A THESIS
FOR THE DEGREE OF MASTER OF SCIENCE

Characterization of genes from abalone cDNA library

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Helani Munasinghe

Department of Marine Biotechnology
GRADUATE SCHOOL
CHEJU NATIONAL UNIVERSITY

2004. 06. 29

Characterization of genes from abalone cDNA library

Helani Munasinghe
(Supervised by Professor Jehee Lee)

A thesis submitted in partial fulfillment of the requirement for the degree of Master of Science

2004. 06.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

This thesis has been examined and approved.

Thesis director, Choon Bok Song, Prof. Marine Biotechnology

In-Kyu Yeo, Prof. of Marine Biotechnology

Jehee Lee, Prof. of Marine Biotechnology

Date

Department of Marine Biotechnology
GRADUATE SCHOOL
CHEJU NATIONAL UNIVERSITY

Contents

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

국문초록

인간 게놈 연구에서 선보였던 expressed sequence taq (EST) 방법은 게놈 서열 분석을 위해 많은 cDNA의 클로닝과 cDNA 서열을 분석 할 수 있다. 본 논문에서는 전복의 소화선에서 cDNA library를 제작하고, 900개의 clone의 염기서열을 분석하여 841개의 expressed sequence tag을 획득하였다. 이들을 유사한 서열의 cluster로 분류한 결과 331개의 EST 서열은 122개의 cluster와 510개의 singleton에 해당하였다. 본 논문에서 제작한 EST의 redundancy는 39%였으며, 이는 전체 전복 cDNA library의 45%에 비해 낮았다. BLAST-X와 BLAST-N 프로그램으로 데이터베이스에 등록된 서열과 비교한 결과 632개의 transcript중 354개는 이미 알려진 서열과 매우 유사하였다. 그러므로, 이 중 278개의 transcripts를 클로닝하고 분석하였다. 이들은 새로운 gene을 확인하기 위한 분자 marker로서 유용하게 이용될 수 있을 것이다. 예상대로 소화선 cDNA library에서 획득한 서열의 대부분은 효소를 암호화하였다. 질병 조건과 세포 해독과 관련된 추정상의 효소인 glutathine transferase, arylsulfatase와 fucosidase는 full length 서열을 E. coli 발현벡터인 pET16b에 클로닝하여 분석되었다. Glutathione transferase는 soluble한 형태로 많은 양이 발현되었으며, 염기서열 분석 결과 N과 C GST를 갖는 Mu protein class에 속하고, conserved domains인 thioredoxin fold와 five-helices를 갖고 있었다. 이와는 달리, arylsulfatase는 insoluble한 형태로 발현되었으며, denaturation 조건하에서의 정제 산물은 inactive하였고, sulfatase domain과 유사한 부분을 갖는 약 54 kDa였다. 컴퓨터 프로그램으로 fucosidase 서열을 분석한 결과 L focosidase conserved domain의 아미노 말단과 prematurely 말단 번역을 유도하는 여러 nonsense mutation이 확인되었다. 이 유전자에 대한 보충 연구는 그 유전자의 활성과 질병에 대한 분자적인 특성을 이해하는데 도움이 될 것이다. Arylsulfatase와 fucosidase는 촉매제로서 산업 분야에 유용하게 이용될 것이며, glutathione transferase는 양식산업에서 생체이물질이 축적되어 있지 않은 독성물질에 저항력을 갖는 어류의 개발을 가능하게 할 것이다.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

List of Figures

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

List of Tables

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

INTRODUCTION

Abalones are univalve marine gastropods, which are highly prized as a food and a source of pearl. The worldwide popularity of abalone has led to the decline of many fisheries through commercial over-fishing. This led to an increase in hatchery rearing during the past decades, which made it possible to improve selected strains for breeding. Research on abalone to date has focused mainly on environmental factors such as culture techniques, tank design and nutrition (Powers *et al.,* 1996). Very little is done on the genetic improvement of abalone due to lack of knowledge on genetic aspects of these animals. Powers *et al.* (1996) developed triploid abalone with enhanced growth showing the importance of genetic improvement of abalone in aquaculture.

Partial cDNA sequencing to generate expressed sequence tags (ESTs) has provided a fast and efficient way of unraveling new genes in various organisms. This method provides a quantitative method to measure specific transcripts within a cDNA library. The low cost of generating ESTs and their usefulness in discovering new genes, genomic mapping, and identifying coding regions in genomic sequences (NCBI staff, 2002) have led to the exponential growth of sequences in EST database (http://www.ncbi.nlm.nih.gov/). To exploit this untapped genetic source, we constructed a cDNA library from disk abalone (*Haliotis discus discus*) digestive gland. Most of the genes identified during this study correspond to enzymes and we further characterized several enzymes implicated in disease conditions and cellular detoxification including fucosidase, sulfatase (ASB) and a glutathione transferase (GST).

Alpha-L-fucosidase is responsible for hydrolyzing the alpha-1,6-linked fucose joined to the reducing-end N-acetylglucosamine of the carbohydrate moieties of glycoproteins (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) and is involved in fucosidosis, an autosomal recessive lysosomal storage disease caused by defective alpha-L-fucosidase with accumulation of fucose in the tissues. Different phenotypes include clinical features such

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

as neurologic deterioration, growth retardation, visceromegaly, and seizures in a severe early form; coarse facial features, angiokeratoma corporis diffusum, spasticity and delayed psychomotor development in a longer surviving form; and an unusual spondylometaphyseoepiphyseal dysplasia in yet another form.

ASB is involved in the desulfation of 4-sulfate groups occuring at the non-reducing terminal N-acetyl galactoseamine-4 sulfate of oligosaccharides derived from dermatan sulfate or chondroitin-4-sulfate (Leznicki and Bailkowski, 1997). Lack of this enzyme results in mucopolysaccharidosis V1. Recent discovery of new sulfated compounds with commercial importance has resulted in a demand for sulfatases, which could be used either as a means of understanding the functional implications of sulfate groups or to remove sulfate groups for better product quality. For example, it has been identified that sulfated homopolysaccharides are more potent than sulfated heteropolysaccharides as antiviral agents and the presence of sulfate group is necessary for anti-HIV activity of these compounds and the potency increases with the degree of sulfation (Schaeffer *et al.*, 2000). On the other hand, sulfate groups in agar adversely affect the strength and melting temperature of commercial agarose (Mollet *et al.,* 1998).

GSTs are major phase II detoxification enzymes which catalyse the conjugation of glutathione (GSH) to various xenobiotic compounds increasing their solubility and excretion. GSTs are dimeric enzymes with each subunit containing a GSH binding domain (G-site) and a variable electrophilic substrate binding domain (H-site). Depending on the H-site, they can catalyze nucleophilic aromatic substitutions, Michael additions to α,β-unsaturated ketones and epoxide ring-opening reactions resulting GSH conjugates. Expression of different class GSTs has been implicated in organism resistance to herbicides, fungicides and insecticides. They are also involed in detoxification of chemicals released during host pathogen interactions (Sheehan *et al.*, 2001).
Considering these commercial and biological values we selected the above genes for further characterization.

Part I

Construction of abalone expressed sequence tag library and sequence analysis

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Part I

Construction of abalone expressed sequence tag library and sequence analysis

1. ABSTRACT

Deciphering the function of proteins is the main task in molecular biology and biochemistry. The increasing flow of genomic information makes the contribution of bioinformatics very relevant for proposing possible new protein functions based on prior information. Identification of coding regions is the first step in functional annotation of proteins they code. Partial cDNA sequencing to generate expressed sequence tags (ESTs) provides a fast and efficient way of unraveling the coding sequences.

As the first step towards functional annotation of unknown genes in abalone, 900 clones from an abalone digestive gland cDNA library were partially sequenced from the 5′ end. A total of 841 clones were successfully sequenced and, after removal of vector and unreliable data, they were assembled into 122 clusters by using TIGR Assembler. Each cluster consisted of at least two ESTs and was considered to be derived from the same gene. The clusters contained a total of 331 sequences, whereas 510 remained as singletons. This resulted in a 39% redundancy, which was much less than the redundancy (45%) of the whole abalone cDNA library we generated. This may be due to the suppression of organ specific mRNA percentages in whole animal tissue by the products of genes which carry out basic functions of all cells. The 632 unique transcripts were compared against the National Center for Biotechnology Information databases by BLAST-X and BLAST-N programs. Out of the 632 unique transcripts, 354 were significantly similar to sequences in the databases with an E-value less than 10^{-5}. Thus, 278 transcripts are cloned and reported for the first time in this study. These may be useful as molecular markers as well as in the identification of novel genes.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

2. MATERIALS AND METHODS

2. 1. Construction of abalone digestive gland cDNA library

The cDNA library was constructed from digestive glands of disk abalone (*Haliotis discus discus*) obtained from an aquarium. Total RNA was extracted from digestive glands (RNeasy Maxi Kit, QIAGEN, Germany) and mRNA was isolated using FastTrack® 2.0 mRNA isolation system (Invitrogen, USA). cDNA synthesis was carried out according to ZAP-cDNA synthesis protocol (Stratagene, USA). The library was constructed by the directional cloning method with *Eco*RI site on the 5′ end and *Xho*I site on the 3′ end. cDNAs larger than 0.5 kb were selected by size fractionation with Sepharose CL-2B column, and packaged with Gigapack III Gold packaging extract. Gel analysis of twenty randomly picked plaques showed 85% cloning efficiency with insert size distribution of 0.5 - 3.0 kb (Fig. 1-1) in the primary lambda phage library. Mass excision resulted in a phagemid library with a yield of 7.0 x 10^5 colony forming unit/mL with 90% white colonies. Randomly picked 900 white colonies were sequenced after plasmid preparation (AccuPrep^{TM} plasmid extraction kit, Bioneer, Korea) using the vector primer AB (5′-CAAAAGCTGGAGCTCCACC-3′).

2. 2. Analysis of cDNA sequences

The ESTs were edited to remove vector sequences and unreliable data using cross_match and PHRED programs. Assembly of the individual ESTs into groups of sequences (clusters) representing unique transcripts was performed by using the TIGR Assembler (Sutton *et al.*, 1995). The resulting data were compared against the National Center for Biotechnology Information (NCBI) nonredundant protein database by using the program BLAST-X on the BLAST network service at NCBI. Sequences that did not match sequences in the protein database were further analyzed by searching for similarities at the nucleotide level by

using the BLAST-N program against the nonredundant nucleotide sequence database (Altschul *et al.*, 1990). Sequences with an E-value less than 10^{-5} were considered significant and top-scoring genes were used to group the transcripts by their putative function.

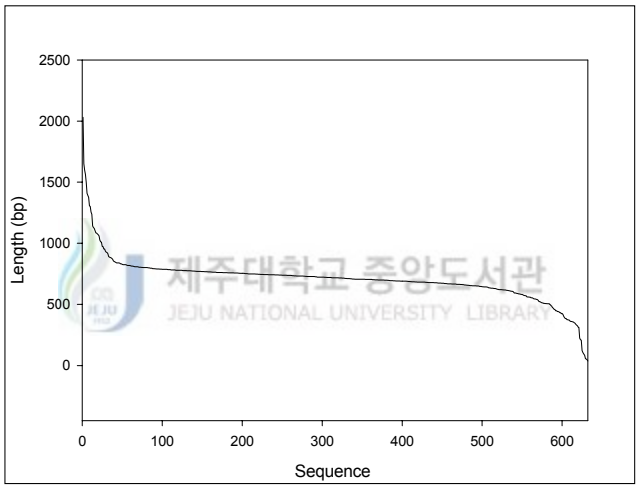
제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

3. RESULTS

3. 1. Abalone digestive gland cDNA library sequence analysis

To gain insight into the abalone transcriptome, we generated 900 EST sequences from disk abalone digestive gland. After removal of low quality data, 841 sequences were selected for further analysis. A total of 122 clusters were formed with these fragments by TIGR Assembler 2.0 (Sutton *et al.,* 1995). Each cluster contained at least two ESTs and was considered to be derived from the same gene by taking sequence ambiguities into account. The clusters contained a total of 331 sequences, whereas 510 remained as singleton ESTs, not similar to any other sequence in the data set. This resulted in 632 unique transcripts but it can be an overestimation since some fragments can be non-overlapping segments of the same transcript.

The digestive gland cDNA library had a 39% redundancy, which was much less than the redundancy (45%) of the whole abalone cDNA library we generated (Unpublished data). Out of the 632 sequences, 354 showed a significant similarity to sequences in the data bases (Table 1-1 and Table 1-2). Out of this, 240 sequences had significant similarity to database sequences with known or putative functions. These included 71 different enzymes/ isoenzymes or different enzyme subunits, 30 nucleic acid binding proteins/ transcription factors, 18 structural/ cytoskeleton proteins and several other functional proteins as summarized in (Fig. 1-2). There were 114 unique sequences similar to sequences in databases with unknown function.



제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Fig.1-1. Graph plotted length of ESTs vs EST sequence.

Table 1-2. ESTs with similarities to sequences in NCBI data bases with an E value less than 10^{-5} in BLAST X analysis [9]

EST	Putative identification	AC	% Identity
2-E10-AB.ab1	2210023K21Rik protein	AAH04572.1	47
pst_2_jhLee_433	2-4-dienoyl-Coenzyme A reductase 2	NP_741993.1	52
06-D04-AB.ab1	40S ribosomal protein	AAO43049.1	84
pst_2_jhLee_95	40S Ribosomal protein S3a	P49395|RS3A_APLCA	85
pst_2_jhLee_155	40S ribosomal protein S5	AAK95187.1|AF402813_1	88
3-D10-AB.ab1	4-nitrophenylphosphatase domain and non-neuronal SNAP25-like	NP_032724.1	46
09-B04-AB.ab1	6030466N05Rik protein	AAH31719.1	58
04-F01-AB.ab1	60S acidic ribosomal protein P1	AAL62466.1	49
10-B02-AB.ab1	acid alpha glucosidase	BAA25884.1	50
pst_2_jhLee_79	acidic ribosomal protein P0, cytosolic	|R5RT10	76
2-H05-AB.ab1	Actin-depolymerizing factor (ADF)	P30175|ADF_LILLO	32
3-F10-AB.ab1	acyltransferase 3 family (5Q785)	NP_507118.1	33
2-G10-AB.ab1	adipose differentiation-related protein	AAF76320.2|AF234676_1	31
pst_2_jhLee_29	ADP/ATP carrier	AAC23561.1	66
4-C05-AB.ab1	alkaline phosphatase, 145K - Synechococcus sp. (strain PCC 7942)	|A47026	36
07-F08-AB.ab1	alpha 1 type XXII collagen [Homo sapiens]	AAN03620.1|AF406780_1	24
4-C12-AB.ab1	Alpha-methylacyl-CoA racemase (2-methylacyl-CoA racemase)	P70473|AMAC_RAT	58
pst_2_jhLee_237	ALR protein - human	|T03455	51
3-D06-AB.ab1	ankyrin repeat protein E4_8 [synthetic construct]	AAO25692.1	45
3-C04-AB.ab1	apextrin [Heliocidaris erythrogramma]	AAC28440.1	26
07-B11-AB.ab1	arginine kinase [Aplysia kurodai]	BAB41095.1	82
08-D07-AB.ab1	aryl sulfotransferase [Rattus norvegicus]	CAA48604.1	44
2-B01-AB.ab1	arylacetamide deacetylase family member (4K352) [Caenorhabditis	NP_501702.1	35
2-C02-AB.ab1	arylsulfatase B [Mus musculus]	XP_283156.1	48
08-E08-AB.ab1	AT3g51050/F24M12_90 [Arabidopsis thaliana]	AAL58934.1|AF462847_1	23
pst_2_jhLee_232	ATPase subunit 6 [Littorina saxatilis]	CAA10596.1	50
06-H05-AB.ab1	B-cell translocation gene 2, anti-proliferative [Mus musculus]	NP_031596.1	59
2-G01-AB.ab1	Beta-glucuronidase precursor	O18835|BGLR_CANFA	61
pst_2_jhLee_401	betaine-homocysteine methyltransferase [Oceanobacillus iheyensis	NP_691612.1	47
1-G02-AB.ab1	beta-tubulin [Meriones unguiculatus]	CAB91644.1	100
05-G02-AB.ab1	calmodulin [Metridium senile]	BAB61794.1	100
06-C11-AB.ab1	calponin homolog [Schistosoma mansoni]	AAB47536.1	65
pst_2_jhLee_252	calreticulin precursor - California sea hare	|JH0795	68
pst_2_jhLee_153	carbonyl reductase [Cricetulus griseus]	BAB07797.1	48
4-C11-AB.ab1	Caspase 10 isoform a preproprotein; FADD-like ICE2; apoptotic	NP_001221.1	30
pst_2_jhLee_410	catalase [Melopsittacus undulatus]	AAO72713.1	76
pst_2_jhLee_476	cathepsin L-like cysteine proteinase A [Rhipicephalus	AAQ16117.1	58
1-F05-AB.ab1	cathepsin Q2 [Rattus norvegicus]	AAO27844.1|AF456460_1	39
2-B07-AB.ab1	CCAAT enhancer-binding protein - California sea hare	|A53066	27
2-F10-AB.ab1	CD109; Gov system alloantigens on platelets [Homo sapiens]	NP_598000.1	40
pst_2_jhLee_268	cellulase [Haliotis discus]	BAC67186.1	91
06-H02-AB.ab1	cellulase [Haliotis discus]	BAC67186.1	47
1-A05-AB.ab1	cellulase [Haliotis discus]	BAC67186.1	47
08-B05-AB.ab1	cellulase EGX [Ampullaria crossean]	AAP31839.1	51
05-A09-AB.ab1	CG1458-PA [Drosophila melanogaster]	NP_651684.1	43
3-B03-AB.ab1	CG14996-PB [Drosophila melanogaster]	NP_647860.1	56
2-B05-AB.ab1	CG16707-PA [Drosophila melanogaster]	NP_729535.1	71
07-D01-AB.ab1	CG16726-PA [Drosophila melanogaster]	NP_648342.1	33
07-A11-AB.ab1	CG18377-PA [Drosophila melanogaster]	NP_610588.2	28
3-A05-AB.ab1	CG3699-PA [Drosophila melanogaster]	NP_569875.2	45
05-D11-AB.ab1	CG4386-PA [Drosophila melanogaster]	NP_611611.1	33
pst_2_jhLee_158	CG4821-PA [Drosophila melanogaster]	NP_648288.1	23

제주대학교 중앙도서관 JEJU NATIONAL UNIVERSITY LIBRARY

pst_2_jhLee_138	CG4928-PB [Drosophila melanogaster]	NP_573179.1	54
06-A05-AB.ab1	CG5167-PA [Drosophila melanogaster]	NP_650190.1	50
1-H06-AB.ab1	CG5446-PA [Drosophila melanogaster]	NP_609565.1	71
pst_2_jhLee_347	CG6847-PA [Drosophila melanogaster]	NP_573259.1	37
pst_2_jhLee_420	Chain A, Fasciculin2 - Mouse Acetylcholinesterase Complex	1MAH|A	29
07-G02-AB.ab1	Chain A, Methionine Adenosyltransferase Complexed With A	1QM4|A	70
1-E08-AB.ab1	checkpoint protein Hus1 [Xenopus laevis]	AAM90260.1|AF516928_1	58
09-E09-AB.ab1	chitin synthase A [Drosophila melanogaster]	CAC83726.1	26
pst_2_jhLee_245	chitinase [Tenebrio molitor]	CAD31740.4	24
07-F04-AB.ab1	chitinase [Tenebrio molitor]	CAD31740.4	28
04-C07-AB.ab1	Cholinesterase 1	Q95000|CHL1_BRALA	45
06-G04-AB.ab1	cholinesterase 1 [Branchiostoma floridae]	AAD05373.1	51
08-B09-AB.ab1	chondroitin sulfate proteoglycan 3 (neurocan) [Homo sapiens]	NP_004377.1	35
1-E12-AB.ab1	Coatomer epsilon subunit (epsilon-coat protein) (epsilon-cop)	Q60445|COPE_CRIGR	66
06-D08-AB.ab1	collagen pro alpha-chain [Haliotis discus]	BAA75668.1	99
pst_2_jhLee_126	collagen pro alpha-chain [Haliotis discus]	BAA75669.1	89
pst_2_jhLee_240	collagen pro alpha-chain [Haliotis discus]	BAA75669.1	87
08-F06-AB.ab1	conserved hypothetical protein [Xanthomonas axonopodis pv. citri	NP_641707.1	49
pst_2_jhLee_186	CRYSTAL PROTEIN PRECURSOR	P21837|CRYS_DICDI	38
06-G03-AB.ab1	cubilin; cubilin (intrinsic factor-cobalamin receptor) [Rattus	NP_445784.1	29
pst_2_jhLee_167	CuZn superoxide dismutase [Apis mellifera ligustica]	AAP93581.1	52
2-D02-AB.ab1	cyclin L1; cyclin ania-6a; cyclin L [Mus musculus]	NP_064321.1	45
4-B03-AB.ab1	Cystatin precursor (Ovarian cystatin) (P12)	P35481|CYT_CYPCA	34
08-F12-AB.ab1	Cyt19 protein; likely ortholog of rat methyltransferase Cyt19;	NP_065733.1	49
05-D05-AB.ab1	cytochrome b [Littorina saxatilis]	CAA10599.1	77
pst_2_jhLee_125	cytochrome b [Penaeus monodon]	NP_038299.1|CYTB_15276	65
pst_2_jhLee_110	cytochrome c oxidase subunit I [Katharina tunicata]	NP_008173.1|COX1_10528	73
pst_2_jhLee_69	cytochrome c oxidase subunit III [Melanocetus murrayi]	NP_739782.1	64
pst_2_jhLee_107	cytochrome c oxidase subunit III [Platynereis dumerilii]	NP_009242.1|COX3_15124	63
pst_2_jhLee_408	cytochrome oxidase subunit II [Littorina saxatilis]	CAA10594.1	59
pst_2_jhLee_214	cytochrome oxidase subunit II [Littorina saxatilis]	CAA10594.1	50
pst_2_jhLee_24	cytochrome-c oxidase (EC 1.9.3.1) chain I - Katharina tunicata	|S50327	74
2-H08-AB.ab1	cytoplasmic actin [Oikopleura longicauda]	BAA86216.1	98
1-B10-AB.ab1	dendritic cell protein [Homo sapiens]	NP_006351.2	54
2-A10-AB.ab1	deoxyribonuclease II alpha; deoxyribonuclease II [Mus musculus]	NP_034192.1	36
pst_2_jhLee_590	EF-9 polyadenylation variant II [Mus musculus]	AAM77638.1|AF517107_1	27
07-B10-AB.ab1	effete CG7425-PA [Drosophila melanogaster]	NP_731941.1	91
4-G10-AB.ab1	effete CG7425-PA [Drosophila melanogaster]	NP_731941.1	90
pst_2_jhLee_121	ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)	P19039|EF1A_APIME	85
pst_2_jhLee_509	embryo cathepsin L-associated protein [Artemia franciscana]	AAP69998.1	29
05-C11-AB.ab1	embryo cathepsin L-associated protein [Artemia franciscana]	AAP69998.1	29
pst_2_jhLee_358	endo-1,4-beta-glucanase [Mytilus edulis]	CAC59694.1	49
1-C10-AB.ab1	endo-1,4-mannanase [Mytilus edulis]	CAC81056.1	44
3-A01-AB.ab1	ENSANGP00000003537 [Anopheles gambiae]	XP_313587.1	48
09-H06-AB.ab1	ENSANGP00000009390 [Anopheles gambiae]	XP_308333.1	39
2-A12-AB.ab1	ENSANGP00000010440 [Anopheles gambiae]	XP_315834.1	79
06-D10-AB.ab1	ENSANGP00000011058 [Anopheles gambiae]	XP_311928.1	72
08-G12-AB.ab1	ENSANGP00000012201 [Anopheles gambiae]	XP_313869.1	41
pst_2_jhLee_154	ENSANGP00000012979 [Anopheles gambiae]	XP_316359.1	48
08-B03-AB.ab1	ENSANGP00000014346 [Anopheles gambiae]	XP_317841.1	39

목포대학교 중앙도서관
MOKPO NATIONAL UNIVERSITY LIBRARY

07-G08-AB.ab1	ENSANGP00000014833 [Anopheles gambiae]	XP_318983.1	52
04-A12-AB.ab1	ENSANGP00000014885 [Anopheles gambiae]	XP_312551.1	57
pst_2_jhLee_87	ENSANGP00000015684 [Anopheles gambiae]	XP_314051.1	92
05-E12-AB.ab1	ENSANGP00000015968 [Anopheles gambiae]	XP_308535.1	61
08-H06-AB.ab1	ENSANGP00000018575 [Anopheles gambiae]	XP_317654.1	58
06-F12-AB.ab1	ENSANGP00000018629 [Anopheles gambiae]	XP_315795.1	41
06-A11-AB.ab1	ENSANGP00000020729 [Anopheles gambiae]	XP_308198.1	52
06-E05-AB.ab1	ENSANGP00000020956 [Anopheles gambiae]	XP_321840.1	72
06-G12-AB.ab1	ergosterol biosynthesis ERG4/ERG24 family protein [Coxiella	NP_820155.1	57
09-D09-AB.ab1	eukaryotic translation initiation factor 5A; eIF5AI [Homo sapiens]	NP_001961.1	66
pst_2_jhLee_161	fatty acid binding protein [Clonorchis sinensis]	AAN04089.1|AF527454_1	28
pst_2_jhLee_47	fatty acid binding protein [Clonorchis sinensis]	AAN04089.1|AF527454_1	28
pst_2_jhLee_265	Fatty acid-binding protein, retina (R-FABP)	Q05423|FABB_CHICK	38
pst_2_jhLee_64	ferritin GF2 [Crassostrea gigas]	AAP83794.1	80
3-G11-AB.ab1	ferritin-like protein [Pinctada fucata]	AAQ12076.1	81
06-A03-AB.ab1	flice [Mus musculus]	CAA04196.1	46
08-C06-AB.ab1	ganglioside M2 activator protein precursor - mouse	|S35613	32
07-D04-AB.ab1	gelsolin-like protein [Lumbricus terrestris]	CAD43405.1	49
05-A03-AB.ab1	glutathione peroxidase	|2204226A	44
05-E07-AB.ab1	glutathione S-transferase, mu 2;	NP_032209.1	55
05-G10-AB.ab1	glyceraldehyde-3-phosphate dehydrogenase [Daphnia pulex]	CAB94909.1	78
08-G06-AB.ab1	GLYcosylation related, UDP-N-acetyl-D-galactosamine:polypeptide	NP_499504.1	28
08-F04-AB.ab1	GM2 ganglioside activator protein [Rattus norvegicus]	NP_758838.1	34
10-A02-AB.ab1	GM2 ganglioside activator protein [Rattus norvegicus]	NP_758838.1	35
pst_2_jhLee_378	GM2 ganglioside activator protein precursor; cerebroside sulfate	NP_000396.1	28
pst_2_jhLee_241	guanine nucleotide-binding protein [Petromyzon marinus]	AAM88904.1	84
10-F03-AB.ab1	guanine nucleotide-binding protein [Petromyzon marinus]	AAM88904.1	84
05-E05-AB.ab1	H3 histone, family 3A [Homo sapiens]	NP_002098.1	100
2-A01-AB.ab1	Hemagglutinin/amebocyte aggregation factor precursor (18K-LAF)	Q01528|HAAF_LIMPO	40
pst_2_jhLee_184	heparanase [Homo sapiens]	AAD54516.1|AF084467_1	35
06-C02-AB.ab1	heparanase-like protein [Bombyx mori]	BAB85191.1	29
3-F12-AB.ab1	Homo sapiens ATP synthase, H+ transporting, mitochondrial F1	AAP36942.1	89
09-H11-AB.ab1	hypothetical protein [Clostridium thermocellum ATCC 27405]	ZP_00060309.1	26
pst_2_jhLee_82	hypothetical protein [Cytophaga hutchinsonii]	ZP_00119937.1	34
pst_2_jhLee_40	hypothetical protein [Cytophaga hutchinsonii]	ZP_00119937.1	31
pst_2_jhLee_46	hypothetical protein [Cytophaga hutchinsonii]	ZP_00119937.1	31
pst_2_jhLee_58	hypothetical protein [Cytophaga hutchinsonii]	ZP_00119937.1	31
09-C11-AB.ab1	hypothetical protein [Cytophaga hutchinsonii]	ZP_00119937.1	33
08-D02-AB.ab1	hypothetical protein [Escherichia coli K12]	NP_417460.1	99
09-D01-AB.ab1	hypothetical protein [Macaca fascicularis]	BAB62213.1	28
4-F05-AB.ab1	hypothetical protein [Macaca fascicularis]	BAB69753.1	63
3-E09-AB.ab1	hypothetical protein [Microbulbifer degradans 2-40]	ZP_00066788.1	32
2-D09-AB.ab1	hypothetical protein [Microbulbifer degradans 2-40]	ZP_00068378.1	32
3-D09-AB.ab1	hypothetical protein [Nostoc punctiforme]	ZP_00106518.1	32
pst_2_jhLee_542	hypothetical protein [Oryza sativa (japonica cultivar-group)]	AAO39872.1	35
07-B07-AB.ab1	hypothetical protein [Plasmodium falciparum 3D7]	NP_473058.1	33
07-G01-AB.ab1	hypothetical protein [Trichodesmium erythraeum IMS101]	ZP_00074867.1	29
06-G02-AB.ab1	hypothetical protein KIAA0684 - human (fragment)	|T00358	49
08-D04-AB.ab1	hypothetical protein XP_243842 [Rattus norvegicus]	XP_243842.1	32

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

1-H12-AB.ab1	Hypothetical protein ZK337. 1b [Caenorhabditis elegans]	CAB05007.2	46
09-F06-AB.ab1	hypothetical protein, putative universal stress protein Usp	AAP06495.1	38
pst_2_jhLee_23	hypothetical protein, putative universal stress protein Usp	AAP06495.1	34
pst_2_jhLee_164	intermediate filament protein A, cytosolic - California sea hare	|S24545	46
3-H08-AB.ab1	intermediate filament protein A, cytosolic - California sea hare	|S24545	39
10-A04-AB.ab1	KIAA0002 [Homo sapiens]	BAA07652.1	67
08-E11-AB.ab1	KIAA0698 protein [Homo sapiens]	BAA31673.2	42
03-H01-AB.ab1	Kruppel-like factor 15 [Rattus norvegicus]	NP_445988.1	74
07-D05-AB.ab1	LD24380p [Drosophila melanogaster]	AAD27865.2|AF132566_1	58
4-B06-AB.ab1	let-23 Fertility Effector/regulator LFE-2,	NP_491503.1	46
pst_2_jhLee_37	L-iditol 2-dehydrogenase (EC 1.1.1.14) - sheep (tentative sequence)	|S10065	44
06-B10-AB.ab1	low density lipoprotein-related protein 1B (deleted in tumors); low	NP_443737.1	37
pst_2_jhLee_189	LP12301p [Drosophila melanogaster]	AAK93491.1	48
4-H04-AB.ab1	lysosomal cofactor/neurotrophic factor prosaposin [Danio rerio]	AAL54381.1|AF276996_1	42
06-D05-AB.ab1	Macrophage receptor MARCO	Q9WUB9|MRCO_MESAU	52
09-F07-AB.ab1	MEGF7 [Homo sapiens]	BAA32468.1	42
1-B04-AB.ab1	metalloproteinase 2 [Hydra vulgaris]	AAD33860.1|AF140020_1	39
pst_2_jhLee_426	methionine adenosyltransferase (EC 2.5.1.6) - mouse	|A47151	75
pst_2_jhLee_100	mitochondrial malate dehydrogenase precursor [Nucella lapillus]	AAG17699.1|AF280052_1	75
09-H09-AB.ab1	msp130 protein [Heliocidaris erythrogramma]	CAC20358.1	34
08-F01-AB.ab1	multicystatin - common sunflower	|JC7333	30
pst_2_jhLee_333	Multicystatin precursor (MC)	P37842|CYTM_SOLTU	25
pst_2_jhLee_317	Myc homolog [Crassostrea virginica]	AAB34577.1	43
pst_2_jhLee_576	Myc homolog [Crassostrea virginica]	AAB34577.1	43
08-D08-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	43
4-A01-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	38
4-B05-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	41
pst_2_jhLee_382	Myc homolog [Crassostrea virginica]	AAB34577.1	32
06-F05-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	43
3-C09-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	43
08-F08-AB.ab1	Myc homolog [Crassostrea virginica]	AAB34577.1	42
06-H04-AB.ab1	myosin:SUBUNIT=essential light chain	|1803425A	66
2-B04-AB.ab1	Na+/Cl- dependent neurotransmitter transporter-like protein	AAM09083.1	54
06-D11-AB.ab1	N-acetyl-beta-glucosaminidase prepro-polypeptide	AAA51828.1	48
07-E03-AB.ab1	NADH dehydrogenase subunit 2 [Loligo bleekeri]	NP_062838.1	33
2-D12-AB.ab1	NADH dehydrogenase subunit 4 [Loligo bleekeri]	NP_062842.1	49
04-H12-AB.ab1	NG5 [Homo sapiens]	AAB47496.1|AAB47496	48
09-H02-AB.ab1	outer arm dynein light chain 4 [Anthocidaris crassispina]	BAA24152.1	74
pst_2_jhLee_284	peptidylprolyl isomerase A (cyclophilin A) [Rattus norvegicus]	NP_058797.1	71
2-C10-AB.ab1	peroxiredoxin V protein [Branchiostoma belcheri tsingtaunese]	AAM18076.1|AF498232_1	58
08-E12-AB.ab1	Peroxisomal sarcosine oxidase (PSO) (L-pipecolate oxidase)	Q9D826|SOX_MOUSE	48
06-E07-AB.ab1	placental protein 11 related [Mus musculus]	NP_032928.1	31
08-D06-AB.ab1	placenta-specific 8 [Homo sapiens]	NP_057703.1	41
06-F11-AB.ab1	pol [Drosophila melanogaster]	CAC16871.1	31
4-D01-AB.ab1	poly(A) binding protein, cytoplasmic 1 [Rattus norvegicus]	NP_599180.1	65
06-H12-AB.ab1	polymerase (RNA) II (DNA directed) polypeptide G [Mus musculus]	AAH05580.2	82
2-E02-AB.ab1	polyubiquitin [Schistosoma mansoni]	AAD02414.1	100

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

2-F06-AB.ab1	probable L-proline 4-hydroxylase [Pirellula sp.]	NP_867845.1	42
06-A10-AB.ab1	PROFILIN	P18321|PROF_CLYJA	30
10-D05-AB.ab1	prostaglandin transporter [Bos taurus]	NP_777254.1	28
pst_2_jhLee_156	protein convertase subtilisin / kexin, type I; prohormone convertase	NP_058787.1	52
3-C03-AB.ab1	protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - horseshoe	|A45321	28
05-G03-AB.ab1	protocadherin-psi1 [Homo sapiens]	AAK51618.1|AF217751_1	28
06-B07-AB.ab1	putative CD209L1 protein [Hylobates lar]	AAL89528.1	27
05-D02-AB.ab1	putative protein kinase [Mus musculus]	CAB61344.1	39
05-B11-AB.ab1	putative RNA-binding protein [Patella vulgata]	AAK32728.1|AF361436_1	44
4-B11-AB.ab1	RAS-like, estrogen-regulated, growth-inhibitor [Homo sapiens]	NP_116307.1	46
pst_2_jhLee_60	Ras-related protein Rab-1A	Q05974|RAB1_LYMST	96
1-B11-AB.ab1	ribosomal protein L10 [Rattus norvegicus]	NP_112362.1	81
07-B08-AB.ab1	ribosomal protein L12 [Argopecten irradians]	AAN05610.1	79
2-G02-AB.ab1	Ribosomal protein L17 [Danio rerio]	AAH55097.1	76
07-E08-AB.ab1	ribosomal protein L34 [Branchiostoma belcheri tsingtaunese]	AAO31772.1	75
pst_2_jhLee_42	ribosomal protein L4, cytosolic [validated] - rat	|JC4277	76
2-F04-AB.ab1	ribosomal protein L5 [Argopecten irradians]	AAN05603.1	70
pst_2_jhLee_325	ribosomal protein L7 [Crassostrea gigas]	CAD89885.1	73
09-G04-AB.ab1	ribosomal protein L7a [Argopecten irradians]	AAN05607.1	81
pst_2_jhLee_32	ribosomal protein S2 [Chlamys farreri]	AAM94271.1	90
10-D01-AB.ab1	ribosomal protein S30 [Argopecten irradians]	AAN05597.1	62
pst_2_jhLee_388	ribosomal protein S4 [Argopecten irradians]	AAN05593.1	83
3-G12-AB.ab1	RIKEN cDNA 6430526J12; LDLR dan [Mus musculus]	NP_766256.2	31
08-F05-AB.ab1	Selenoprotein M precursor (SelM protein)	Q8VHC3|SELM_MOUSE	37
pst_2_jhLee_118	SHG [Littorina littorea]	AAM20843.1|AF369699_1	25
pst_2_jhLee_38	SHG [Littorina littorea]	AAM20843.1|AF369699_1	24
pst_2_jhLee_187	SHG [Littorina littorea]	AAM20843.1|AF369699_1	22
pst_2_jhLee_105	SHG [Littorina littorea]	AAM20843.1|AF369699_1	23
pst_2_jhLee_370	similar to 30 kDa adipocyte complement-related protein [Rattus	XP_139092.2	31
06-E02-AB.ab1	Similar to actin related protein 2/3 complex, subunit 1A, 41kDa	AAH41267.1	55
08-B07-AB.ab1	Similar to aldehyde dehydrogenase 7 family, member A1 [Danio rerio]	AAH44367.1	72
pst_2_jhLee_556	similar to alpha-L-fucosidase [Schistosoma japonicum]	AAP05896.1	62
2-G04-AB.ab1	Similar to anaphase promoting complex subunit 5 [Homo sapiens]	AAH34243.1	32
09-A01-AB.ab1	similar to brain-specific protein p25 alpha [Homo sapiens] [Rattus	XP_217738.1	56
09-D11-AB.ab1	Similar to chaperonin containing TCP1, subunit 5 (epsilon) [Xenopus	AAH44997.1	70
07-D09-AB.ab1	Similar to COP9 constitutive photomorphogenic homolog subunit 3	AAH45415.1	52
07-A12-AB.ab1	Similar to deleted in malignant brain tumors 1 [Mus musculus]	AAH49835.1	28
4-G02-AB.ab1	similar to elicitor-like mating protein M81 [Phytophthora	XP_225445.1	31
07-H02-AB.ab1	Similar to high density lipoprotein binding protein (vigilin)	AAH44314.1	46
07-A04-AB.ab1	Similar to high density lipoprotein binding protein (vigilin) [Homo	AAH14305.1|AAH14305	50
4-D11-AB.ab1	Similar to hydroxyprostaglandin dehydrogenase 15-(NAD) [Danio	AAH52123.1	43
08-G05-AB.ab1	similar to hypothetical protein FLJ20375 [Homo sapiens] [Rattus	XP_233144.1	38
1-E11-AB.ab1	similar to hypothetical protein FLJ40597 [Homo sapiens] [Rattus	XP_221401.1	42
pst_2_jhLee_439	Similar to integrin beta 4 binding protein [Danio rerio]	AAH49488.1	86
1-G10-AB.ab1	similar to repeat organellar protein-related [Plasmodium yoelii	XP_219238.1	22
05-B06-AB.ab1	Similar to RIKEN cDNA 4430402G14 gene [Xenopus laevis]	AAH41528.1	48

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

07-H05-AB.ab1	similar to sulfotransferase family 1A, phenol-preferring, member 2	XP_065757.2	35
1-D02-AB.ab1	Similar to yolk sac gene 2 [Danio rerio]	AAH49448.1	45
pst_2_jhLee_174	SLC25A3 protein [Homo sapiens]	AAH51367.1	67
3-D05-AB.ab1	Sodium/potassium-transporting ATPase alpha chain (Sodium pump)	P17326|AT1A_ARTSF	74
pst_2_jhLee_177	Soma ferritin	P42577|FRIS_LYMST	70
07-H04-AB.ab1	Sorbitol dehydrogenase-2 CG4649-PA [Drosophila melanogaster]	NP_524311.1	62
09-G09-AB.ab1	succinate dehydrogenase complex, subunit D precursor; succinate	NP_002993.1	42
09-F08-AB.ab1	sulfatase 1 precursor [Helix pomatia]	AAF30402.1|AF109924_1	49
pst_2_jhLee_375	sulfatase 1 precursor [Helix pomatia]	AAF30402.1|AF109924_1	42
06-E08-AB.ab1	sulfatase 1 precursor [Helix pomatia]	AAF30402.1|AF109924_1	38
pst_2_jhLee_522	sulfotransferase family, cytosolic, 1A, phenol-preferring, member	NP_001045.1	35
08-C02-AB.ab1	SWI/SNF-related matrix-associated actin-dependent regulator of	NP_003070.3	88
1-G07-AB.ab1	target of Jun 3 [Coturnix coturnix]	AAG16624.1	75
2-G06-AB.ab1	tetraspanin-CD63 receptor [Geodia cydonium]	CAA77025.1	33
10-A03-AB.ab1	translation Elongation FacTor (94.8 kD) (eft-2) [Caenorhabditis	NP_492457.1	77
09-G10-AB.ab1	transposase homolog [Haemonchus contortus]	AAD34306.1|AF099908_1	32
pst_2_jhLee_337	transposase, Tc1/Tc3 and Integrase, catalytic domain containing	NP_497684.1	36
05-B05-AB.ab1	tubulin, beta, 2 [Homo sapiens]	AAH29529.1	90
09-G05-AB.ab1	tumor rejection antigen 1gp96 [synthetic construct]	AAQ02595.1	68
05-G05-AB.ab1	ubiquitin/ribosomal protein S27a fusion protein [Branchiostoma	AAL55470.1	75
08-H07-AB.ab1	Unknown (protein for IMAGE:3544292) [Homo sapiens]	AAH03577.1|AAH03577	92
05-F05-AB.ab1	Unknown (protein for IMAGE:5269996) [Homo sapiens]	AAH28610.1	48
06-G07-AB.ab1	Unknown (protein for IMAGE:6881027) [Xenopus laevis]	AAH53826.1	53
pst_2_jhLee_440	Unknown (protein for MGC:53465) [Xenopus laevis]	AAH46271.1	71
08-G03-AB.ab1	Unknown (protein for MGC:53465) [Xenopus laevis]	AAH46271.1	79
pst_2_jhLee_469	Unknown (protein for MGC:55617) [Danio rerio]	AAH44200.1	78
3-F01-AB.ab1	Unknown (protein for MGC:9625) [Homo sapiens]	AAH16295.1|AAH16295	79
3-D02-AB.ab1	unknown [Homo sapiens]	AAC19158.1	60
04-H09-AB.ab1	unnamed protein product [Homo sapiens]	BAB15189.1	46
09-D05-AB.ab1	unnamed protein product [Homo sapiens]	BAC05050.1	33
1-D08-AB.ab1	unnamed protein product [Macaca fascicularis]	BAB01686.1	40
2-H07-AB.ab1	unnamed protein product [Mus musculus]	BAB25726.1	51
4-G09-AB.ab1	unnamed protein product [Mus musculus]	BAB25726.1	49
pst_2_jhLee_266	upregulated in colorectal cancer gene 1 protein precursor;	NP_060019.1	29
05-A08-AB.ab1	upregulated in colorectal cancer gene 1 protein precursor;	NP_060019.1	32
05-C09-AB.ab1	variable region-containing chitin-binding protein 2 [Branchiostoma	AAN62849.1	38
pst_2_jhLee_213	variable region-containing chitin-binding protein 5 [Branchiostoma	AAN62911.1	43
06-A09-AB.ab1	Xrcc5 [Rattus norvegicus]	BAB83859.1	36
09-A08-AB.ab1	xylose isomerase [Pirellula sp.]	NP_865078.1	57
2-B11-AB.ab1	Yolk sac gene 2 [Mus musculus]	AAH07136.1	34
3-G06-AB.ab1	ZP2 [Carassius auratus]	CAA96576.1	40
2-D07-AB.ab1	ZP2 [Carassius auratus]	CAA96576.1	46

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Table 1-2. ESTs with similarities to sequences in NCBI data bases with an E value less than 10^{-5} in BLAST N analysis

EST	Putative identification	AC	% Identity
4-H03-AB.ab1	Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MDC8	AP000373.1	86
07-F03-AB.ab1	Chromobacterium violaceum strain ATCC 12472 section 6 of 16 of the	AE016915.1	100
pst_2_jhLee_264	Cloning vector TLF97-3, phage lambda lacZ translational fusion vector,	U39286.1 |CV U39286	100
pst_2_jhLee_293	Cyprinus carpio clone cL41a ribosomal protein L41 mRNA, complete	AY117540.1	89
10-F02-AB.ab1	H.rubra mRNA for putative abalone protein (628bp)	X92692.1 |HRMRNA628	94
07-B05-AB.ab1	Haliotis discus discus DNA, CA repeat region	AB025396.1	86
10-C01-AB.ab1	Haliotis discus discus DNA, CA repeat region	AB025396.1	91
pst_2_jhLee_563	Haliotis discus discus DNA, CA repeat region	AB025369.1	93
07-D02-AB.ab1	Haliotis discus discus DNA, CA repeat region	AB025369.1	88
3-E11-AB.ab1	Haliotis discus discus gene, microsatellite Hd201	AB085642.1	97
3-H07-AB.ab1	Haliotis discus hdcel-1 gene for cellulase, partial cds	AB092979.1	88
pst_2_jhLee_298	Haliotis diversicolor 16S ribosomal RNA gene, mitochondrial gene	U51989.1 |HDU51989	90
pst_2_jhLee_54	Haliotis diversicolor 16S ribosomal RNA gene, mitochondrial gene	U51989.1 |HDU51989	90
pst_2_jhLee_92	Haliotis diversicolor 16S ribosomal RNA gene, mitochondrial gene	U51989.1 |HDU51989	89
09-C10-AB.ab1	Haliotis fulgens fertilization protein precursor, gene, exon 1 and	AF076836.1 |HFGAFP1	88
3-C01-AB.ab1	Haliotis fulgens lysin precursor, gene, exon 4 and partial cds	AF076835.1 |HFGALP3	88
07-A07-AB.ab1	Haliotis kamtschatkana clone Hka48 microsatellite sequence	AY013578.1	92
1-E07-AB.ab1	Haliotis kamtschatkana clone Hka56 microsatellite sequence	AY013579.1	83
1-B01-AB.ab1	Haliotis rubra clone 2.20 microsatellite VNTR sequence	AF302830.1 |AF302830	91
09-D07-AB.ab1	Haliotis rubra clone 2.20 microsatellite VNTR sequence	AF302830.1 |AF302830	94
05-E09-AB.ab1	Haliotis rubra clone 2.20 microsatellite VNTR sequence	AF302830.1 |AF302830	90
4-F02-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	94
4-G05-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	95
07-A09-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	88
05-B12-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	89
06-H07-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	88
05-C02-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	84
07-C05-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	82
1-E09-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	84
2-G03-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	90
05-F08-AB.ab1	Haliotis rufescens from La Jolla, CA, G-alpha signal transducing	AF070959.1 |HRGALJ1	86
09-C01-AB.ab1	Haliotis rufescens from Mendocino, CA, G-alpha signal transducing	AF070957.1 |HRGAME1	93
05-C03-AB.ab1	Haliotis rufescens from Mendocino, CA, G-alpha signal transducing	AF070958.1 |HRGAME2	95
04-E03-AB.ab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	93
04-A05-AB.ab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	90
04-H07-AB.ab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	95

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

pst_2_jhLee_459	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	95
pst_2_jhLee_2	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
05-B07-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	92
08-H09-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
03-B05-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
07-C06-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
pst_2_jhLee_10	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
pst_2_jhLee_9	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
02-G12-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
08-A08-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
pst_2_jhLee_18	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
07-D06-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
08-A02-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	91
08-G07-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	89
08-C07-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	92
03-H06-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	92
04-G08-ABab1	Haliotis rufescens lustrin A mRNA,	AF023459.1 |AF023459	92
06-A01-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	89
2-A07-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	88
06-D06-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	88
4-E10-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	90
08-C08-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	84
06-H08-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	86
2-A09-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	87
2-E05-ABab1	Haliotis rufescens lysin precursor, gene, exons 4 and 5 and partial	AF076824.1 |AF076824	90
3-C02-ABab1	Haliotis tuberculata partial H1 gene for hemocyanin, exons 1-17	AJ252741.1 |HTU252741	86
4-D12-ABab1	Homo sapiens laminin receptor 1 (ribosomal protein SA, 67kDa), mRNA	BC013827.2	87
05-E02-ABab1	Mouse DNA sequence from clone RP23-201H16 on chromosome 13, complete	AL590870.14	97
pst_2_jhLee_202	Nucella emarginata cytochrome b (cytb) gene, mitochondrial gene	U69726.1 |NEU69726	81
05-C01-ABab1	Rana tigrina ranavirus, complete genome	AF389451.1	96
3-H10-ABab1	Suberites domuncula mRNA for gelsolin (gels gene)	AJ344135.1 |SDO344135	93
4-C06-ABab1	Zebrafish DNA sequence from clone DKEY-248K5, complete sequence	BX005440.3	100

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

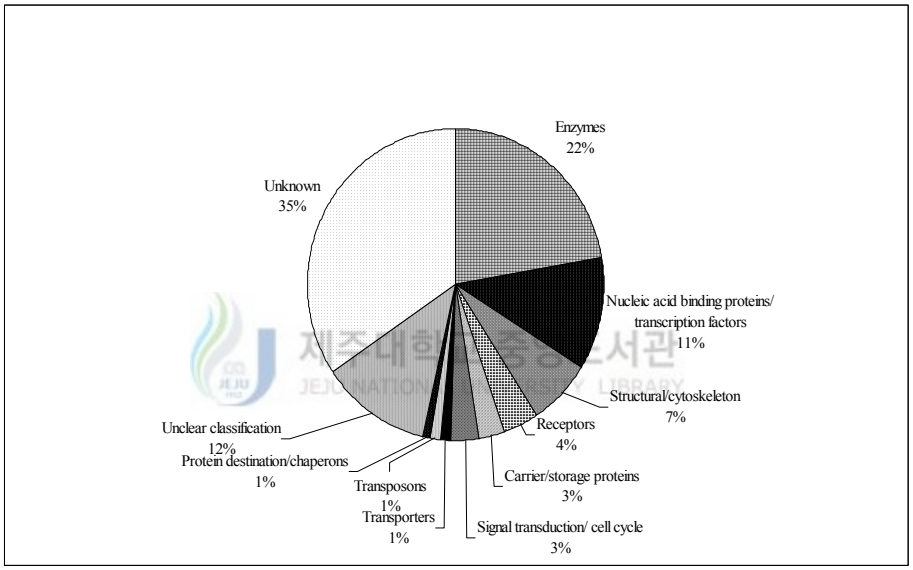


Fig. 1-2. Classification of disk abalone digestive gland ESTs, showing the proportions of transcripts from different genes according to their putative biological role.

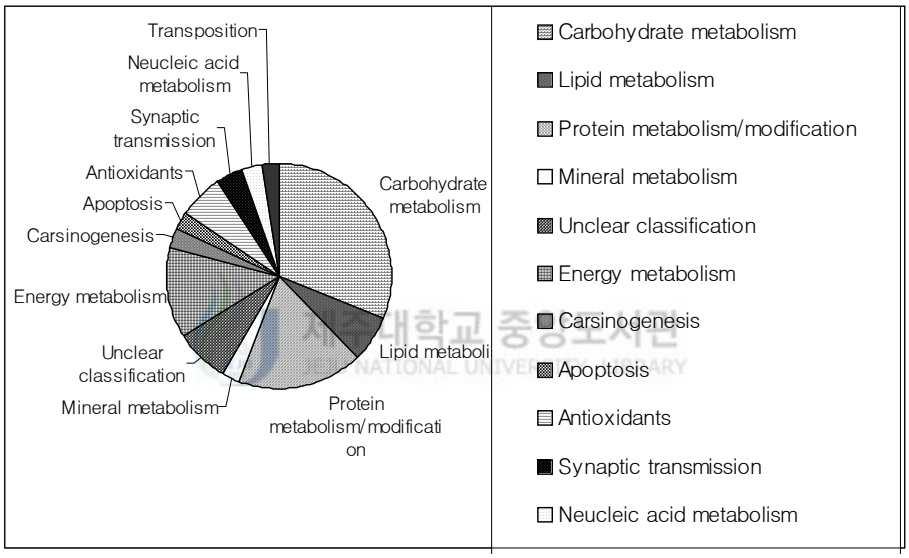


Fig. 1-3. Classification of disk abalone digestive gland ESTs with significant similarity to known enzymes.

4. DISCUSSION

Despite their fragmentary and inaccurate nature, ESTs were found to be an invaluable resource for the discovery of new genes (Sikela and Auffray, 1993). To gain information on the composition of expressed transcripts in abalone digestive gland, we analysed 841 high quality ESTs.

Digestive gland is a mass of branching tubules bathed in blood in the visceral haemocoel. Ducts from these glands open into the stomach. They are not known to secrete digestive juices but they do take up food particles from the fluid that flows up the ducts into the glands and digest them in food vacuoles (Villee *et al.*, 1968). The intimate juxtaposition of digestive gland and gonad makes it difficult to study either organ separately (Carefoot *et al.*, 1998). We encountered several gonad specific gene transcripts in digestive gland cDNA library, indicating its contamination by gonad tissue. There were two clones of egg membrane protein ZP2, which forms zona pellucida with ZP1 and ZP3 in mammals. ZP2 acts as a secondary sperm receptor. It is proteolytically cleaved after fertilization, and this modification, along with presumed changes in ZP3 is thought to play an important role in the postfertilization block to polyspermy (Liang and Dean, 1993). So far, ZP1, ZP2, or ZP3 like proteins have not been detected in invertebrate vitelline envelopes. Instead, invertebrate vitelline envelopes are composed of completely different proteins, such as VERL in the abalone (Evans, 2000). This unexpected sequence similarity could be due to inaccuracy of ESTs and further studies involving cloning of full-length cDNA and analyzing its product could verify this. We also found cyclins, cathepsin, lysine and a fertilization protein. Cyclins play an important role in egg maturation, in meiosis as well as in the normal cell cycle (Casas *et al.*, 1999). Cathepsins are likely to be involved in processing of egg yolk protein precursor, vitellogenins (Carnevali *et al.*, 1999).

Lysin is a protein released by the sperm acrosome reaction and it nonenzymatically and species selectively creates a hole in the egg vitelline envelope (Vacquier *et al.*, 1997) by binding to a high molecular weight glycoprotein (Swanson and Vacquier, 1997). This species

specificity is important to prevent cross-fertilization in heterospecific mixtures of sperm and eggs since many marine invertebrate species, including abalone, spawn gametes into seawater. The 18-kDa fertilization protein also released from the sperm acrosome, appears to mediate fusion of the sperm with the egg membrane (Swanson and Vaquier, 1995; Evans, 2000).

In the digestive gland cDNA library we generated, most of the unique sequences coded for enzymes (Fig. 1-3). Enzymes, the workhorses of the cell, are often overlooked in analysis of tissue expression patterns in favor of other groups, such as transcription factors and receptors (Wistow *et al*., 2002). However they are clearly of great importance and, for the digestive gland and gonad, they have special significance in some key areas. Morse *et al*., 1977 found that highly reactive and short-lived free radical oxidants could cause the induction of spawning in mollusks. Presence of these chemicals in excess can also cause gametes to be non-viable. Being a high metabolically active organ, as indicated by a large proportion of mRNA expressed being related to metabolism, digestive gland generates a large number of reactive oxygen species and it needs to detoxify these harmful by-products of aerobic respiration. Although cytochrome *c* oxidase and other proteins that reduce O_2 are remarkably successful in not releasing intermediates, small amounts of superoxide anion and hydrogen peroxide are unavoidably formed during respiration. Superoxide dismutase scavenges superoxide radicals by catalyzing the conversion of two of these radicals into hydrogen peroxide and molecular oxygen. The hydrogen peroxide formed by superoxide dismutase and by other processes is scavenged by catalase, a ubiquitous heme protein that catalyzes the dismutation of hydrogen peroxide into water and molecular oxygen (Berg *et al*., 2002). Glutathione peroxidase is another enzyme involved in detoxification of hydrogen peroxide and organic peroxides by catalyzing the reaction between these reactive species and glutathione (Siegel *et al.,* 1999). Peroxiredoxin-5 is a recently discovered mitochondrial, peroxisomal and cytosolic thioredoxin peroxidase able to reduce hydrogen peroxide and alkyl hydroperoxides (Tien and Knoops, 2003). All of these enzymes were found in the cDNA library of digestive gland, which reflects the

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

importance of these antioxidants in the proper functioning of these organs.

As expected, most of the sequences were coding for digestive enzymes with a ratio of 6 : 3 : 1 for carbohydrate : protein : lipid. This can be an indication of the composition of their diet. The enzymes involved in carbohydrate metabolism included cellulase, chitin synthase A, alpha-L-fucosidase, beta-agarase and chitinase. Because kelp and other macroalgae comprise the bulk of the natural diet, abalone may be expected to digest complex carbohydrates such as agar, fucoidan and cellulose (Monje and Viana, 1998). The origin of the invertebrate cellulases was initially explained as products of symbiotic microorganisms in the intestine or contamination by foods (Cleveland 1924; Martin and Martin, 1978). However, those cellulases have been considered to be the products of invertebrates themselves, as animals bred in the presence of antibiotics could produce cellulases (O'Brien *et al*., 1979). Present study supports the idea that cellulases are products of abalone itself, since it had higher homology to *Haliotis discus hannai* cellulase (Suzuki *et al*., 2003) than to any other cellulase. Chitinases cleave the β-1,4-glycosidic bonds of chitin, a β-1,4-linked, unbranched polymer of *N*-acetylglucosamine, which is a major component of insect exoskeletons, shells of crustaceans, and fungal cell walls. These enzymes have been detected in a variety of organisms, including organisms that do not contain chitin as a structural component, such as bacteria, plants, and animals. The production of chitinases by plants is thought to be involved in defense reactions against chitin-containing pathogens. Bacteria utilize chitinases for assimilation of chitin as a carbon and nitrogen source, and these enzymes play an important ecological role in the degradation of chitin (Wu *et al*., 2001). They have been used in a number of applications, including biocontrol of pests and as agents for the control of phytopathogenic fungi. Other useful enzymes identified in this study include several sulfatases, proteases and phosphatases.

Among the most abundant cDNAs in the data set are several nucleic acid binding proteins and transcription factors. These included Kruppel-like factor-15, which is suggested to be involved in type I collagen synthesis and tissue fibrosis (Uchida *et al*., 2000) as well as

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Myc homologs and poly(A)-binding protein. Poly(A)-binding protein is well known for its ability to bind with high affinity to poly(A) tails of mRNAs, prerequisite for mRNA stabilization and stimulation of translational initiation (Mohr *et al.,* 2001). Proto-oncogene product c-Myc has a direct role in both metazoan cell growth and division. RNA polymerase III (pol III) is involved in the generation of transfer RNA and 5S ribosomal RNA, and these molecules must be produced in bulk to meet the need for protein synthesis in growing cells. c-Myc binds to TFIIIB, a pol III-specific general transcription factor, and directly activates pol III transcription (Gomez-Roman *et al.*, 2003). Other sequences identified with clinical value are several genes involved in apoptosis, acid alpha glucosidase and RAS related proteins.

Assembly of ESTs into clusters with overlapping fragments revealed a high redundancy in whole abalone cDNA library than in digestive gland cDNA library we generated. Since whole abalone cDNA library contains sequences from genes expressed in almost all tissues except shell, it is expected to give an indication of average expression level of genes in the whole body. Therefore products of genes that carry out basic functions in every cell type are expected to be more prevalent in the whole abalone library than in digestive gland cDNA library.

To our knowledge, this is the first description of an abalone digestive gland cDNA library. The results show that the library contains clones for several mRNAs related to metabolism as expected. The occurrence of several transcripts (35%) without any significant similarity to known genes within the library makes it a valuable resource with high sequence complexity for unraveling new genes related to digestion and reproduction.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Part II

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Characterization of genes

Part II

Characterization of genes

1. ABSTRACT

Putative functions could be assigned manually to 65% of the 354 sequences which had significant similarity to known genes from the abalone digestive gland cDNA library. Three clones from this coding for a putative glutathione transferase, sulfatase and fucosidase were selected for further characterization for their predicted functions. Glutathione transferases are a family of multifunctional enzymes involved in detoxification of xenobiotic agents, drug biotransformation and protection of cells against peroxidative damage. Sulfatase and fucosidase are implicated in disease conditions and are useful as catalysts in industry. Glutathione transferase is expressed in *E. coli* in the soluble form with the theoretically expected molecular weight of 24.6 kDa with high catalytic activity towards 1-chloro-2,4-dinitrobenzene, the universal substrate of all glutathione transferases except theta class. Comparison of its amino acid sequence with pdb protein structure database reveals it to contain the GST N and C conserved domains with the thioredoxin fold and five α-helices and sequence homology with other known glutathione transferases asignes it to Mu class. Unlike glutathione transferase, sulfatase expressed in the insoluble form with an expected molecular weight of 54 kDa and purification under denaturing conditions resulted in an inactive form towards *p*-nitrocatechol sulfate, the chromogenic substrate of most of the arylsulfatases. This could be due to several reasons including lack of posttranslational modifications, different substrate specificity, or its being coded for a different product than sulfatase. An interesting finding of sulfatase was the presence of a very short 3′ end composing only of 41 bp. Fucosidase did not express in *E. coli* and sequence analysis showed it to contain several nonsense mutations in

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

the coding sequence with the cloned fragment coding only for the amino end of L-fucosidase domain. Further knowledge on these enzymes would help to elucidate their biological significance and glutathione transferase will provide a means of improving resistance of plants and animals to environmental pollutants and less accumulation of environmental toxins in food products and will also help in prevention of oxidative damage.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

2. MATERIALS AND METHODS

2. 1. Selection of clones for further characterization and full length sequencing

Sequences with a significant similarity ($E< 10^{-60}$) to known useful enzymes were compared with full length coding sequences available at NCBI Genebank for the genes with corresponding putative function. The presence of full length coding sequence was determined by the alignment of each EST with the 5′ end of coding sequence of known gene, including the start codon. Availability of assay systems for further characterization was also a main determining factor in the selection process. Three clones with putative GST, ASB and L-fucosidase functions were selected from the digestive gland cDNA library. XLI-Blue MRF′ cells were transformed with plasmid DNA of each clone and plasmids were isolated by Accuprep^{TM} plasmid extraction kit (Bioneer Co., Korea). Insert size of each clone was determined by restriction digestion of 2 ng miniprep DNA with *Xho*1 and *EcoR*I followed by analysis on a 1% agarose gel.

Full length of each clone was determined by several sequencing reactions carried out with the primers given in Table 2-1.

Table 2-1. Primers used to sequence full lengths.

Putative gene name	First sequencing primer	Second sequencing primer
Glutathione transferase	5'agaataatgcagcctggc3'	5'gattctgtcttagacgtgctg3'
Arylsulfatase B	5'tacctggcgtaccaagctg3'	5'gtacggcaatgtctggtc3'
L-fucosidase	5'agcgagaagtacaagcctg3'	

2. 2. Amplification of the coding sequences and sub cloning into pBlueScript SK-I vector

Full length sequences derived were compared with coding sequences of known genes to identify the start sites. This was further verified by the presence of an open reading frame with a length of expected size range as compared to known sequences. After considering the restriction sites of each insert, primers containing *Xho*I for L-fucosidase and *Nde*I sites for both GST and ASB were synthesized (Table 2-2).

Table 2-2. Primers used to amplify coding sequences

Putative gene name	Primer for the amplification of 5'end	Primer for the amplification of 3'end
Glutathine S transferase	5'gacatatgcctactcttggatactggg3'	5'gacatatgtcacttgaacagtgcactcttg'
Arylsulfatase B	5'gacatatgtttgtccagttattatgc3'	5'gacatatgtcaacaccagccagg3'
L-fucosidase	5'gactcgagatgaagaggttccttgaagg3'	5'gactcgagctattgaagtgcagtaagtttg3'

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

These were used to amplify the coding sequences of each clone. Each 50 μL PCR reaction mixture contained 2.5 units Takara Ex Taq polymerase (Takara Korea Biomedical Inc., Korea), 5 μL 10x Ex Taq buffer, 4 μL of 2.5 mM each dNTP, 50 ng of template and 50 pmol of each primer. After initial denaturation at 94°C for 10 min, each reaction was subjected to cycling conditions of danaturation at 94°C for 30 s followed by 30 s of annealing at 51°C for GST, 49°C for L-fucosidase and 45°C for ASB. The extension was carried out at 72°C for 30 s for GST and 1 min for both L-fucosidase and ASB. After 20 cycles, the final extension was carried out at 72°C for 5 min. The PCR products were purified with AccuPrep^{TM} gel purification kit (Bioneer Co., Korea) and were phosphorylated at 37°C for 1 hr in a 50 μL reaction by adding 5 μL 10x kinase buffer, 5 μL of 10 mM ATP, and 1 μL Takara kinase (Takara Korea Biomedical Inc., Korea) into 39 μL of purified PCR product. The phosphorylated products were dissolved in 28.5 μL distilled water each and were trimmed with 2.5 units of T4 DNA polymerase (Takara Korea Biomedical Inc., Korea) in 40 μL reaction mixtures

containg 4 μL of 10 × T4 DNA polymerase buffer, 0.1% BSA and 3 μL 2.5 mM dNTP at 37°C for 5 min. The products were then purified from a 1% agarose gel using QiaexII gel purification kit (QIAGEN Inc., USA). The vector pBlueScript SK-I was prepared by digesting 2 μg of vector with 20 units of *Hinc*II enzyme (Takara Korea Biomedical Inc., Korea) in a 40 μL reaction containing 4 μL of 10x H buffer.

Each PCR product was ligated to the prepared pBlueScript SK-I in a 10 μL reaction each containing 10 units of Takara DNA ligase (Takara Korea Biomedical Inc., Korea), 1x T4 DNA ligase buffer, 0.7 μg of PCR product and 100 ng of the vector at 16°C overnight. The ligated products were transformed into XLIBlue cells and the transformants were identified by blue white screening.

2. 3. Cloning into pET16b expression vector and protein expression.

The inserts digested with respective restriction enzymes were ligated to dephosphorylated pET16b vector. The dephosphorylation of the vector was carried out twice at 37°C for 15 min followed by heat inactivation of the enzyme at 50°C for 15min. After gel purification it was ligated with the gel purified inserts. The ligated products were transformed into XLI-BlueMRF′ cells and transformants were selected by ampicilin. Transformants containing inserts were identified by colony cracking and the orientation of insert was determined by restriction digestion with the enzymes as given in Table 2-3. Plasmids purified from clones containing the insert in correct orientation were transformed into *E. coli* BL21(DE3) cells and each protein was induced at 15°C and 37°C. The amount of expressed GST protein was higher at 37°C than 15°C and was found in both soluble fraction and insoluble fraction of cell lysate. ASB was in inclusion bodies at both temperatures. GST was purified from soluble fraction using His Bind Kits (Novagen, USA) using mechanical disruption method from columns containing 500 μL of settled bed volume His binding resin under native conditions. ASB was purified from a similar way but from inclusion bodies under denaturing conditions containing urea. After purification, GST was

dialysed against PBS (pH5.2) buffer and ASB with Tris.HCl (pH 7.2) buffer.

.

Table 2-3. Restriction enzymes used for the determination of insert orientation

Putative gene name	Restriction enzyme	Expected size if the insert is in the correct orientation (bp)	Expected size if the insert is in the opposite orientation (bp)
Glutathione transferase	*NCo*I	368	400
Arylsulfatase	*Hind*III	420	1624
L-fucosidase	*Xba*1	228	752

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

3. RESULTS

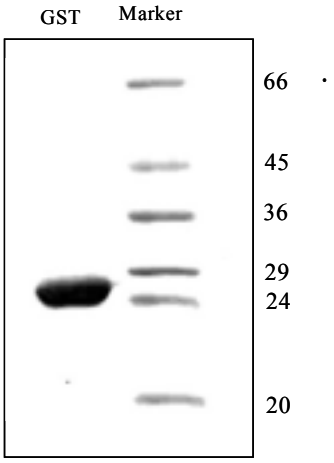


Fig. 2-1. Analysis of glutathione transferase (GST) protein expressed in *Escherichia coli* (BL21) cells following purification in a 12% denaturing polyacrylamide gel. Cells were grown at 15°C and induced with 1 mM IPTG. Recombinant protein was purified under native conditions by Ni column.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

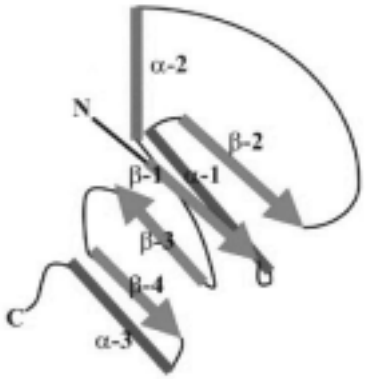


Fig. 2-2. A schematic diagram depicting thioredoxin fold. α-helices are shown as cylinders and β-sheets as arrows. In glutathione transferases, domain 2 is connected to the C-terminus via a short linker peptide (Sheehan *et al.*, 2000).

Table 2-4. The Highest and lowest % identities seen within coding regions of GSTs from same class and with abalone GST. Some classes had same identity between different sequences and zeta class had only one sequence for comparison.

Class	Alpha	Beta	Insect	Kappa	Microsomal	Omega	Pi	Tau	Theta	Zeta	Mu
Lowest % Identity (with abalone GST)	7	1	3	5	2	3	9	2	3	3	54
Highest % Identity (with abalone GST)	14	4	9	13	7	5	40	6	9		59
Lowest % Identity (within class)	59	5	12	89	3	3	30	75	82		84
Highest % Identity (within class)	91	30	44		79	80	93				91

Table 2-5. The Highest and lowest % identities seen within amino acid sequences of GSTs from same class and with abalone GST. Some classes had same identity between different sequences and zeta class had only one sequence for comparison.

Class	Alpha	Beta	Insect	Kappa	Microsomal	Omega	Pi	Tau	Theta	Zeta	Mu
Lowest % Identity (with abalone GST)	23	9	8	5	2	9	30	8	14	8	48
Highest % Identity (with abalone GST)	27	20	14	8	6	13	31	13			50
Lowest % Identity (within class)	53	12	9	78	2	14	47	64	80		82
Highest % Identity (within class)	88	36	38		82	77	97				86

제주대학교 중앙도서관

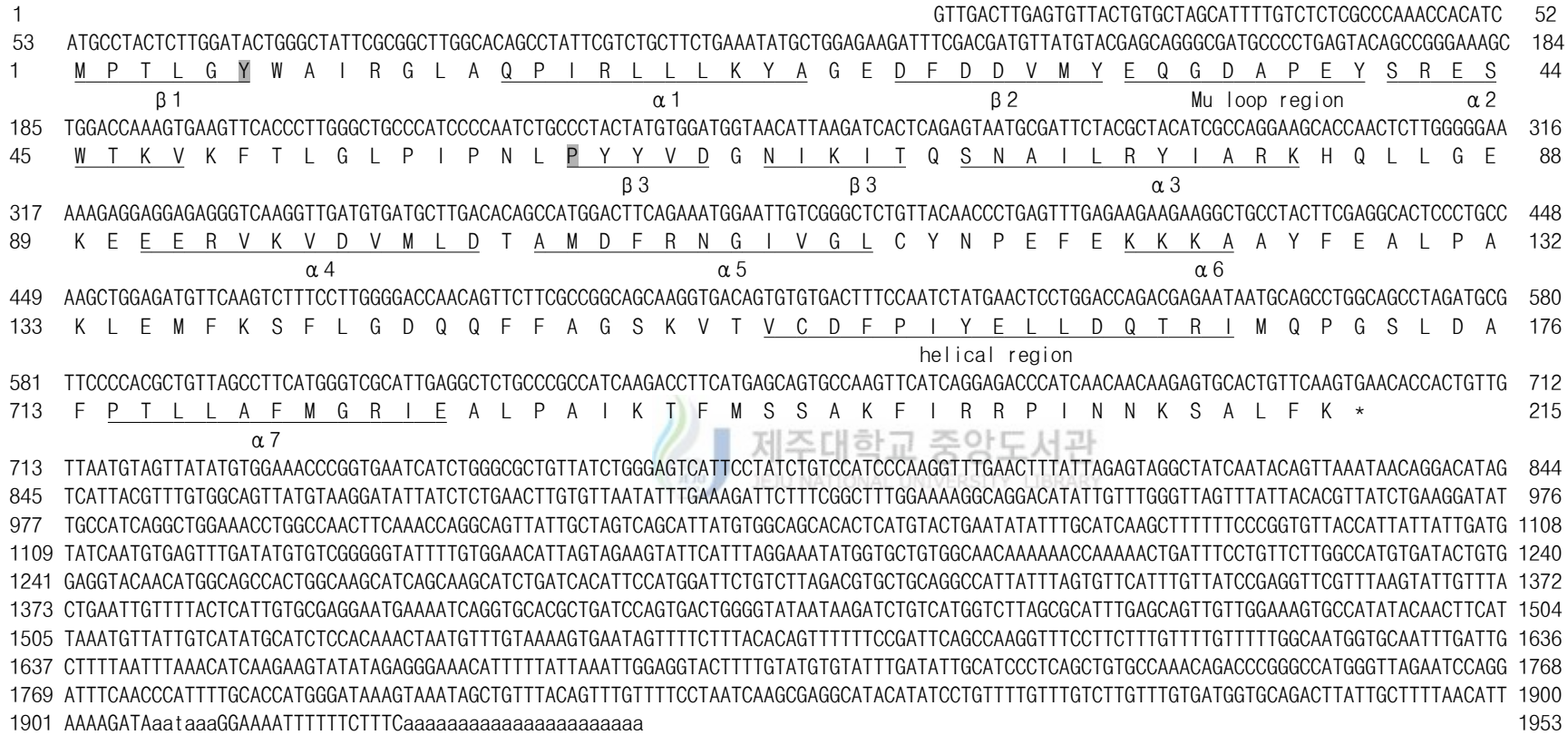


Fig. 2-3. Nucleotide and deduced amino acid sequence of glutathione transferase. The secondary structures composing two domains are indicated and the active site Tyr-6 and *cis* proline in *cis*-Pro loop are highlighted. The poly(A) tail and the poly(A) signal sequence are shown in simple case letters. The stop codon is marked with an asterisk.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

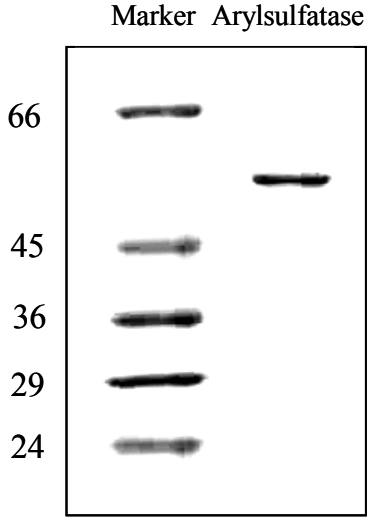


Fig. 2-4. Analysis of arylsulfatase protein expressed in *Escherichia coli* BL21(DE3) cells following purification in a 12% denaturing polyacrylamide gel. Cells were grown at 15℃ and induced with 1 mM IPTG. Recombinant protein was purified under denaturing conditions by Ni column.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

```
1                                                                                                                    AGAAATAGGAAACCAACGTGTAATTCCTGTCCACGGAAGGAGAG    44
45   ATGTTTGTCCAGTTATTATGCACAGTTTTGGTCATCATCAACCTCTGTGATGACGTTTCTGCAGCAGGACGTCCACGCCATATTGTGTTCATCGTGGCGGATGATCTCGGATGGAACGACATTGGCTTTCAC  176
1     M  F  V  Q  L  L  C  T  V  L  V  I  I  N  L  C  D  D  V  S  A  A  G  R  P  R  H  I  V  F  I  V  A  D  D  L  G  W  N  D  I  G  F  H    44
177  AACCCCGATATAATCACACCCAACATCGACAAGCTGGCAAGAGAAGGCTTGCTTCTGAATCATCACTATGTTCAACCACTCTGCAGTCCATCGAGAGCTGCCTTTATGTCCGGCTACTACCCCTTCAAGACA  308
45    N  P  D  I  I  T  P  N  I  D  K  L  A  R  E  G  L  L  L  N  H  H  Y  V  Q  P  L  C  S  P  S  R  A  A  F  M  S  G  Y  Y  P  F  K  T    88
309  GGTCTGCAGCACTCGGTCATTCTGGAGAACCAGCCCGTCTGTCTACCCCTGAATATCACAATCCTGCCACAGAAACTGAAGGAGCTTGGATATGCAACACACATTGTCGGCAAGTGGCACAATGGGTTCTGT  440
89    G  L  Q  H  S  V  I  L  E  N  Q  P  V  C  L  P  L  N  I  T  I  L  P  Q  K  L  K  E  L  G  Y  A  T  H  I  V  G  K  W  H  N  G  F  C   132
441  AGTTGGAATTGCACCCCGACGTACCGTGGCTTTGACAGCTTCTTTGGCTACTACGGCGCCATGGAAGACTACTACACCCACGTCATTCGTGGCTTCCTTGACTACCGTAACAACACCACCCCCGTTTGGACC  572
133   S  W  N  C  T  P  T  Y  R  G  F  D  S  F  F  G  Y  Y  G  A  M  E  D  Y  Y  T  H  V  I  R  G  F  L  D  Y  R  N  N  T  T  P  V  W  T   176
573  GACAACGGCACTTACTCAACGCTTCGGTTTACTGACGTAGCCACTGACATCATCGAGCGTCACAACCAGAGTCAGCCATTGTTTCTGTACCTGGCGTACCAAGCTGTCCACGGACCTATTGAGGTTCCCGCA  704
177   D  N  G  T  Y  S  T  L  R  F  T  D  V  A  T  D  I  I  E  R  H  N  Q  S  Q  P  L  F  L  Y  L  A  Y  Q  A  V  H  G  P  I  E  V  P  A   220
705  AAGTATGAAGCAATGTATCCAAACATTAAATCAGAAAATCGTCGAAAGTTTTCGGGAATGGTCTCTGCTCTTGATGAAGCAGTTGGTAACGTAACTAAAACGTTAAGACAAAGAGGGTTAATGGACGACACG  836
221   K  Y  E  A  M  Y  P  N  I  K  S  E  N  R  R  K  F  S  G  M  V  S  A  L  D  E  A  V  G  N  V  T  K  T  L  R  Q  R  G  L  M  D  D  T   264
837  CTGATTCTGTTCACTGCCGATAATGGCGGCGGGGTCGACGAATCTGGGAACAACTACCCTCTGCGTGGAAGCAAGTTTACCGTGTACGAAGGCGGAACGAGAGCTGTGGGCTTCATGTATGGATCGGGTCTC  968
265   L  I  L  F  T  A  D  N  G  G  G  V  D  E  S  G  N  N  Y  P  L  R  G  S  K  F  T  V  Y  E  G  G  T  R  A  V  G  F  M  Y  G  S  G  L   308
969  CAAAAGACTGGAACTGTATTTGACGGGATGATCCACGCCGTGGACTGGCTGCCCACCCTGACAGCAGCTGCCGGGGGGACCCCAGTGTCCGACCGTGACGGCATCAATCTGTGGCCTAGTCTCAGCACAGCC 1100
309   Q  K  T  G  T  V  F  D  G  M  I  H  A  V  D  W  L  P  T  L  T  A  A  A  G  G  T  P  V  S  D  R  D  G  I  N  L  W  P  S  L  S  T  A   352
1101 TCCCCGTCCCCCCGCACTGAGGTCGTCTACAACTACGACTCGCACCCCCAGCCCGTTCAAGGACACGCTGCCATCAGAGTGGGTGACTACAAACTGATCGATGGCTACCCGGGACCCTTCCCTGATTGGTAC 1232
353   S  P  S  P  R  T  E  V  V  Y  N  Y  D  S  H  P  Q  P  V  Q  G  H  A  A  I  R  V  G  D  Y  K  L  I  D  G  Y  P  G  P  F  P  D  W  Y   396
1233 AAGCCTGAACAAGTCACATCTAGTTTGAACACCAGATTCAGCAGGGATTCGGCCAATCAGTATCAGCTGTTCAATTTGAAAGATGACCCCAATGAGCGCAACGACCTCTCCAACTTTCGTCCGGACATGGTA 1364
397   K  P  E  Q  V  T  S  S  L  N  T  R  F  S  R  D  S  A  N  Q  Y  Q  L  F  N  L  K  D  D  P  N  E  R  N  D  L  S  N  F  R  P  D  M  V   440
1365 AAGAAGCTTGCTGCCAGACTGGCCTGGTATAAGAAGCAGGCAGTACCACCCAACTTCCCTGAGACCCCCGACGACCTGAGCAACCCTGCACTGTACGGCAATGTCTGGTCTCCTGGCTGGTGTTGAGAGCTT 1496
405   K  K  L  A  A  R  L  A  W  Y  K  K  Q  A  V  P  P  N  F  P  E  T  P  D  D  L  S  N  P  A  L  Y  G  N  V  W  S  P  G  W  C  *        481
1497 CTTGTTGTACTGTCACTGaataaaGTCGATATGTGaaaaaaaaaaaaaaaaaaaaaaaaaa                                                                          1557
```

Fig. 2-5. Nucleotide and deduced amino acid sequence of arylsulfatase B. The putative signal peptide sequence is underlined. .Poly(A) tail and the poly(A) signal sequence are shown in simple case letters. The stop codon is indicated with an asterisk.

```
Abalone         --------------------------------------------------ATGTTTGTC- 9

AF109924        --------------------------------ATGTGCAAGTGTTTGCTCGTTTTGATCG 28
M32373          ---------ATGGGTCCGCGCGGCGCGGCGAGCTTGCCCCGAGGCCCCGGACCTCGGCGG 51
NM_198709       ---------ATGGGTCCGCGCGGCGCGGCGAGCTTGCCCCGAGGCCCCGGACCTCGGCGG 51
NM_009712       ATGGGCAAGCTTAGCCCGTGCACTGGCAGGAGCCGGCCTGGAGGTCCCGGGCCGCAGCTG 60


Abalone         CAGTTATTATGCACAGTTTTGGTCAT-CATCAACC-TCTGTGATGACGTTTCTGCAGCAG 67
AF109924        CCATCATCACCGCCTGTGCTGTGGCT-GACCAGTCGTCTGCATCAGCAGGAACGAGGCAA 87
M32373          CTGCTCCTCCCCGTCGTCCTCCCGCTGCTGCTGCTGCTGTTGTTGGCGCCGCCGGGCTCG 111
NM_198709       CTGCTCCTCCCCGTCGTCCTCCCGCTGCTGCTGCTGCTGTTGTTGGCGCCGCCGGGCTCG 111
NM_009712       CCACTCTTGC------TGCTGCTGCTGCAGCTGCTGCTGCTGCTGTCGCCGGCCCGG 114
                *               *  *     *    *                      *

Abalone         GACG------------TCCACGCCATATTGTGTTCATCGTGGCGGATGATCTCGGATGG 114
AF109924        GATGCCGG---------TCAACCCAACATTGTGTTTGTTCTAGCAGACGACTTCGGTTTC 138
M32373          GGCGCCGGGGCCAGCCGGCCGCCCCACCTGGTCTTCTTGCTGGCAGACGACCTAGGCTGG 171
NM_198709       GGCGCCGGGGCCAGCCGGCCGCCCCACCTGGTCTTCTTGCTGGCAGACGACCTAGGCTGG 171
NM_009712       GCCTCCGGAGCTACCCAGCCTCCGCATGTGGTCTTCGTGCTGGCGGACGACCTGGGCTGG 174
                *                 *  *   *  * ** **  *  * ** ** **  * ** *

Abalone         AACGACATTGGCTTTCACAACCCCGATATAATCACACCCAACA-TCGACAAGCTGGCAAG 173
AF109924        CATGACGTCGGTTACCATGGTTCCGAGATCCACAC-CCCAACACTTGATGCGCTGTCCGC 197
M32373          AACGACGTCGGCTTCCACGGCTCCCGCATCCGCACGCCGCACC-TGGACGCGCTGGCGGC 230
NM_198709       AACGACGTCGGCTTCCACGGCTCCCGCATCCGCACGCCGCACC-TGGACGCGCTGGCGGC 230
NM_009712       AACGACTTGGGTTTTCACGGCTCGGTCATCCGCACGCCGCACC-TGGATGCGCTGGCGGC 233
                 * *** * ** *  **     *    **   *** **  **  * **   **** *

Abalone         AGAAGGCTTGCTTCTGAATCATCACTATGTTCAACCACTCTGCAGTCCATCGAGAGCTGC 233
AF109924        CAGCGGCGTCAGGCTGGAGAACTACTACGTGCAGCCGATTTGCACCCCGACCAGGAGTCA 257
M32373          CGGCGGGGTGCTCCTGGACAACTACTACACGCAGCCGCTGTGCACGCCGTCGCGGAGCCA 290
NM_198709       CGGCGGGGTGCTCCTGGACAACTACTACACGCAGCCGCTGTGCACGCCGTCGCGGAGCCA 290
NM_009712       CGGCGGCGTGGTCCTGGACAACTACTACGTGCAGCCGCTGTGCACGCCCTCGCGGAGCCA 293
                    **  *    *** *  *  ****    ** **  * ****  **  *  *

Abalone         CTTTATGTCCGGCTACTACCCCTTCAAGACAGGTCTGCAGCACTCGGTCATTCTGGAGAA 293
AF109924        GCTCATGTCCGGGAGATATCAGATACATACTGGGCTTCAACATGGTATTATTAATAGTTG 317
M32373          GCTGCTCACTGGCCGCTACCAGATCCGTACAGGTTTACAGCACCAAATAATCTGGCCCTG 350
NM_198709       GCTGCTCACTGGCCGCTACCAGATCCGTACAGGTTTACAGCACCAAATAATCTGGCCCTG 350
NM_009712       GCTGCTCACCGGCCGCTATCAGATCCATTTGGGCTTACAGCATTACCTCATCATGACCTG 353
                  *  *  * **    ** *   *       **  * ** **     * **

Abalone         CCAGCCCGTCTGTCTACCCCTGAATATCACAATCCTGCCACAGAAACTGAAGGAGCTTGG 353
AF109924        TCAGCCAAACGCTCTTCCAAATGACAGTCCGACTTTGGCAGATAAACTGAAAGAATCTGG 377
M32373          TCAGCCCAGCTGTGTTCCTCTGGATGAAAAACTCCTGCCCCAGCTCCTAAAAGAAGCAGG 410
NM_198709       TCAGCCCAGCTGTGTTCCTCTGGATGAAAAACTCCTGCCCCAGCTCCTAAAAGAAGCAGG 410
NM_009712       TCAGCCCAGCTGTGTTCCCCTAGACGAAAAACTCCTGCCCCAGCTTTTAAAGGAAGCAGG 413
                 *****   *  * * **     *           ** *  *     * ** **    **

Abalone         ATATGCAACACACATTGTCGGCAAGTGGCACAATGGGTTCTGTAGTTGGAATTGCACCCC 413
AF109924        CTACGCCACACACATGGTTGGCAAGTGGCACCTAGGATTCTACAAACAGGAATACTTGCC 437
M32373          TTATACTACCCATATGGTCGGAAAATGGCACCTGGGAATGTACCGGAAAGAATGCCTTCC 470
NM_198709       TTATACTACCCATATGGTCGGAAAATGGCACCTGGGAATGTACCGGAAAGAATGCCTTCC 470
NM_009712       CTATGCTACGCACATGGTCGGGAAATGGCACCTGGGGATGTACCGGAAAGAATGTCTGCC 473
                 **  * ** ** ** ** ** ******   **  * *         * *     **

Abalone         GACGTACCGTGGCTTTGACAGCTTCTTTGGCTACTACGGCGCCATGGAAGACTACTACAC 473
AF109924        TTGGAACCGAGGCTTCGACACTTACTTTGGTTATCTTAATGCTGCTGAGGACTACTTTAA 497
M32373          AACCCGCCGAGGATTTGATACCTACTTTGGATATCTCCTGGGTAGTGAAGATTATTATTC 530
NM_198709       AACCCGCCGAGGATTTGATACCTACTTTGGATATCTCCTGGGTAGTGAAGATTATTATTC 530
NM_009712       GACACGCCGAGGATTCGATACCTACTTCGGATACCTCCTAGGCAGTGAAGACTATTACAC 533
                      *** ** ** ** *  * *** ** **       *     ** ** ** *

Abalone         CCACGT----CATTC-----------------GTG---------GCTTCCTTGACTACCG 503
AF109924        CCACAA----TGTACCTTGGAGA------CAAGTGC--------GCTACCTTGACCTCAG 539
M32373          CCATGAACGCTGTACATTAATTGACGCTCTGAATGTCACACGATGTGCTCTTGATTTTCG 590
NM_198709       CCATGAACGCTGTACATTAATTGACGCTCTGAATGTCACACGATGTGCTCTTGATTTTCG 590
NM_009712       CCACGAGGCCTGTGCGCCGATTGAGTCTTTGAATGGTACCCGCTGTGCGCTTGACTTAAG 593
                ***         * *                  **         *    *****     *

Abalone         TAACAACACCACCCCCGTTTGGACCGACAACGGCACTTACTCAACGCTTCGGTTTACTGA 563
AF109924        AGATAACAACGGACCCGTGAGAAACGAAACTGGGCAGTATTCAGCTCACCTGTTCACCGG 599
M32373          AGATGGCGAAGAAGTTGCAACAGGATATAAAAATATGTATTCAACAAACATATTCACCAA 650
NM_198709       AGATGGCGAAGAAGTTGCAACAGGATATAAAAATATGTATTCAACAAACATATTCACCAA 650
NM_009712       AGATGGTGAAGAACCTGCAAAAGAGTATAACAATATTTACTCAACAAACATATTTACCAA 653
                  *             *         * *        ** ***  *      **  **

Abalone         CGTAGCCACTGACATCATCGAGCGTCACAACCAGAGTCAGCCATTGTTTCTGTACCTGGC 623
AF109924        CAAAGCCATTGATGTTGTACAGTCACACAACACGTCCAAGCCGCTATTCTTGTACCTGGC 659
M32373          AAGGGCTATAGCCCTCATAACTAACCATCCACCAGAGAAGCCTCTGTTTCTCTACCTTGC 710
NM_198709       AAGGGCTATAGCCCTCATAACTAACCATCCACCAGAGAAGCCTCTGTTTCTCTACCTTGC 710
NM_009712       AAGGGCTACAACCGTCATAGCTAACCACCCACCGGAGAAACCACTGTTCCTCTACCTTGC 713
                    ** *      *  *       **           * **  * **  * ***** **

Abalone         GTACCAAGCTGTCCACGGACCTATTGAGGTTCCCGCAAAGTATGAAGCAATGTATCCAAA 683
AF109924        CTACCAGTCCGTGCACGCCCCACTTGAGGTGCCAGAAAAGTACGAACACAAGTATAGAAA 719
M32373          TCTCCAGTCTGTGCATGAGCCCCTTCAGGTCCCTGAGGAATACTTGAAGCCATATGACTT 770
NM_198709       TCTCCAGTCTGTGCATGAGCCCCTTCAGGTCCCTGAGGAATACTTGAAGCCATATGACTT 770
NM_009712       TTTCCAGTCTGTCCACGATCCCCTGCAGGTCCCTGAGGAGTACATGGAACCATATGGCTT 773
                   ***  * ** ** *  **  *  **** ** *   * **         ***

Abalone         CATTAAATCAGAAAATCGTCGAAAGTTTTCGGGAATGGTCTCTGCTCTTGATGAAGCAGT 743
AF109924        CATCACAGACAAAAACAGGCGCACTTTCGCGGGTATGGTCTCAGCCCTGGATGAGGGGGT 779
M32373          TATCCAAGACAAGAACAGGCATCACTATGCAGGAATGGTGTCCCTTATGGATGAAGCAGT 830
NM_198709       TATCCAAGACAAGAACAGGCATCACTATGCAGGAATGGTGTCCCTTATGGATGAAGCAGT 830
NM_009712       CATCCAAGATAAGCACAGACGTATTTATGCAGGGATGGTGTCCCTCATGGATGAAGCAGT 833
                 **   *    *  *  * *     *   * ** ***** **     * ***** *  **
```

Fig. 2-6. CLUSTAL W (1.82) multiple sequence alignment of arylsulfatase coding sequences. Identical residues are indicated with an asterisk.

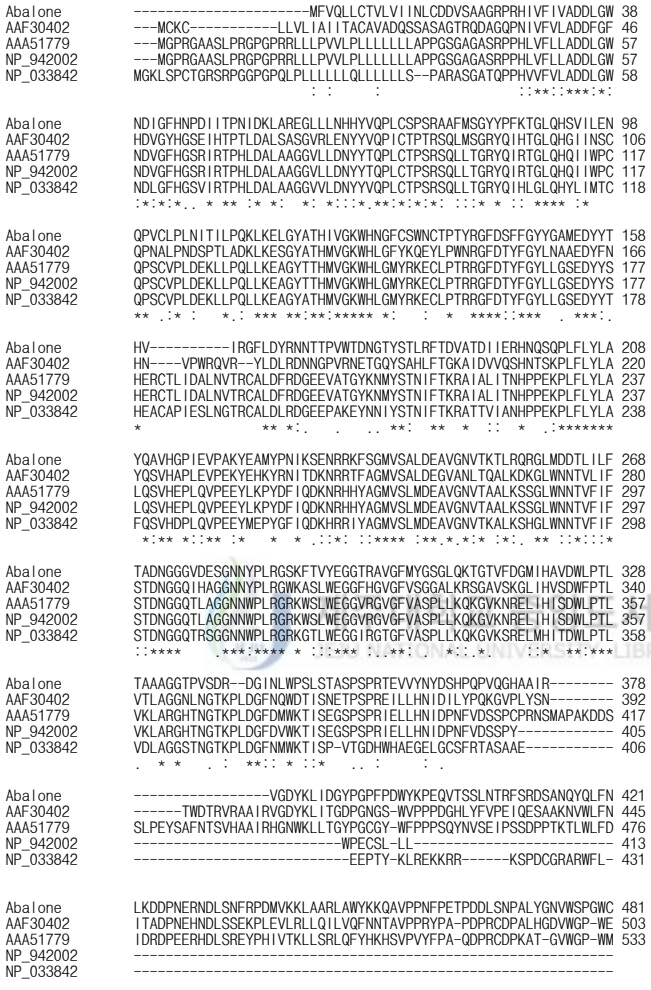


Fig. 2-7. CLUSTAL W (1.82) multiple sequence alignment of arylsulfatase amino acid sequences. Identical residues are shown in asterisks. Conserved substitutions depending on functionality are indicated with colon and the semi-conserved residues with a dot.

```
1                                                                                                                    AG    2
3    ATGAAGAGGTTCCTTGAAGGCTCTCTGCTTTTTGCCTTTGTCTGTCTGTGTGTAGCAACACCAACAGCAAAGGACAAAGCAAAATTCACTCGATATGAACCTAACTGGGCTTCTATAGATTCTAGACCTCTG  134
1     M  K  R  F  L  E  G  S  L  L  F  A  F  V  C  L  C  V  A  T  P  T  A  K  D  K  A  K  F  T  R  Y  E  P  N  W  A  S  I  D  S  R  P  L    44
135  CCAGCATGGTACGATGAAGCCAAACTTGGAATCTTCATTCACTGGGGTATCTTTTCCGTGCCCAGTTATGGTTCTGAGTGGTTCTGGTGGCTCTGGCAAGGGCAGAAGGTTCCAGACGTTGTTGCTTTCATG  266
45    P  A  W  Y  D  E  A  K  L  G  I  F  I  H  W  G  I  F  S  V  P  S  Y  G  S  E  W  F  W  W  L  W  Q  G  Q  K  V  P  D  V  V  A  F  M    88
267  AAAGACAATTACCGACCAGACTGGACATATGCTGATTTTGCCAGGGATTTTACAGCTGAGTTTTTCGATCCAGTACAGTGGGCCAACATCTTCAATGCATCCGGTGCACAATACGCTGTGTTAGTCAGCAAG  398
89    K  D  N  Y  R  P  D  W  T  Y  A  D  F  A  R  D  F  T  A  E  F  F  D  P  V  Q  W  A  N  I  F  N  A  S  G  A  Q  Y  A  V  L  V  S  K  132
399  CATCATGAAGGTTTCTGCAACTGGCCAACGAATGTCTCCTTCAACTGGAACTCCCAGATGCTTGGACCCAACAGAGATCTCGTGGGTGAACTAGCAGCGGCGATACGAGGAAACACTAACATCCGCTTTGGT  530
133   H  H  E  G  F  C  N  W  P  T  N  V  S  F  N  W  N  S  Q  M  L  G  P  N  R  D  L  V  G  E  L  A  A  A  I  R  G  N  T  N  I  R  F  G  176
531  CTCTACCACTCCCTGTTTGAATGGTTCCATCCCCTGTACCTTCAAGACAAGGCCAACAACTTTTCTACATCACGCTTTGTTGATGAGAAAACAATTCCTGAGTTGATTGAAATTAGCGAGAAGTACAAGCCT  662
177   L  Y  H  S  L  F  E  W  F  H  P  L  Y  L  Q  D  K  A  N  N  F  S  T  S  R  F  V  D  E  K  T  I  P  E  L  I  E  I  S  E  K  Y  K  P  220
663  GAAGTAATTTGGTCTGACaGtGACTGGGAgGCACCTTcCTcCTCTGGAAATCTGCAGAATTTCTGGCCTGGCTATATAATGAAAGTCCCACCAAGGACTCTGTGGTGACCAACGACCGATGGGGAAAAGATG  794
221   E  V  I  W  S  D  S  D  W  E  A  P  S  S  S  G  N  L  Q  N  F  W  P  G  Y  I  M  K  V  P  P  R  T  L  W  *                          255
795  CCACATGTCATCATGGAGGTTTCCTCACCTGCATGGACAGATACAACCCAGGAACACTACAGCCCCGTAAGTTTGAGAATGCCATGACAATAGACAAGAAGTCATGGGGTTTCCGGCGTAATGCTGATCTGG  926
927  CAGCCTACCTCAACATGGAAGAAATCTTAAGTACATTTGCCGAGACCATCAGTTGTGGAGGCAACATGCTCATCAATGTAGGGCCAACCAAATACGGTATGATCAGTGCTCTCTATGAGGAAAGACTGCGCC 1058
1059 AACTAGGCGGATGGTTGAGTGTCAATGGTGAGGGCATCTATTCCACAAAACCATGGACCTCCCAGAATGACACAGTGACCAAAGGAATCTGGTATGTACAGAAGCAGAGCAATGTCTATGCCATTGTACTGA 1190
1191 ACTGGCCGGATGCAGAGCTTCAGCTCGGGGCACCCAAGACAACTGCAGCTACTACAGTCAGTTTGCTCGGATACGACGGTAACTTCAGCTTTCAAGAACGAGCCGGAGGGGGCATAACCATAGAGATTCCTC 1322
1323 CAATCCCCATCAACAAAATGCCATGTGAATGGGCATGGATCTTCAAACTTACTGCACTTCAATAGACTAGATACTAGTATGCCAACTGGTTCTTGCAATATTTGTGTATAGATTTCATGTACAGATGTGAAG 1454
1455 TTACTGAATGTGAACATATAATACTCAAAATGTATATACAGAGCTACTGAAATAAAATATTTTGAAaataaaA                                                                  1527
```

Fig. 2-8. Nucleotide and deduced amino acid sequence of L-fucosidase. The putative signal peptide sequence is underlined. The poly(A) signal sequence is shown in simple case letters. The stop codon in the cloned fragment is indicated with an asterisk.

Fig. 2-9. CLUSTAL W (1.82) multiple sequence alignment of abalone fucosidase putative amino acid sequence with known L-fucosidase sequences. Identical residues are shown in asterisks. Conserved substitutions depending on functionality are indicated with colon and the semi-conserved residues with a dot.

제주대학교 중앙도서관 LIBRARY

```
Abalone         ------AGATGAAGAGGTTCCTTGAAGGCTCTCTGCTTTTTGCCTTTGTCTGTCTGTGTG 54
NM_012562       ATGTGGGACCTGAAAAGTGAGTGGTGGG--CGGTGGGCTTCGGGCTTCTGCTGCTACTGG 58
NM_000147       --------ATGAGGTCGCGGCCGGCGGGTCCCGCGCTGTTGCTGCTGCTGCTCTTCCTCG 52
NM_024243       --------------------------------ATGCTGCTGCTGCTGCTGCTGCTGCTAG 28
                                                *   *    * *    *   *

Abalone         TAGCAAC--ACCAACAGCAAAGGACAAAGCAAAATTCAC-TCGATATGAACCTAACTGGG 111
NM_012562       CAGCTTC--GGCCCAGGCAGGGGGCTTGGC----TCCGCACCACTACACTCCGGACTGGC 112
NM_000147       GAGCGGCCGAGTCGGTGCGTCGGGCCCAGCC---TCCGCGCCGCTACACCCCAGACTGGC 109
NM_024243       TGGCTGC--GGCCCAGGCAGTGGCCCTGGC----TCCGCGCCGCTTCACTCCGGACTGGC 82
                  **  *         **   ** *   **    * * *  *  *     **  *****

Abalone         CTTCTATAGATTCTAGACCTCTGCCAGCATGGTACGATGAAGCCAAACTTGGAATCTTCA 171
NM_012562       CGAGCTTGGACTCGCGACCACTGCCGAGATGGTTCGACGAGGCCAAGTTCGGGTTGTTCG 172
NM_000147       CGAGCCTGGATTCTCGGCCGCTGCCGGCCTGGTTCGACGAAGCCAAGTTCGGGGTGTTCA 169
NM_024243       AGAGCTTGGACTCGCGGCCACTGCCGAGCTGGTTCGATGAGGCCAAGTTCGGGGTGTTCG 142
                      * ** **  * ** *****    **** *** ** *****  * **  * ***

Abalone         TTCACTGGGGTATCTTTTCCGTGCCCAGTTATGGTTCTGAGTGGTTCTGGTGGCTCTGGC 231
NM_012562       TGCATTGGGGGGTGTACTCGGTGCCCGCCTGGGGCAGCGAGTGGTTCTGGTGGCACTGGC 232
NM_000147       TCCACTGGGGCGTGTTCTCGGTGCCCGCCTGGGGCAGCGAGTGGTTCTGGTGGCACTGGC 229
NM_024243       TGCACTGGGGGGTGTTCTCGGTGCCCGCCTGGGGCAGTGAATGGTTCTGGTGGCACTGGC 202
                * ** *****  * *  ** ******   *  **    ** ************* *****

Abalone         AAGGGCAGAAGGTTCCAGACGTTGTTGC-TTTCATGAAAGACAATTACCGACCAGACTGG 290
NM_012562       A-GGGCGAGCAGTCGTCGGCGTACGTGCGCTTCATGAAAGAAAACTACCCGCCCGGCTTC 291
NM_000147       A-GGGCGAGGGGCGGCCGCAGTACCAGCGCTTCATGCGCGACAACTACCCGCCCGGCTTC 288
NM_024243       A-GGGCGATCGGATGCCGGCCTACCAGCGCTTCATGACAGAAAACTACCCGCCCGGCTTC 261
                * ****     *    *   *    **  ******   ** ** ****  ** * **

Abalone         ACATATGCTGATTTTGCCAGGGATTTTACAGCTGAGTTTTTCGATCCAGTACAGTGGGCC 350
NM_012562       AGCTACGCCGACTTTGCACCGCAGTTTACAGCGCGCTTCTTCCATCCGGAGGAGTGGGCA 351
NM_000147       AGCTACGCCGACTTCGGACCGCAGTTCACTGCGCGCTTCTTCCACCCGGAGGAGTGGGCC 348
NM_024243       AGCTACGCCGACTTCGCACCGCAGTTCACAGCGCGCTTCTTCCACCCGGATCAGTGGGCC 321
                *  ** ** ** ** *    * * ** ** **    ** *** * ** *  *******

Abalone         AACATCTTCAATGCATCCGGTGCACAATACGCTGTGTTAGTCAGCAAGCATCATGAAGGT 410
NM_012562       GACCTCTTCCAGGCTGCCGGGGCCAAGTATGTCGTCCTGACCGCAAAGCATCACGAAGGC 411
NM_000147       GACCTCTTCCAGGCCGCGGGCGCCAAGTATGTAGTTTTGACGACAAAGCATCACGAAGGC 408
NM_024243       GAACTCTTTCAGGCTGCCGGGGCCAAGTACGTCGTCCTGACCACAAAGCATCATGAAGGC 381
                 *  ****  * **  * ** **  * ** *  **  *       ******** *****

Abalone         TTCTGCAACTGGCCAACGAATGTCTCCTTCAACTGGAACTCCCAGATGCTTGGACCCAAC 470
NM_012562       TTCACAAACTGGCCAAGCGCCGTGTCTTGGAACTGGAACTCGAAGGACGTGGGGCCCCAC 471
NM_000147       TTCACAAACTGGCCGAGTCCTGTGTCTTGGAACTGGAACTCCAAAGACGTGGGGCCTCAT 468
NM_024243       TTCACAAACTGGCCAAGCCCTGTGTCTTGGAACTGGAACTCGAAGGACGTGGGGCCCCAC 441
                ***   ******** *     ** ** *  **********  *     * ** **  *

Abalone         AGAGATCTCGTGGGTGAACTAGCAGCGGCGATACGAGGAAACACTAACATCCGCTTTGGT 530
NM_012562       CGTGATTTGGTCGGTGAGTTGGGAGCAGCTGTGCGGAAGAGGA---ACATACGATACGGC 528
NM_000147       CGGGATTTGGTTGGTGAATTGGGAACAGCTCTCCGGAAGAGGA---ACATCCGCTATGGA 525
NM_024243       CGTGATTTGGTCGGTGAGTTGGGAGCAGCTGTGCGGAAGAGGA---ACATACGCTACGGC 498
                 * *** * ** *****  * * * * **  * **    *  *   **** ** *  **

Abalone         CTCTACCACTCCCTGTTTGAATGGTTCCATCCCCTGTACCTTCAAGACAAGGCCAACAAC 590
NM_012562       CTCTACCACTCGCTCTTTGAGTGGTTCCATCCACTCTACCTACTAGATAAGAAAAATGGC 588
NM_000147       CTATACCACTCACTCTTAGAGTGGTTCCATCCACTCTATCTACTTGATAAGAAAAATGGC 585
NM_024243       CTCTACCACTCGCTCTTGGAATGGTTCCATCCACTCTACCTACTTGATAAGAAAAATGGC 558
                ** ******** ** ** ** *********** ** ** ** *  ** ***   **   *

Abalone         TTTTCTACATCACGCTTTGTTGATGAGAAAACAATTCCTGAGTTGATTGAAATTAGCGAG 650
NM_012562       CTCAAGACTCAGCATTTCGTCAGTACAAAAACAATGCCAGAGCTGTACGACCTCGTGAAT 648
NM_000147       TTCAAAACACAGCATTTTGTCAGTGCAAAAACAATGCCAGAGCTGTACGACCTTGTTAAC 645
NM_024243       TTCAAAACTCAGCATTTCGTCAGGGCAAAAACAATGCCAGAGCTGTATGACCTTGTTAAC 618
                 *    **    *  ** **       ******** ** *** **   **  *    *

Abalone         AAGTACAAGCCTGAAGTAATTTGGTCTGACAGTGACTGGGAGGCACCTTCCTCCT-CTGG 709
NM_012562       AGGTACAAGCCTGACCTAATCTGGTCTGATGGGGAGTGGGAGTGTCCCGATTCGTACTGG 708
NM_000147       AGCTATAAACCTGATCTGATCTGGTCTGATGGGGAGTGGGAATGTCCTGATACTTACTGG 705
NM_024243       AGCTACAAGCCTGACCTGATCTGGTCCGACGGGGAGTGGGAGTGTCCTGACACGTACTGG 678
```

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

```
                *  ** ** *****  * ** ***** **  * ** *****    **     * * ****

Abalone         AAATCTGCAGAATTTCTGGCCTGGCTATATAATGAAAGTCCCACCAAGGACTCTGTGGTG 769
NM_012562       AACTCCACGGAGTTCCTTGCTTGGCTCTACAATGAAAGCCCGGTCAAGGATCAGGTGGTA 768
NM_000147       AACTCCACAAATTTTCTTTCATGGCTCTACAATGACAGCCCTGTCAAGGATGAGGTGGTA 765
NM_024243       AACTCCACCAGCTTCCTTGCTTGGCTCTACAACGATAGCCCTGTCAAGGATGAGGTGATA 738
                ** **  *    ** **  * ***** ** ** ** ** **   ******    *** *

Abalone         ACCAACGACCGATGGGGAAAAGATGCCACATGTCATCATGGAGGTTTCCTCACCTGCATG 829
NM_012562       GTGAATGACCGGTGGGGTCAGAACTGCTCCTGTCGTCACGGAGGGTACTACAACTGTGAA 828
NM_000147       GTAAATGACCGATGGGGTCAGAACTCTTCCTGTCACCATGGAGGATACTATAACTGTGAA 825
NM_024243       GTGAATGACCGGTGGGGCCAGAACTGCTCCTGTCATCATGGAGGGTACTACAACTGTCAA 798
                   ** ***** *****  *  *     * ****  ** ***** * *   * ***

Abalone         GACAGATACAACCCAGGAACACTACAGCCCCGTAAGTTTGAGAATGCCATGACAATAGAC 889
NM_012562       GACAAATACAGACCACACAGCCTGCCAGACCACAAGTGGGAGATGTGCACCAGCGTTGAC 888
NM_000147       GATAAATTCAAGCCACAGAGCTTGCCAGATCACAAGTGGGAGATGTGCACCAGCATTGAC 885
NM_024243       GACAAATACAAGCCACAGAGCTTGCCAGACCACAAGTGGGAGATGTGCACCAGCATGGAC 858
                ** * ** **  ***   *   * *     *  ****  ****    **  *   * ***

Abalone         AAGAAGTCATGGGGTTTCCGGCGTAATGCTGATCTGGCAGCCTACCTCAACATGGAAGAA 949
NM_012562       AAGGCGTCCTGGGGCTATCGGAGAGACATGAGCATGTCTACCATCGTCGACGAAAATGAA 948
NM_000147       AAGTTTTCCTGGGGCTATCGTCGTGACATGGCATTGTCTGATGTTACAGAAGAATCTGAA 945
NM_024243       AGAGCATCCTGGGGCTATCGAAAAGACATGACCATGTCGACCATCGCCAAGGAAAATGAA 918
                *     ** ***** *  **     *        ** *          *       ***

Abalone         ATCTTAAGTACATTTGCCGAGACCATCAGTTGTGGAGGCAACATGCTCATCAATGTAGGG 1009
NM_012562       ATCATTGAGGAATTGGTTCAGACAATAAGTCTGGGAGGCAACTATCTTCTCAACATCGGA 1008
NM_000147       ATCATTTCGGAACTGGTTCAGACAGTAAGTTTGGGAGGCAACTATCTTCTGAACATTGGA 1005
NM_024243       ATCATCGAGGAATTGGTTCAGACGGTAAGTTTGGGAGGCAACTATCTTCTCAACATTGGA 978
                *** *      * * *   ****  * ***   *********   **  * **  * **

Abalone         CCAACCAAATACGGTATGATCAGTGCTCTCTATGAGGAAAGACTGCGCCAACTAGGCGGA 1069
NM_012562       CCGAATAAAGACGGCGTGATCGTCCCCATCTTCCAGGAAAGGCTCCTTGCTGTTGGCAAG 1068
NM_000147       CCAACTAAAGATGGACTGATTGTTCCCATCTTCCAAGAAAGGCTTCTTGCTGTTGGGAAA 1065
NM_024243       CCAACTAAAGATGGTCTGATCGTCCCCATCTTCCAAGAAAGGCTTCTTGCTGTCGGCAAG 1038
                ** *  *** * **  ****     *  ***   * ***** ** *      * **

Abalone         TGGTTGAGTGTCAATGGTGAGGGCATCTATTCCACAAAACCATGGACCTCCCAGAATGAC 1129
NM_012562       TGGCTGCAGATCAACGGGGAGGCCATCTATGCCTCCAAACCATGGAGGGTGCAGTCTGAA 1128
NM_000147       TGGCTGAGCATCAATGGGGAGGCTATCTATGCCTCCAAACCATGGCGGGTGCAATGGGAA 1125
NM_024243       TGGCTGCAGATCAACGGGGAGGCCATCTATGCCTCCAAACCCTGGAGGGTGCAGTCGGAA 1098
                *** **    **** ** ****  ****** ** * ***** ***      **    **

Abalone         ACAGTGACCAAAGGAATCTGGTATGTACAGAAGCAGAGCAATGTCTATGCCATTGTACTG 1189
NM_012562       AGGAACAAGACAGTTGTGTGGTACACCACTAAAGACTCGGCTGTTTATGCCACTTTTCTG 1188
NM_000147       AAGAACACAACATCTGTATGGTATACCTCAAAGGGATCGGCTGTTTATGCCATTTTTCTG 1185
NM_024243       AAGAACAAGACGGTTGTGTGGTACACTACTAAAAACACAACTGTTTACGCCACTTTCCTG 1158
                *     *  *      * *****       **         *** ** **** * * ***

Abalone         AACTGGCCGGA---TGCAGAGCTTCAGCTCGGGGCACCCAAGACAACTGCAGCTACTACA 1246
NM_012562       CACTGGCCAGAAGATGGAGTGGTAAACCTCCAATCTCCCAAAATGACATCGGCCACAAAG 1248
NM_000147       CACTGGCCAGAAAATGGAGTCTTAAACCTTGAATCCCCCATAACTACCTCAACTACAAAG 1245
NM_024243       TACTGGCCAGAAAATGGGATCGTCAACCTCAAATCCCCCAAAACGACCTCGGCCACAAAG 1218
                 ******* **   **      *  * **     * **** *  **  *  * ** *

Abalone         GTCAGTTTGCTCGGATACGACGGTAACTTCAGCT---TTCAAGAACGAGCCGGAGGGGGC 1303
NM_012562       ATAACGATGCTAGGAATGGAAGGAGAACTGCACTGGACCCAGGACCCACTGGAAGGCGTC 1308
NM_000147       ATAACAATGCTGGGAATTCAAGGAGATCTGAAGTGGTCCACAGATCCAGATAAAGGTCTC 1305
NM_024243       ATAACAATGCTAGGACTAGAAGGAGACCTGAGCTGGACCCAGGATCCACTGGAGGGCGTC 1278
                 * *   **** ***    * **  *  *    *        ** * *      **   *

Abalone         ATAACCATAGAGATTCCTCCAATCCCCATCAACAAAATGCCATGT-GAATGGGCATGGAT 1362
NM_012562       CTCATCACT----CTGCCCCAGTTGCCACCAGGCACTTTCCCAGTGGAGTCTGCCTGGAC 1364
NM_000147       TTCATCTCT----CTACCCCAGTTGCCACCCTCTGCTGTCCCCGCAGAGTTTGCTTGGAC 1361
NM_024243       CTCATCTCT----CTGCCACAGTTGCCACCTACCGTTCTGCCGGTGGAGTTTGCGTGGAC 1334
                 * * *        * *  ** *  *** *          *  *  ** *  ** ****

Abalone         CTTCAAACTTACTGCACTTCAATAGACTAGATACTAGTATGCCAACTGGTTCTTGCAATA 1422
NM_012562       TCTAAAGCTGACAAAAGTGAACTGA---------------------------------- 1389
```

한국해양대학교 KOREA MARITIME UNIVERSITY LIBRARY

```
NM_000147       TATAAAGCTGACAGGAGTGAAGTAA---------------------------------- 1386
NM_024243       TCTGAAGCTGACAAAGGTGAACTGA---------------------------------- 1359
                  * ** ** **     *  * *

Abalone         TTTGTGTATAGATTTCATGTACAGATGTGAAGTTACTGAATGTGAACATATAATACTCAA 1482
NM_012562       ------------------------------------------------------------
NM_000147       ------------------------------------------------------------
NM_024243       ------------------------------------------------------------

Abalone         AATGTATATACAGAGCTACTGAAATAAAATATTTTGAAAATAAAA 1527
NM_012562       ---------------------------------------------
NM_000147       ---------------------------------------------
NM_024243       ---------------------------------------------
```

Fig. 2-10. CLUSTAL W (1.82) multiple sequence alignment of abalone fucosidase full length sequence with coding sequences of other L-fucosidases. Possible nonsense mutations identified in the full length sequence of abalone fucosidase which align with the coding sequence of other fucosidases are underlined. The codon which may have functioned as the stop codon in an ancient active form is highlighted. Identical residues are shown in asterisks. Conserved substitutions depending on functionality are indicated with a colon and the semi-conserved residues with a dot.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

4. DISCUSSION

4. 1. Glutathione transferase

GSTs are a family of enzymes responsible for the metabolism of a broad range of xenobiotics and carcinogens (Mannervik, 1985). These enzymes are also known as Glutathione S-alkyltransferases. glutathione S-aryltransferases, S-(hydroxyalkyl) glutathione lyases and glutathione S-aralkyltransferases. They catalyze the reaction

RX + glutathione <=> HX + R-S-glutathione

where R can be an aliphatic, aromatic or heterocyclic group and X can be a sulfate, nitrite or halide group. They also catalyze the addition of aliphatic epoxides and arene oxides to glutathione; the reduction of polyol nitrate by glutathione to polyol and nitrite; certain isomerization reactions and disulfide interchange.

Most of the GSTs exist as soluble enzymes but a small family of microsomal GSTs has been reported (Anderson *et al.*, 1994; Jakobsson *et al.*, 1996). Cytosolic and membrane-bound forms of GST are encoded by two distinct supergene families. At present, eight distinct classes of the soluble cytoplasmic mammalian GSTs have been identified: alpha, kappa, mu, omega, pi, sigma, theta and zeta. The abalone GST isolated during this study was in the soluble fraction of the cell lysate indicating it to be a soluble form. The soluble forms generally exist as dimeric proteins, with subunit molecular weights of approximately 25 kDa (Eato and Bammler, 1999). The abalone GST isolated coded for a 215 amino acid protein with the theoretically expected molecular weight of 24.6 kDA (Fig. 2-1) as revealed by SDS-PAGE and a theoretical isoelectric point of 8.39 with high enzymatic activity towards 1-chloro-2,4-dinitrobenzene, the universal substrate of all GSTs except theta class. Phylogenetic tree analysis group abalone GST with other mu class GSTs. Identity of subunits is based mainly on primary DNA sequence homology. In general, members of the same class share more than 40-50% sequence identity but less than about 25-30% sequence identity with GSTs in other classes except in theta class which share only 30% homology across species (Hayes and Pulford, 1995; Coggan *et al.*, 1998). Abalone GST shares 54-59% identity with other known Mu class GST coding sequences and 48-50% with other Mu class proteins

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

at amino acid level (Table 2-4 and Table 2-5) and share less than 25% with other class coding sequences except Pi class. One feature that is readily apparent from the above tables is that the aforementioned theory is valid only for the extensively studied classes such as alpha, pi and mu. Lack of sequence information regarding other classes makes it difficult to find relationships.

Alignment of the putative amino acid sequence with known GST structures in pdb data bank via NCBI BLASTP server and Cn3D structure viewer (Marchler-Bauer *et al.,* 2003) revealed it to contain the GST N, the glutathione binding domain or G-site and C terminal domain, the variable electrophilic substrate binding site or H-site as expected. The N terminal had the highest sequence similarity to chicken GST, cGSTM1-1, with a topology similar to thioredoxin fold (Fig. 2-2 and 2-3). This fold appears in several proteins of limited sequence identity from other enzyme families, which appear to have evolved to bind cysteine or GSH. This fold is characterized by an N terminal βαβ motif linked by an α-helix to a C terminal ββα motif. Interestingly, the C domain had the highest similarity to the second domain of chain A, *Saccharomyces cerevisiae* Ure2p protein, a prion protein involved in nitrogen catabolism. Despite results indicating that Ure2p lacks GST activity, it is proposed that Ure2p is a member of the GST superfamily that may describe a novel GST class (Bousset *et al.,* 2001). The region in abalone GST corresponding to the short α6-helix is replaced by a long α-helix composed of 28 amino acids in Ure2p. The P개 60 found in the loop connecting α-2 to β-3 (cis-Pro loop) is thought to be in the less favoured *cis* conformation and is highly conserved in all GSTs. It is not directly involved in catalysis but has shown to be important in maintaining the protein in a catalytically competent structure (Allocati *et al*., 1999). Tyr 6 on the other-hand is thought to be directly involved in catalysis by stabilizing the thiolate group in glutathione (Atkins *et al*., 1993). A SNAIL/TRAIL motif observed in many GSTs (Sheehan *et al*., 2001) is seen in α-3 helix of abalone. This is involved in the formation of GSH binding site. The C domain contains five helices as seen in Pi and Mu classes and as discussed earlier the primary DNA sequence homology assigns it to Mu class. This is further supported by the presence of a region corresponding to the Mu-loop between β-2

and □-2. Considering all these factors abalone GST could be assigned to class mu.

The Alpha, Mu, Pi and Theta class GST genes differ markedly from each other in size and in intron/exon structure, and there is a trend for human GST genes to be found in class specific clusters (Webb *et al*., 1996). It is also demonstrated that polyclonal antisera raised against a particular GST class will often cross react with same class GSTs but not with GSTs from different classes found even within the same species (Hayes and Mantle, 1986). GST subunits can dimerize only with subunits within the same class. Kinetic properties such as substrate and inhibitor sensitivities also could be used to find clues about the class of a given GST even though this is a poor criteria to extend the mammalian classes into non-mammalian sources. Therefore further characterization of abalone GST at immunological, biochemical and genomic level is required to verify its assignment to Mu class. The trans isomer of stilbene oxide (TSO) is uniquely conjugated by hGSTM1-1, and thus this epoxide substrate serves as a selective marker for this polymorphic mu class GST (Seidegard *et al*., 1989). Biochemical tests with this substrate would lead to the confirmation of the identity of the putative GST we isolated.

The mu class of enzymes functions in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. The genes encoding the mu class of enzymes are organized in a gene cluster on chromosome 1p13.3 in human and are known to be highly polymorphic. These genetic variations can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of certain drugs. Null mutations of this class mu gene have been linked with an increase in a number of cancers, likely due to an increased susceptibility to environmental toxins and carcinogens. Multiple protein isoforms are encoded by transcript variants of this gene in human (http://au.expasy.org/sprot/)

Differences in expression of specific isoforms of GST is thought to be an important determinant of target organ and species sensitivity. For example, mice are remarkably resistant to the hepatocarcinogenic effects of aflatoxin B1 (AFB) because they constitutively express

mGSTA3-3, which has high activity toward the reactive AFB-*exo*-epoxide (Eaton and Gallagher, 1994).

In mammals, hepatic GSTs often accompany cytochrome P450 enzymes where they quench highly reactive electrophilic species formed during the oxidative metabolism of aromatics, drugs, and other xenobiotics. These phase I (oxidating) and phase II (conjugating) enzymes together appear to be necessary for the maintainance of normal metabolic and liver fuction. (Dhar *et al*, 2003). Digestive gland of abalone represents the liver in mammals and several ESTs in abalone digestive gland cDNA library had a significant similarity to cytochrome P450 enzymes. Therefore further studies on abalone GST is expected to be useful in elucidating its biological significance in cellular detoxification, which will lead to the improvement of fisheries by over expression of this gene resulting less accumulation of toxic wastes in the organisms.

4. 2 . Arylsulfatase B

we amplified the full-length cDNA of a sulfatase with highest sequence similarity to *Helix pomatia* sulfatase 1 from the abalone EST library. Expression of this in *E. coli* BL21 (DE3) cells under pET16b vector resulted in an expression band corresponding to a 54 kDa protein on an SDS-PAGE (Fig. 2-4). The protein was mainly found in inclusion bodies and was purified under denaturing conditions using urea. As expected, the amount of protein expressed decreased with decreasing temperature from 37°C to 15°C. Sequence analysis revealed the presence of seven Cys residues in a coding sequence of 1446 bp coding for a peptide with 481 amino acids with a putative peptide cleavage site at 21-22 amino acids (Fig. 2-5). Surprisingly, the sequence obtained revealed a very short 3′ untranslated region comprising only 41bp followed by the poly(A) tail.

Even though it was expressed at a very high level in *E. coli,* the purified protein didn't shaw any activity towards *p*-nitrocatechol sulfate, the chromogenic substrate which is used to characterize arylsulfatases. This loss of activity could be due to many factors including erroneous folding of the protein during purification under denaturing conditions

and lack of posttranslational modifications in the *E. coli* system. Most sulfatases are found in the lysosomes where they perform, in an acidic medium, the desulfation reactions required in the degradative pathways of glycosaminoglycans. Some newly identified members of this family are found in the endoplasmic reticulum and golgi apperatus. The physiological roles and the natural substrates of these recently discovered members have not been defined. It has been shown however that some of them are active on fluorogenic substrate 4-methylumbelliferryl. Therefore different substrate specificity also could be a reason for its inability to catalyse the above reaction. Conversion of a crucial cysteine residue to a serine semialdehide as a co- or post-translational modification is required for catalytic activity and lack of this results in multiple sulfatase deficiency (Parenti *et al.,* 1997). Therefore expression of this ASB gene in an expression system where post-translational modifications could take place might result in the formation of active form. Sequence comparisons with other known arylsulfatases revealed it to lack a few amino acids from the N-terminal which includes a leucin conserved in several mammalian and *Helix pomatia* arylsulfatases and this also could have contributed to its inactivity (Fig. 2-6 and 2-7). From the multiple alignment of several aryulsulfatases and abalone ASB, it is apparent that there is a trend to have several deletions/ insertions at N and C termini while preserving the middle region of the sulfatases. Comparison of abalone putative ASB with known sulfatase structures in pdb database unveiled the presence of sulfatase conserved domain.

4. 3 L-fucosidase

We isolated the full length cDNA of a sequence with highest similarity to other α-L-fucosidases containing an open reading frame of 765 bp coding for a 255 amino acid protein with a putative signal cleavage site between 19-20 (CVA-TP) (Fig. 2-8). Eventhough we were able to clone it into pET16b expression vector in the correct orientation as unraveled by restriction enzyme digestion, we were unable to express it in *E. coli*. Sequence analysis revealed it to contain

part of the α-L-fucosidase conserved domain. Comparison of its coding region and full length sequence with other known fucosidases revealed it to contain several nonsense mutations thus prematurely ending the translation process which will result in the production of a truncated protein lacking ~200 amino acids from its C-terminus (Fig. 2-9) which is validated by the presence of a stop codon at the expected position when compared its full length sequence with coding regions of other known α-L-fucosidases (Fig. 2-10). Phisico-chemical analysis via ProtParam program (http://au.expasy.org/tools/protparam.html) predicts it to code for a stable protein with a half life of > 10 hrs in *E. coli* and therefore its unlikely that any unstability caused by the lack of C terminus to be the reason for its undetection. As expected, the sequence identity it shares with other mammalian α-L-fucosidase coding regions and protein sequences (~55%) is much less than the sequence identity found within mammmalian α-L-fucosidases (~85%).

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

SUMMARY

The expressed sequence tag approach, first demonstrated in the human genome project, is powerful in massive cloning of cDNAs as well as in large scale characterization of cDNA sequences for deciphering genome sequence. In the present study, a cDNA library from abalone digestive gland was constructed and a total of 841 expressed sequence tags were generated after sequencing 900 clones. Assembly of these into clusters resembling sequences derived from the same gene resulted in 122 clusters including 331 sequences and 510 singletons. This correspond to a redundancy of 39%, which was less than the redundancy we encountered in the whole abalone cDNA library (45%) we generated. Comparison of these 632 unique transcripts with known database sequences via BLAST-X and BLAST-N programs revealed 354 sequences to share significant similarity with known sequences. Thus, 278 transcripts are cloned and reported for the first time in this study. These may be useful as molecular markers as well as in the identification of novel genes.

As expected, most of the unique sequences in the digestive gland cDNA library coded for enzymes. Three putative enzymes implicated in disease conditions and cellular detoxification including a glutathione transferase (GST), arylsulfatase-B (ASB) and a fucosidase were further characterized by cloning full length sequences in an *E. coli* expression vector, pET16b. GST was expressed in substantial amounts in the soluble form with high catalytic activity towards 1-chloro-2,4-dinitrobenzene and sequence analysis revealed it to be a class Mu protein with the N and C GST conserved domains containing thioredoxin fold and five α-helices. Unlike GST, ASB expressed in insoluble form and purification under denaturing conditions resulted in inactive form with an apparent molecular weight of 54 kDa containing a region similar to sulfatase domain. Computational analysis of fucosidase sequence unraveled the presence of amino terminal of α-L focosidase conserved domain and several nonsense mutations which prematurely ends translation.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

Further studies on these genes would facilitate the understanding of molecular nature of their activity and diseases involved. ASB and fucosidase will be useful as catalysts in industry and constitutive expression of GST in commercial fisheries would result in toxic resistant fish with less accumulation of xenobiotics.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

REFERENCES

Allocati, N., E. Casalone, M. Masulli, I. Ceccarelli, E. Carletti, M. W. Parker and C. Dillio. 1991. Functional analysis of the evolutionarily conserved proline 53 residue in *Proteus mirabilis* glutathione S-transferase B1-1. Febbs Lett. 445: 347-350.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215 : 403-410.

Anderson, C., E. Mosialou, R. Weinander and R. Morgenstern. 1994. Enzymology of microsomal glutathione S transferase. Adv. Pharmacol. 27: 19-35.

Atkins, W. M., R. W. Wang, A. W. Bird, D. J. Newton and A. Y. H. Lu. 1993. The chatalytic mechanism of glutathione S-transferase (GST). Spectroscopic determination of the pKa of Tyr-9 in rat α 1-1 GST. J. Biol. Chem. 268: 19188-19191.

Berg, J. M., J. L. Tymoczko and L. Stryer. 2002. Biochemistry. W. H. Freeman and Co., New York.

Bousset, L., H. Belrhali, R. Melki and S. Morera.2001. Crystal structures of the yeast prion Ure2p functional region in complex with glutathione and related compounds. Biochemistry 40: 13564-13573.

Carefoot, T. H., B. E. Taylor and D. A. Donovan. 1998. Seasonality in digestive-gland size and metabolism in relation to reproduction in *Haliotis kamtschatkana*. J. Shellfish Res. 17: 713-716.

Carnevali, O., R. Carletta, A. Cambi, A. Vita and N. Bromage. 1999. Yolk formation and degradation during oocyte maturation in seabream *Sparus aurata*: involvement of two lysosomal proteinases. Biol. Reproc. 60: 140-146.

Casas, E., M. Betancourt, E. Bonilla, Y. Duculomb, H. Zayas and R. Trejo. 1999. Changes in cyclin B localisation during pig oocyte in vitro maturation. Zygote 7: 21-26.

Cleveland, L. R. 1924. The physiological and symbiotic relationships between the intestinal protozoa of termites and their host, with special reference to *Reticulitermes flavipes* Kollar. Biolog. Bull. Mar. Biol. Lab. 46: 117-227.

제주대학교 중앙도서관 JEJU NATIONAL UNIVERSITY LIBRARY

Coggan, M., L. Whitbread, A. Whittington and P. Board. 1998. Structure and organization of the human theta class glutathione S transferase and D-dopachrome tautomerase gene complex. Biochem. J. Interact. 111: 377-388.

Dhar, K., A. Dhar and J. P. N. Rosazza. 2003. Glutathione S-transferase isoenzymes from Streptomyces griseus. Appl. Environ. Microbiol. 69: 707-710.

Eaton, D. L. and E. P. Gallagher. 1994. Mechanisms of aflatoxin carcinogenesis. Annu. Rev. Pharmacol. Toxicol. 34: 135-172.

Eaton, D. L. and T. K. Bammler. 1999. Concise review of the glutathione S transferases and their significance to toxicology. Toxicol. Sci. 49: 156-164.

Evans, J. P. 2000. Getting Sperm and Egg Together: Things Conserved and Things Diverged. Biol. Reproc. 63: 3558-3560.

Gomez-Roman, N., C. Grandori, R. N. Eisenman and R. J. White. 2003. Direct activation of RNA polymerase III transcription by c-Myc. Nature 421: 290-294.

Hayes, J. D. and T. J. Mantle. 1986. Use of immunoblot techniques to discriminate between the glutathione S-transferase Y_f, Y_k, Y_a, Y_n/Y_b and Y_c subunits and to study their distribution in extra-hepatic tissue. Evidence for three immunochemically distinct groups of transferases in the rat. Biochem. J. 233: 779-788.

Hayes, J. D. and D. J. Pulford. 1995. The glutathione S transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. Crit. Rev. Biochem. Mol. Biol. 30: 445-600.

Jakobsson, P. J., J. A. Mancini and A. W. Ford-Hutchinson. 1996. Identification and characterization of a novel human microsomal glutathione S-transferase with leukotriene C4 synthase activity and significant sequence identity to 5-lipoxygenase-activating protein and leukotriene C4 synthase. J. Biol. Chem. 271: 22203-22210.

Leznicki, A. J. and K. Bailkowski. 1997. Improved high-performance liquid chromatographic method for N-acetylgalactoseamine-4-sulfate sulfatase (arylsulfatase B) activity determination using uridine diphospho-N-acetylgalactosamine-4-sulfate. J. Chromatogra. B 696:

193-202.

Liang, L. F. and J. Dean. 1993. Conservation of mammalian secondary sperm receptor genes enables the promoter of the human gene to function in mouse oocytes. Dev. Biol. 156: 399-408.

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, and Bryant SH (2003), "CDD: a curated Entrez database of conserved domain alignments", Nucleic Acids Res. 31:383-387.

Mannervik, B., Y. C. Awasthi, P. G. Board, J. D. Hayes, C. Di Ilio, B. Ketterer, I. Listowsky, R. Morgenstern, M. Maramatsu, W.R. Pearson. 1992. Nomenclature for human glutathione transferases [letter]. Biochem. J. 282: 305-306.

Mannervik, B.: 1985. The isozymes of glutathione transferase. Adv. Enzym. Relat. Areas Molec. Biol. 57: 357-417,

Martin, M. M. and J. S. Martin. 1978. Cellulose digestion in the midgut of the fungus-growing termite *Macrotermes natalensis*: The role of acquired digestive enzymes. Science 199: 1453-1455.

Mohr, E., N. Prakash, K. Vieluf, C. Fuhrmann, F. Buck and D. Richter. 2001. Vasopressin mRNA localization in nerve cells: Characterization of cis-acting elements and trans-acting factors. Proc. Natl. Acad. Sci. USA. 98: 7072-9.

Mollet, J., A. Rahaoui and Y. Lemoine. 1998. Yield, chemical composition and gel strength of agarocolloids of *Gracilaria gracilis longissima* and the newly reported *Gracilaria* cf. *vermiculophylla* from Roscoff (Brittany, France). J. Appl. Phycol. 10: 59-66.

Monje, H. and M. T. Viana. 1998. The effect of cellulose on the growth and cellulolytic activity of abalone *Haliotis fulgens* when used as an ingredient in formulated artificial diets. J. shellfish Res. 17: 667-671.

Morse, D. E., H. Duncan, N. Hooker and A. Morse. 1977. Hydrogen peroxide induces spawning in mollusks, with activation of prostaglandin endoperoxide synthetase. Science 196: 298-300.

NCBI staff. 2002. The NCBI handbook. National Library of Medicine, Bethesda

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

O'Brien, G. W., P. C Veivers, S. E. McEwen, M. Slaytor and R. W. O'Brien. 1979. The origin and distribution of cellulase in the termites, *Nasutitermes exitiosus* and *Coptotermes lacteus*. Insect Biochem. 9: 619-625.

Parenti, G. G. Meroni and A. Ballabio. 1997. The sulfatase gene family. Curr. Opin. Genet. Med. 7: 386-391.

Powers, D. A., V. Kirby and M. Gomez-Ghiarri. 1996. Genetic engineering abalone: Gene transfer and ploidy manipulation. J. .Shellfish Res. 15 : 477

Schaeffer, D. J. and V. S. Krylov. 2000. Anti-HIV activity of extracts and compounds from algae and cyanobacteria. Ecotoxicol. Environ. Saf. 45: 208-227.

Seidegard, J., R. W. Pero and B. Stille. 1989. Identification of the trans-stilbene oxide-active glutathione transferase in human mononuclear leukocytes and in liver as GST1. Biochem. Genet. 27: 253-261.

Sheehan, D., G. Meade, V. M. Foley and C. A. Dowd. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem. J. 360: 1-16.

Siegel, G. J., B. W. Agranoff, R. W. Albers, S. K. Fisher, M. D. Uhler, editors. 1999. Basic neurochemistry, molecular, cellular and medical aspects. Lippincott, Williams and Wilkins, Philadelphia.

Sikela, J. M. and C. Auffray. 1993. Finding new genes faster than ever. Nat. Genet. 3 : 189-191.

Sutton, G. G., O. White, M. D. Adams and A.R. Kerlavage. 1995. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. Genome Sci. Technol. 1: 9-19.

Suzuki, K. I., T. Ojima and K. Nishita. 2003. Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai.* Eur. J. .Biochem. 270 : 771-778.

Swanson, W. J., V. D. Vacquier. 1995. Liposome fusion induced by a M(r) 18,000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoa. Biochemistry. 34: 14202-14208.

Swanson, W. J. and V. D. Vacquier. 1997. The abalone egg vitelline

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

envelope receptor for sperm lysin is a giant multivalent molecule. Proc. Natl. Acad. Sci. USA 94: 6724-6729.

Tien, N. N. N. and B. Knoops. 2003. Mitochondrial and cytosolic expression of human peroxiredoxin 5 in Saccharomyces cerevisiae protect yeast cells from oxidative stress induced by paraquat. FEBS Lett. 544: 148-52.

Uchida, S., Y. Tanaka, H. Ito, F. Saitoh-Ohara, J. Inazawa, K. K. Yokoyama, S. Sasaki and F. Marumo. 2000. Transcriptional Regulation of the CLC-K1 Promoter by myc-Associated Zinc Finger Protein and Kidney-Enriched Krüppel-Like Factor, a Novel Zinc Finger Repressor. Mol. Cell. Biol. 20: 7319-7331.

Vacquier, V. D., W. J. Swanson, Y. H. Lee. 1997. Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J. Mol. Evol. 44 Suppl 1: S15-22.

Villee, C. A., W. F. Walker and F. E. Smith. 1968. General Zoology. 3rd Edition, W.B. Saunders Company, London.

Webb, G., V. Vaska, M. Coggan and P. Board. 1996. Chromosomal localization of the gene for the human theta class glutathione transferase (GSTT1). Genomics. 33: 121-123.

Wistow, G., S. L. Bernstein, M. K. Wyatt, A. Behal, J. W. Touchman, G. Bouffard, D. Smith, K. Peterson. 2002. Expressed sequence tag analysis of adult human lens for the NEIBank Project: over 2000 non-redundant transcripts, novel genes and splice variants. Mol. Vis. 8: 171-84.

Wu, M. L., Y. C. Chuang, J. P. Chen, C. S. Chen and M. C. Chang. 2001. Identification and Characterization of the Three Chitin-Binding Domains within the Multidomain Chitinase Chi92 from *Aeromonas hydrophila* JP101. Appl. Environ. Microbiol. 67: 51005106.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

ACKNOWLEDGEMENTS

I am grateful to my parents and Prof. Jehee Lee for their guidance and encouragement given throughout the project. I thank Prof. Choon Bok Song for helpful discussions. The support given by Prof. You-Jin Jeon, Prof. In-Kyu Yeo, Prof. Keun-Tae Park, Prof. Moon-Soo Heo and their laboratory staffs are also acknowledged. I particularly appreciate Miss. Young Mi Park for her help in writing the manuscript and Mr. Chang Nam Gin, Kyoung Im Kang, Ho Jin Park, Hyoung Gun Kim, Bong Ryong Cho, Sang Bong Kim, Jin Young Choi, Jung-A Kim, Yeun-young Ko, Ji-eun Kim, Soon Mi Ko, Young Ju Moon, Prashani, Nalin, Yasantha, Mahinda, Herath, Rohan for their helpful suggestions and encouragement. My special thanks goes to Mrs. Whang Ilson, Oh Cheal Hong and Kang Hyun Sil for the help given during difficulties. The financial assistance given by The Korea Science and Engineering Foundation and the support given by the staff of the Graduate School of Cheju National University is gratefully acknowledged. A debt of gratitude is also owed to my teachers, specially the academic staff of University of Colombo, Sri Lanka from whom I learnt the basics of life sciences.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY