FEBS Letters 444 (1999) 38–42 FEBS 21500

Complete exon-intron organization of the mouse fibulin-1 gene and its comparison with the human fibulin-1 gene

Te-Cheng Pan[a], Günter Kostka[c], Rui-Zhu Zhang[a], Rupert Timpl[b], Mon-Li Chu[a,b,*]

[a] *Department of Dermatology and Cutaneous Biology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, 233 South 10th Street, Philadelphia, PA 19107, USA*
[b] *Department of Biochemistry and Molecular Pharmacology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA 19107, USA*
[c] *Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany*

Received 20 November 1998

Abstract Fibulin-1 is a 90 kDa calcium-binding protein present in the extracellular matrix and in the blood. Two major variants, C and D, differ in their C-termini as well as the ability to bind the basement membrane protein nidogen. Here we characterized genomic clones encoding the mouse fibulin-1 gene, which contains 18 exons spanning at least 75 kb of DNA. The two variants are generated by alternative splicing of exons in the 3′ end. By searching the database we identified most of the exons encoding the human fibulin-1 gene and showed that its exon-intron organization is similar to that of the mouse gene.

Key words: Fibulin-1; Gene structure; Extracellular matrix; Alternative splicing

1. Introduction

Fibulin-1 is a 90 kDa calcium-binding protein found in the extracellular matrix and in the blood plasma [1,2]. Cloning of cDNAs has revealed at least four variant fibulin-1 polypeptides in humans, designated A–D, and two major variants, C and D, in mice and zebrafish [3–6]. These variants differ in the C-terminal domain III but share a common core structure consisting of three anaphylatoxin-like repeats in the N-terminal domain I, and nine consecutive epidermal growth factor-like modules (EG modules) in domain II, eight of which possess a consensus sequence for calcium binding (Fig. 1A). A homologous protein, fibulin-2, contains domains I, II, and III similar to fibulin-1C and an additional globular domain at the N-terminus [7]. The extracellular protein, S1-5, recently was suggested as the third member of the fibulin family, even though it does not possess domain I [6].

Fibulin-1 expression is widespread in the basement membrane and stroma of most organs [4,7,8]. It colocalizes with elastic microfibrils in the skin and with fibronectin fibrils deposited by cultured cells [8,9]. The biochemical evidence that fibulin-1 binds fibronectin and nidogen [10,11], the major components of basement membranes and interstitial connective tissues, is consistent with the expression pattern. The C and D variants differ in the ability to bind the basement membrane protein nidogen [11]. The expression pattern of fibulin-1 during embryonic development suggests a critical role in organogenesis, particularly in the development of cardiac septa and valves [12–14]. As a blood protein, fibulin-1 binds fibrinogen and may serve a function in hemostasis and thrombosis [15,16]. Moreover, fibulin-1 appears to be involved in tumor formation and invasion [17–19].

The fibulin-1 gene maps to human chromosome 22q13 and mouse chromosome 15, band E–F [20], but its genomic structure has not been reported to date. In this study, we isolated genomic clones containing the entire coding region of the mouse fibulin-1 gene, determined the exon-intron organization of the gene, and showed that the C and D variants are encoded by alternative exons in the 3′ end of the gene. We also identified most of the corresponding exons for the human fibulin-1 gene and alternative exons for the A and B variants by searching the genome database. Comparison of the mouse and human genes showed a high degree of conservation in the exon-intron organization.

2. Materials and methods

2.1. Isolation of genomic clones

Genomic clones were isolated from a λFIXII phage genomic library constructed from DNA of mouse strain 129 (Stratagene, CA) and a cosmid library constructed from D3 embryonic stem cells (a gift from John S. Mudgett and Reinhard Fässler). The libraries were screened with [^{32}P]dCTP labeled mouse fibulin-1C and D cDNAs [4] using standard methods [21]. Positive clones were characterized by restriction enzyme mapping. Exons were localized by Southern blotting using cDNA fragments encoding different regions of fibulin-1 as probes.

2.2. DNA sequencing and sequence analysis

Exon-containing fragments from the phage clones were subcloned into pBluescript vector and subjected to DNA sequencing, while the cosmid clone was used directly as a template for sequencing without further subcloning. DNA sequencing was performed by the dideoxy-chain termination method using [^{33}P]dATP and the Sequenase kit (Amersham), or by the cycle sequencing method with Taq polymerase and fluorescence labeled dideoxynucleotides on an automatic sequencer (Applied Biosystems). Sequencing was performed with T3 and T7 primers or with primers derived from the cDNA sequence. DNA sequences were analyzed using GCG software (Genetics Computer Group, Madison, WI). Database comparisons were performed using BLAST computer program [22].

2.3. Polymerase chain reaction (PCR)

The intron sizes were determined by restriction enzyme mapping, DNA sequencing, or PCR amplification of genomic clones with primer pairs from two adjacent exons. PCR was performed using AmpliTaq according to the protocols provided by the manufacturer (Perkin-Elmer). The conditions were an initial denaturation at 94°C for 2 min, followed by 30 cycles of 94°C for 45 s, annealing at 60°C for 45 s, and extension at 72°C for 1 min, and then a final extension at 72°C for 10 min. For longer PCR, amplification was carried out with rTth DNA polymerase and AmpliWax beads (Perkin-Elmer) as sug-

*Corresponding author. Fax: (1) (215) 503-5788.
E-mail: mon-li.chu@mail.tju.edu

Abbreviations: EG module, epidermal growth factor-like module; PCR, polymerase chain reaction

PII: S0014-5793(99)00024-1

gested by the manufacturer. Amplification conditions were an initial denaturation at 94°C for 2 min, followed by 30 cycles of 94°C for 45 s and 72°C for 2–4 min, and then a final extension of 72°C for 10 min. The PCR products were analyzed by agarose gel electrophoresis.

3. Results and discussion

3.1. Mouse fibulin-1 gene

A total of six phage clones were isolated by screening 3×10^6 clones from the mouse λFIXII genomic library with the full-length fibulin-1 cDNA, a 0.6 kb *Eco*RI fragment from the 5′ end of the cDNA, and two fragments specific for the C and D variants in three separate experiments. Southern blot analysis showed that these clones contained the entire coding region of fibulin-1 except for part of the D variant-specific sequences (Fig. 1B). The genomic cosmid library constructed from D3 mouse embryonic stem cells was then screened with the cDNA probe specific for the D variant. The cosmid clone C10 was isolated and was found to contain the coding regions for the C-termini of both C and D variants (Fig. 1B).

A comparison between the full-length cDNA and the genomic sequence showed that the gene consists of 18 exons (Fig. 1B, Table 1). Exon 1 corresponds with the 5′ end of the gene since its 5′-flanking sequence confers promoter activity (Castoldi and Chu, in preparation). Southern blotting and restriction mapping showed that clones P9 and P13 did not overlap, suggesting that intron 1 is larger than 8 kb in size. This is supported by the finding that long-range PCR of mouse genomic DNA using primers in exons 1 and 2 failed to amplify a specific fragment. Thus, the gene spans more than 75 kb of genomic DNA with an exon/intron ratio of over 1/25. All introns begin with GT and end with AG, conforming to the consensus sequences of the splice donor and the acceptor sites (Table 1). Almost all introns disrupted condons.

The first exon contains the 5′-untranslated region, the

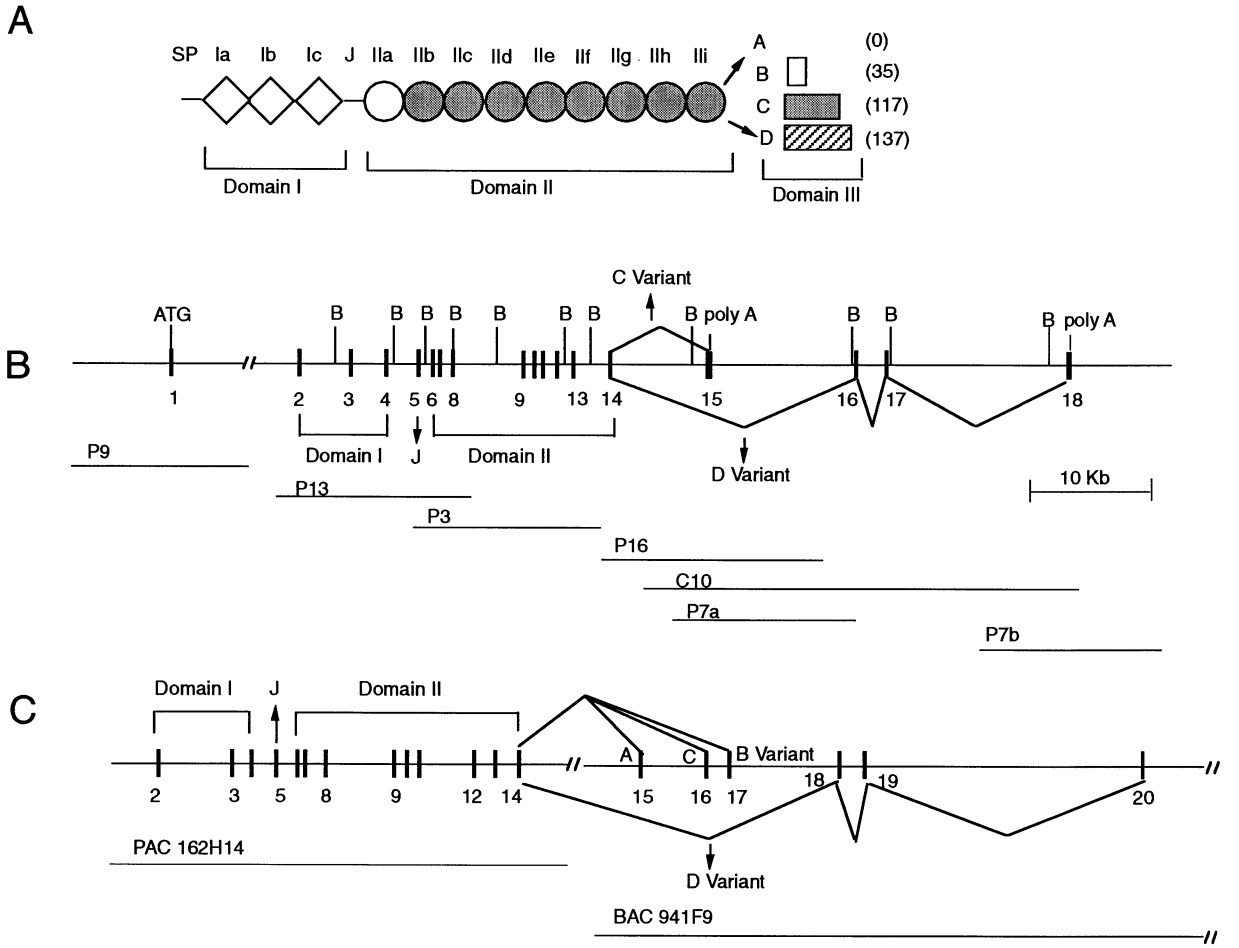

Fig. 1. A: Schematic diagram of the fibulin-1 variants, A–D. The A and B forms are less abundant and are found only in humans. The common region is composed of domains I and II connected by a short junctional segment (J). Domain I consists of three anaphylatoxin-like motifs (Ia–Ic), and domain II contains nine EG modules (IIa–IIi), of which IIb–IIi possess a consensus sequence for calcium binding. The A variant contains only the common sequence, while domains III of the B, C, and D variants are 35, 117 and 137 amino acids in size. SP: signal peptide. B: Genomic organization of the mouse fibulin-1 gene. The positions of the 18 exons (filled boxes), the *Bam*HI restriction sites (B), the cosmid (C10) and phage clones (P9, P13, P3, P16, P7a and P7b) covering the genomic region are shown. Exon 1 encodes the SP; exons 2–4 encode the three anaphylatoxin-like motifs (Ia–Ic) in domain I; exons 6–14 encode the nine EG modules (IIa–IIi) in domain II. Exon 15 is alternatively spliced and encodes domain III of the C variant plus its 3′-untranslated sequence, while exons 16–18 encode domain III of the D variant. The positions of the translation start site (ATG) and the polyadenylation sites (poly A) are indicated. C: Genomic organization of the human fibulin-1 gene (FBLN1) according to DNA sequences of a PAC clone (162H14) and a BAC clone (941F9) deposited in the database (accession numbers Z98047 and 95331). Note that two additional and alternatively spliced exons encode human fibulin-1 A and B variants.

Table 1
Intron-exon junction sequences of the mouse fibulin-1 gene

Exon No.	Intron/Exon Junctions	Exon Size (bp)	Exon/Intron Junctions	Intron size (kb)	Protein Domain
1[a]	cccgcccccgccgcg	CCTCCTCCGGG	...(202)..GCC CGA G A R	>8	5′-UT + SP
2	ctcttccctacacag	CG AAT GCA	...(106)..GAG TGC AG (A) N A E C (R)	4.5	Ia
3	ctgctctgcccacag	G ATG GTC	....(138).....ATA AAG M V I K	3.6	Ib
4	tctttctgttcatag	AGG TGC TGC	..(170)...GAC CCA G R C C D P	2.8	Ic
5	actctctacccctag	CT AAG ATT	....(57)...TGT CGA G (A) K I C R	1.6	Junction
6	ggactctggttgcag	GT GGC GGG	...(102)...TGC GAA G (G) G G C E	0.8	IIa
7	ccttaatctctctag	AT ATC AAT	...(138)...TGC AAA G (D) I N C K	1.1	IIb
8	tttacgatgcaccag	AT ATT GAC	...(138)...TGC ATT G (D) I D C I	6.5	IIc
9	ctcttttctctgtag	AT ATC AAT	...(144)...TGT GTT G (D) I N C V	1.0	IId
10	tttcttcttcacaag	AT GTG GAT	...(129)...TGC GTG G (D) V D C V	0.8	IIe
11	cccctctgttttcag	AT ATC AAC	...(126)...TGT GAA G (D) I N C E	1.4	IIf
12	cctcttctcttctag	AT GTG AAC	...(120)...TGC GAA G (D) V N C E	1.3	IIg
13	ctggggctgttgcag	AT ATT GAT	...(132)...TGC CAA G (D) I D C Q	2.3	IIh
14	gccgcaactgccaag	AC ATT GAT	...(124)..GCA GAC AC (D) I D A D (T)	8.0	IIi
15[b]	ccacttgtctttcag	C CGC TGT	....(440)..TCCCAAGAGC R C	atgcctcttgtacct ~12	III + 3′-UT (C variant)
16	tgtgtttctctgcag	C TTC CGC	....(143)...CCT GAG G F R P E	gtgagtggatcaagt 2.2	III (D variant)
17	tcatccttgctgcag	AG ATC ATC	...(131)...ACT GTG G (E) I I T V	gtgagtgacccttga ~14	III (D variant)
18[b]	ctgctctctccccag	GT GTC GTG	...(604)....AATTCTCC (G) V V	ccaaaccaactgcca	III + 3′- UT (D variant)

Capital letters represent exon sequences; a space is inserted between codons and the amino acids encoded are shown underneath. Lowercase letters represent intron sequences. UT: untranslated sequences; SP: signal peptide.
[a]The 5′ end of exon 1 corresponds with the transcription start site.
[b]The 3′ ends of exons 15 and 18 correspond with the polyadenylation sites.

translation start site and the coding sequence for the signal peptide. Exons 2, 3, and 4 encode the three anaphylatoxin-like repeats in domain I. The exon-intron junctions, however, do not correspond with the boundaries of the anaphylatoxin-like repeats. Exon 5 encodes a short junctional region connecting domain I and domain II. Exon 6 encodes the first EG module but the first of the six cysteines in this motif is encoded by the preceding exon. Exons 7–14 encode the eight calcium-binding EG modules in domain II, and these exons begin at the second nucleotide of a codon. Each exon defines a single EG module with six cysteines just as the genes for fibrillin-1 and S1-5, which also contain tandem arrays of EG modules [23,24]. Exon 15 encodes domain III of the C variant and its 3′-untranslated region, while domain III of the D variant is divided into three exons (16–18), the last of which contains its 3′-untranslated sequence. Overall, there is a good correlation between the exon organization and the three structural domains. However, in the 5′ end of the gene the splice junctions appear to interrupt protein motifs that likely will prove to be independently folded structures. For instance, the three introns in domain I occur at different positions in the three consecutive anaphylatoxin motifs. This could be an evolutionary mechanism to prevent crossing over that leads to different numbers of the repetitive motifs.

3.2. *Human fibulin-1 gene*

The genomic sequence for most of the human fibulin-1 gene recently was deposited in the database by the Sanger Center

Table 2
Intron-exon junction sequences of the human fibulin-1 gene*

Exon No.	Intron/Exon Junctions	Exon Size (bp)	Exon/Intron Junctions	Intron size (kb)	Protein Domain
2	ctcttccctacacag TG GAC GCG	(106)	GAA TGC AG gtacgtttgccagtg	6.8	Ia
	(V) D A		E C (R)		
3	acccctcacccacag G ATG GTG	(136)	GTG AAG gtgagagccaaagac	2.2	Ib
	M V		V K		
4	cacctgtgtttgcag AGG TGC TGC	(163)	GAA ACG G gtaactttccccctt	3.3	Ic
	R C C		E T		
5	ctttttcccccttag AT AAG ATC	(50)	TGC CGA G gtgagactcgggcgt	1.7	Junction
	(D) K I		C R		
6	acgctgtgcttccag GA GGC GGG	(102)	TGT GAA G gtaatgtccctatcc	0.6	IIa
	(G) G G		C E		
7	ctcggtctctcctag AT GTC AAT	(138)	TGC AAA G gtacagcatgcgctc	1.3	IIb
	(D) I N		C K		
8	tttatgatgtaccag AT ATT GAC	(138)	TGT ATT G gtaagaggtgtgccg	5.9	IIc
	(D) I D		C I		
9	cttttcccgctgtag AT ATC AAT	(144)	TGT GTT G gttggtattaagaaa	0.8	IId
	(D) V N		C V		
10	tttctcctttgcaag AT GTG GAC	(129)	TGT GTC G gtgcgtgggggggccc	1.1	IIe
	(D) V D		C V		
11	acccctcactttcag AT GTC AAC	(126)	TGT GAA G gtgaggctggggccc	3.6	IIf
	(D) V N		C E		
12	cgttttgtatttcag AC ATC AAT	(120)	TGT GAA G gtgcggacgcccctg	1.4	IIg
	(D) I N		C E		
13	gctttgccgttgcag AC ATC GAC	(132)	TGC CAA G gtgagcaggagggat	1.7	IIh
	(D) I D		C Q		
14	tggggtctcttgcag AC ATT GAT	(124)	GCA GCC AC gtaagtcccttggac	NK	IIi
	(D) I D		A A (T)		
15[a]	ctcttcctctgtcag ATGATCGT	(641)	TAGGCCCA ctaggcgttgtgtct	4.8	3′-UT (A variant)
16	ccacttttcttgcag C CGC TGT	(449)	CCACACAGT gagcctcgcgtgcct	1.5	III + 3′-UT (C variant)
	R C				
17[a]	tgacactgtttccag G CAG AAA	(818)	GGTTGATGG atggatggacagacc	9.6	III_B + 3′-UT (B variant)
	Q K				
18	tgtggttcccctcag G CTC CAG	(143)	CCT GAA G gtgagtgggatgggt	2.3	III (D variant)
	L Q		P E		
19	tctgtgcctctgcag AG ATC ATC	(131)	ACC GTG G gtgagtggctgggaa	23.2	III (D variant)
	(E) I I		T V		
20[a]	ctgctctctccgcag GT GTC GTG	(812)	AATAAACAA ctttgtgatcctcct		III + 3′-UT (D variant)
	(G) V V				

*According to DNA sequences in the database, accession numbers: Z98047 and Z95331.
[a]The 3′ ends of exons 15, 16, 17, and 20 correspond with the polyadenylation sites.
Capital letters represent exon sequences; a space is inserted between codons and the amino acids encoded are shown underneath. Lowercase letters represent intron sequences. UT: untranslated sequences; SP: signal peptide, NK: not known.

Chromosome 22 Mapping Group. Analysis of the sequence shows that a 36 kb region in a PAC clone (accession number Z98047) contains exons 2–14 of the human gene, and a non-overlapping 127 kb BAC clone (accession number Z95331) contains exons encoding the A, C, B, and D variants in the 3′ end of the gene (Fig. 1C). Except for two additional exons encoding the human fibulin-1 A and B variants, the exon organizations of the human and mouse genes are identical (Table 2). The intron sizes also appear to be conserved. For instance, the last intron and the intron between the C and D variants are over 10 kb in size in both the human and mouse genes. Since introns in the 3′ end of the mouse gene have not been sequenced completely, it is possible that there are additional exons encoding the mouse counterparts of the A and B variants.

The exon-intron structure of the fibulin-1 gene is remarkably similar to that of the fibulin-2 gene, which is homologous to the fibulin-1C variant (Grässel et al., in preparation). The similarity in the gene structure suggests that there is a common ancestral gene encoding the C variant, and that exons encoding the D variant are acquired after duplication of the ancestral gene. The finding that exons specific for the D var-

iant are located distal to the exon for the C variant is consistent with this hypothesis. Previous in vitro binding studies showed a substantial difference between the C and D variants in nidogen binding [11], but their functional distinction in vivo is still unclear. Characterization of the fibulin-1 gene structure demonstrates that the two variants are generated by alternative splicing of a single gene, and represents one of the first steps in elucidating the regulation of the expression of the two fibulin-1 isoforms.

Acknowledgements: We thank Drs. John S. Mudgett and Reinhard Fässler for the cosmid library from D3 mouse embryonic stem cells, and Loretta Renkart for excellent technical assistance. This work was supported by NIH Grant GM55625, and by the Deutsche Forschungsgemeinschaft (SFB 266).

References

[1] Argraves, W.S., Dickerson, K., Burgess, W.H. and Ruoslahti, E. (1989) Cell 58, 623–629.
[2] Kluge, M., Mann, K., Dziadek, M. and Timpl, R. (1990) Eur. J. Biochem. 193, 651–659.
[3] Argraves, W.S., Tran, H., Burgess, W.H. and Dickerson, K. (1990) J. Cell Biol. 111, 3155–3164.
[4] Pan, T.-C., Kluge, M., Zhang, R.-Z., Mayer, U., Timpl, R. and Chu, M.-L. (1993) Eur. J. Biochem. 215, 733–740.
[5] Zhang, H.Y., Lardelli, M. and Ekblom, P. (1997) Dev. Genes Evol. 207, 340–351.
[6] Tran, H., Mattei, M., Godyna, S. and Argraves, W.S. (1997) Matrix Biol. 15, 479–493.
[7] Pan, T.-C., Sasaki, T., Zhang, R.-Z., Fässler, R., Timpl, R. and Chu, M.-L. (1993) J. Cell Biol. 123, 1269–1277.
[8] Roark, E.F., Keene, D.R., Haudenschild, C.C., Godyna, S., Little, C.D. and Argraves, W.S. (1995) J. Histochem. Cytochem. 43, 401–411.
[9] Godyna, S., Mann, D.M. and Argraves, W.S. (1995) Matrix Biol. 14, 467–477.
[10] Balbona, K., Tran, H., Godyna, S., Ingham, K.C., Strickland, D.K. and Argraves, W.S. (1992) J. Biol. Chem. 267, 20120–20125.
[11] Sasaki, T., Kostka, G., Gohring, W., Wiedemann, H., Mann, K., Chu, M.-L. and Timpl, R. (1995) J. Mol. Biol. 245, 241–250.
[12] Spence, S.G., Argraves, W.S., Walters, L., Hungerford, J.E. and Little, C.D. (1992) Dev. Biol. 151, 473–484.
[13] Zhang, H.Y., Chu, M.-L., Pan, T.-C., Sasaki, T., Timpl, R. and Ekblom, P. (1995) Dev. Biol. 167, 18–26.
[14] Zhang, H.Y., Timpl, R., Sasaki, T., Chu, M.-L. and Ekblom, P. (1996) Dev. Dyn. 205, 348–364.
[15] Tran, H., Tanaka, A., Litvinovich, S.V., Medved, L.V., Haudenschild, C.C. and Argraves, W.S. (1995) J. Biol. Chem. 270, 19458–19464.
[16] Godyna, S., Diaz-Ricart, M. and Argraves, W.S. (1996) Blood 88, 2569–2577.
[17] Clinton, G.M., Rougeot, C., Derancourt, J., Roger, P., Defrenne, A., Godyna, S., Argraves, W.S. and Rochefort, H. (1996) Proc. Natl. Acad. Sci. USA 93, 316–320.
[18] Qing, J., Maher, V.M., Tran, H., Argraves, W.S., Dunstan, R.W. and McCormick, J.J. (1997) Oncogene 15, 2159–2168.
[19] Hayashido, Y., Lucas, A., Rougeot, C., Godyna, S., Argraves, W.S. and Rochefort, H. (1998) Int. J. Cancer 75, 654–658.
[20] Mattei, M.-G., Pan, T.-C., Zhang, R.-Z., Timpl, R. and Chu, M.-L. (1994) Genomics 22, 437–438.
[21] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[22] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[23] Pereira, L., D'Alessio, M., Ramirez, F., Lynch, J.R., Sykes, B., Pangilinan, T. and Bonadio, J. (1993) Hum. Mol. Genet. 2, 961–968.
[24] Ikegawa, S., Toda, T., Okui, K. and Nakamura, Y. (1996) Genomics 35, 590–592.