Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 318 (2004) 38–45
BBRC
www.elsevier.com/locate/ybbrc

Characterization of a novel human gene containing ANK repeats and ARM domains

Gianluca Occhi, Massimiliano Olivieri, Alessandra Rampazzo, and Gian Antonio Danieli*

Department of Biology, University of Padua, Padua, Italy

Received 30 March 2004

Abstract

We report here on characterization of a novel human gene on chromosome 2q32, containing several ankyrin repeats and ARM domains. The gene FLJ25415, including at least 31 exons and spanning about 90 kb of a gene-rich genomic region, is present in the human genome as a single copy. It seems highly conserved along the mammalian phylogeny. By analysing FLJ25415 gene expression in different human tissues, we demonstrated a ubiquitously expression pattern with extensive alternative splicing. In silico prediction of promoter regions revealed the presence of a TATA box and some hypothetical transcription factor binding sites, such as TMF, C/EBPα, LE-1, and NF-ATp.

Keywords: Ankyrin repeat; ARM domain; Multiple alternative splicing; Heart; Eukaryotic polymerase II promoter regions

After publication of human genome finished sequence, many predicted genes are currently investigated in order to determine their potential relevance to clinical phenotypes. By using a combination of computer-based analysis and canonical molecular biology methods we attempted to characterize the FLJ25415 gene (GenBank Accession No. AJ549812), mapped to human chromosome 2q32.2, where arrhythmogenic right ventricular dysplasia-4 (ARVD4) [1], wrinkly skin syndrome [2], cleft palate [3], and mesomelic dysplasia [4] critical regions were previously assigned.

Preliminary bioinformatic analysis of putative aminoacidic structure of FLJ25415 protein revealed an unusual domain combination: ankyrin repeats (ANK) and armadillo/β-catenin-like repeats (ARM). Ankyrin repeats are 33-residue sequence motifs found in a large variety of proteins, which functions range from transcription initiation, cell cycle regulation, cytoskeletal integrity to cell–cell signaling [5]. Armadillo/β-catenin-like repeats are approximately 40 amino acid long tandemly repeated sequence motif identified in the *Drosophila* segment polarity gene armadillo, but also present in several mammalian proteins such as the junctional plaque protein plakoglobin and the adenomatous polyposis coli (APC) tumor suppressor protein [6].

In this paper, we report on the genomic structure and expression profile of the FLJ25415 gene, as well as on its transcription pattern in cardiac tissue. We also show open reading frames, significant similarity to mouse, rat or other human proteins, and hypothetical promoter elements, which were obtained by a computational extensive analysis performed on gene promoter region.

Materials and methods

Sequencing of IMAGE clones [7]. Human spliced ESTs were identified by scanning the region of interest at Human Genome Browser (July 2003 assembly, http://genome.ucsc.edu/) [8]. cDNA clones corresponding to selected ESTs were obtained from the MRC Geneservice of the UK Human Genome Mapping Project Resource Centre (HGMP-RC). DNA sequencing reactions were performed using the BIG DYE dideoxy-terminator chemistry (Perkin–Elmer) on an ABI 3700 DNA sequencer (PE Applied Biosystems). Chromas 1.5 software (Technelysium), LASERGENE package computer programs (DNASTAR) were used to edit and assemble sequences.

DNA and protein sequence analysis. To determine the genomic structure of FLJ25415, mRNA, and EST sequences retrieved from GenBank and sequenced IMAGE clones were compared with genomic sequences, by using the NCBI tool Spidey (http://

* Corresponding author. Fax: +39-049-827-6209.
E-mail address: danieli@bio.unipd.it (G.A. Danieli).

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.03.186

www.ncbi.nlm.nih.gov/ieb/research/ostell/spidey/) [9]. Open reading frames (ORF) were identified by the ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Orthologues of the human FLJ25415 gene were searched using BLAST algorithm (www.ncbi.nlm.nih.gov/BLAST) against mouse and rat RefRNA and EST databases. A BLAST search for related proteins was performed with the deduced amino acid sequence against the protein non-redundant database. A search for known protein sequence motifs was performed using SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de/) [10] and ScanProsite (http://au.expasy.org/tools/scanprosite/) [11], whereas secondary structure predictions were made by HNN (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html) [12]. The molecular weight was calculated with ProtParam tool (http://ca.expasy.org/tools/protparam.html). Transmembrane domains were predicted by TMHNN (http://www.cbs.dtu.dk/services/TMHMM/) and TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html) [13]. The prediction of protein sorting signals and of protein localization sites in cells was carried out with PSORT (http://psort.nibb.ac.jp/). Polyadenylation signals were determined using Polyadq software (http://rulai.cshl.org/tools/polyadq/polyadq_form.html) [14] with parameters set to the following: requested approximate specificity 50%, approximate expected sensitivity 97%, and approximate expected correlation coefficient 0.190. Comparisons of related aminoacidic sequences and the display of the resulting alignments were performed using the CLUSTAL W program (http://www.ebi.ac.uk/clustalw/index.html).

Expression. This method, which has already been described elsewhere [15], was based on the amplification on the human multiple tissue cDNA (MTC) Panel I (Clontech). The set of primer used is listed in Table 1.

Molecular characterization of FLJ25415 mRNA, cardiac isoform. PCR primers were designed based on EST sequences and Twinscan predictions [16] (see Table 1) in order to determine the sequence of the FLJ25415 cardiac isoform mRNA through RT-PCR. The RT-PCRs were performed as described elsewhere [15]. Amplified DNA fragments were excised and recovered from a 2% agarose gel using the QIAEX II gel extraction kit (QIAGEN), following the manufacturer's instructions. Purified fragments were sequenced as described above.

5′ and 3′-rapid amplification of cDNA ends. 5′ and 3′ RACE were performed using the Marathon Ready cDNA from human heart (Clontech) and Platinum *Taq* DNA polymerase High Fidelity (Invitrogen). Adaptor primer AP1 was used in the primary RACE reaction with 3′ and 5′ specific primer (all the gene-specific RACE primers are listed in Table 1). The primary reaction was carried out using 1 μl template cDNA with 0.2 μM of each gene-specific primer and AP1, 0.2 mM dNTPs, 1× High Fidelity PCR buffer, 2 mM magnesium sulphate, and 1 U Platinum *Taq* DNA polymerase High Fidelity (Invitrogen) in a final reaction volume of 15 μl. One microliter of the 1:100 diluted first round RACE reaction products, was used as template for secondary RACE reactions using the AP2 adaptor primer and nested specific primers. The PCRs were carried out using the following cycling conditions: initial denaturation at 94 °C for 1 min 30 s; five cycles of 30 s at 94 °C and 4 min at 72 °C; five cycles of 30 s at 94 °C and 4 min at 70 °C; and 23 cycles for the first round, and 15 for the second round, of 30 s at 94 °C and 4 min at 72 °C. The RACE products were cloned using the TOPO TA cloning kit (Invitrogen) according to the manufacturer's recommendations. The clones were sequenced as described above.

Promoter prediction. Several web-tools were used in order to identify the putative transcription starting site (TSS), gene promoter, and transcriptional factor binding sites, such as Promoter 2.0 (http://www.cbs.dtu.dk/services/Promoter/) [17], NNPP 2.2 (Neural Network Promoter Prediction; http://www.fruitfly.org/seq_tools/promoter.html) [18], Promoter Scan 1.7 (http://bimas.dcrt.nih.gov/molbio/proscan/), and TESS (maximum allowable string mismatch 0%, minimum string length 7 bp, factors filter: mammalia) (http://www.cbil.upenn.edu/tess).

Table 1
Primer sequences used for analysing the cardiac isoforms and the transcription pattern of the human FLJ25415 gene

Name	Sequence 5′ → 3′
ex4F	5′-aacagccctgaagaaacagc-3′
ex8R	5′-caacagcggacttcatacca-3′
ex8F	5′-tgtgcaatgctaacttcaagg-3′
ex10R	5′-ttttccacactgggagttca-3′
ex10F	5′-tcggtgctaactggagaaaaa-3′
ex17R	5′-tagtgtttgtcctgccaagg-3′
ex17F	5′-aaggagctgttgcactttgg-3′
ex24R	5′-gagccgttggattgttccta-3′
R5′ex17R	5′-gcccaaagtgcaacagctccttgt-3′
R5′ex4R	5′-cagctgtttcttcagggctgttgg-3′
R3′ex17F	5′-tgttaaggaacaaggagctgttgcactt-3′
R3′ex26F	5′-tctggagaagaatggaggaccatcc-3′
FLJ-24F	5′-agcaagcctggctcattcta-3′
FLJ-26R	5′-tggatggtcctccattcttc-3′

Results and discussion

Human FLJ25415 genomic organization

In the early phase of this study, by Human Genome Browser scanning, two clusters of overlapping spliced ESTs were identified on chromosome region 2q32.2. They were predicted to be part of a single transcript. This prediction was confirmed both by RT-PCR analysis with a primer pair designed to cover two predicted exons, and by sequencing one IMAGE cDNA clone for each cluster (AA634472, BG721498). One DNA sequence (BG721498) was already in GenBank, while the second one (AJ605761) was submitted by our group. Such sequences were used to perform a BLAST search against dbEST and nucleotide nr (non-redundant) database in order to check the genomic organization of the putative gene and to detect alternative splice variants. As shown in Fig. 1, several mRNAs and spliced ESTs were found. Sequence alignment revealed how alternative splicing generates different FLJ25415 gene products.

According to these data, FLJ25415 gene contains at least 31 exons, the last five of which serve as alternative last exon, and span about 91 kb of human chromosome 2 positive strand. The size of introns and exons and the sequence of the exon–intron junctions are reported in Table 2. All donor and acceptor sites follow the GT–AG rule [19]. Exon sizes range from the 47 bp of exon 26 to the hypothetical 3674 bp of exon 29. Introns are generally short, ranging from 93 bp of intron located between exons 4a and 4b, and 20,068 bp between exons 26 and 31, where the gene for *O*-sialoglycoprotein endopeptidase-like 1 (OSGEPL1) is located. Moreover OSGEPL1, which lies on the reverse strand, shares part of its sequence with the FLJ25415 gene, since exons 2 and 4 of the former partially overlap with exons 28 and 29 of the latter.

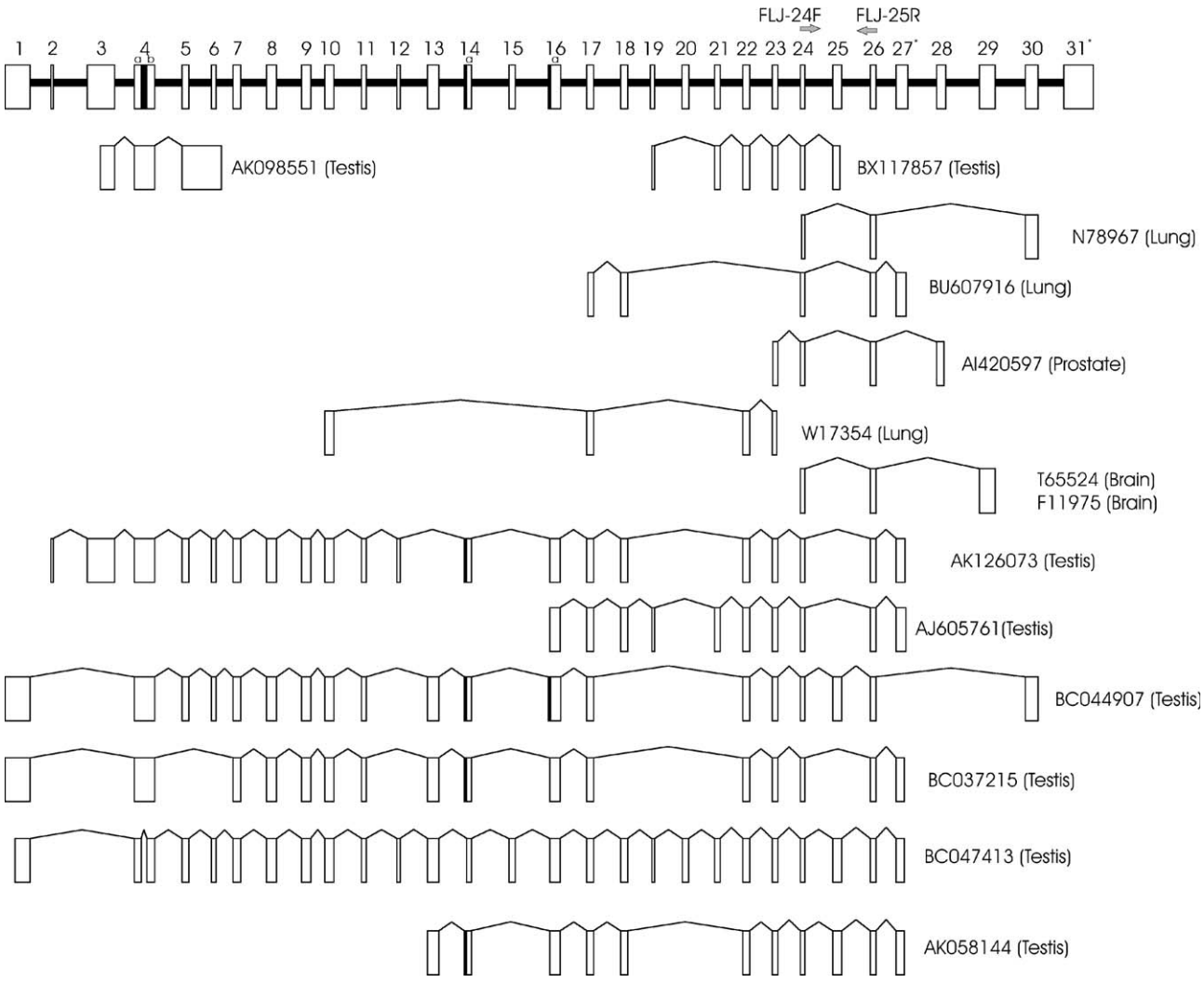

Fig. 1. Organization of human FLJ25415 gene. Boxes represent exons. Horizontal arrows indicate primers used for expression analysis. In the bottom part of the figure mRNA and ESTs belonging to FLJ25415 gene cluster. *Exons identified by 3′ RACE. In this figure, ESTs adding no information concerning alternative splicing were omitted.

Transcription of different messengers starts alternatively from two potential non-coding starting exons, defined as exons 1 and 2. Transcripts beginning on exon 1 spliced to exon 4, where their ATG-starting codon is located, while those starting on exon 2 are translated from exon 3.

Translation start site on exon 3 is located in an adequate context (GAA**ATG**T) since it conforms to the consensus sequence at least in one of the two key positions (RNNatgG, where R is a purine) [20]. On the contrary, the first ATG start codon on exon 4 lies in a weak context (TTT**ATG**A). Since in very few cases a weak ATG codon has been recognized as the true start site, we consider the second ATG occurring on this exon as the primary site of initiation (AT-T**ATG**C).

Exon 4, mostly used as single exon, seems to include an additional splicing site (based on mRNA BC047413), which generates two exons we refer to as 4a and 4b. The mRNA BC047413 is also the only transcript, which includes exons 12, 15, 20, and a shorter exon 14 we called 14a. Exon 16 as well contains two possible alternative 5′ splice sites, which generate a shorter exon we refer to as exon 16a (based on mRNA BC044907). As mentioned above, the FLJ25415 transcripts are extensively alternative spliced: just exons 4, 8, 10, 17, 24, and 26 seem to be used in every isoform (Fig. 1).

The polyadq program was used in order to identify polyadenylation signals (PAS) in the five last exons of the FLJ25415 human gene. For exons 27, 30, and 31, where a polyA tail was observed either by RACE or in the GenBank mRNA sequence, the identified 3′UTR sequence was submitted for PAS search. For exons 28 and 29, for which not enough information on their size was available, whole intron sequences (2143 and 5779 nucleotide, respectively) were looked up.

Table 2
Exon–intron junctions

Exon number	Exon size (bp)	3′ exon junction	Splicing donor	Intron size (bp)	Splicing acceptor	5′ exon junction	Exon number
1	541[a]	CCATTCTAGA	gtaggttcct	n/a	n/a	n/a	n/a
2	53[b]	CTGCAGAATT	gtgagccaaa	420	ttcattgcag	AAGCTTCAAA	3
3	636	AGTTCAGCAG	gtaagagaat	12,437	tatcatgaag	GTTTGACTGA	4
4a	166	AAATTCAACA	gtatgaaaat	93	tacctaccag	AGACTACAGC	4b
4b	181						
4	438	CCTTTTTCAG	gtaagagtat	2292	atttttggag	ATGACAAGGT	5
5	164	AGATTTTCAG	gtttaaatga	657	catattttag	AAAAATTTAG	6
6	104	ATGGAAAAAG	gtacggagct	1805	ttttttatag	CTACTATGTG	7
7	181	GAAATGCAAG	gtatttcagt	990	ttaattttag	AGTATTCCAT	8
8	220	TTCAGCCAAG	gtaccataaa	8655	tccattctag	GTCCAACACC	9
9	202	CAACAGCTGA	gtaagtcatt	1715	taatttttag	GAATCAGTGC	10
10	209	ACTTTGGTAG	gtgagtataa	3904	cacttaacag	AAATGTTACA	11
11	105	TTGGATGCAG	gtgatgcaaa	483	ccgttccaag	ATGGCCAAAT	12
12	83	TCCAACTGAG	gtacctgctt	7856	gtcattacag	GCACCATTCC	13
13	242	TGCCAAATAT	gtaagttcct	807	ttatttttag	AATGGAATCC	14
				898	ttgtgtatag	GAAATGAAAA	14a
14a	78						
14	169	TCTGATTCAG	gtgagcttct	5094	tgcattccag	CCTGGGTGAC	15
15	138	AGAGAATAAG	gtagggctct	1807	ttttcataag	GTTTATTTAA	16
				1840	tcatttgaag	ATGTGTTGAA	16a
16a	242						
16	275	ACTCCTAAAG	gtaggaattt	171	ttaaaaatag	GCATTTCAAA	17
17	154	CAGTATGTTG	gtaagttatt	241	tctttatcag	GAGGTGAAGC	18
18	162	ATTTGTATTG	gtttgtatac	1675	actcttacag	GTGTAGCCCA	19
19	107	TAACATTAAG	gtataaaggt	339	acctccccag	ACATCAGTTG	20
20	139	TTGGCTGGAG	gtaatagttt	2031	acatgtttag	GTTGAAGTGG	21
21	123	AACAGAAAAG	gtaatacctt	4455	ctatttttag	GATATTTGCT	22
22	159	AGCATTTCAG	gtataaaatt	725	ctgttttcag	ATTGTTGTAC	23
23	118	GTCTTGACAG	gtaagaaatg	2661	taatttatag	GGAATTTAAT	24
24	110	AATAGAAGAG	gtaaaaacaa	1825	cttcatttag	GTTCGTGCAG	25
25	200	CTTCCTTAAG	gtatggtcct	1269	ttccttttag	TCTGGAGAAG	26
26	47	TTTAAAAGAG	gtaattgatt	1593	taatttttag	GGAAGGAGCA	27
				8148	aaatttccag	AGCTCTGCGG	28
				10,395	ccttgtcaag	CATGTCACCT	29
				16,311	tgactaacag	TTTTCCTAGG	30
				20,068	tttttttag	ATGGAGTCTC	31
27	265						
28	462[c]						
29	3674[c]						
30	94						
31	698						

[a] The size refers to the mRNA BC044907.
[b] The size refers to the mRNA AK126073.
[c] The size is calculated up to the more reliable polyadenylation signal predicted by polyadq software.

Exons 27, 30, and 31 contain one canonical PAS (AAUAAA), whereas exons 28 and 29 show several positive signals, the most likely of which are shown in Table 3. The average size of the 3′ UTRs is about 1027 bp.

Depending on the considered isoform, the stop codon could be located either on one of the last five exons, exon 14 or 22.

Table 3
Polyadenylation signals in the terminal exons of human FLJ25415 gene with the highest score, identified by polyadq software

Exon	PolyA signal	Score	Site[a]
27	AAUAAA	0.131995	247
28	AAUAAA	0.114355	463
29	AAUAAA	0.231403	3675
30	AAUAAA	0.390224	76
31	AAUAAA	0.075051	676

[a] The site position is calculated from to the first nucleotide of each terminal exon.

Expression pattern of human FLJ25415 gene

We examined the tissue mRNA expression of FLJ25415 gene by RT-PCR on Human Multiple cDNA Tissue panel I using a set of primers (FLJ-24F, FLJ-26R), which allow the detection of all so far described isoforms (Fig. 1). A representative amplification after 38 cycles is shown in Fig. 2A.

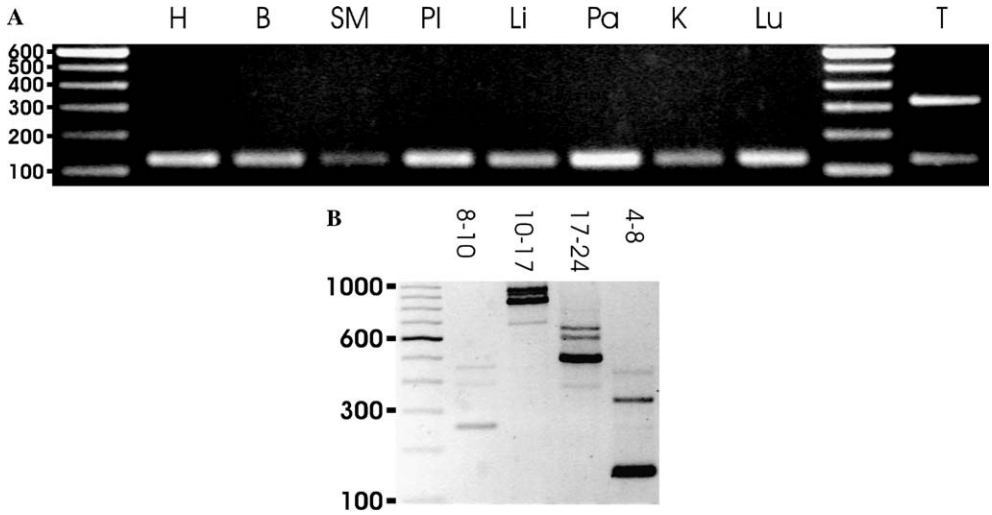

Fig. 2. (A) Expression pattern of FLJ25415 gene lines from 2 to 9 represents the RT-PCR of human cDNA multiple-tissue panel I, whereas line 11 represents the amplification using testis cDNA. (B) Evidence of alternative splicing in the heart. The numbers above each line represent the exons on which the primers used in the RT-PCR are located. (H, heart; B, brain; SM, skeletal muscle; Pl: placenta; Li, liver, Pa, pancreas; K, kidney; Lu, lung; and T, testis.)

FLJ25415 gene was ubiquitously expressed with the highest level in pancreas and the lowest in skeletal muscle.

Exon 25 lacks in each of the 8 tissues considered, suggesting a tissue specific expression. BLAST analysis revealed the presence of exon 25 only in testis libraries, that was confirmed by RT-PCR analysis on testis cDNA (Fig. 2A).

Identification of FLJ25415 cardiac isoform

To verify the possible existence of other splicing variants several RT-PCRs were carried out on heart cDNA using primers located within two regions of amplification, in which we divided the FLJ25415 gene: the N-terminal domain ranging from exon 1 to 17, and the C-terminal domain from exon 18 to 31. DNA sequencing of the RT-PCR products confirms an extensive alternative splicing, which appears more diffused in the 5′ end of the gene (Fig. 2B).

To obtain sequence information of the 5′ and 3′ ends of the transcript, two rounds of RACE were performed and the products were cloned. No specific 5′ sequences were obtained, whereas in the 3′ end two alternative products were found, which correspond to exons 27 and 31.

Only 18 out of the 31 exons characterizing the FLJ25415 transcripts have been detected in the cardiac tissue. They are exons 4, 6, 7, 8, 9, 10, 11, 13, 14, 16a, 17, 18, 22, 23, 24, 26, 27, and 31. However, due to the complexity of the splicing pattern it is possible that exons used less frequently, were more difficult to detect. No heart specific exons were found.

Human FLJ25415 predicted protein

Since the complex transcription pattern of FLJ25415 two main transcripts differing in the starting exons (exons 1 and 2, respectively) were considered. The more 3′ portion was identical in both cases, with exons 5–11, 13, 14, 16a, 17–19, and 21–27. The transcripts are, respectively, 3576 and 4305 bp long, encoding for a 1191 aminoacid (1P) and 1434 aminoacid (2P) protein with a molecular weight of 134 and 162 kDa, respectively.

HNN showed a similar secondary structure organization between the two submitted proteins with about 50% of α-helices, 35% coiled-coil, and 15% β-sheet. The SMART analysis revealed five ankyrin repeats (ANK) and seven armadillo/β-catenin-like repeats (ARM) common to both isoforms and a single transmembrane domain (from residue 69 to 86) typical of the 1P product. TMHNN suggested for 1P a short N-terminal tail followed by a long extracellular portion. The same analysis performed with TMPRED suggested 5 and 6 strong transmembrane helices for 1P and 2P, respectively, indicating FLJ25415 products as membrane proteins. Besides the domains described above ScanProsite found in both sequences also a eukaryotic RNA recognition motif (RRM) RNP-1.

The ankyrin repeat takes its name from one of the proteins in which it was first found, the human erythrocyte protein ankyrin [21]. ANK are present in a variety of proteins of various structures, functions, and subcellular distribution and have been implicated in many protein–protein interactions. These proteins include: cytoskeletal proteins, toxins, transcription factors, channels, enzymes, development regulators, cell cycle control proteins, and others [22].

Armadillo repeats [6] were first identified in the *Drosophila* segment polarity gene [23]. Subsequent studies revealed the presence of ARM in proteins with diverse cellular functions, such as the junctional plaque protein plakoglobin, β-catenin, the tumour suppressor adenomatous polyposis coli, and the nuclear transport factor importin-α. Analogous to ANK, armadillo repeats mediate specific protein–protein interactions.

PSORT analysis identified a peroxisomal targeting signal at position 228 for 1P and an endoplasmic reticulum membrane retention signal in the N-terminus for 2P, besides a transport motif from cell surface to Golgi, common to both isoforms (position 42 and 285, respectively). This leads to the conclusion that 1P is likely a plasma membrane protein (47.8%), although an endoplasmic reticulum localization could not be ruled out (30.4%). On the contrary, 2P is more likely an endoplasmic reticulum protein (44.4%) rather than a plasma membrane one (22.2%).

Computational identification of the promoter regions of the human FLJ25415 genes

The 5′-upstream 2 kb sequences of the exons 1 and 2 were analysed by an in silico approach in order to find potential transcription starting sites (TSS) and regions with promoter activity. The sequences were submitted to Promoter Scan 1.7, Promoter 2.0, and Neural Network Promoter Prediction.

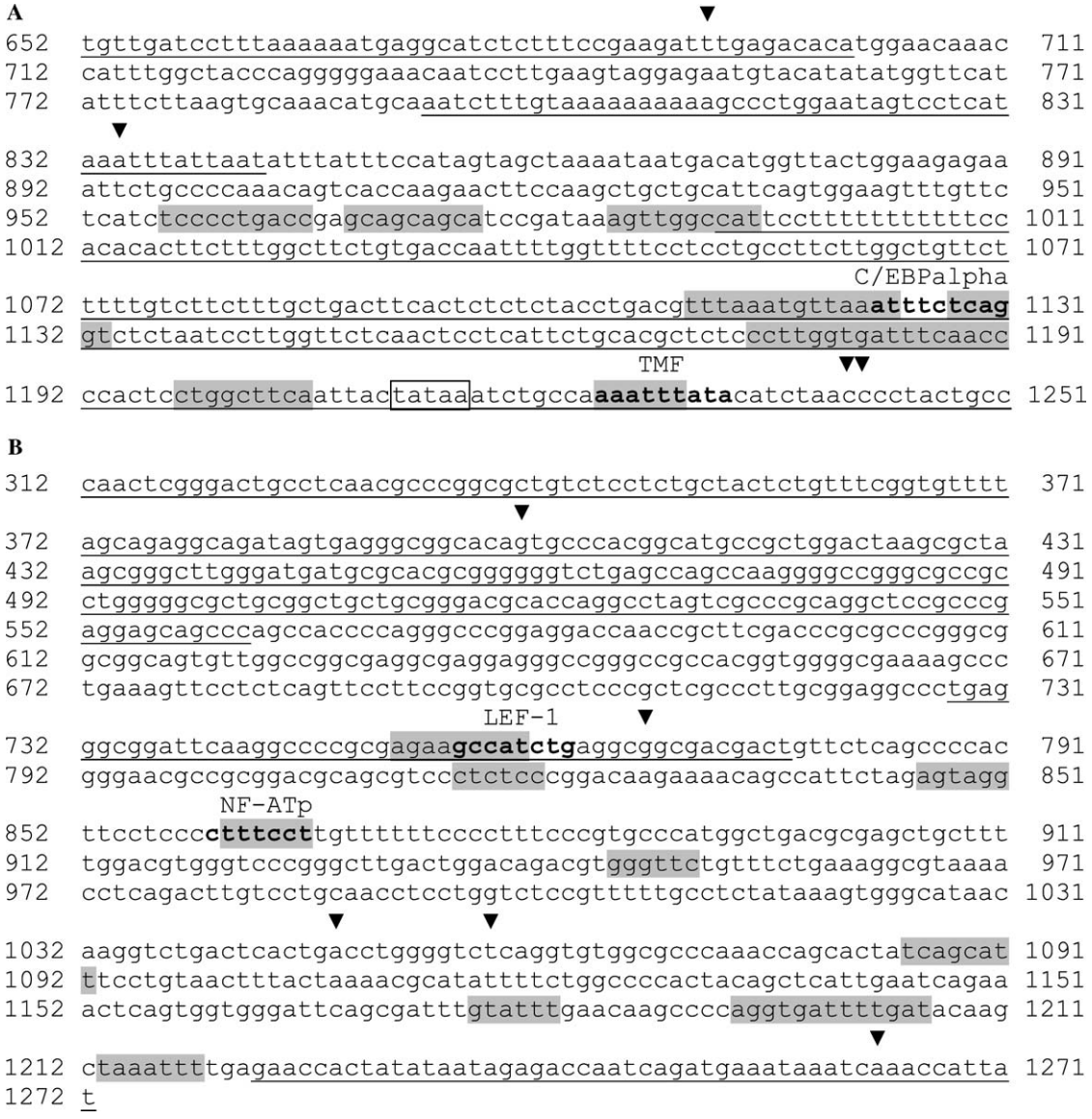

Fig. 3. Analysis of the regulatory region of FLJ25415 gene. The regions upstream to exon 1 and 2 are shown, respectively, in A and B. Predicted promotorial regions are underlined. Regions conserved between human and rodents are shadowed. Bold-faced letters indicate TF binding sites. Triangles point to putative transcription start sites. TATA-box is boxed.

In exon 1a hypothetical TATA-box was identified at position 1212 about 30 bp upstream two contiguous potential TSS (see Fig. 3A). Different potential promoter regions were identified, none of them outside the interval between nucleotides 652 and 1251. Therefore, this region was submitted to TESS and Transfac to detect mammalian transcription factors (TFs). Due to the large number of false positive, even if an high stringency were used, we decided to consider only TFs binding DNA fragments for which we hypothesized a selective constraint among human, mouse, and rat (when the sequence is available). Two elements known as C/EBPα and TMF (TATA element modulatory factor) were identified.

Exon 2 (see Fig. 3B) lacks a typical TATA-box, while several TSS have been found. LEF-1 (lymphoid enhancer binding factor 1) and NF-ATp binding sites were identified, respectively, at position 756 and 860.

Mouse and rat homologous

The sequence of the human FLJ25415 gene was used as query to perform a BLAST search against mouse and rat RefRNA and EST databases, in order to identify rodent homologous genes. Mouse RefSeq sequence NM_176980, based on mRNA AK030065, contains exons corresponding to 3–11, 13, 14, 16a, 17–19, and 21–27 of the human homologue, whereas NM_028091 (derived from AK045669), which represents *O*-sialoglycoprotein endopeptidase-like 1 matches, as expected, just with exons 28 and 29. The same analysis carried out for *Rattus norvegicus* reveals a single hit for the putative sialoglycoprotease type 2-gene (XM_234772). A more accurate surveying of the rat and mouse genome assembly (UCSC version rn3 and mm4) allows one to hypothesize that also in the rodents genome FLJ25415 and *O*-sialoglycoprotein endopeptidase-like gene have a comparable chromosomal organization.

Alignment of the predicted human FLJ25415 protein product and the mouse orthologues NP_795954 demonstrates an identity of 81%, followed by an almost identical domain composition: except the transmembrane domain, which is absent in the mouse protein, all the repeats characterizing the human polypeptide were also detected in the murine one. The protein alignment shows the presence of 13 additional amino acids between exon 8 and exon 9 in the murine sequence. Thus, we hypothesize this new exon might be present also in the human transcripts. Further analysis in the human genomic sequence demonstrates the high level of conservation between the human and the murine sequence at this level except for the lacking of a canonical splice acceptor site in the human sequence, which could explain the reason why this exon would have never been found in human transcripts (see Fig. 4).

```
Mouse   AGCAGACTCATCCAAAGTGGATATGAAAAAAGAAAGAAATGGT 43
Human   AACAGAGTCATCCAAAGTAGAGGTAAAAAAGGAAAGAAATGGT 43
        * **** ********** **  * ***** **********  **
```

Fig. 4. Alignment between additional exon identified in mouse transcript NM_176980 and human genomic sequence. The acceptor and donor splicing sites are shaded.

In addition nucleic and protein sequences of FLJ25415 gene showed no similarities with other organism sequences available so far, suggesting an evolution of this gene in the mammalian lineage.

Conclusion

On the basis of the so far knowledge on FLJ25415 gene we can only speculate on its possible function. FLJ25415 gene shows a complex pattern of alternative splicing, with several exons contributing to a large number of possible isoforms. Alternative splicing may be most important in complex systems where information must be processed differently at different times (such as immune tolerance or development) or very high level of diversity is required (such as axonal guidance). A high percentage of spliced genes is involved in signalling and regulation (such as receptors, signal transduction, transcription factors, and so on). Interesting examples of how alternative splicing generates high complexity have recently been discovered in the nervous system, including Dscam (axonal guidance receptor in *Drosophila*) and neurexins (neuropeptide receptors) [24]. Moreover, we showed that FLJ25415 is potentially located on the plasma membrane and it contains several domains involved in protein–protein interactions, characteristic of a protein which can work as transmembrane receptor, the stimulation of which can lead to the activation or silencing of a specific biological pathway.

Acknowledgments

This publication is based on the project performed pursuant to Baylor College of Medicine Grant 1 U01 HL 65652 from the National Institutes of Health and to ARVC EC contract QLG1-CT-2000-01091.

References

[1] A. Rampazzo, A. Nava, M. Miorin, P. Fonderico, B. Pope, N. Tiso, B. Livolsi, R. Zimbello, G. Thiene, G.A. Danieli, ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy, maps to chromosome 2 long arm, Genomics 45 (1997) 259–263.

[2] F.R. Kreuz, B.H. Wittwer, Del(2q)—cause of the wrinkly skin syndrome? Clin. Genet. 43 (1993) 132–138.

[3] C.M. Brewer, J.P. Leek, A.J. Green, S. Holloway, D.T. Bonthron, A.F. Markham, D.R. FitzPatrick, A locus for isolated, cleft palate located on human chromosome2q32, Am. J. Hum. Genet. 65 (1999) 387–396.

[4] M. Fujimoto, P.N. Kantaputra, S. Ikegawa, Y. Fukushima, S. Sonta, M. Matsuo, T. Ishida, T. Matsumoto, S. Kondo, H. Tomita, H.X. Deng, M. D'urso, M.M. Rinaldi, V. Ventruto, T. Takagi, Y. Nakamura, N. Niikawa, The gene for mesomelic dysplasia Kantaputra type is mapped to chromosome 2q24-q32, J. Hum. Genet. 43 (1998) 32–36.
[5] P. Bork, Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally?, Proteins 17 (4) (1993) 363–374.
[6] M. Peifer, S. Berg, A.B. Reynolds, A repeating amino acid motif shared by proteins with diverse cellular roles, Cell 76 (5) (1994) 789–791.
[7] G. Lennon, C. Auffray, M. Polymeropoulos, M.B. Soares, The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression, Genomics 33 (1) (1996) 151–152.
[8] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, D. Haussler, The Human Genome Browser at UCSC, Genome Res. 12 (6) (2002) 996–1006.
[9] S.J. Wheelan, D.M. Church, J.M. Ostell, Spidey: a tool for mRNA-to-genomic alignments, Genome Res. 11 (11) (2001) 1952–1957.
[10] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains, Proc. Natl. Acad. Sci. 95 (1998) 5857–5864.
[11] A. Gattiker, E. Gasteiger, A. Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, Appl. Bioinform. 1 (2002) 107–108.
[12] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, NPS@: Network Protein Sequence Analysis. TIBS, 2000 March Vol. 25, No. 3 [291]:147–150.
[13] K. Hofmann, W. Stoffel, TMbase—A database of membrane spanning proteins segments, Biol. Chem. Hoppe-Seyler 374 (1993) 166.
[14] J.E. Tabaska, M.Q. Zhang, Detection of polyadenylation signals in human DNA sequences, Gene 231 (1–2) (1999) 77–86.
[15] G. Occhi, A. Rampazzo, G. Beffagna, G.A. Danieli, Identification and characterization of heart-specific splicing of human Neurexin 3 mRNA (NRXN3), Biochem. Biophys. Res. Commun. 298 (1) (2002) 151–155.
[16] I. Korf, P. Flicek, D. Duan, M.R. Brent, Integrating genomic homology into gene structure prediction, Bioinformatics 17 (2001) S140–S148.
[17] S. Knudsen, Promoter 2.0: for the recognition of PolII promoter sequences, Bioinformatics 15 (1999) 356–361.
[18] M.G. Reese, Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome, Comput. Chem. 26 (1) (2001) 51–56.
[19] M. Burset, I.A. Seledtsov, V.V. Solovyev, Analysis of canonical and non-canonical splice sites in mammalian genomes, Nucleic Acids Res. 28 (21) (2000) 4364–4375.
[20] M. Kozak, Point mutation define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, Cell 44 (1986) 283–292.
[21] S.E. Lux, K.M. John, V. Bennett, Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins, Nature 344 (6261) (1990) 36–42.
[22] A. Hryniewicz-Jankowska, A. Czogalla, E. Bok, A.F. Sikorsk, Ankyrins, multifunctional proteins involved in many cellular pathways, Folia Histochem. Cytobiol. 40 (3) (2002) 239–249.
[23] B. Riggleman, E. Wieschaus, P. Schedl, Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a *Drosophila* segment polarity gene, Genes Dev. 3 (1) (1989) 96–113.
[24] B.R. Graveley, Alternative splicing: increasing diversity in the proteomic world, Trends Genet. 17 (2) (2001) 100–107.