Korean Journal of Malacology, Vol. 22(2): 109-113, December 30, 2006

생물정보학을 이용한 연체동물의 NLS (Nuclear Localization Signals) 포함 단백질의 분석

이용석, 강세원, 조용훈[1], 곽희철[2], 채성화[3], 최상행[3], 안인영[4], 박홍석[3], 한연수[1], 고원규

인제의대 기생충학교실 및 백인제임상의학연구소 말라리아연구부,
[1]전남대학교 농업생명과학대학 농생물학과, [2]인제의대 부산백병원 정형외과,
[3]한국생명공학연구원 유전체연구센터, [4]한국해양연구원 부설 극지연구소

Bioinformatic Analysis of NLS (Nuclear Localization Signals)-containing Proteins from Mollusks

Yong Seok Lee, Se Won Kang, Yong-Hun Jo[1], Heui-Chul Gwak[2], Sung-Hwa Chae[3], Sang-Haeng Choi[2], In-Young Ahn[3], Hong-Seog Park[2], Yeon Soo Han[1] and Weon-Gyu Kho

Department of Parasitology, Inje University College of Medicine and Department. of Malariology, PICR, Busan 614-735, Korea
[1]Department of Agricultural Biology, Chonnam National University, Gwangju 500-757, Korea
[2]Department of Orthopedic Surgery, Pusan Paik-Hospital, Inje University College of Medicine, Busan, Korea
[3]Genome Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea
[4]Polar Sciences Laboratory, Korea Ocean Research and Development Institute, Ansan 425-170, Korea

ABSTRACT

Subcellular localization of a protein containing nuclear localization signals (NLS) has been well studied in many organisms ranging from invertebrates to vertebrates. However, no systematic analysis of NLS-containing proteins available from Mollusks has been reported. Here, we describe *in silico* screening of NLS-containing proteins using the mollusks database that contains 22,138 amino acids. To screen putative proteins with NLS-motif, we used both predict NLS and perl script. As a result, we have found 266 proteins containing NLS sequences which are about 1.2% out of the entire proteins. On the basis of KOG (The eukaryotic orthologous groups) analysis, we can't predict the precise functions of the NLS-containing proteins. However, we found out that these proteins belong to several types of proteins such as chromatin structure and dynamics, translation, ribosomal structure, biogenesis, and signal transduction mechanism. In addition, we have analysed these sequences based on the classes of mollusks. We could not find many from the species that are the main subjects of phylogenetic studies. In contrast, we noticed that cephalopods has the highest number of NLS-containing proteins. Thus, we have constructed mollusks NLS database and added these information and data to the mollusks database by constructing web interface. Taken together, these information will be very useful for those who are or will be studying NLS-containing proteins from mollusks.

Keywords: NLS (nuclear localization signals), Mollusks, KOG analysis.

Received October 25, 2006; Accepted December 8, 2006
Corresponding author: Kho, Weon-Gyu
Tel: +82 (51) 890-6731 e-mail: wgkho@inje.ac.kr
1225-3480/22202
© The Malacological Society of Korea

서 론

핵 속에서 필요한 단백질이 세포질에서 핵 속으로 들어가는 메카니즘에 대하여 많은 연구가 진행되고 있다. 핵 속에서 사

용되어지는 아미노산 잔기 몇 개를 NLS (nuclear localization signal) 라고 부른다. 이 시그널을 가진 단백질은 핵공을 통해 이동되어지는데 그 과정은 다음과 같다. 첫째, NLS를 가진 단백질은 우선 importin (receptor protein) 과 결합한 후 세포질에서 핵공으로 이동되어진다. 둘째, importin + NLS 콤플렉스 단백질은 NPC (nuclear pore complex) 를 통과한 후 discharge된다. 셋째, RAN-GTP와 importin 이 결합 한 후 NLS포함 단백질은 핵 속에서 떨어뜨린다. 넷째, Ran-GTP-importin complex 는 NPC를 통해 다시 세포질로 이동되어진다. 다섯째, 세포질에 도달한 후 importin과 분리되어지면서 GTP에서 GDP로 가수분해된다 (Wayne *et al.,* 2002).

이러한 NLS의 기본 작용기작과 원리를 기초하여 최근에 다음과 같은 연구가 활발히 진행되고 있다. 첫째, NLS-like motif를 갖고 있는 새로운 유전자/단백질에 대한 기능 연구가 활발한 상황이다. 즉, 세포 내에서 합성된 NLS 영역을 갖고 있는 단백질의 최종 subcellular localization 기작에 대한 기초 연구가 진행되고 있다. 또한 척추 및 무척추동물의 대량유전자발굴 사업의 결과로 얻게 된 많은 유전자중 NLS- 포함 단백질의 기능연구로 인하여 NLS 분야의 기초연구는 더욱더 활발히 진행될 것으로 생각된다. 둘째, 이와 같은 NLS의 기초연구는 더욱더 나아가 다양한 응용연구를 통하여 산업화를 추진하고 있는 실정이다. 그 예로 우선, NLS 원리를 응용하여 transgenesis 방법을 개선하는 일이다. 일반적으로 형질전환 동물을 만들기 위하여 plasmid DNA를 주사하게 된다. Zebrafish의 경우 pasmid DNA를 알의 세포질에 주사하지만 전이유전자가 F1 세대로 전이되는 비율이 매우 낮기 때문에 이를 개선하여 하고자 plasmid DNA-NLS complex를 형성하여 주사함으로서 transgene expression의 효율을 증진된 사례가 보고되고 있다 (Collas and Alestrom, 1997; Arenal *et al.,* 2004). 또한, 약물전달 (drug delivery) 방법을 NLS domain의 특성을 활용하여 핵 내에 존재하는 목표단백질이나 DNA에 대한 접근성과 선택성 및 효율성을 개선하여 약제의 효율성을 증가시키고자 하는 노력이 있다 (Li *et al.,* 2006).

이러한 관점에서 본 연구팀은 현재까지 연체동물에서 연구된 유전자를 생물정보학적인 방법을 이용하여 NLS-like motif를 갖고 있는 유전자 혹은 단백질을 체계적으로 조사하고자 하였다. 이러한 기초 연구는 향후 육상동물과 달리 특수한 환경에서 서식하는 다양한 연체동물에서 NLS-포함 단백질에 대한 연구를 *in silico* 상에서 수행함으로서 연체동물의 유전자 연구를 효과적으로 할 수 있는 NLS 기초자료를 확보하고자 수행되었다.

실험재료 및 방법

1. 분석서버 구축

사용된 분석서버는 Intel server platform KP470P에 tetra Xeon MP 3.0 GHz CPU 시스템을 사용하였으며, 운영체제는 Linux Enterprise AS-3 를 사용하였다. 운영체제 설치 후 cgi (common gate interface) 및 perl을 사용할 수 있도록 환경설정을 한 후 NCBI-BLAST를 설치하여 서열 분석을 할 수 있도록 하였다 (Altschul *et al.,* 1990).

2. NLS 데이터베이스 구축

연체동물에서 NLS (nuclear localization signal) 를 포함한 아미노산 서열을 추출하기 위해 연체동물 전용 서열 BLAST (basic local alignment search tool) 서버에 구축되어진 데이터베이스를 활용하였다. 2006년 11월 현재까지 구축되어진 연체동물 유래 아미노산은 모두 22,138 개였다. NLS 영역을 예측하기 위해 predict NLS 프로그램 (Cokol *et al.,* 2000) 을 주축 엔진으로 하는 perl script를 만들고 script에서 활용할 수 있도록 데이터베이스를 변형하여 query로 활용하였으며, 나온 결과물에서 NLS를 포함한 서열들만을 선택하여 데이터베이스를 만들고, 이를 활용 할 수 있도록 웹용 인터페이스를 제작하여 연체동물 전용 BLAST 서버에 추

Table 1. The number and percentage of NLS in mollusks amino acids registered in Genbank.

Class	Registered Amino acid	NLS containing Amino acid	Percent (%)	NLS containing specimen No.
Aplacophora	29	0	0.00	0
Bivalvia	5,756	78	1.36	19
Cephalopoda	1,801	52	2.89	10
Gastropoda	14,334	133	0.93	28
Monoplacophora	0	0	0.00	0
Polyplacophora	122	1	0.82	1
Scaphopoda	96	2	2.08	2
total	22,138	266	1.2%	61

가하였다 (Lee *et al.*, 2004).

3. 아미노산 기능 분석을 위한 KOG 분석

연체동물의 아미노산의 기능을 분석하기 위하여 KOG (the eukaryotic orthologous groups) 데이터를 이용하였다 (Tatusov *et al.*, 2003). NCBI에 등록되어진 KOG 데이터를 모두 다운받은 후 BLAST 수행 시 자동으로 KOG를 판단할 수 있도록 데이터베이스를 만들고 E-value < e^{-10} 조건에서 BLASTP를 사용하여 BLAST를 수행 한 후 perl script를 사용하여 데이터를 정리하였다.

결과 및 고찰

연체동물 유래 아미노산 서열 22,138 개에서 NLS가 예측되는 서열은 266 개로 전체 아미노산 중 1.2% 정도가 NLS 서열부위를 포함하고 있었다. 현재 등재되어 있는 연체동물 8,314 종 중 NLS를 포함한 아미노산이 밝혀진 생물은 60여 종에 불과 했다. 60 종의 연체동물에서 발견된 NLS 서열을 포함하고 있는 아미노산 서열들의 전체 평균길이는 330이었다. 이 결과를 각 강 (class) 별로 살펴보면 전반적으로 복족류

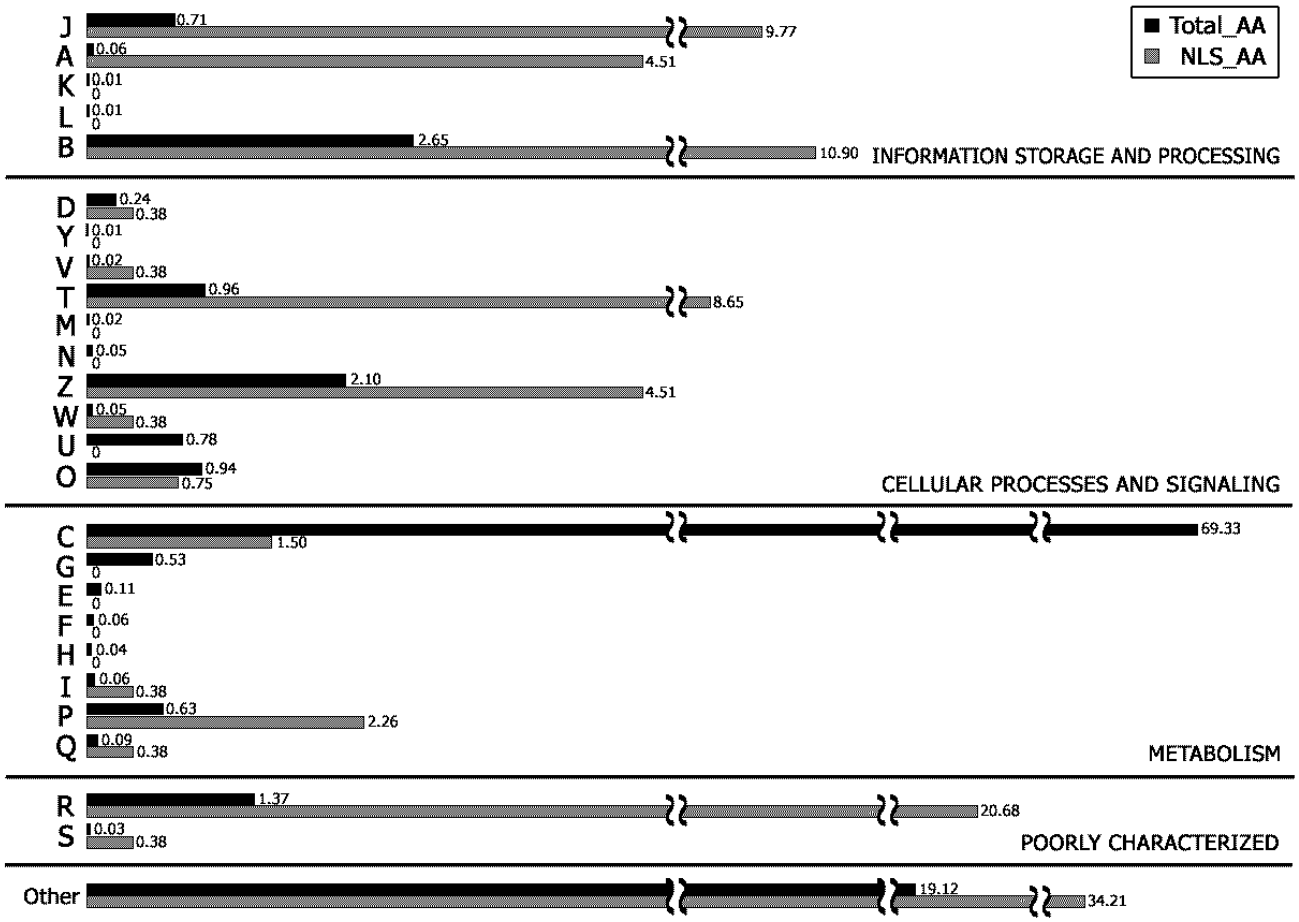

Fig. 1. Comparative functional prediction of NLS containing proteins against the entire Mollusks amino acids registered in Genbank. Code descriptions of KOG: J (translation, ribosomal structure and biogenesis), A (RNA processing and modification), K (transcription), L (replication, recombination and repair), B (chromatin structure and dynamics), D (cell cycle control, cell division, chromosome partitioning), Y (nuclear structure), V (defense mechanisms), T (signal transduction mechanisms), M (cell wall/membrane/envelope biogenesis), N (cell motility), Z (cytoskeleton), W (extracellular structures), U (intracellular trafficking, secretion, and vesicular transport), O (posttranslational modification, protein turnover, chaperones), C (energy production and conversion), G (carbohydrate transport and metabolism), E (amino acid transport and metabolism), F (nucleotide transport and metabolism), H (coenzyme transport and metabolism), I (lipid transport and metabolism), P (inorganic ion transport and metabolism), Q (secondary metabolites biosynthesis, transport and catabolism), R (general function prediction only), S (function unknown).

(gastropods) 가 유전자 연구가 가장 활발하게 이루어 진 것으로 보이며 NLS를 포함하는 아미노산이 가장 높은 비율로 나타나는 강은 두족 강 (class Cephalopoda) 이었으며, 그 다음은 굴족 강 (class Scaphopoda) 으로 나타났다.

현재 연체동물 서열 BLAST 데이터베이스에 등록되어진 모든 아미노산을 기능에 따라 분류하기 위해 KOG 방법을 사용한 결과 KOG-C (energy production and conversion) 가 69.3%, 기능을 알 수 없는 단백질이 19.12%로 나타났으며 KOG-B (chromatin structure and dynamics) 가 2.65% 정도로 나타났다. 그 수가 가장 많은 KOG-C에 속하는 단백질은 대부분 NADH와 cytochrome oxidase 종류로서 일반적으로 계통분류 및 진화연구에 많이 사용되어지는 유전자다. 이러한 분석 결과는 그간 연체동물 유전자 연구의 방향이 기능연구가 아닌 분류의 목적으로 더 많이 연구되어 왔음을 말해준다 (Table 1).

PredNLS 프로그램을 통해 결과를 얻은 NLS를 포함하는 연체동물의 아미노산은 모두 266 개였다. 이 서열들에 대한 KOG 분석을 실시한 결과 기능을 예측하기 힘든 유전자가 모두 34% 정도였으며, KOG-R (general function prediction only) 이 20.68%로 가장 많이 나왔으며, 기능이 알려져 있는 경우에는 histone 단백질 등이 포함된 KOG-B (chromatin structure and dynamics) 가 10.9%, ribosomal protein 등이 포함된 KOG-J (translation, ribosomal structure and biogenesis) 가 9.77%, synaptotagmin 및 Wnt 등을 포함하고 있는 KOG-T (signal transduction mechanisms) 가 8.65%로 조사되었으며 그밖에 RNA binding protein 등이 포함된 KOG-A (RNA processing and modification) 및 troponin, myosin 등을 포함하는 KOG-Z (cytoskeleton) 의 경우 각각 4.51%가 조사되었다 (Fig. 1).

이러한 성향은 정보저장 및 프로세싱 그리고 signaling에 관여하는 단백질이 주로 NLS를 많이 가지고 있다고 해석할 수 있으며, 이는 일반적으로 잘 알려진 모델생물의 NLS 포함 단백질들의 성향과 잘 일치함을 알 수 있었다.

좀 더 자세한 분석을 위해 등록되어진 아미노산 숫자에 NLS 포함률이 높은 생물 순으로 나열하고 각각 포함하는 유전자를 조사해 본 결과, 청자고둥 과에 속하는 *Conus ebraeus*의 경우 등록되어진 23 개의 유전자 중 20 개의 유전자가 NLS를 포함하고 있었다. 하지만 실재로 등록되어진 모든 아미노산은 모두 conotoxin 관련 유전자로 청자고둥류에서 분비되는 독소 단백질의 전구체였다. 청자고둥은 전 세계적으로 약 500여 종이 서식하는 것으로 보고되어진 종으로서 독소 물질에 약 100여 종의 peptide가 포함되어 있다고 하며, 니코틴 아세틸콜린 수용체 (nicotinic acetylcholine receptor) 를 차단하여 진통효과를 낸다고 한다. 니코틴 아세틸콜린 수용체는 통증전달에 참여하는 척추주변 신경세포의 그룹인 배근신경절을 포함하여 인체의 여러 부위에 존재한다 (Ekberg *et al.*, 2006).

그리고 군소 과에 속하는 *Aplysia parvula*의 경우에는 6 개 아미노산 중 5 개가 NLS를 가지고 있었는데, 모두 egg-laying hormone 단백질이었으며 (Nambu and Scheller, 1986), 또 다른 군소 과에 속하는 종인 *Aplysia californica* 의 경우에는 신호전달에 관련된 많은 단백질들이 NLS를 포함하고 있는 것으로 조사되었다. 그 외 삿갓조개류에 속하는 *Patella vulgata*와 가리맛조개 종류인 *Ensis minor*의 경우 주로 histone 단백질, 계통분류와 관련된 *hox* 유전자가 주종을 이루고 있었으며, 가리비 종류인 *Argopecten irradians, Mizuhopecten yessoensis*의 경우는 EST 연구 결과와 마찬가지로 주로 myosin 관련 유전자가 주종을 이루고 있었는데 NLS를 포함한 유전자는 주로 ribosomal 단백질 종류들이었다 (Roberts and Goetz, 2003).

진주조개인 *Pinctada fucata*의 경우에는 진주 생산과 관계가 있는 외투막과 관련된 유전자가 주로 연구되어 있었으며,

Table 2. List of mollusk species that have NLS-containing protein according to its percentage.

Species	NLS No.	Total AA	Percent
Conus ebraeus	20	23	86.96
Aplysia parvula	5	6	83.33
Ensis minor	8	11	72.73
Patella vulgata	13	43	30.23
Conus imperialis	11	45	24.44
Euprymna scolopes	16	103	15.53
Mytilus californianus	6	52	11.54
Sepia officinalis	9	89	10.11
Pinctada fucata	6	63	9.52
Ilyanassa obsoleta	5	55	9.09
Argopecten irradians	8	108	7.41
Mizuhopecten yessoensis	5	75	6.67
Haliotis asinina	5	85	5.88
Lymnaea stagnalis	23	428	5.37
Aplysia californica	21	450	4.67
Crassostrea gigas	10	307	3.26
Mytilus trossulus	5	157	3.18
Mytilus edulis	22	745	2.95
Mytilus galloprovincialis	5	269	1.86
Biomphalaria glabrata	5	612	0.82

그중 NLS를 포함하는 유전자에는 glycine이 풍부한 구조 유전자인 Shematrin이 포함되어 있었다 (Yano *et al.,* 2006). 그 외의 다른 종들은 모두 분류에 관계되어진 NADH 및 cytochrome oxidase가 주로 연구가 되어 있었으며, NLS 포함 단백질은 histone 및 ribosomal 단백질이었다. 만손주혈흡충 (*Schistosoma mansoni*) 의 중간숙주로 의학적으로 매우 중요한 종이라 연구가 많이 되어진 *Biomphalaria glabrata*의 경우에는 분류 및 면역관련 유전자인 fibrinogen-related protein이 주종을 이루고 있었으며 (Zhang *et al.,* 2004), NLS 포함 서열에는 nuclear receptor superfamily에 속하는 RXR-like protein 및 ribosomal protein의 일종인 ubiquitin이 포함되어 있었다 (Bouton *et al.* 2005) (Table. 2).

KOG 분석 및 NLS 예측결과 유전자 연구가 많이 되어 있을수록, 또 분류의 목적이 아닌 특수 목적을 가지고 연구가 진행되어진 생물종일수록 다양한 기능의 유전자 연구가 진행되어 왔으며 이에 따라 NLS 포함 유전자도 더 많이 나오는 양상을 볼 수가 있었다.

요 약

연체동물 유래 아미노산 서열 22,138 개에서 NLS가 예측되는 아미노산 서열은 266 개였으며 이는 연체동물 전체 아미노산 중 1.2% 정도였다. 또한 현재 등재되어 있는 연체동물 8,314 종 중 NLS를 포함한 아미노산이 밝혀진 생물은 60여 종에 불과 하였다. 현재 알려진 연체동물 서열 중에는 두족 강의 경우가 NLS를 포함한 아미노산이 많을 것으로 예측되었다.

ACKNOWLEDGEMENT

이 연구과제는 1999년도 인제대학교 학술연구조성연구사업의 일환으로 수행되었습니다.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biolology,* **215**: 403-410.

Arenal, A., Pimentel, R., Garcia, C., Pimentel, E. and Alestrom, P. (2004) The SV40 T antigen nuclear localization sequence enhances nuclear import of vector DNA in embryos of a crustacean (*Litopenaeus schmitti*). *Gene,* **337**: 71-77.

Bouton, D., Escriva, H., de Mendonca, R.L., Glineur, C., Bertin, B., Noel, C., Robinson-Rechavi, M., de Groot, A., Cornette, J., Laudet, V. and Pierce, R.J. (2005) A conserved retinoid X receptor (RXR) from the mollusk *Biomphalaria glabrata* transactivates transcription in the presence of retinoids. *Journal of Molecular Endocrinology,* **34**: 567-582.

Cokol, M., Nair, R. and Rost, B. (2000) Finding nuclear localization signals. *EMBO Reports*, **1**: 411-415.

Collas, P. and Alestrom, P. (1997) Rapid targeting of plasmid DNA to zebrafish embryo nuclei by the nuclear localization signal of SV40 T antigen. *Molecular Marine Biology and Biotechnology*, **6**: 48-58.

Ekberg, J., Jayamanne, A., Vaughan, C.W., Aslan, S., Thomas, L., Mould, J., Drinkwater, R., Baker, M.D., Abrahamsen, B., Wood, J.N., Adams, D.J., Christie, M.J. and Lewis, R.J. (2006) {micro}O-conotoxin MrVIB selectively blocks Nav1.8 sensory neuron specific sodium channels and chronic pain behavior without motor deficits. *Proceedings of the National Academy of Sciences of the United States of America*, **103**: 17030-17035.

Lee, Y.S., Jo, Y.H., Kim, D.S., Kim, D.W., Kim, M.Y., Choi, S.H., Yon, J.O., Byun, I.S., Kang, B.R., Jeong, K.H. and Park, H.S. (2004) Construction of BLAST server for mollusks. *The Korean Journal of Malacology,* **20**: 165-169.

Li, H.Y., Guo, A.H., Liu, Z.F., Liu, Y., Liu, J.H., Deng, P., Li, Z.J., Liu, Y.W. and Jiang, Y. (2006) Construction and functional study of a cell penetrating peptide-based expression vector for targeted delivery of proteins into the cell nuclei. *Journal of Southern Medical University,* **26**(10): 1394-1399, 1407.

Nambu, J.R. and Scheller, R.H. (1986) Egg-laying hormone genes of *Aplysia*: evolution of the ELH gene family. *Journal of Neuroscience*, **6**: 2026-2036.

Roberts, S.B. and Goetz, F.W. (2003) Expressed sequence tag analysis of genes expressed in the bay scallop, *Argopecten irradians*. *The Biological Bulletin,* **205**: 227-228.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. and Natale, D.A. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics,* **4**: 41.

Yano, M., Nagai, K., Morimoto, K., and Miyamoto, H. (2006) Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster *Pinctada fucata*. *Comparative Biochemistry and Physiology. B. Biochemistry and Molecular Biology*, **144**: 254-262.

Wayne M.B., Lewis J.K. and Jeff H. (2002) The World of the Cell. 5th Edition, 912 p, Benjamin Cummings.

Zhang, S.M., Adema, C.M., Kepler, T.B. and Loker, E.S. (2004) Diversification of Ig superfamily genes in an invertebrate. *Science,* **305**: 251-254.