FEBS Letters 434 (1998) 300–304 FEBS 20759

Expression profile and structural divergence of novel human annexin 31

Reginald O. Morgan*, Maria-Pilar Fernandez

Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Oviedo, E-33006 Oviedo, Spain

Received 13 July 1998

Abstract Systematic analysis of expressed sequence tags in dbEST yielded an expression profile of the ten known human annexins and led to the discovery of a novel subfamily expressed mainly in differentiating tissues. Full-length cDNAs encoded a 338-amino acid protein with less than 40% identity to other annexins, an atypical amino acid composition, and an insertion and deletion in internal repeat 3. The most striking feature was a complete ablation of all four type II calcium-binding sites in the conserved tetrad core. Annexin 31 thus constitutes a unique, natural probe for investigating the role of membrane binding in annexin function.

Key words: Annexin multigene family; Calcium-binding site; Evolutionary divergence; Expressed sequence tag; Protein structure; Tissue distribution

1. Introduction

The biological function(s) and phenotypic profile(s) of annexins remain unresolved, despite an eclectic record of basic research on their molecular biology, biochemical properties and pharmacological interactions [1]. Type II calcium-binding sites in each of the four internal repeat domains have been identified as the principal mechanism by which annexins aggregate and cooperatively bind anionic phospholipids and extracellular matrix proteins [2,3] to perform various membrane-related functions. Characterization of the highly conserved tetrad core structure by crystallographic analyses [4,5] further attests to an associated role as calcium ion channel. These two canonical features of the annexin structure have been formulated into an integral mechanism of action [6–8] which can be interpreted to account for a myriad of membrane interactions involving signal transduction, vesicular transport and nuclear regulation of cell growth and differentiation.

It follows that if the bidentate calcium-binding sites on the external surface and the tertiary pore structure are both intrinsic to annexin function, then strategic deviations that disrupt either of these features should be selected against in evolution [9,10] unless our concept of annexin structure-function requirements is imprecise or there exists some positive selection force for adaptation to modified functions. Individual calcium-binding motifs are indeed altered in certain repeats of both plant and animal annexins [8,11,12] and other important, calcium-dependent, membrane-binding proteins such as phospholipase A_2, protein kinase C and synaptogamins are now known to include calcium-independent isoforms [13–15]. If this property is not an obligatory hallmark of annexin function, then the discovery of calcium-independent annexins could significantly alter our concept of the mechanisms by which these proteins modulate membrane-related processes. A systematic molecular analysis of annexins in the database of expressed sequence tags (dbEST) was therefore undertaken to identify novel genes that could help to define the full range of regulatory, structural and functional properties of the vertebrate annexin multigene family.

2. Materials and methods

2.1. DNA sequencing

IMAGE (Integrated Molecular Analysis of Genome Expression) Consortium (Lawrence Livermore National Laboratory) cDNA clones [16] were obtained from the Human Genome Mapping Project Resource Centre (Hinxton, UK). Clone IDs #212319 and 111373 from Soares 1NFLS normalized library of 20-week, human fetal liver-spleen were sequenced on both strands by the dideoxy chain termination method with [α-^{35}S]dATP from Amersham (Little Chalfont, UK) and the Sequenase 2.0 kit from US Biochemical (Cleveland, OH, USA).

2.2. Molecular systematics

The current database of approximately 1 million human expressed sequence tags (dbEST) [17] represented cDNA clones from more than 100 different tissue libraries. It was cyberscreened using the NCBI-BLAST netserver [18] with protein and DNA sequences of known, consensus and ancestral annexins and high-scoring matches were verified by pairwise comparison using SSEARCH and PRSS computer programs [19]. Consensus sequences were compiled from a spreadsheet alignment of more than 100 unique annexin proteins, and ancestral protein sequences were derived by the ANCESTOR program [20]. Predictions of protein secondary structure were made by the EMBL email server predictprotein@embl-heidelberg.de [21] and protein isoelectric points were calculated by the netserver in Geneva (http//www.expasy.ch/ch2d/pi_tool.html).

3. Results

3.1. Identification and classification of annexin expressed sequence tags

BLAST searches of dbEST with annexin sequences led to the assignment of more than 1850 human ESTs to the 10 known annexin subfamilies, with a median nucleotide identity of 98% (range 70–100%) (Fig. 1, upper panel). The distribution reflected overall expression levels of the individual subfamilies since they originated from a broad range of tissues in large-scale, random-sequencing projects. Annexin 2 was detected over five times more frequently than any other annexin and displayed the ubiquitous tissue distribution typical of most annexins. Annexins 11, 5, 1 and 4 were also abundantly and ubiquitously expressed, while annexins 6, 7, 3, 8 and 13

*Corresponding author. Fax: (34) (98) 510 3157.
E-mail: mpff@dwarf1.quimica.uniovi.es

Abbreviations: aa, amino acid(s); *ANX*/ANXn-Gsp, human annexin gene/protein (suffixed by subfamily number and genus-species abbreviation)

The sequence reported in this paper has been deposited in the EMBL/GenBank/DDBJ databases under accession no. AJ009985.

PII: S0014-5793(98)00997-1

showed progressively restricted frequency and tissue distribution, despite library normalization. Occasional weak sequence matches were attributed to rarer, alternatively spliced isoforms of annexins 2, 6 and 7, incompletely processed transcripts containing introns, occasional recombination artifacts, or poor sequence traces.

3.2. Expression profile of a novel human annexin

One group of 24 partially overlapping human ESTs (Fig. 1, upper panel) had a modest but significant 52–56% nucleotide identity to the coding regions of known annexins and greater than 90% identity to each other. Despite their comparatively low expression level, these transcripts were especially abundant in one cDNA library of human fetal liver-spleen (74% of matching ESTs). This expression pattern resembled annexin 8 (Fig. 1, lower panel) and suggested some regulatory or functional association with differentiating tissues and/or these particular organs. Other related clones originated from placenta and prostate ($n=2$ each), whole embryo, HepG2 liver cells, macrophage, thyroid and parathyroid ($n=1$ each), so their tissue distribution was not so restricted as intestine-specific annexin 13. An abrupt alignment break from known annexins in untranslated regions pointed to a novel, paralogous subfamily, but reading frame shifts due to EST sequencing error and non-contiguous segments made it uncertain whether these ESTs represented only one new cDNA.

3.3. Human annexin 31 cDNA

IMAGE Consortium (LLNL) cDNA clones #212319 and 111373 with the longest 5′ partial sequences were obtained from the UK-HGMP for detailed analysis. DNA sequencing revealed a unique cDNA structure containing 4 internally homologous repeats (Fig. 2) with 93–98% identity to all 24 unassigned human ESTs and 51.6–55.6% nucleotide identity

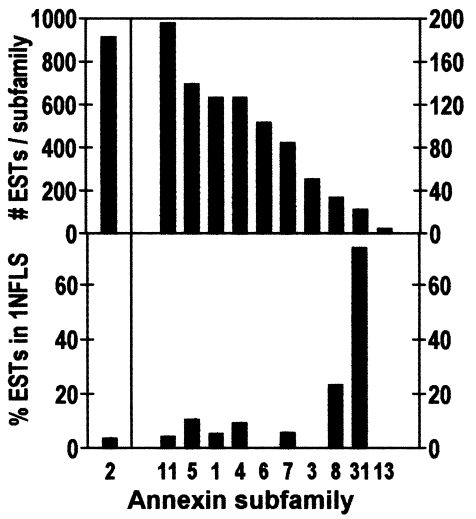

Fig. 1. Classification of annexin expressed sequence tags. The upper panel tabulates the distribution of 1850 human ESTs from more than 1 million entries in dbEST release 070398 that were confirmed with 98% median nucleotide identity to be homologs of the respective annexin subfamilies. ESTs ultimately assigned to the novel annexin 31 subfamily had 50–55% nucleotide identity to other vertebrate annexins and 90–100% identity to each other and a fully sequenced cDNA. The lower panel shows the percentage of ESTs from each subfamily that were detected in one cDNA library of normalized, human fetal liver-spleen (1NFLS).

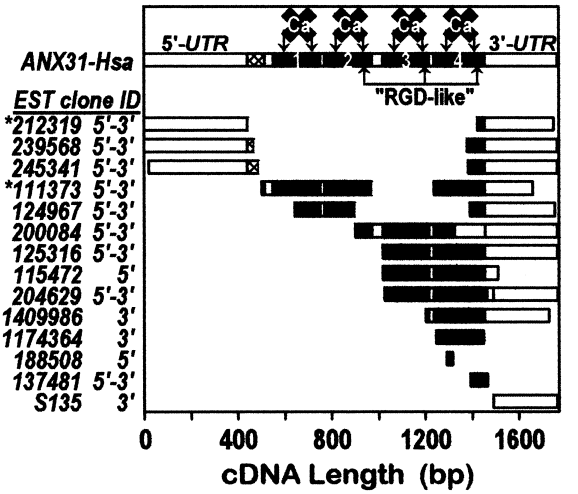

Fig. 2. Human annexin 31 cDNA structural features and EST overlap. The top outline depicts the 1775-bp cDNA sequence of human annexin 31 obtained by complete sequencing of IMAGE Consortium (LLNL) cDNA clone IDs 212319 and 111373 (asterisks) and deposited in the sequence databanks under accession no. AJ009985. The 338-codon open-reading frame contained a 3′ core tetrad of four internally homologous repeats (solid-filled segments) within which arrows point to sites coding for mutated type II calcium-binding residues and 'RGD-like' cell-attachment motifs. Unique features included a 439-bp 5′-untranslated region, a 25-codon 5′-coding sequence (hatched block), and insertion/deletion in repeat 3. Partially overlapping sequences of matching ESTs are shown below the full-length cDNA for *ANX31-Hsa*.

to the core regions of vertebrate annexins. Annexin homology was statistically confirmed using PRSS [19] and the novel representative was formally classified in order of discovery [11,12] as annexin subfamily 31, with abbreviated symbols *ANX31* and ANX31 for the corresponding gene and protein. Its unusually long, 439-bp 5′-untranslated region was followed by a 338-codon open reading frame initiated by the only in-frame ATG codon. This site exclusively fulfilled other critical requirements for initiation of translation, being flanked by nucleotides A at position −3 and G at position +4 [22] and preceded by an in-frame stop codon 15 bp upstream. A single codon insertion in the repeat 2–3 linker segment resembled annexins 2, 8 and 10 and a 3-codon deletion in mid-repeat 3 was unique.

3.4. Protein structure divergence

The 338-amino acid (aa) deduced protein sequence of human annexin 31 (Fig. 3) had only 31.8–40.4% aa identity to known vertebrate annexin tetrads, well below the 50% average between other annexin paralogs [11]. Its 25-aa amino-terminus resembled annexin 2 in size and in the presence of sites adjacent to the hinge region for potential *N*-myristoylation (position −3 relative to the core region), phosphorylation by protein kinase C (position +2) and *N*-glycosylation (position +8), as defined in the PROSITE database of protein motifs [23]. Smith-Waterman alignment scores [19] reflected greatest phenetic similarity to annexin 2 (SW = 842), annexin 11 (SW = 714) and annexin 4 (SW = 707) and its closest phylogenetic relatives were annexins 2 and 1 (not shown). Annexin 31 showed 38.7% aa identity to the predicted common ancestor [20] of all vertebrate annexins and 51.3% aa identity to the most recent common ancestor of annexins 1, 2 and 31 (Fig. 3). This divergence was uniformly diffuse as there were no iden-

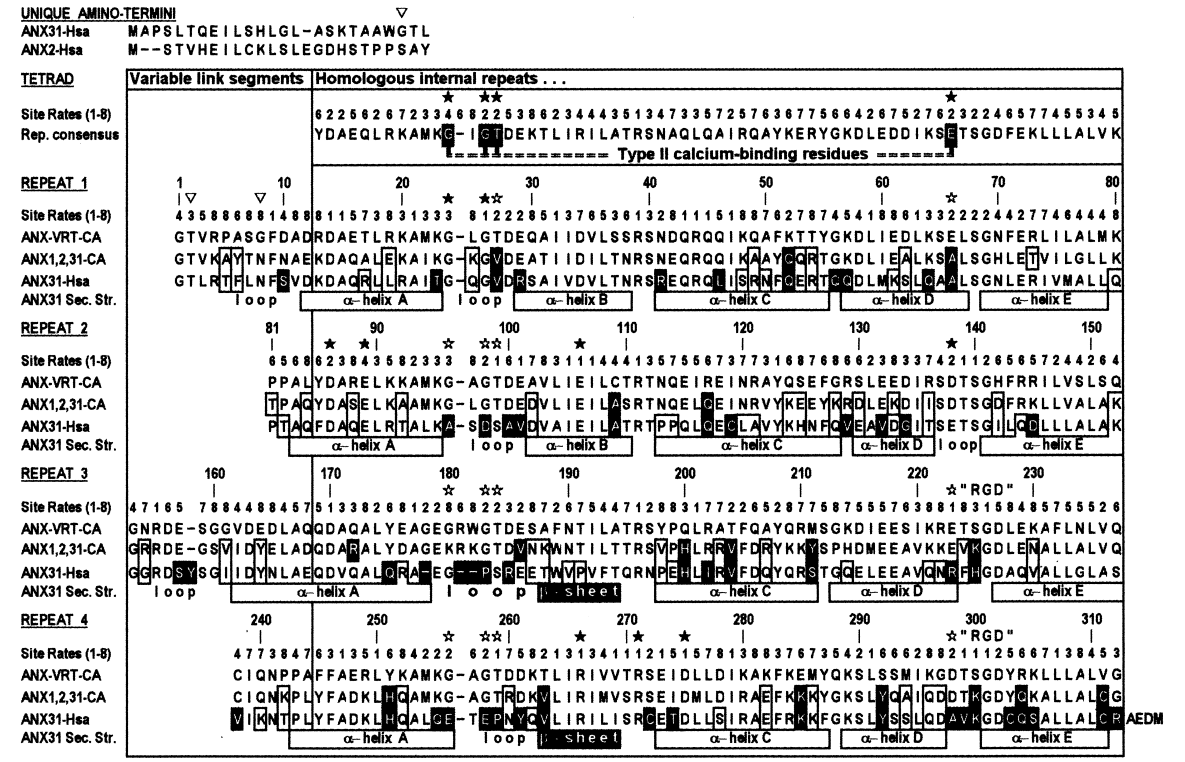

Fig. 3. Human annexin 31 protein structure and divergence pattern. The unique 25-aa amino-terminus of deduced human annexin 31 is contrasted with annexin 2 (upper left) and its 309-aa tetrad core is vertically aligned across the 4 internally homologous repeats (right rectangle). Inverted triangles near the amino-terminal hinge (positions −3, 2 and +8 relative to the tetrad) denote potential myristoylation, protein kinase C phosphorylation and *N*-glycosylation sites, respectively, and stars in the core region mark key residues known to be involved in calcium binding or ion channel function. The consensus sequence line above the homologous repeats identifies the 'GxGTde-36 aa-D/E' motif regarded as canonical for vertebrate annexin type II calcium-binding sites. Additional sequences in the core aa alignment represent the derived putative common ancestors of all vertebrate annexins (ANX-VRT-CA) and for the annexins 1-2-31 subclade only (ANX1,2,31-CA). Non-conservative aa replacements with respect to the primary ancestor are boxed at variable sites or reverse-shaded at highly conserved sites. Numbers on the 'site rates' line specify the relative rate categories 1–8 for the observed number of aa replacements at each residue position (i.e. 0.1973–2.2945 in discrete steps of 0.125) according to aa variability in an alignment of vertebrate annexins. Secondary structure regions denote α-helices, loops and β-sheets predicted by the EMBL PredictProtein email server [21].

tical matches longer than 4 residues. However, charged residues with a critical role in annexin ion channel function [4] were all fully conserved in repeats 2 and 4 (Fig. 3, D-86, E-89, E-106, R-112, E-115, R-266, R-271 and D-275).

A closer examination of slowly evolving, conserved sites revealed the extraordinary finding that all type II calcium-binding sites in annexin 31 were replaced by dysfunctional residues (Fig. 3). These canonical features of annexin internal repeats have been precisely defined on the external convex protein surface by crystallographic analysis [8] and are indispensable for the calcium-mediated, high-affinity binding to membranes [6]. Functional inactivation of these coordination sites in the first repeat of annexins 1 and 2 has previously been attributed to the replacement of polar and acidic residues in the type II calcium-binding motif 'GxGTde-36 aa-E' with non-polar, basic and/or bulkier residues in 'T/VxGVdr-36 aa-A' [8]. More extensive changes in repeats 2 and 3 of plant and *C. elegans* annexins provided additional evidence for the natural mutation of these calcium-binding sites [12]. The complete replacement of these critical residues in all repeats of annexin 31, together with a 3-codon proximal deletion in repeat 3, further implied its evolutionary selection for calcium independent function. The PROSITE database was searched for alternative protein motifs that might mediate annexin membrane binding. The important cell attachment motif 'RGD' (Arg-Gly-Asp) was detected in all species of plant annexins 23 and 24 in the carboxy-terminal region of repeat 4 immediately adjacent to the type II calcium-binding half-site. Analogous 'R/K/H-G-D' motifs were detected in members of the annexin 1-2-31 subclade (Fig. 3) at equivalent locations in repeats 4, 3 and 2 (i.e. core positions 300–303, 225–227 and 140–142), at proximal positions in repeat 4 of annexins 4, 5, 16, 17, and in other repeats of annexins 3, 9 and 13.

Secondary structure predictions [21] were made for human annexin 31 (Fig. 3) and a three-dimensional model (not shown) was visualized by protein threading through crystallography coordinates of annexin homologs. These indicated only minor deviations from the well-preserved, α-helical tetrad structure and suggested that the putative ion channel remained ostensibly intact. In view of the extensive aa replacements observed in annexin 31 relative to annexin 2, its aa composition and isoelectric point were compared with other annexins (Fig. 4) to assess any general character change that might influence its biochemical properties or evolutionary analysis. Annexin 31 showed significant deviation in its aa composition, with very low levels of acidic Asp and Glu and basic Lys being replaced by less polar Gln, and with elevated levels of non-polar Ala, Val and Leu. The calculated p*I* of 5.53 (from the ExPASy netserver) for the unmodified,

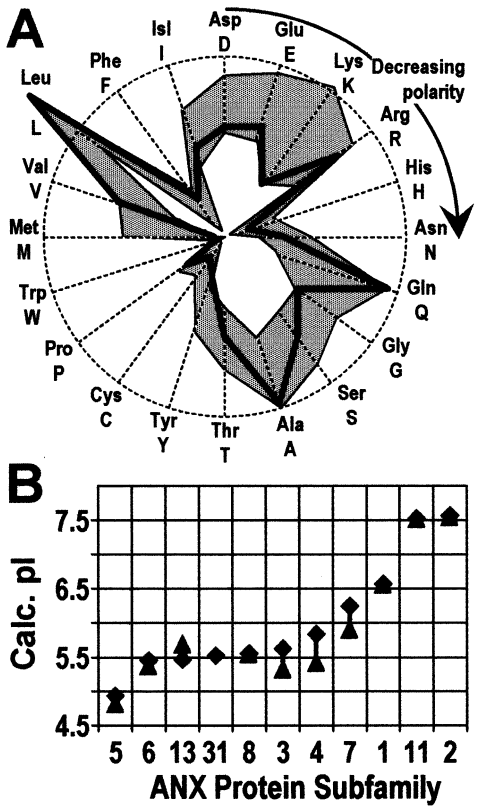

Fig. 4. Amino acid composition and isoeletric points of vertebrate annexins. A: The shaded outline gives the upper and lower limits of aa percentages in the homologous core regions of vertebrate (human and mouse) annexin proteins and the bold line traces these values for human annexin 31. B: Isoelectric points (p*I*) were predicted by the ExPASy network server (Geneva) for all human (diamond) and mouse (triangles) annexin full-length proteins in their common, unmodified isoforms.

37 645-Da protein was similar to other annexins, but significantly lower than its nearest structural neighbors annexins 1, 2 and 11.

4. Discussion

Human annexin 31 represents a new subfamily of paralogous, full-length annexins and a special subclass by virtue of its eroded type II calcium-binding sites. It was culled from among homologous cDNA transcripts in the dbEST by probing with a battery of divergent annexins and characterized by sequencing and systematic analysis. Though well-represented in actively differentiating tissues (Fig. 1) as well as a host of other cDNA libraries (Fig. 2), it showed both extensive sequence divergence (Fig. 3) and atypical aa composition (Fig. 4) compared to other annexins. Differential control of annexin expression in developmental stages [24,25] and cell growth phases [26] has previously suggested their contributory role in tissue growth and differentiation [1], and annexin polarity and isoelectric point are known to critically affect their ability to intercalate membranes at acidic pH [27]. The distinct expression pattern and protein properties of annexin 31 could therefore endow it with a unique regulatory and ontogenic role.

The discovery of human annexin 31 adds a new dimension to annexin research, primarily because the conserved tetrad core containing the putative internal ion channel can now be dissociated from the external type II calcium-binding sites by which other annexins self-aggregate and attach to membrane anionic phospholipids and cytoskeletal components [2,3,6]. Modification of these sites in nature [4] or targeted by site-directed mutagenesis has been well-established to drastically reduce annexin calcium affinity [28], membrane-binding and self-aggregation [2] and phospholipase A_2 inhibition [29]. The ablation of all type II calcium-dependent binding sites in annexin 31 thus represents an evolutionary landmark for this multigene family, combining advantageous mutation with positive selection. Calcium-independent isoforms of phospholipase A_2, protein kinase C and synaptogamins have altered properties but similar basic actions as their calcium-dependent, phospholipid-binding counterparts [13–15]. By analogy, such selective adaptation of annexin 31 may challenge some of the basic concepts of annexin action, such as the importance of calcium sensitivity and the role of membrane binding. The distinct expression and structural features described here emphasize the need for a broader concept of annexin action that encompasses their full diversity before their truly common properties can be assessed. We have recently identified partial mouse cDNAs that corroborate the structural deviation of annexin 31, and preliminary attempts to purify recombinant protein by calcium-dependent binding to phosphatidylserine-containing liposomes appear to confirm its loss of this property. Alternate strategies for expression, purification and functional characterization will undoubtedly be required to study annexin 31, and any contribution of RGD-like cell attachment motifs to membrane interactions may help to clarify a possible extracellular role for annexins.

Acknowledgements: This work was supported by Grant PM95-0152 from DGES of Spain.

References

[1] Raynal, P. and Pollard, H.B. (1994) Biochim. Biophys. Acta 1197, 63–93.
[2] Nelson, M.R. and Creutz, C.E. (1995) Biochemistry 34, 3121–3132.
[3] von der Mark, K. and Mollenhauer, J. (1997) Cell. Mol. Life Sci. 53, 539–545.
[4] Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H., Römisch, J. and Paques, E. (1992) J. Mol. Biol. 223, 683–704.
[5] Liemann, S., Benz, J., Burger, A., Voges, D., Hofmann, A., Huber, R. and Göttig, P. (1996) J. Mol. Biol. 258, 555–561.
[6] Swairjo, M.A., Concha, N.O., Kaetzel, M.A., Dedman, J.R. and Seaton, B.A. (1995) Nat. Struct. Biol. 2, 968–974.
[7] Hofmann, A., Benz, J., Liemann, S. and Huber, R. (1997) Biochim. Biophys. Acta 1330, 254–264.
[8] Liemann, S. and Huber, R. (1997) Cell. Mol. Life Sci. 53, 516–521.
[9] Morgan, R.O. and Fernandez, M.P. (1997) Cell. Mol. Life Sci. 53, 508–515.
[10] Morgan, R.O., Bell, D.W., Testa, J.R. and Fernandez, M.P. (1998) Genomics 48, 100–110.
[11] Morgan, R.O. and Fernandez, M.P. (1995) Mol. Biol. Evol. 12, 967–979.
[12] Morgan, R.O. and Fernandez, M.P. (1997) J. Mol. Evol. 44, 178–188.
[13] Larsson, P.K., Claesson, H.E. and Kennedy, B.P. (1998) J. Biol. Chem. 273, 207–214.
[14] Kruse, M., Gamulin, V., Cetkovic, H., Pancer, Z., Muller, I.M. and Muller, W.E.G. (1996) J. Mol. Evol. 43, 374–383.
[15] von Poser, C., Ichtchenko, K., Shao, X., Rizo, J. and Südhof, T.C. (1997) J. Biol. Chem. 272, 14314–14319.

[16] Lennon, G.G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) Genomics 33, 151–152.
[17] Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) Nat. Genet. 4, 332–333.
[18] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[19] Pearson, W.R. (1990) Methods Enzymol. 183, 63–98.
[20] Zhang, J. and Nei, M. (1997) J. Mol. Evol. 44, S139–S146.
[21] Rost, B. (1996) Methods Enzymol. 266, 525–539.
[22] Kozak, M. (1986) Cell 44, 283–292.
[23] Bairoch, A., Bucher, P. and Hofmann, K. (1997) Nucleic Acids Res. 25, 217–221.
[24] Hofmann, C., Gropp, R. and von der Mark, K. (1992) Dev. Biol. 151, 391–400.
[25] Ivanenkov, V.V., Weber, K. and Gerke, V. (1994) FEBS Lett. 352, 227–230.
[26] Schlaepfer, D.D. and Haigler, H.T. (1990) J. Cell Biol. 111, 229–238.
[27] Köhler, G., Hering, U., Zschörnig, O. and Arnold, K. (1997) Biochemistry 36, 8189–8194.
[28] Jost, M., Weber, K. and Gerke, V. (1994) Biochem. J. 298, 553–559.
[29] Mira, J.P., Dubois, T., Oudinet, J.P., Lukowski, S., Russo-Marie, F. and Geny, B. (1997) J. Biol. Chem. 272, 10474–10482.