FEBS 27050 FEBS Letters 538 (2003) 183–191

EST analysis of gene expression in the tentacle of *Cyanea capillata*

Yanzhen Yang, Shujian Cun, Xiaojin Xie, Jianghai Lin, Jianwen Wei, Wenli Yang, Chunyan Mou, Cuiling Yu, Lanting Ye, Yang Lu, Zhiyan Fu, Anlong Xu*

Department of Biochemistry, State Open Laboratory for Marine Functional Genomics of State High-Tech Development Program, State Key Laboratory of Genetic Engineering of MOE, College of Life Sciences, Sun Yat-sen (Zhongshan) University, Guangzhou 510275, PR China

Received 11 November 2002; revised 4 February 2003; accepted 7 February 2003

First published online 24 February 2003

Edited by Gunnar von Heijne

Abstract **Jellyfish, *Cyanea capillata*, has an important position in head patterning and ion channel evolution, in addition to containing a rich source of toxins. In the present study, 2153 expressed sequence tags (ESTs) from the tentacle cDNA library of *C. capillata* were analyzed. The initial ESTs consisted of 198 clusters and 818 singletons, which revealed approximately 1016 unique genes in the data set. Among these sequences, we identified several genes related to head and foot patterning, voltage-dependent anion channel gene and genes related to biological activities of venom. Five kinds of proteinase inhibitor genes were found in jellyfish for the first time, and some of them were highly expressed with unknown functions.**

Key words: Jellyfish tentacle; Expressed sequence tag analysis; Head patterning; Voltage-dependent anion channel; Toxin; Proteinase inhibitor

1. Introduction

Cnidarians, including jellyfish, anemones, and corals, are the earliest existing organisms possessing a neuromuscular system [1]. Instead of having a brain, jellyfish has a primitive neuro-network, which consists of receptors capable of detecting light, odor and other stimuli and coordinating appropriate responses to stimuli. Recent molecular analyses have shown that an apical patterning is present in jellyfish, so head patterning is perhaps much older than previously thought [2]. Jellyfish also occupies an important position in the evolution of voltage-dependent Na^+, Ca^{2+} channels that are responsible for the electrical excitability of nerve and muscle in diverse species [3].

Jellyfish is composed of two cell layers and equipped with a specialized venom cnidoblast for defense and feeding. Nematocysts, concentrated on the tentacles or oral arms, contain a variety of toxins and a stinging structure. The stings of jellyfish can induce persistent urticaria, local edema, muscle weakness, paresthesia in extremities, transient dyspnea and shock [4]. *Cyanea capillata* is generally considered a moderate stinger. The pain caused by the stinger is relatively mild and often described as burning rather than stinging.

To study the active components in the nematocyst venom and to provide a molecular path of head patterning and voltage-dependent anion channel (VDAC) evolution, we analyzed the expressed sequence tags (ESTs) derived from the tentacle cDNA library of *C. capillata*. A total of 1016 clusters composed of 2153 ESTs were generated in the present study, and most of them were reported in jellyfish for the first time.

2. Materials and methods

2.1. cDNA library construction

The tentacles of *C. capillata* were collected from Qingzhou, Guangxi, China, by South China Sea. After being frozen with liquid nitrogen quickly, tissues of tentacles were triturated into fine particles for total RNA extraction using Trizol reagent (Gibco BRL) according to the supplier's method. cDNA was prepared with SMART™ cDNA library construction kit (Clontech) following the manufacturer's instruction. All cDNAs were ligated into pcDNA3.0 and electroporated into *Escherichia coli* DH5α cells using a Gene Pulser II electroporation system under standard conditions. The library contained approximately 5.1×10^6 independent clones, and a total of 7000 independent cDNA clones were picked out randomly and stored at −80°C for further analysis.

2.2. EST sequencing

cDNA clones were thawed, inoculated directly into 96-well plates containing 1 ml Luria–Bertani (LB) broth and cultured in 37°C overnight. DNAs were extracted using Vitagene 96-easy plasmid Miniprep Kit (Vitagene Biochemical Technique Co., Ltd). The 5′-end sequencing of each cDNA was conducted in an automated ABI Prism 3700 sequencer (Perkin-Elmer), using ABI Prism Big-Dye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems) and T7 primer.

2.3. Sequence data analysis

Before being assembled into clusters, groups of sequences, sequencing outputs were edited by computer programs to remove vector sequences and ambiguous regions. The consensus sequence of each cluster was used as query to search homologous sequences against GenBank DNA sequences with FASTA and BLAST X [5]. Annotations of possible protein-coding genes were performed and assembled for future research.

3. Results and discussion

3.1. Overall distribution of sequences from the tentacle of C. capillata

We determined 2153 sequences from cDNA clones of the tentacle library of *C. capillata*. The length of inserts was 650 bp on average and mainly in the range from 0.2 to 2 kb. The average readable sequence length, on which the following

*Corresponding author. Fax: (86)-20-84038377.
E-mail address: ls36@zsu.edu.cn (A. Xu).

Abbreviations: EST, expressed sequence tag; Xp8, *Xenopus laevis* homolog of p8; TSG, twisted gastrulation; VDAC, voltage-dependent anion channel

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00164-9

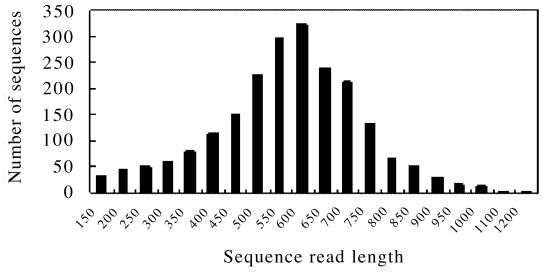

Fig. 1. Distribution of sequence read length. The number of sequence reads is shown for each group of 50 sequence lengths between 150 and 1200 bp. The total number of high quality sequence reads was 2153. The average length of the sequence reads was 530 bp. A large fraction of sequences is between 500 and 700 bp. Failed sequencing attempts were not shown.

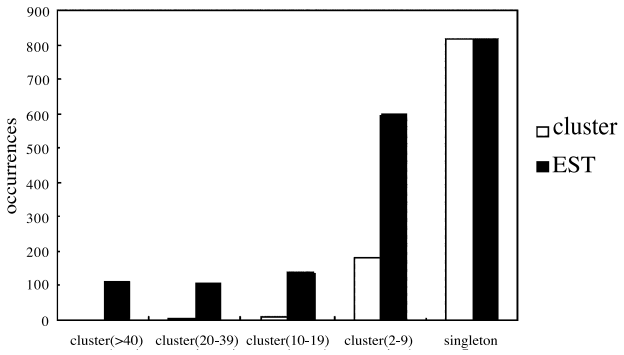

Fig. 2. Prevalence distribution of the cluster size. The initial 2153 ESTs were grouped into 1016 clusters, consisting of two clusters (113 of 2153 ESTs) comprised of more than 40 ESTs each, four clusters comprised of 20–39 ESTs (107 of 2153 ESTs) each, 10 clusters comprised of 10–19 ESTs (139 of 2153 ESTs) each, 182 clusters comprised of two to nine ESTs (596 of 2153 ESTs) each and 818 unique sequences. The complexity of mRNA stored in the tentacle of *C. capillata* could be observed for most ESTs belonging to medium-sized clusters and singletons.

analysis was based, was approximately 530 bp, and the distribution of the readable sequence lengths was shown in Fig. 1.

The initial ESTs were grouped into 1016 consensus sequences, among which 198 contained more than two ESTs per consensus sequence and 818 were singletons. The abundance, or say the cluster size, could reflect the relative mRNA population since a non-normalized primary cDNA library was used [6]. The distribution of cluster size was indicated in Fig. 2 and all clusters were divided into five classes according to the number of ESTs they contained:

1. There were only two clusters containing more than 40 ESTs, constituting 5.25% of the total clones (113 of 2153 clones) and 0.197% of the total clusters (two of 1016 clusters). They could be regarded as the most abundant transcripts, and one of the two was cystatin (cysteine proteinase inhibitor) that contained 68 clones, and the other was EPPT protein that contained 45 clones.
2. There were four clusters containing 20–39 ESTs, constituting 4.97% of the total clones (107 of 2153 clones) and 0.394% of the total clusters (four of 1016 clusters). These highly prevalent ones encoded proteins such as ribosomal protein L30, ferritin, dickkopf homolog 4 and thrombospondin.
3. There were 10 clusters containing 10–19 ESTs, constituting 6.46% of the total clones (139 of 2153 clones) and 0.984% of the total clusters (10 of 1016 clusters). These ones were represented by peptidylprolyl isomerase F (cyclophilin F), NADH:ubiquinone oxidoreductase MLRQ subunit, proteinase inhibitor PAPI I and several ribosomal proteins. More information about the clusters containing more than 10 ESTs was summarized in Table 1.
4. There were 182 clusters containing two to nine ESTs, constituting 27.68% of the total clones (596 of 2153 clones) and 17.91% of the total clusters (182 of 1016 clusters). These clusters belonging to the lowly prevalent class encoded the proteins involved in cell metabolism, such as cytochrome *c* oxidase, ubiquitin and the proteins with special functions such as coagulant enzyme, tumor-specific transplantation antigen, senescence-associated protein and so on.
5. There were 818 clusters containing only one EST, which we called singletons, constituting 37.99% of the total clones (818 of 2153 clones) and 80.51% of the total clusters (818 of 1016 clusters). The majority in total clusters, these singletons, encoded signal receptors, regulatory proteins and other proteins without significant similarities to public databases.

The GC contents of all the ESTs obtained were calculated for each terminus [6]. The 5′-terminus sequences, which

Table 1
Assembled clusters that contain more than 10 ESTs

Cluster	Gene description	Organism	ESTs
0127	cystatin-type cysteine proteinase inhibitor	*Cyprinus carpio*	68
0010	EPPT protein	*Podocoryne carnea*	45
0067	ribosomal protein L30	*Spodoptera frugiperda*	37
0058	ferritin	*Ornithodoros moubata*	25
0040	dickkopf homolog 4	*X. laevis*	23
0178	thrombospondin	*Haemonchus contortus*	22
0004	peptidylprolyl isomerase F (cyclophilin F)	*H. sapiens*	18
0007	NADH:ubiquinone oxidoreductase MLRQ subunit	*H. sapiens*	17
0086	ribosomal protein L35	*M. musculus*	17
0084	ribosomal S29-like protein	*Arabidopsis thaliana*	15
0013	proteinase inhibitor PAPI I	*Pacifastacus leniusculus*	13
0072	ribosomal protein L32	*S. frugiperda*	13
0186	cytochrome *c*	*Hippopotamus amphibius*	12
0176	dickkopf-1	*D. rerio*	12
0177	elongation factor 1A	*Geodia cydonium*	11
0206	ribosomal protein S20	*X. laevis*	11

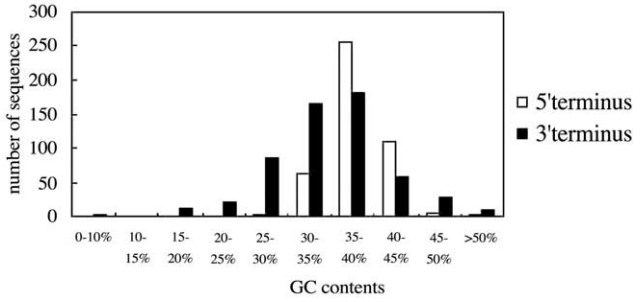

Fig. 3. Distribution of GC contents of the ESTs. The GC contents of all the 2153 ESTs were calculated for each terminus. Poly(A) tails at the 3′-terminus sequences were discarded. The number of sequences with the GC content from 0 to 50% was shown. Most 3′-terminus GC contents were concentrated on 30–40%. Whereas, those of 5′-terminus were mostly concentrated on 35–45%.

mainly consisted of open reading frames (ORFs), had 87 597 GC bases and 141 645 AT bases (38.21% GC). In contrast, the 3′-terminus sequences without poly(A) tails, most of which were considered to be 3′-untranslated regions (UTRs), had 64 073 GC bases and 117 457 AT bases (34.68% GC). The distribution of GC contents was shown in Fig. 3. The correlations between gene expression levels and GC contents were further studied. The clusters including more than 10 ESTs per cluster contained 42.21% GC content on average. However, the clusters including less than 10 ESTs per cluster contained 36.10% GC content on average, suggesting that higher GC levels could result in higher expression of genes. This phenomenon found in jellyfish genes could indicate a selective advantage for translational efficiency [7].

3.2. Overall identification and categorization of the sequences

The sequences of 1016 clusters were submitted to BLAST X for homologous analysis. According to the analytic results, 282 clusters (amount to 27.8% of the total) were highly matched ($P < 10^{-7}$) to the proteins that had been previously identified with distinct functions. These 282 proteins could be categorized into three major classes, basically according to previous classification [8] shown in Table 2. Class A (AI–AX), which contained 242 proteins (711 ESTs), was associated with housekeeping proteins; Class B (BI–BIII), which contained 30 proteins (93 ESTs), was associated with cell–cell communication; Class C (CI–CIII), which contained 10 proteins (21 ESTs), was associated with transcription factors and other gene regulatory proteins. The average number of hits/protein was 2.93 (825/282).

There were still 163 clusters (approximately 16.0% of the total) classified into Class DI. These clusters were matched to the ESTs and reported proteins mostly from *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*, whose functions had not been fully defined due to lacking of enough information. The remaining 571 clusters (56.2% of the total) were categorized into Class DII with no significant similarities to any known sequences. mRNA transcripts as represented in each category were shown in Table 3, and most of them were reported in jellyfish for the first time.

3.3. ESTs related to head patterning and neurogenesis

Jellyfish is one of simple animals that are composed of two cell layers and have nerve cells but not organs. Recent analyses have indicated that some of the developmental pathways leading to anterior patterning in bilaterians were already

Table 2
The number of different genes expressed in different classes

Class		Number of clusters	Number of clones
A	Housekeeping proteins		
AI	Transportation and binding proteins for ions and other small molecules	20	83
AII	RNA processing, polymerizing, splicing and binding proteins, and enzymes	10	16
AIII	Cell replication, histones, cyclins and allied kinases, DNA polymerases, topoisomerases, DNA modification	6	7
AIV	Cytoskeleton and membrane proteins	26	54
AV	Protein synthesis co-factors, tRNA synthetases, ribosomal proteins	69	317
AVI	Intermediary synthesis and catabolism enzymes	51	87
AVII	Stress response, detoxification and cell defense proteins	25	66
AVIII	Protein degradation and processing, proteases	23	63
AIX	Transportation and binding proteins for proteins and other macromolecules	8	11
AX	Apoptosis-related proteins	4	7
Total		242	711
B	Cell–cell communication		
BI	Signaling receptors, including cytokine and hormone receptors, and signaling ligands	12	17
BII	Intracellular signal transduction pathway molecules including kinases and signal intermediates	16	74
BIII	Extracellular matrix proteins and cell adhesion	2	2
Total		30	93
C	Transcription factors and other gene regulatory proteins		
CI	Sequence-specific DNA-binding proteins	8	18
CII	Proteins involved in gene regulation	1	2
CIII	Chromatin proteins other than AIII with regulatory function	1	1
Total		10	21
D	Miscellaneous		
DI	Not enough information to classify	163	457
DII	Not significant similarities to known proteins	571	871
Total		734	1328
Total		1016	2153

Table 3
EST sequence similarities, gene description and probability of occurrence by chance

Class	Cluster ID	Accession number	Database entry name	Organism	Probability
AI	0058	AAC19132	ferritin	*O. moubata*	2.00E-60
	0123	Q24439	ATPase oligomycin sensitivity conferral protein	*D. melanogaster*	2.00E-31
	0125	NP_445811	copper transport protein ATX1	*R. norvegicus*	4.00E-13
	0209	P55277	H^+-ATPase V-type subunit	*Heliothis virescens*	2.00E-53
	0256	MCJZR	calmodulin	*Renilla reniformis*	3.00E-10
	0280	P35317	Na,K-ATPase alpha subunit	*Hydra vulgaris*	2.00E-62
	0644	AAF78964	VDAC	*S. scrofa*	5.00E-10
	0700	P05044	sorcin 22-kd protein	*Cricetulus longicaudatus*	5.00E-22
	0791	NP_077128	ethanol induced 6	*M. musculus*	1.00E-27
AII	0139	XP_012145	small nuclear ribonucleoprotein polypeptide F	*H. sapiens*	2.00E-32
	0200	NP_057394	DNA directed RNA polymerase III polypeptide K	*H. sapiens*	2.00E-46
	0263	NP_067000	U6 snRNA-associated Sm-like protein	*H. sapiens*	1.00E-42
	0367	AAH05603	TBP-associated factor I	*M. musculus*	2.00E-26
	0422	NP_524416	elongin B	*D. melanogaster*	3.00E-26
	0896	NP_006223	RNA polymerase II	*H. sapiens*	4.00E-60
AIII	0003	P35059	histone H4	*Acropora formosa*	3.00E-44
	0485	BAB12726	DNA polymerase epsilon subunit B	*X. laevis*	1.00E-58
	0742	P14381	transposable element Txlc protein 2	*X. laevis*	2.00E-11
	0874	NP_006182	proliferation-associated 2G4	*H. sapiens*	2.00E-78
	0878	NP_032605	meiosis expressed gene 1	*M. musculus*	1.00E-29
AIV	0027	AAF44064	calponin-like protein Chd64	*D. melanogaster*	8.00E-19
	0044	CAB65408	profilin	*Suberites domuncula*	4.00E-39
	0128	NP_080832	dynein light chain 2	*H. sapiens*	2.00E-42
	0134	CAB40910	ribosome-associated membrane protein RAMP4	*R. norvegicus*	7.00E-18
	0148	P80585	tubulin-folding cofactor A	*G. gallus*	2.00E-28
	0159	AAG31472	cryptophyte-like actin	*Pyrenomonas helgolandii*	3.00E-16
	0253	BAB24240	dynein-associated protein HKM23	*M. musculus*	5.00E-34
	0313	NP_005710	actin-related protein 2/3 complex, subunit 3	*H. sapiens*	7.00E-39
	0356	AAF05616	dynactin subunit p27	*M. musculus*	1.00E-49
	0439	XP_030355	membrane component, surface marker 1	*H. sapiens*	1.00E-09
	0484	Q9N2R3	heat stable allergen tropomyosin	*Charybdis feriatus*	2.00E-16
	0507	XP_009501	myosin regulatory light chain 2	*H. sapiens*	2.00E-75
	0559	AAK58683	alpha tubulin	*Chironomus tentans*	1.00E-126
	0667	NP_002466	myosin alkali light chain 1 slow-twitch muscle A	*H. sapiens*	1.00E-29
	0761	BAB22582	microtubule-associated proteins 1A/1B light chain	*M. musculus*	2.00E-26
	0925	NP_001093	actinin, alpha 1	*H. sapiens*	7.00E-83
AV	0038	AAF64457	ribosomal protein L18	*Oreochromis niloticus*	4.00E-64
	0049	AAK92184	ribosomal protein S15	*S. frugiperda*	1.00E-40
	0069	P10160	translation initiation factor eIF-5A	*O. cuniculus*	6.00E-59
	0080	I51237	translation elongation factor EF-1 gamma	*X. laevis*	1.00E-43
	0153	AAF61073	ribosomal protein large P2	*Paralichthys olivaceus*	2.00E-14
	0177	CAA70221	elongation factor 1A	*G. cydonium*	9.00E-77
	0213	NP_523462	overgrown hematopoietic organs at 23B	*D. melanogaster*	4.00E-24
	0249	A53221	acidic ribosomal protein P1-hydromedusa	*Polyorchis penicillatus*	7.00E-25
	0258	AAF31449	60S acidic ribosomal protein P0	*Sarcophaga crassipalpis*	4.00E-58
	0448	NP_079750	mitochondrial ribosomal protein S14	*M. musculus*	3.00E-20
	0521	XP_056380	mitochondrial ribosomal protein L24	*H. sapiens*	2.00E-20
	0522	2IF1	human translation initiation factor Eif1	*H. sapiens*	1.00E-33
	0631	NP_061269	eukaryotic translation initiation factor 3, subunit 2	*M. musculus*	6.00E-60
	0828	P09445	translation elongation factor eEF-2	*Cricetus cricetus*	4.00E-36
AVI	0007	NP_064527	NADH:ubiquinone oxidoreductase MLRQ subunit	*H. sapiens*	4.00E-14
	0098	NP_031775	cytochrome *c* oxidase subunit VIIc	*M. musculus*	5.00E-68
	0141	BAB31369	cytochrome *c* oxidase	*M. musculus*	2.00E-35
	0202	AAK51137	nucleoside diphosphate kinase	*H. vulgaris*	2.00E-53
	0302	NP_172540	phosphoribosyl diphosphate synthase	*A. thaliana*	8.00E-20
	0317	Q95123	succinate dehydrogenase cytochrome B subunit	*B. taurus*	2.00E-15
	0355	AAH02668	peroxisomal D3,D2-enoyl-CoA isomerase	*H. sapiens*	7.00E-24
	0372	Q9W719	hypoxanthine guanine phosphoribosyl transferase	*G. gallus*	3.00E-51
	0378	P23935	NADH dehydrogenase complex I 13K-B chain	*B. taurus*	2.00E-29
	0396	NP_115990	methylmalonyl-CoA epimerase	*H. sapiens*	4.00E-49
	0412	NP_056414	*N*-acetylglucosamine-phosphate mutase	*H. sapiens*	8.00E-38
	0509	NP_003869	gamma-glutamyl hydrolase	*H. sapiens*	3.00E-18
	0526	BAB22060	glyoxalase-1 (EC 4.4.1.5)	*M. musculus*	5.00E-67
	0556	XP_054951	similar to polypeptide GalNAc transferase T1	*H. sapiens*	1.00E-21
	0587	A59047	phospholipase A2 (EC 3.1.1.4) conodipine M	*Conus magus*	7.00E-09
	0593	NP_536693	beta 1,3-galactosyltransferase	*M. musculus*	3.00E-18
	0634	AAH05270	NADH dehydrogenase Fe-S protein 4	*H. sapiens*	2.00E-43
	0647	NP_037466	GDP-mannose pyrophosphorylase B, isoform 1	*H. sapiens*	2.00E-23
	0649	NP_081105	5,10-methenyltetrahydrofolate synthetase	*M. musculus*	4.00E-45
	0682	CAC87049	mitochondrial NADH:ubiquinone oxidoreductase	*B. taurus*	8.00E-28
	0733	O14463	thioredoxin II	*S. pombe*	9.00E-22

Table 3 (*Continued*).

Class	Cluster ID	Accession number	Database entry name	Organism	Probability
	0771	AAF57793	glycosyltransferase	*D. melanogaster*	3.00E-17
	0778	NP_000572	glutathione peroxidase 1	*H. sapiens*	1.00E-12
	0928	BAA34050	endoglucanase 2	*Reticulitermes speratus*	1.00E-48
	0937	NP_057184	N-terminal acetyltransferase complex ard1 subunit	*H. sapiens*	2.00E-69
	0948	NP_000273	propionyl-coenzyme A carboxylase	*H. sapiens*	2.00E-19
	0968	XP_010849	sulfotransferase 1C2	*H. sapiens*	5.00E-27
AVII	0051	NP_036151	peroxiredoxin V	*M. musculus*	4.00E-43
	0090	BAB22368	microsomal glutathione S-transferase	*M. musculus*	3.00E-21
	0146	O15532	selenoprotein W	*H. sapiens*	2.00E-22
	0186	P00007	cytochrome *c*	*H. amphibius*	2.00E-46
	0211	AAB86581	heat shock protein 10	*G. gallus*	7.00E-30
	0216	P81926	superoxide dismutase [Cu-Zn]	*Halocynthia roretzi*	5.00E-54
	0261	1BCC	chain J, cytochrome *Bc1* complex	*G. gallus*	2.00E-12
	0305	CAC38768	cytochrome *P*450 17alpha-hydroxylase	*Pimephales promelas*	5.00E-18
	0354	NP_495431	superoxide dismutase	*C. elegans*	2.00E-39
	0690	CAA04673	heat shock protein 70	*Oreochromis mossambicus*	4.00E-27
	0720	P79382	glutathione S-transferase	*S. scrofa*	9.00E-29
	0872	JC7132	heat shock protein 70	*Rhizopus stolonifer*	2.00E-23
AVIII	0004	NP_005720	peptidylprolyl isomerase F (cyclophilin F)	*H. sapiens*	3.00E-62
	0091	NP_003338	ubiquitin-conjugating enzyme E2L 3	*H. sapiens*	6.00E-60
	0121	XP_029886	ubiquitin	*H. sapiens*	1.00E-30
	0269	O04438	signal recognition particle subunit 9	*Zea mays*	7.00E-20
	0275	S38529	multicatalytic endopeptidase complex chain XC8	*Xenopus* sp.	7.00E-48
	0283	P28687	signal peptidase 23 kDa subunit	*G. gallus*	7.00E-46
	0447	Q9PTW9	proteasome subunit alpha type 7	*Carassius auratus*	4.00E-81
	0466	AAA92361	metalloproteinase 1	*H. vulgaris*	8.00E-47
	0666	NP_179569	clathrin assembly protein	*A. thaliana*	6.00E-47
	0889	AAA74445	cathepsin B-like protease	*Urechis caupo*	3.00E-35
AIX	0119	BAB22381	translocon-associated protein	*M. musculus*	1.00E-58
	0219	Q9PRL8	acyl-coenzyme A-binding protein	*G. gallus*	1.00E-24
	0306	NP_003896	amyloid beta precursor protein-binding protein 1	*H. sapiens*	3.00E-27
	0583	P56384	ATP synthase lipid-binding protein	*M. musculus*	3.00E-20
	0984	XP_032443	importin beta-3 subunit	*H. sapiens*	8.00E-54
AX	0229	Q29036	DAD1, defender against cell death 1	*S. scrofa*	1.00E-42
	0549	NP_035181	programmed cell death 6	*M. musculus*	6.00E-52
	0633	AAD32953	DNA fragmentation factor DFF35	*H. sapiens*	4.00E-09
BI	0228	NP_067494	endothelial differentiation-related factor 1	*M. musculus*	1.00E-49
	0235	NP_009209	GABA-A receptor-associated protein	*H. sapiens*	2.00E-60
	0316	XP_045613	EH domain containing 1	*H. sapiens*	2.00E-23
	0343	XP_028263	adaptor-related protein complex 2, sigma 1 subunit	*H. sapiens*	2.00E-71
	0514	AAH02131	ARF-like 2 binding protein BART1	*M. musculus*	2.00E-38
	0539	BAB13535	Hym-323	*Hydra magnipapillata*	5.00E-12
	0728	NP_009216	ganglioside expression factor 2 (GEF-2)	*H. sapiens*	6.00E-44
	0744	CAA56454	dopamine receptor	*Takifugu rubripes*	2.00E-20
	0749	O42248	receptor for activated protein kinase C	*D. rerio*	1.00E-92
	0870	AAK27327	TSG protein	*X. laevis*	3.00E-27
BII	0176	BAA82135	dickkopf-1	*D. rerio*	4.00E-08
	0212	AAC18962	GTPase cRhoA	*G. gallus*	3.00E-66
	0215	AAF55974	RAS small monomeric GTPase (EC:3.6.1.47)	*D. melanogaster*	1.00E-71
	0408	P40122	cyclase-associated protein	*Chlorohydra viridissima*	3.00E-14
	0511	BAA95482	glia maturation factor beta	*C. carpio*	1.00E-37
	0527	P17713	tyrosine-protein kinase STK (P57-STK)	*H. vulgaris*	6.00E-65
	0614	NP_005767	cornichon protein	*H. sapiens*	1.00E-32
	0624	BAB22727	WD domain, G-beta repeat containing protein	*M. musculus*	5.00E-14
	0710	AAD48958	similar to Pfam families (protein kinase)	*A. thaliana*	6.00E-11
	0959	BAA75197	membrane guanylyl cyclase	*Brissus agassizii*	4.00E-17
	0886	NP_079595	homolog to MBP-1 interacting protein-2A	*M. musculus*	3.00E-47
	0817	NP_492636	tetraspanin	*C. elegans*	2.00E-18
BIII	0328	NP_524645	extracellular matrix protein papilin	*D. melanogaster*	2.00 E-08
	0414	JH0675	restrictin precursor	*G. gallus*	3.00E-48
CI	0149	XP_038290	transcription factor BTF3	*H. sapiens*	4.00E-45
	0230	BAB33387	Xp8 (encodes a transcription factor)	*X. laevis*	9.00E-12
	0435	AAB27565	metallothionein-I gene transcription activator	*Mus* sp.	2.00E-21
	0496	AAK52090	Tax-responsive element binding protein 107	*G. gallus*	1.00E-42
	0674	AAC32813	universal minicircle sequence binding protein	*Crithidia fasciculata*	6.00E-09
	0936	XP_002435	prolactin regulatory element binding protein	*H. sapiens*	4.00E-15
CII	0150	NP_004441	enhancer of rudimentary homolog	*H. sapiens*	3.00E-36
CIII	0535	T30534	chromosome segregation protein SMC1	*T. rubripes*	2.00E-08

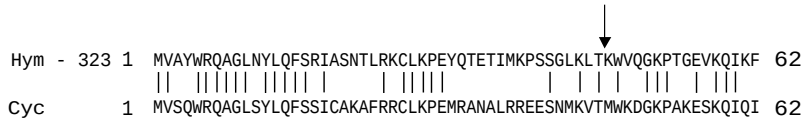

Fig. 4. Alignment of Hym-323 and jellyfish Cyc. Numbering of amino acid sequence began at the first residue of the protein precursor. Vertical lines indicated residues identical in Hym-323 and Cyc. Arrow designated a possible processing site at the N-terminus of the peptide.

present in their common ancestor and this observation was also found in the rudimentary cnidarian head [2].

Five proteins associated with head patterning and neurogenesis were found in the tentacle cDNA library of *C. capillata* (Table 4). Among these proteins, dickkopf-1 (dkk-1), which acted as a potent inhibitor of *Wnt* signaling [9], was identified as a head inducer and played an important role in patterning the anterior head [10]. Glial maturation factor β (GMF β), which was found exclusively in the nervous system, could cause differentiation of brain cells and stimulation of neural regeneration as well as what Glia could [11–14]. Some other proteins that expressed at high levels in the central nervous system [15–17] and brain [18,19] were also found in our study such as ganglioside expression factor 2, *Xenopus laevis* homolog of p8 (Xp8) and IMPACT.

The presence of those mentioned proteins supported the hypothesis [2,20] that the ancestral head organizer activity might not have required a competent mesoderm layer, and it was conceivable that diploblasts and triploblasts (animals developed from two- and three-layered embryos, respectively) might share common mechanisms for anterior specification. Whereas, it was very interesting to be noticed that the proteins related to head patterning and neurogenesis were found in jellyfish tentacles (but not the head), since tentacles did not obviously share the functional purpose of head. The further study about functions of these proteins in tentacles would reveal the significance of cnidarians in studying the process of head patterning.

3.4. ESTs relevant to foot patterning and gastrulation

Commonly the signaling molecules regulating pattern-forming processes in animals have been shown to be proteins [21]. However, some evidences have recently indicated that some small peptides could also affect patterning processes in cnidarians [21–23]. In *Hydra*, two functionally similar peptides, Hym-323 and Hym-346, which were produced and localized in epithelial cells, could enhance foot formation by increasing foot activation potential (or lowering positional value) in the peptide-treated tissue [23]. The peptide, Hym-323, is 16 amino acids long but its precursor protein is 62 amino acids long. A novel protein that had never been found in jellyfish before was identified from our library, designated as Cyc (*C. capillata*), with 50% amino acid identity to Hym-323, suggesting that the protein probably participated in foot patterning. The alignment of Cyc precursor and Hym-323 precursor was shown in Fig. 4, where the possible processing site at the N-terminus was indicated.

As we all know, jellyfish is a diploblast organism. Therefore, it has no dedicated mesodermal germ layer, which is the third cell layer characteristic for triploblasts [24,25]. Recently questions were raised on the phylogenetic position of mesoderm and of Cnidaria since a Twist homolog has been identified in jellyfish [25]. Besides that, a crucial protein for proper gastrulation and mesoderm formation [26], twisted gastrulation (TSG) protein, was found in our library. A relevant study showed that Twist in jellyfish, acting a likely function with that in flies, was targeting genes required for the processes typical in gastrulation, such as cell shape changes, cell migration and cell proliferation [24].

3.5. ESTs homologous to VDAC

Members of the phylum Cnidaria are the lowest organisms to possess a neuromuscular system in terms of evolution, and they are thought to be the first ones that contain voltage-dependent Na^+, Ca^{2+} channels so far [1,3]. Therefore, jellyfish occupies a critical position in the evolution of Na^+, Ca^{2+} channels. However, the VDACs, which play a key role in regulatory voltage decrease [27], had never been reported in jellyfish before. From the library we constructed, a VDAC was found with more than 50% amino acid identity to the similar channels in a great variety of organisms, which span the phylogenetic tree. VDACs belonging to a multigene family are widely represented in many cell types. Their permeability to small molecules such as adenosine triphosphate (ATP) and adenosine diphosphate (ADP) [28] is regulated by NADH and NADPH, suggesting that VDACs are very important in the mitochondrial respiration [29]. Furthermore, VDACs control the release of cytochrome *c* into the cytoplasm, suggesting that they act as a crucial regulator in the release of death-promoting factors in the cytosol [29]. To evaluate the phylogenetic position of jellyfish's voltage-dependent anion channel in the VDAC family, we analyzed the relationship between 16 VDAC sequences and the jellyfish VDAC using the neighbor-joining methods (Fig. 5) [30]. According to the resulting tree, the VDAC derived from jellyfish, *C. capillata*, resided on a separate evolutionary branch of the tree and showed a closer relationship to the ones of *Saccharomyces cerevisiae* and *Caenorhabditis elegans* than to that of the others. It could be concluded that there would exist a distinct functional difference between the VDAC originated from jellyfish and those

Table 4
Proteins related to head patterning and neurogenesis

Cluster ID	Accession number	Gene description	Organisms	Probability
0176	BAA82135	dickkopf-1	*D. rerio*	4.00E-08
0230	BAB33387	Xp8	*X. laevis*	9.00E-12
0511	BAA95482	glia maturation factor beta	*C. carpio*	1.00E-37
0424	AAG35736	IMPACT	*H. sapiens*	3.00E-14
0731	NP_009216	ganglioside expression factor 2	*H. sapiens*	6.00E-44

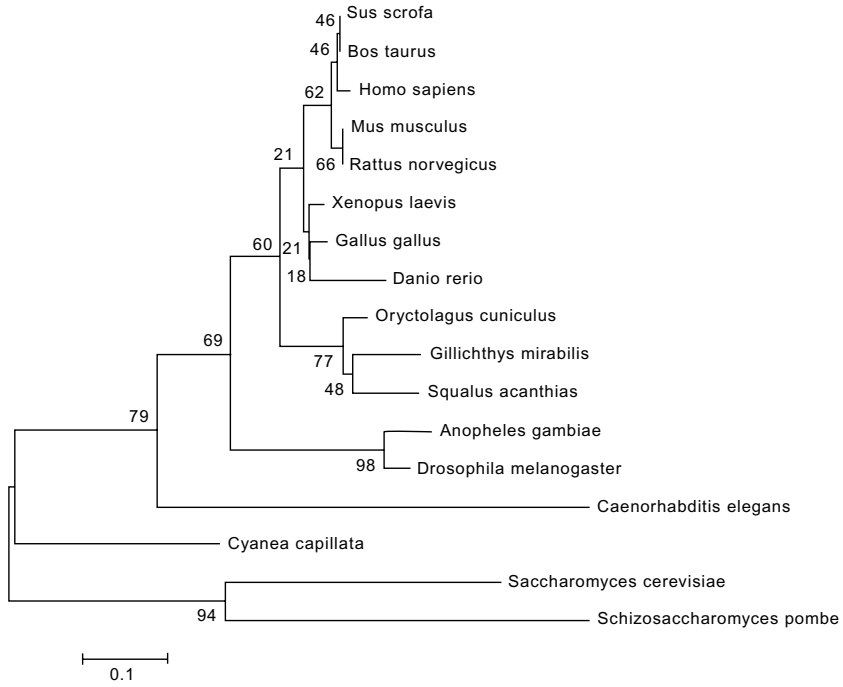

Fig. 5. Phylogenetic analysis of VADC sequences. Amino acid sequences were aligned using Clustal W. The aligned sequences were used to construct phylogenetic trees using the neighbor-joining method, as implemented in MEGA version 2.1. The numbers at nodes indicated bootstrap values. Bar showed genetic distance of 0.1. The tree shown in the figure was identical in branching pattern to a tree produced using a maximum parsimony algorithm. Accession numbers of published sequences: *Sus scrofa* (AF268462), *Bos taurus* (AAF80102), *H. sapiens* (XP_016625), *M. musculus* (AAC13321), *Rattus norvegicus* (BAB13474), *X. laevis* (P81004), *Gallus gallus* (AAF73513), *Danio rerio* (AAH42329), *Oryctolagus cuniculus* (AAF22837), *Gillichthys mirabilis* (AAL24505), *Squalus acanthias* (AAF65254), *Anopheles gambiae* (AAN16031), *D. melanogaster* (AAC02635), *C. elegans* (NP_501211), *S. cerevisiae* (NP_014343), *Schizosaccharomyces pombe* (NP_594661).

from other species, and the former might hold a lower position in the evolution.

3.6. ESTs relevant to venom active components and homologous to disease-related genes

Nematocytes are the stinging cells in the ectoderm of cnidarians that contain a unique and very large secretory vesicle, the capsular nematocyst, which can eject its contents in a spectacular discharge process. Jellyfish is a phylum of marine invertebrates equipped with nematocysts [31]. Many toxins were isolated from the nematocyst venom such as CAH1 [4], SNTX [32], CrTX [33] and so on. But most of the toxins

Table 5
Proteins related to disease

Cluster ID	Accession number	Gene description	Organisms	Probability
0033	BAB72100	leucine zipper and ICAT homologous protein LZIC	*H. sapiens*	1.00E-59
0074	A44367	tumor-specific transplantation antigen P198 homolog p23	*B. taurus*	7.00E-70
0097	Q9PSN3	bilineobin, coagulant enzyme	*Agkistrodon bilineatus*	7.00E-35
0105	AAK84394	translationally controlled tumor protein	*Branchiostoma belcheri*	2.00E-21
0232	BAB33421	senescence-associated protein	*Pisum sativum*	2.00E-43
0387	NP_060934	hematopoietic stem/progenitor cells protein MDS0	*H. sapiens*	5.00E-21
0538	BAB22734	HBV X interacting protein	*M. musculus*	5.00E-11
0547	I51694	transforming protein bmi-1	*X. laevis*	1.00E-43
0596	CAC38780	allograft inflammatory factor 1	*S. domuncula*	6.00E-45
0760	P48594	squamous cell carcinoma antigen 2	*H. sapiens*	7.00E-24
0835	AAK74192	coagulation factor VII	*D. rerio*	2.00E-27
0844	NP_000382	lysosomal pepstatin insensitive protease	*H. sapiens*	1.00E-08
0894	NP_005972	sarcoma amplified sequence	*H. sapiens*	1.00E-09
0913	NP_033866	Bcl2-associated athanogene 1	*M. musculus*	2.00E-16

Table 6
Proteins homologous to proteinase inhibitor

Cluster ID	Accession number	Gene description	Organisms	Probability
0013	S45677	proteinase inhibitor PAPI I	*P. leniusculus*	3.00E-28
0026	AAL60465	antithrombin	*Struthio camelus*	3.00E-14
0104	NP_507993	Kunitz/bovine pancreatic trypsin inhibitor domain	*C. elegans*	5.00E-10
0127	AAB69857	cystatin-type cysteine proteinase inhibitor	*C. carpio*	2.00E-09
0226	UDHUB	cystatin B	*H. sapiens*	3.00E-16

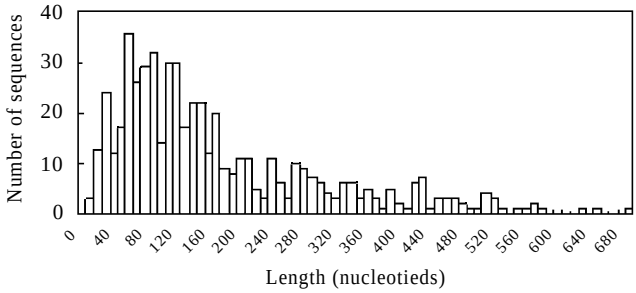

Fig. 6. Statistical analysis of the sequences showing no matches to known genes. We analyzed 571 consensus sequences without matches in GenBank. The largest ORF was determined. The nucleotide lengths between start codons and stop codons were calculated and classified into 10 bp bins. The number of clusters was calculated accordingly. Of those, 33 sequences without stop codons were not shown in this histogram and 538 sequences were statistically estimated. The peak distribution of the lengths is between 30 and 180 bp.

belonged to three classes: cytolytic proteins, phospholipase A2 and acid or alkaline proteases. They were the basic active components identified in the venom of jellyfish [34]. In the tentacle cDNA library of *C. capillata*, each representation of the three classes, hemolysin C, phospholipase A2 (EC 3.1.1.4) and cysteine proteinase, was found respectively. In general, the hemolysin C and phospholipase A2 were derived from the nematocysts, and the proteinase might be derived from the cells surrounding the nematocysts or from the intracapsular content of the nematocysts [34]. In addition, an inflammatory factor and an antithrombin were also found in the library, which perhaps could explain the hurting mechanism caused by jellyfish stinging.

Besides the active components in venom, some proteins related to diseases were also detected in the library (Table 5). All of them were reported in jellyfish for the first time. Among these proteins, seven were tumor-related, including pro-oncogene bmi-1, tumor-specific transplantation antigen, translationally controlled tumor protein and so on. Some coagulant enzymes and senescence-associated proteins were also expressed in jellyfish, and further studies on these proteins could generate potential applications in medicine.

3.7. ESTs homologous to proteinase inhibitor

Interestingly, five kinds of proteinase inhibitors were found in jellyfish (Table 6), and cystatin, one of them, showed the highest expression level among all clusters (Table 1). According to the previous studies, these inhibitors acted as a regulator to protect cells from unfavorable proteolysis by intracellular and external proteinases [35–38]. Cystatin played as a biological defender against outer invaders, for instance, bacteria [39–42], fungi [43] and viruses [44,45]. It could be inferred that the identified cystatin might participate in the jellyfish defense system for the possibility of an antimicrobial protein.

3.8. ESTs identified no significant matches to known genes

ORFs of 571 clusters (56.2% of the total clusters) without significant GenBank matches were determined and analyzed according to the start codons and stop codons. The lengths of such ORFs were classified into 10 bp bins [46], and the statistical analysis results were shown in Fig. 6. The data indicated that the peak distribution of the ORF lengths was between 30 and 180 bp, and a large fraction lay between 60 and 90 bp. The high abundance of these sequences might correspond to the complexity of mRNA transcripts stored in the tentacle of *C. capillata*.

In conclusion, a total of 1016 clusters consisting of 2153 ESTs derived from the tentacle of *C. capillata* were determined in the present work. A series of genes related to head patterning and neurogenesis were identified in our study, which supported the hypothesis that head organization should occur in cnidarians. VDAC was also found, which might provide valuable insights into the molecular mechanism of ion channel evolution. Besides the active components of nematocyst venom, disease-related proteins were also detected in the study along with five kinds of proteinase inhibitors with high expression levels.

Further functional analyses on genes identified from the tentacle cDNA library of *C. capillata* will provide more information about the molecular mechanism that is involved in the evolution and development of cnidarians.

Acknowledgements: We thank Guijun Zhao, Ning Zhang, Bing Xu and Junjie Zhu for their help with the data analysis. This work was supported by the State High-Tech Research and Development Program (863 Program) of The Ministry of Science and Technology of China (2001AA626010), (819-06-06), and the key project from the National Natural Science Foundation of China (69935020).

References

[1] Jeziorski, M.C., Greenberg, R.M., Clark, K.S. and Anderson, P.A.V. (1998) J. Biol. Chem. 273, 22792–22799.
[2] Galliot, B. and Miller, D. (2000) Trends Genet. 16, 1–5.
[3] Anderson, P.A.V., Holman, M.A. and Greenberg, R.M. (1993) Proc. Natl. Acad. Sci. USA 90, 7419–7423.
[4] Chung, J.J., Ratnapala, L.A., Cooke, I.M. and Yanagihara, A.A. (2001) Toxicon 39, 981–990.
[5] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L. (2000) Nucleic Acids Res. 28, 15–18.
[6] Makabe, K.W., Kawashima, T., Kawashima, S., Minokawa, T., Adachi, A., Kawamura, H., Ishikawa, H., Yasuda, R., Yamamoto, H., Kondoh, K., Arioka, S., Sasakura, Y., Kobayashi, A., Yagi, K., Shojima, K., Kondoh, Y., Kido, S., Tsujinami, M., Nishimura, M., Takahashi, M., Nakamura, T., Kanehisa, M., Ogasawara, M., Nishikata, T. and Nishida, H. (2001) Development 128, 2555–2567.
[7] Konu, O. and Li, M.D. (2002) J. Mol. Evol. 54, 35–41.
[8] Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., Hamaguchi, M., Ishikawa, H., Chiba, S., Imai, K., Kano, S., Murakami, S.D., Nakayama, A., Nishino, A., Sasakura, Y., Satoh, G., Shimotori, T., Shin-I, T., Shoguchi, E., Suzuki, M.M., Takada, M.,

Utsumi, N., Yoshida, M., Saiga, H., Kohara, Y. and Satoh, N. (2001) Development 128, 2893–2904.
[9] Monaghan, A.P., Kioschis, P., Wu, W., Xuniga, A., Bock, D., Poustka, A., Delius, H. and Niehrs, N. (1999) Mech. Dev. 87, 45–56.
[10] Krupnik, V.E., Sharp, J.D., Jiang, C., Robison, K., Chickering, T.W., Amaravadi, L., Brown, D.E., Guyot, D., Mays, G., Leiby, K., Chang, B., Duong, T., Goodearl, A.D.J., Gearing, D.P., Sokol, S.Y. and McCarthy, S.A. (1999) Gene 238, 301–313.
[11] Fujiki, K., Shin, D.H., Nakao, M. and Yano, T. (2000) Fish Shellfish Immun. 10, 643–650.
[12] Asai, K., Fujita, K., Yamamoto, M., Hotta, T., Morikawa, M., Kokubo, M., Moriyama, A. and Kato, T. (1998) Biochim. Biophys. Acta 1396, 242–244.
[13] Nishiwaki, A., Asai, K., Tada, T., Ueda, T., Shimada, S., Ogura, Y. and Kato, T. (2001) Mol. Brain Res. 95, 103–109.
[14] Liu, L.X., Xu, H., Weller, P.F., Shi, A. and Debnath, I. (1997) Gene 186, 1–5.
[15] Shindler, K.S. and Roth, K.A. (1996) Dev. Brain Res. 92, 199–210.
[16] Pyo, H., Joe, E.H., Jung, S., Lee, S.H. and Jou, I. (1999) J. Biol. Chem. 274, 34584–34589.
[17] Igarashi, T., Kuroda, H., Takahashi, S. and Asashima, M. (2001) Dev. Growth Differ. 43, 693–698.
[18] Sigala, S., Missale, C. and Spano, P. (1997) Eur. J. Pharmacol. 336, 107–112.
[19] Kosaki, K., Suzuki, T., Kosaki, R., Yoshihashi, H., Itoh, M., Goto, Y. and Matsuo, N. (2001) Mol. Psychiatry 6, 87–91.
[20] Mackie, G.O. and Meech, R.W. (2000) J. Exp. Biol. 203, 1797–1807.
[21] Grens, A., Shimizu, H., Hoffmeister, S.A.H., Bode, H.R. and Fujisawa, T. (1999) Development 126, 517–524.
[22] Hoffmeister, S.A. (1996) Development 122, 1941–1948.
[23] Harafuji, N., Takahashi, T., Hatta, M., Tezuka, H., Morishita, F., Matsushima, O. and Fujisawa, T. (2001) Development 128, 437–446.
[24] Castanon, I. and Baylies, M.K. (2002) Gene 287, 11–22.
[25] Spring, J., Yanze, N., Middel, A.M., Stierwald, M., Groger, H. and Schmid, V. (2000) Dev. Biol. 228, 363–375.
[26] Mason, E.D., Williams, S., Grotendorst, G.R. and Marsh, J.L. (1997) Mech. Dev. 64, 61–75.
[27] Lemonnier, L., Vitko, Y., Shuba, Y.M., Vanden Abeele, F., Prevarskaya, N. and Skryma, R. (2002) FEBS Lett. 521, 152–156.
[28] Hinsch, K.D., Asmarinah, Hinsch, E. and Konrad, L. (2001) Biochim. Biophys. Acta 1518, 329–333.
[29] Roosens, N., Al Bitar, F., Jacobs, M. and Homble, F. (2000) Biochim. Biophys. Acta 1463, 470–476.
[30] David-Watine, B., Goblet, C., de Saint Jan, D., Fucile, S., Devignot, V., Bregestovski, P. and Korn, H. (1999) Neuroscience 90, 303–317.
[31] Rottini, G., Gusamni, L., Parovel, E., Avian, M. and Patriarca, P. (1995) Toxicon 33, 315–326.
[32] Cao, C.J., Eldefrawi, M.E., Eldefrawi, A.T., Burnett, J.W., Mioduszewki, R.J., Menking, K.E. and Valdes, J.J. (1998) Toxicon 36, 269–281.
[33] Nagai, H., Takuwa, K., Nakao, M., Ito, E., Miyake, M., Node, M. and Nakajima, T. (2000) Biochem. Biophys. Res. Commun. 275, 582–588.
[34] Gusmani, L., Avian, M., Galil, B., Patriarca, P. and Rottini, G. (1997) Toxicon 35, 637–648.
[35] Jiang, H. and Kanost, M.R. (1997) J. Biol. Chem. 272, 1082–1087.
[36] Stubbs, M.T., Morenweiser, R., Sturzebecher, J., Bauer, M., Bode, W., Huber, R., Piechottka, G.P., Matschiner, G., Sommerhoff, C.P., Gritz, H. and Auerswald, E.A. (1997) J. Biol. Chem. 272, 19931–19937.
[37] Saunders, D.N., Buttigieg, K.M.L., Gould, A., McPhun, V. and Baker, M.S. (1998) J. Biol. Chem. 273, 10965–10971.
[38] Sun, J., Rose, J.B. and Bird, P. (1995) J. Biol. Chem. 270, 16089–16096.
[39] Agarwala, K.L., Kawabata, S., Hirata, M., Miyagi, M., Tsunasawa, S. and Iwanaga, S. (1996) J. Biochem. (Tokyo) 119, 85–94.
[40] Blankenvoorde, M.F., van't Hof, W., Walgreen-Weterings, E., van Steenbergen, T.J., Brand, H.S., Veerman, E.C. and Nieuw Amerongen, A.V. (1998) Biol. Chem. 379, 1371–1375.
[41] Kasprzykowski, F., Schalen, C., Kasprzykowska, R., Jastrzebska, B. and Grubb, A. (2000) APMIS 108, 473–481.
[42] Takahashi, M., Tezuka, T., Korant, B. and Katunuma, N. (1999) Biofactors 10, 339–345.
[43] Soares-Costa, A., Beltramini, L.M., Thiemann, O.H. and Henrique-Silva, F. (2002) Biochem. Biophys. Res. Commun. 296, 1194–1199.
[44] Korant, B.D., Brzin, J. and Turk, V. (1985) Biochem. Biophys. Res. Commun. 127, 1072–1076.
[45] Bjorck, L., Grubb, A. and Kjellen, L. (1990) J. Virol. 64, 941–943.
[46] Lee, Y.H., Huang, G.M., Cameron, R.A., Graham, G., Davidson, E.H., Hood, L. and Britten, R.J. (1999) Development 126, 3857–3867.