Volume 336, number 3, 467–471 FEBS 13460 December 1993

Organisation of the chicken annexin V gene and its correlation with the tertiary structure of the protein

Eva Pfannmüller, Javier Turnay**, Wolf Bertling***, Klaus von der Mark*

Max-Planck-Society, Clinical Research Units for Rheumatology at the Medical Clinic III of the University of Erlangen-Nürnberg, Schwabachanlage 10, 91054 Erlangen, Germany

Received 8 November 1993

Chicken annexin V (anchorin CII) is a collagen binding, membrane-associated molecule with Ca^{2+} channel activity. Here we report on the coding sequences, promotor region, size and distribution of exons, and exon–intron junctions of the chicken annexin V gene. It is about 25 kb long and codes for 13 short exons between 50 and 581 bp length. Exon sizes and locations of splice sites are almost completely homologous to those of the human and mouse annexin II or pigeon annexin I genes, although there is only 50–60% homology in the sequence of the corresponding proteins. The four repeat structure and symmetry of the annexin V as evident from sequence and X-ray analysis studies is only partially reflected in this highly conserved exon distribution. In the first two repeats of chicken annexin V the exons correlate with protein domains containing one, two, or three α-helices, while in the repeats 3 and 4 exon junctions and α-helical domains do not correlate. The analysis of the promotor structure revealed the absence of a typical TATA-box, but a GC-rich region which may possibly promote transcription from several start sites.

Chick annexin; Anchorin CII; Gene structure; Exon–intron splice junction; Endonexin II; Intron size

1. INTRODUCTION

Annexins are a family of at least 12 membrane-associated proteins with high affinity to calcium and phospholipids with diverse and partially unknown functions (for review see [1,2,3]). Although soluble in water, annexins associate intimately with cell membranes owing to their affinity for phospholipids in the presence of Ca^{2+} and are therefore classified a third class of amphiphatic membrane proteins [3]. They are characterized by a conserved structure consisting of four repeats in annexins I–V and VII–XII, and eight repeats in annexin VI (p68 or calelectrin). Each repeat of about 70 amino acid residues contains a consensus sequence of 17 residues which is involved in Ca^{2+} binding [4,5].

While different annexins are more than 50% homologous in the repeat domains, they differ significantly in length and sequence of the amino terminal domains which are thought to be responsible for the different functions of annexins [6].

*Corresponding author. Fax: (49) (9131) 20-6951.
***Present address:* Universidad Complutense, Department of Biochemistry, Madrid, Spain.
****Present address:* Paul-Ehrlich-Institut, Langen, Germany.

Abbreviations: Aa, amino acid; bp, base pairs; kb, kilobases; M-MLV, Moloney murine leukemia virus; PCR, polymerase chain reaction; RACE-PCR, rapid amplification of cDNA ends by polymerase chain reaction; UTR, untranslated region. *Enzymes:* MMLV reverse transcriptase, cloned (Superscript II) (EC 2.7.7.49); sequenase (EC 2.7.7.7); Taq DNA polymerase (EC 2.7.7.7); terminal transferase (EC 2.7.7.31).

Human and chicken annexin V (= endonexin II = PP4 or IBC) have recently been crystallized in the presence of calcium [4,7,8]. X-Ray analysis at 2.0–2.3 Å resolution revealed a symmetric, convex-shaped molecule consisting of four similar domains, each containing 5 α-helices. A hydrophilic pore was identified in the center of the molecule which is consistent with the concept of a calcium channel protein; voltage-gated calcium channel activity was confirmed for annexin V, VI and VII [9,10,11].

Chicken annexin V (anchorin CII) was originally isolated from chondrocyte membranes by affinity chromatography on native collagen type II [12,13]. In contrast to annexin II which remains strictly intracellular [14], annexins I and V are also secreted from cells by a yet unknown mechanism, although the proteins contain neither a signal peptide nor a transmembrane sequence [15,16]. Chicken annexin V was located on the outer surface of chondrocytes and fibroblasts by cell surface iodination, immunofluorescence and electron microscopy [15]. In matrix vesicles budding from microvilli of hypertrophic chondrocytes where binding of type II and X collagen seems to regulate Ca^{2+} uptake anchorin CII and annexin VI were identified as major collagen binding membrane proteins ([17,18,19]; T. Kirsch and R. Wuthier, in preparation).

Expression of annexin V is highly regulated in the developing chicken cartilage, with a steep increase in hypertrophic cartilage [20], but nothing is known on regulatory mechanisms of annexin gene expression. In this paper we report on the gene structure including the

presumptive promotor region of chicken annexin V. The crystallographic data on the protein [4] are compared to its gene structure to assess if exons encode functional or topological modules. Furthermore, we show that the presumptive promoter does not contain a TATA box, but a GC rich region which may possibly promote transcription from several start sites.

2. MATERIALS AND METHODS

The chick annexin V gene was isolated from a genomic library from liver DNA in EMBL-3 (Clonetech). The following cDNA probes were used to screen the library (numbers refer to positions in the cDNA. *Note:* In this work we refer to the start of the annexin V sequence as published by Genge et al. [22] as +1 of the cDNA — this numbering does not, however, refer to the transcription start site which has not yet been identified unequivocally). The clones ACIIc2–129, ACIIc78–268 and ACIIc78–628 were prepared according to the sequence published [21]. The cDNA clones ACIIc295–567, ACIIc568–953 and ACIIc954–1254 are described in [20]. Five overlapping clones were obtained that span the entire gene. The clones were analysed by restriction digestion with the enzymes *Eco*RI, *Hin*dIII and *Pst*I (New England Biolabs) in single and double digestions followed by gel electrophoresis. Exon containing fragments were identified by southern blot hybridization with the cDNA probes labelled with ^{32}P (Amersham) by random priming, cloned in pUC19 and sequenced with the Sequenase kit (USB). Splice sites were determined on the basis of intron/exon junction consensus sequences. To confirm the poly adenylation site, first strand cDNA was prepared by reverse transcription (with Superscript II, Gibco BRL) of total chondrocyte RNA with an oligo(dT)-primer and used as template in a PCR reaction with the annexin V specific upstream primer ACIIup31 (cDNA positions 10055–1064) and the oligo(dT)-primer. Specificity of the product was confirmed by nested PCR with the primer ACIIup32 (cDNA positions 1260–1278). The PCR product was cloned and sequenced as described above. 5'RACE PCR was carried out to assess the 5' extension of the first exon. First strand cDNA obtained by reverse transcription of total chick chondrocyte RNA with the downstream primer ACII1212 (cDNA positions 94–64) was poly(A)-tailed using terminal transferase (Gibco BRL) and subjected to PCR with the primer ACII1212 and an oligo(dT)-primer as upstream primer. To increase specificity of the product rePCR was carried out with the oligo(dT)-primer and the downstrean primer ACII1253 (cDNA positions 53–23). Total PCR products were cloned and sequences as described above.

3. RESULTS

The chick annexin V gene is approximately 25 kb long and composed of 13 exons (Fig. 1). Exons containing only protein coding sequences range between 57 and 123 bp in length, while exons 2 and 13 which additionally contain 5'- or 3'untranslated sequences span 50 and 581 bp, respectively. Intron sizes are variable, ranging from 92 bp for the smallest to 8 kb for the largest intron.

Splice sites of the annexin V gene follow the known consensus sequences with frequencies of 100%, 100%, 100%, 67%, 75%, and 33% for each base in the GTR-AGT consensus at the 5' ends of introns and a frequency of 100% at all positions of the 3' consensus N-YAG. Exon 13 of the annexin V gene contains an AATAAA sequence. Sequencing of a PCR product specific for the 3' end of the annexin v cDNA confirmed that this site is recognized as polyadenylation signal.

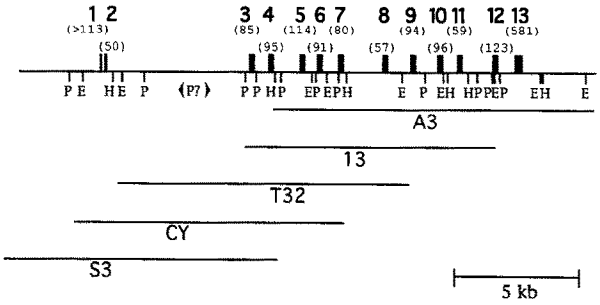

Fig. 1. Organization of the anchorin CII gene. The overlapping genomic clones A3, 13, T32, CY, and S3 cover more than 25 kb of genomic chicken DNA containing the anchorin CII gene. Numbered bars indicate the exon locations as determined by restriction analysis, Southern blotting and sequencing. Exon sizes in bp are written in parenthesis. Restriction sites for *Eco*RI, *Hin*dIII and *Pst*I are marked E, H and P, respectively. The restriction mapping for *Pst*I is incomplete between exons 2 and 3, because the large second intron contains many *Pst*I sites and has not yet been analyzed completely.

The poly(A) tail starts 21 nucleotides downstream of the AATAAA site (Fig. 2B).

The translation start site is located within the 2nd exon (Fig. 2A). In addition to the 5'-untranslated sequences in exon 2, another exon containing only 5'-untranslated sequences (exon 1) is located only 92 bp upstream of exon 2. It was not possible to determine a single defined transcription start site by S1 nuclease experiments or primer extension. 5' RACE-PCR indicated, however, that genomic sequences upstream from the published cDNA 5'end [22] are still part of the mRNA (Fig. 2A). Two out of seven positive clones that were sequenced contained sequences reaching as far as position −60 upstream of the sequence published in [21] and [22]. A PCR reaction using a downstream primer located in the second exon and an upstream primer corresponding to the newly defined 5'end of the annexin V cDNA confirmed the existence of these sequences in the cDNA and excluded the possibility that our findings might be due to amplification of genomic rather than cDNA sequences (Fig. 3A and B).

The entire 5' end of the gene around exon 1 and 2 is extremely GC-rich with a GC-content as high as 81% over a stretch of 300 bp. Eight Sp1 binding sites are found in this region. There is no TATA box in the vicinity of the first exon.

Comparison of the organization of the annexin V gene with the human and mouse annexin II [23,24] and the pigeon annexin I gene [25] revealed a striking homology: although the genes vary considerably in size (mouse annexin II gene 21 kb, human annexin II gene 40 kb, and chicken annexin V gene 25 kb), the differences are due almost entirely to differences in intron length. Exon sizes are conserved, and the localization of exon borders within the cDNA sequence of chicken annexin V and annexin I and II are identical except for the untranslated regions of the cDNA and the regions

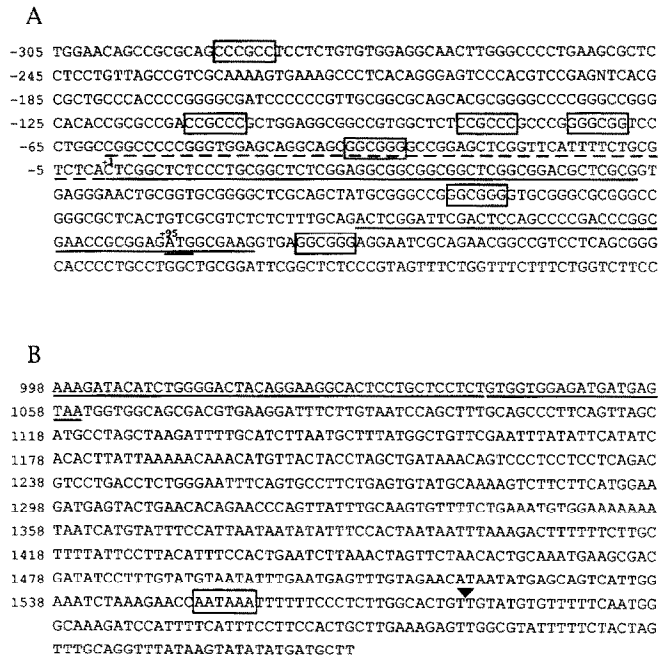

Fig. 2. (A) Sequence of the 5′ region of the annexin V gene. Exon one and two as determined by comparison to the cDNA sequence are underlined. Numbering refers to positions relative to the 5′ end of the cDNA sequence published by Genge et al. [22]. The dashed line marks the 5′ extension of the first exon as determined by RACE-PCR. Sp1 sites are boxed; of the second intron only the first Sp1 site is shown. The ATG used as translation start site lies in the second exon at position +95 of the cDNA sequence and is underlined twice. (B) Sequence of the 3′ region of the annexin V gene starting at the begin of the 13th exon. Coding sequences are underlined; numbers refer to positions in the cDNA sequence. The polyadenylation signal is boxed; the site of polyadenylation is marked by a triangle.

coding for the unique amino ends of the proteins (Table I).

Fig. 4 shows the α-helical structure and Ca^{2+} binding sites of the annexin V molecule, which were determined by X-ray crystallography first for the human protein [4,7] and later confirmed for the chicken protein [8]. The exon distribution seems to correlate well with the α-helical structure in the first two repeats where exons code for either one two or three whole α-helices. This pattern is interrupted in repeats three and four, follow-

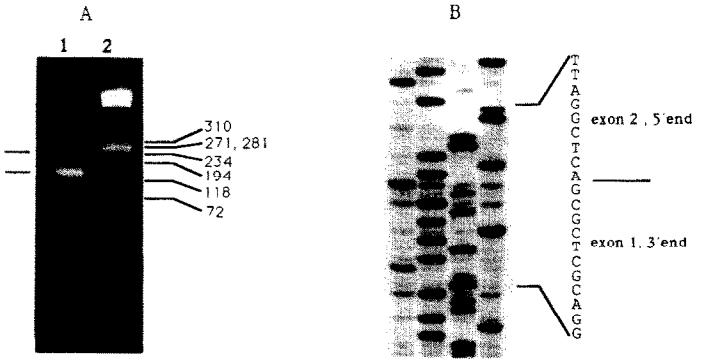

Fig. 3. To exclude the possibility that a genomic contamination of the RNA preparation might have been amplified in the RACE-PCR, a PCR reaction using a downstream primer located in the second exon and an upstream primer corresponding to the newly defined 5′ end of the anchorin cDNA was carried out. The same poly(A)-tailed cDNA preparation as used in the RACE-PCR was used as template. In case of genomic contaminations two PCR products would be expected corresponding to amplified cDNA and amplified genomic DNA, differing by the 92 bp contained in the first intron. Only one product was observed, corresponding in size to the amplified cDNA region (Fig. 3A); subsequent sequencing of the product showed that the first intron was correctly spliced out and that the PCR template had therefore been cDNA and not genomic DNA (Fig. 3B). (A) PCR of the first and second exon of the annexin V cDNA. The downstream primer was located in the second exon, the upstream primer was designed according to results of the 5′ RACE-PCR and is located 58 bp upstream of the published 5′ end of the annexin V cDNA. Lane 1 = PCR using first strand cDNA as template. The product correponds to amplified cDNA containing exon 1 and 2 but no intron sequences and is about 150 bp in size. No band could be detected at 240 bp, the size expected for the amplification of the corresponding genomic fragment. Lane 2 = *PhiX/HaeIII* marker. (B) Sequence of the cloned RACE-PCR fragment showing the correct transition from exon 1 to exon 2 without intervening intron sequences as a proof that the template for amplification had been indeed cDNA not genomic DNA.

Table I

Splice junctions of the anchorin CII gene as determined by alignment with the consensus sequence bases matching with the consensus are printed in bold. Exon borders are shown on the protein level for annexin V, chicken and annexin II, human, borders with homologous locations in both proteins are printed in bold. Note that the numbering of the annexin II amino acids differs from annexin V by 18 residues for the borders of exon 3 to 8, due to the 18 residue longer N-terminus of annexin II, and by 19 amino acids for the borders of exon 8 to 13, due to a one amino acid insertion in the annexin II exon 8

Exon x----3´	5´---------Intron----------3´	5´---Exon x+1	Exon border in chicken Annexin V	Corresponding border in human Annexin II
	Consensus Sequence			
...$^{C}_{A}$AG	GT$^{A}_{G}$ AGT..... $(^{T}_{C})_{n>10}$ N $^{C}_{T}$AG	G.....		
1..CTC GCG	**GTG AGG**.....CGTCTCTCTTTG **CAG**	ACT CGG...2	5´-UTR	*5`-UTR, no corresp.*
2...GC GAA G	**GTG AGG**.....TCCCTTTTTCTC **TAG**	TA TAC A.3	L/ys^{3}	*Asp^{16}/His^{17} no corresp.*
3...GA ATG G	**GTA AGA**.....TCTTTCTAACTC **CAG**	GG ACT G.4	**G/ly^{32}**	**G/ly^{50}**
4..GGC AGG	**GTG AGG**.....TTTGTTCTGCAC **TAG**	GAT CTT...5	**Arg^{63}/Asp^{64}**	**Lys^{81}/Glu^{82}**
5..ATC AAG	**GTA AGA**.....CTACTTGCACCT **TAG**	GGA GCA...6	**Lys^{101}/Gly^{102}**	**Lys^{119}/Gly^{120}**
6...TG CAA G	**GTA AAA**.....ATTGCTCTTTTG **CAG**	AG TAT G.7	**G/lu^{132}**	**M/et^{150}**
7..CTG CAG	**GTG TGT**.....AAAATTGTTCTT **TAG**	GCA AAT...8	**Gln^{158}/Ala^{159}**	**Lys^{176}/Gly^{177}**
8..GCT CAG	**GTG GGT**.....TTCAATTTACAA **CAG**	GTC TTG...9	**Gln^{177}/Val^{178}**	**Arg^{196}/Asp^{197}**
9...GG AGG G	**GTA GGT**.....TTTCTATTTTC **CAG**	TG TTT G.10	**V/al^{209}**	**V/al^{228}**
10...CA GTT G	**GTA ATT**.....TTAGCCCTTTTA **CAG**	TG AAG T.11	**V/al^{241}**	**V/al^{260}**
11..ATG AAA	**GTA GGG**.....AACTCTTCTCCT **TAG**	GGG GCT...12	**Lys^{260}/Gly^{261}**	**Lys^{279}/Gly^{280}**
12..ATT CAG	**GTA AAA**.....TTGTTGTATTAA **CAG**	AAA GAT...13	**Gln^{301}/Lys^{302}**	**Gln^{320}/Gln^{321}**
13..**AATAAA**TTTTTTCCCTCTTGGCACTGTpolyA			Polyadenylation site	*Polyadenylation site, no corresp.*

ing the connector region. Exon borders here are also found in the middle of helices. Interestingly, the coordinating residues of the three major Ca^{2+} binding sites are coded for by three different exons at each site.

4. DISCUSSION

Sequence analysis and X-ray crystallography of annexin V has revealed a symmetrical, convex-shaped molecule build by four domains, each consisting of five short α-helices [4]. By comparing the exon borders with the position of the α-helical domains, a certain extent of correlation can be seen: in repeats one and two, domains containing one, two, or three complete α-helices are encoded by single exons. This pattern is changed in repeats three and four where exon borders are also found within α-helices. The Ca^{2+} binding sites always contain an exon border between the three closely spaced coordination sites. In exon three which contains no Ca^{2+} binding site, exon borders are shifted. However, the lack of a Ca^{2+} binding site is not correlated with this shift because annexin II, which has the same exon distribution as annexin V in this area, has a binding site in this repeat.

Comparison of the gene structure of annexin V with those of annexin I [25] and annexin II [23,24] revealed a high degree of conservation. Although homologous in different annexins, most of the exon borders are not found at homologous locations in the four repeats of a given annexin. These findings imply that the tetrade repeat structure of an ancient annexin gene must have

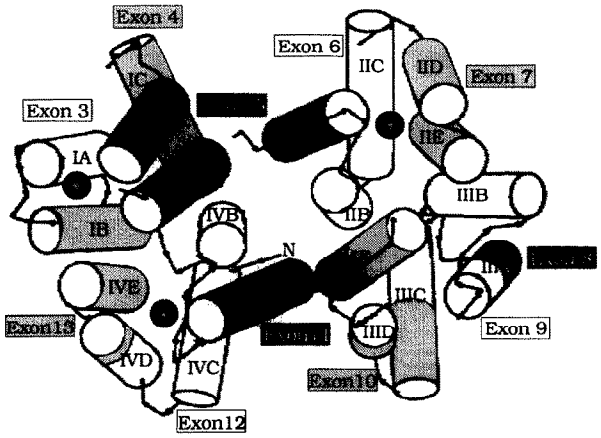

Fig. 4. Exon borders of the annexin V projected onto its tertiary structure. N marks the amino end of the sequence (Tertiary structure taken from [4] with kind permission from Dr. R. Huber). The α-helices of annexin V are indicated as cylinders and numbered IA through IVE. Exons coding for α-helical structures are shown in different colors and numbered 3 through 13. The three major calcium binding sites in annexin V are shown as spheres. The coordinating residues are located in the loops connecting helices A and B and at the C-terminal end of helix D of repeat I, II, and IV.

evolved and been selected as a useful structure long before diversification of this protein family into different annexins and before phylogenetic development.

Given the diversity of functions of the different annexins allocated to the non-conserved amino ends, which apparently arose much later than the tetrade structure in the evolution of annexins, it is likely that the highly conserved tetrade domains have evolved as a more general and basic membrane-binding module. Due to their amphiphatic character, annexins can modulate the geometry and structure of phospholipid membranes, e.g. the formation of hydrophilic pores or channels and vesicle budding in endocytosis or exocytosis, e.g. the budding of matrix vesicles from microvilli of hypertrophic chondrocytes [15,17].

Analysis of the promoter structure of the annexin V gene revealed the absence of a typical TATA box, but a GC-rich region containing eight potential Sp1 binding sites, possibly promoting transcription initiation from several start sites, which may account for the difficulties in defining a single transcription start site. The structure of this promotor thus resembles that of a 'housekeeping gene', which is consistent with the finding that annexin V is a major component of cell membranes from chondrocytes, fibroblasts and osteoblasts and many other cells. Yet, it is not evenly expressed in the vertebrate organism and probably not a housekeeping protein [20]. Specific regulatory sequences in the annexin V gene responsible for its regulation in development remain to be identified.

Acknowledgements: We gratefully acknowledge Dr. R. Huber for the use of his figure of the annexin V α-helical structure. We also thank Ms. Christine Matzner for expert technical assistance. This work was supported by a grant from the Deutsche Forschergruppe (Hi 291/5–1) and a grant by the Commission of the European Communities to J.T. (S/BMH/293001).

REFERENCES

[1] Geisow, M. (1988) in: Molecular Mechanisms in Secretion (Thorn, Treiman and Peterson eds.) Munksgaard, Copenhagen, pp. 596–608.
[2] Zaks, W.J. and Creutz, C.E. (1990) J. Bioenerg. Biomembr. 22, 97–120.
[3] Smith, V.L., Kaetzer, M.A. and Dedman, J.R. (1990) Cell Regul. 1, 165–172.
[4] Huber, R., Römisch, J. and Paques, E.-P. (1990) EMBO J. 9, 3867–3874.
[5] Thiel, C., Weber, K. and Gerke V. (1991) J. Biol. Chem. 266, 14732–14739.
[6] Barton, G.J., Newman, R.H. Freemont, P.S. and Crumpton, M.J. (1991) Eur. J. Biochem. 198, 749–760.
[7] Huber, R., Schneider, M., Mayer, J., Römisch, J. and Paques, E.-P. (1990) FEBS Lett. 275, 15–21.
[8] Bewley, M.C., Boustead, C.M., Walker, J.H. and Waller, D.A. (1993) Biochemistry 32, 3923–3929.
[9] Rojas, E., Pollard, H.B., Haigler, H.T., Parra, C. and Burns, A.L. (1990) J. Biol. Chem. 34, 21207–21215.
[10] Diaz-Munoz, M. and Hamilton S.L. (1990) J. Biol. Chem. 265, 15894–15899.
[11] Berendes, R., Burger, A., Voges, D., Demange, P. and Huber, R. (1993) FEBS Lett. 317, 131–134.
[12] Mollenhauer, J. and von der Mark, K. (1983) EMBO J. 2, 45–50.
[13] Mollenhauer, J., Bee, J.A., Lizarbe, M.A. and von der Mark, K. (1984) J. Cell Biol. 98, 1572–1578.
[14] Nigg, E.A., Cooper, J.A. and Hunter, T. (1983) J. Cell Biol. 96, 1601–1609.
[15] Pfäffle, M., Ruggiero, F., Hofmann, H., Fernandez, M.P., Selmin, O., Yamada, Y., Garrone, R. and von der Mark, K. (1988) EMBO J. 7, 2335–2342.
[16] Christmas, P., Callaway, J., Fallon, J., Jones, J. and Haigler, H.T. (1991) J. Biol. Chem. 266, 2499–2507.
[17] Genge, B.R., Wu, L.N.Y., Adkisson, H.D. and Wuthier, R.E. (1991) J. Biol. Chem. 266, 10678–10685.
[18] Wu, L.N.Y., Genge, B.R., Lloyd, G.C. and Wuthier, R.E. (1991) J. Biol. Chem. 266, 1195–1203.
[19] Kirsch, T. and Pfäffle, M. (1992) FEBS Lett. 310, 143–147.
[20] Hofmann, C., Gropp, R. and von der Mark, K. (1992) Dev. Biol. 151, 391–400.
[21] Fernandez, M.P., Selmin, O., Martin, G.R., Yamada, Y., Pfäffle, M., Deutzmann, R., Mollenhauer, J. and von der Mark, K. (1988) J. Biol. Chem. 263, 5921–5925.
[22] Genge, B.R., Cao, X., Wu, L.N., Buzzi, W.R. Showman, R.W., Arsenault, A.L., Ashikana, Y. and Wuthier, R.E. (1992) J. Bone Mineral Res. 7, 807–819.
[23] Amiguet, P., D'Eustachio, P., Kristensen, T., Wetsel, R.A., Saris, C.J.M., Hunter, T., Chaplin, D.D. and Tack, B.F. (1990) Biochemistry 29, 1226–1232.
[24] Spano, F., Raugei, G., Palla, E., Colella, C. and Melli, M. (1990) Gene 95, 243–251.
[25] Hitti, Y.S. and Horseman, N.D. (1991) Gene 103, 185–192.