug-binding pocket: the missense substitutions affecting the TM7 residues F728 and A729 with Cys residues resulted in significant decrease in the drug substrate affinity for Verapamil;²⁸ in ABCB4 it is likely that TM7 is part of the PC-translocation pathway. As a matter of fact, our results show that (i) TM7, encoded by exon 17, is subject to frequent mutational events (Table 2) and (ii) the four less conserved amino-acids are all clustered in exon 17 (green bars, Figure 2). Moreover, this exon is the part of the *ABCB4* most involved in PFIC3 mutations (20.7% in our study and 14.9% considering also data from other published mutations^{1,2,16}).

In summary, for 18 of the 25 new mutations described in this study, the predicted impairment effect on ABCB4 is quite evident; six mutations resulting in truncated proteins and 12 consisting in missense mutations. Among the latter, six mutations are located in predicted functional domains with involved residues fully conserved in eukaryotic

P-glycoproteins and in prokaryotic ABC proteins, and six mutations result in radical amino-acid substitutions in TMDs and linker region. The structural or functional relevance of the remaining seven amino-acid substitutions (p.A250P, p.A286V, p.F357L, p.A364V, p.M630V, p.T715I and p.A737V) is less clear and the induced effect on ABCB4 is less evident. Further characterization is required to understand better their causative role.

The low percentage of identified ABCB4-mutated alleles in our cohort of patients with PFIC3 phenotype (26.5%) suggests that the direct sequencing technique used in this study has a high sensitivity for point mutations and small insertions/deletions, but it misses the large allelic deletions; in addition, genetic heterogeneity underlying the PFIC3 phenotype can also be present. Finally, although all other known causes of liver disease were excluded, some degree of overlap in clinical phenotype could be present.

According to the category ABCB4 mutations (Table 3), our 13 patients with two ABCB4-mutated alleles can be divided into two groups:

- (1) Six patients with more severe genotype which predicts substantial impairment of ABCB4 floppase activity (IC2, IC7, IC8, IC10, IC13 and IC18); patients in this group present on the first allele one mutation of the category X (truncated proteins) or category A (mutations affecting fully conserved residues) and on the second allele one mutation of the categories X or A or one mutations with predicted structural perturbations in TMDs of the category B.
- (2) Seven patients with less severe genotype (IC3, IC4, IC5, IC9, IC12, IC14 and IC17) who present on the first allele one mutation of the category X (truncated proteins) or category B (mutations affecting conserved residues in functional domains of mammals ABCB4 and human ABCB1) and on the second allele one mutation of the categories C (mutations affecting conserved residues only in functional domains of mammals ABCB4) or B.

Although detailed clinical data on long-term follow up are needed to define genotype–phenotype correlation, it is of interest that two of the three patients (IC2 and IC8) with the more severe clinical picture at presentation belong to the more severe genotype group.

In conclusion, our data indicate that (a) the large majority of the new mutations of the ABCB4 gene are found in functional domains of the protein, and likely impair ABCB4 floppase activity; (b) ABCB4 deficiency is characterized by wide allelic heterogeneity, with private mutations in each affected family, with the N-terminal NBD as preferred hot spot for causative missense mutations and with more than half mutations located in exons 6, 9, 15 and 17; (c) frequent and advantageous mutational events in exon 17, and in TM7 segment in particular, of the

ancestral *ABCB1* gene could have very likely contributed to definition of the new PC floppase activity in the mammalian genome.

Acknowledgements

We thank the families of the patients for their collaboration. This study was supported in part by the Italian Ministry of University and Research (MIUR), PRIN contract number 2005068307; grant Regione Lombardia, DG Sanità n. 12298 (2004), DG Sanità n.19081 (2005) and by Eurogentest-Network of Excellence, contract n.o. FP6-512148-2004. We are grateful to the participants in the 'Multicenter study of Italian Society for Pediatric Gastroenterology, Hepatology and Nutrition': Prof G Maggiore, Clinica Pediatrica Università di Pisa; Dr G Torre, Centro Trapianto di Fegato, Bergamo; Dr R Iorio, Clinica Pediatrica, Università di Napoli; Dr M Resti, Azienda Ospedaliera Meyer, Firenze; Dr M Calacoci, Clinica Pediatrica, Università di Ferrara; Dr E Castellano (U. O. Pediatria III) and MG Marazzi (U. O. Malattie Infettive), IRCCS G. Gaslini, Genova; Prof F Balli, Clinica Pediatrica, Università di Modena; Prof P Vajro, Dipartimento di Pediatria, Università di Napoli; Dr E Ferretti, ASL 1 Napoli est; Prof L Zancan, Clinica pediatrica, Università di Padova; Dr S Martellosi, Clinica Pediatrica, IRCCS Burlo Garofalo, Trieste; Dr F Oliveri, U.O. Gastroenterologia ed Epatologia, Ospedale S Chiara, Pisa. We are grateful to Dr. Giandomenico Russo for oligonucleotides sequences.

References

- De Vree JML, Jacquemin E, Sturm E *et al*: Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. *Proc Natl Acad Sci USA* 1998; **95**: 282–287.
- Jacquemin E, De Vree JML, Cresteil D *et al*: The wide spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. *Gastroenterology* 2001; **120**: 1448–1458.
- Higgins CF: ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* 1992; **8**: 67–113.
- Van der Blik AM, Baas F, Ten Houte de Lange T, Kooiman PM, Van der Velde-Koerts T, Borst P: The human *mdr3* gene encodes a novel P-glycoprotein homologue and gives rise to alternatively spliced mRNAs in liver. *EMBO J* 1987; **6**: 3325–3331.
- Kast C, Canfield V, Levenson R, Gros P: Membrane topology of P-glycoprotein as determined by epitope insertion: transmembrane organization of the N-terminal domain of *mdr3*. *Biochemistry* 1995; **34**: 4402–4411.
- Kast C, Canfield V, Levenson R, Gros P: Transmembrane organization of mouse P-glycoprotein determined by epitope insertion and immunofluorescence. *J Biol Chem* 1996; **271**: 9240–9248.
- Klein I, Sarkadi B, Varadi A: An inventory of the human ABC proteins. *Biochim Biophys Acta* 1999; **1461**: 237–262.
- Holland IB, Blight MA: ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol* 1999; **293**: 381–399.
- Dawson RJ, Locher KP: Structure of a bacterial multidrug ABC transporter. *Nature* 2006; **443**: 180–185.
- Sankatsing SU, Beijnen JH, Schinkel AH *et al*: P glycoprotein in human immunodeficiency virus type 1 infection and therapy. *Antimicrob Agents Chemother* 2004; **48**: 1073–1081.
- Burk O, Arnold KA, Nussler AK *et al*: Antimalarial artemisinin drugs induce cytochrome P450 and MDR1 expression by activation of xenosensors pregnane X receptor and constitutive androstane receptor. *Mol Pharmacol* 2005; **67**: 1954–1965.
- Germann UA, Pastan I, Gottesman MM: P-glycoproteins: mediators of multidrug resistance. *Semin Cell Biol* 1993; **4**: 63–76.
- Annilo T, Chen ZQ, Shulenin S *et al*: Evolution of the vertebrate ABC gene family: Analysis of gene birth and death. *Genomics* 2006; **88**: 1–11.
- Oude Elferink RP, Paulusma CC: Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). *Pflugers Arch* 2006; **453**: 601–610.
- Smit JJ, Schinkel AH, Oude Elferink RP *et al*: Homozygous disruption of the murine *mdr2* P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. *Cell* 1993; **75**: 451–462.
- Chen HL, Chang PS, Hsu HC *et al*: Progressive familial intrahepatic cholestasis with high gamma-glutamyltranspeptidase levels in Taiwanese infants: role of MDR3 gene defect? *Pediatr Res* 2001; **50**: 50–55.
- Lincke CR, Smit JJ, van der Velde-Koerts T, Borst P: Structure of the human MDR3 gene and physical mapping of the human MDR locus. *J Biol Chem* 1991; **266**: 5303–5310.
- Kopp J, Schwede T: The SWISS-MODEL repository: new features and functionalities. *Nucleic Acids Res* 2006; **34**: D315–D318.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M: Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 1993; **50**: 157–163.
- Tatusova TA, Madden TL: BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 1999; **174**: 247–250.
- Shyamala V, Baichwal V, Beall E, Ames GF: Structure–function analysis of the histidine permease and comparison with cystic fibrosis mutations. *J Biol Chem* 1991; **266**: 18714–18719.
- Ambudkar SV, Kim IW, Xia D, Sauna ZE: The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett* 2006; **580**: 1049–1055.
- Urbatsch IL, Julien M, Carrier I, Rousseau ME, Cayrol R, Gros P: Mutational analysis of conserved carboxylate residues in the nucleotide binding sites of P-glycoprotein. *Biochemistry* 2000; **39**: 14138–14149.
- Loo TW, Clarke DM: Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein. *J Biol Chem* 2002; **277**: 44332–44338.
- Loo TW, Bartlett MC, Clarke DM: Processing mutations located throughout the human multidrug resistance P-glycoprotein disrupt interactions between the nucleotide binding domains. *J Biol Chem* 2004; **279**: 38395–38401.
- Loo TW, Clarke DM: Identification of residues in the drug-binding domain of human P-glycoprotein. *J Biol Chem* 1999; **274**: 35388–35392.
- Hrycyna CA, Airan LE, Germann UA, Ambudkar SV, Pastan I, Gottesman MM: Structural flexibility of the linker region of human P-glycoprotein permits ATP hydrolysis and drug transport. *Biochemistry* 1998; **37**: 13660–13673.
- Loo TW, Bartlett MC, Clarke DM: Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket. *Biochem J* 2006; **399**: 351–359.