Chapter 4

RESULTS

Result of the molecular identification, characterization and expression analysis of genes encoding antioxidant genes, pattern recognition receptor protein genes and the genes involved in the pearl formation in *Pinctada fucata* was carried out as part of the present study are presented below.

4.1 Molecular Identification and Characterization of Functional Genes

4.1.1 Antioxidant Genes

Molecular identification and characterization of three antioxidant enzyme genes viz. SOD, GPX and GST were carried out. The initial partial characterization was followed by RACE to develop the full length nucleotide sequence. The amino acid sequence of the proteins being encoded by them were deduced, and their characteristic features were worked out as detailed below.

4.1.1.1 Superoxide Dismutase (SOD)

Polymerase Chain Reaction (PCR) carried out to amplify the SOD gene using the primers designed from the conserved region gave positive amplicon. A single PCR product of 471 bp was visualized by agarose gel electrophoresis. This product was purified and sequenced. This partial cDNA sequence provided the necessary information to obtain an additional 368 bp sequence by 3’RACE, and 87 bp sequence by 5’RACE. Finally, the full-length sequence information of the SOD cDNA was obtained by overlapping the three cDNA sequences. The nucleotide sequence and the deduced amino acid sequence are shown in Fig.4.1. The full-length SOD cDNA is comprised of 924 bp,
containing 87 bp in the 5'-terminal untranslated region (UTR), 471 bp in the ORF, 366 bp in 3'-terminal UTR with a poly(A) tail of 30 bp and a putative polyadenylation consensus signal (AATAAA). Deduction of amino acid sequence from the complete coding region using online software Transeq of European Molecular Biology Laboratory (EMBL) revealed that 156 amino acids are encoded by the ORF. Analysis of the amino acid sequence using InterProscan online software to find out specific domains revealed that it is a Cu/Zn SOD with copper/zinc binding domains within the SOD sequence. The SOD cDNA sequence and its deduced amino acid sequence were submitted to the NCBI GenBank under accession no (JX013537). No signal peptide was identified in the deduced amino acid sequence of Cu/Zn SOD by the signal P program, indicated that this SOD was cytoplasmic Cu/Zn SOD.

In order to determine the variation and similarity, BLAST searches were conducted in NCBI. The BLAST search showed that the deduced amino acid sequence of SOD has extremely high identity with the Cu/Zn SOD of Crassostrea hongkongensis, C. gigas and Mytilus chilensis (99%). Similarly, it was found to have high identity with Cu/Zn SOD of Crassostrea ariakensis (98%) and Haliotis discus discus (97%).

Multiple alignment of the deduced amino acid sequences (Fig.4.2) with other closely related cytoplasmic Cu/Zn SOD sequences showed the presence of three cysteines (Cys 9, Cys 60 and Cys 149) in the mature SOD. The Cys 60 and Cys 149 were conserved in all Cu/Zn SODs and it was believed that they form an intramolecular disulfide bond. The other amino acids which were required for binding of copper (His-49, -51, -66, and -123) and zinc (His-66, -74 and -83 and Asp-86) were also conserved. Two Cu/Zn SOD family signature sequences were found in the deduced amino acid sequence of SOD; signature 1 (consensus sequences: [GA]-[IMFAT]-H-[LIVF]-H-{S}]-
x-[GP]-[SDG]-x-[STAGDE], and signature 2 (consensus sequences: G-[GNHD]-[SGA]-[GR]-x-R-x-[SGAWRV]-C-x(2)-[IV]).

Phylogenetic relationships of Cu/Zn SOD from pearl oyster and other invertebrates and vertebrates were estimated. Cu/Zn SOD of Candida albicans was used as the out-group. As shown in Fig.4.3, Cu/Zn SOD of P. fucata formed a single cluster with the Cu/Zn SODs from oyster C. gigas and mussel M. edulis indicated closer evolutionary relationship than that with other aquatic invertebrates. Vertebrates were evolutionarily distinctly separated.

Fig.4.1 Nucleotide sequence of SOD cDNA from P. fucata and its deduced amino acid sequence. Two Cu/Zn SOD family signatures are underlined (_). The start codon is in bold and termination codon is indicated with asterisk (*). The amino acids required for binding of copper (His-49 -51, -66, and -123) and zinc (His-66, -74, and -83 and Asp-86) are shaded. Polyadenylation consensus signal sequences are shaded black with white lettering. Two cysteines (Cys60 and Cys149) predicted to be engaged in the disulfide bond formation were boxed.
Fig. 4.2 Alignment of the deduced amino acid sequence of SOD of different species. The characteristic conserved regions are shaded.
Molecular Genetic Profiling of Functional Genes of Pearl Oyster, Pinctada fucata

Fig. 4.3 Phylogentic tree using SOD amino acid sequences from pearl oyster \textit{P. fucata} and other species. Note: Numbers represent the bootstrap values. GenBank accession numbers are shown next to each species.

4.1.1.2 Glutathione Peroxidase (GPX)

PCR amplification of the GPX gene segment using the cDNA and primer designed from the conserved regions successfully produced amplicon (<450 bp) as visualized through agarose gel electrophoresis. Optimization of PCR parameters helped to get specific amplicon free from nonspecific PCR product. The PCR product was sequenced for identification and characterization of the gene. Sequencing of the purified amplicon in both directions gave the sequence information of 436 bp length. The sequences were aligned by using Bio Edit software. The nucleotide sequence obtained were subjected to similarity search in NCBI using BLAST and confirmed to be GPX gene fragment.

RACE-PCR using the above 436 bp product revealed the sequence of a 1201 bp long segment of the GPX gene which was used for translation to amino acids in Transq (EMBL). It consisted of the ORF region and part of UTR (Fig.4.4). The ORF encodes a polypeptide of 198 amino acids. No signal peptide was present in the GPX sequence. The active site of the enzyme contains a selenocysteine encoded by a TGA, which is the main characteristic feature of the selanoprotein family. Se-GPX
contains a characteristic GPX signature sequence motif 2, consisting of 8 amino acids at $^{75}$LGFCNQF$^{82}$. Also, it showed an extra active site motif located at $^{163}$WNFEKF$^{168}$ and glutamine-Gln (Q$_{81}$,Q$_{85}$), tryptophan-Trp (W$_{153}$,W$_{163}$) residues, which was responsible for fixation of selenium. Additionally, Se-GPX showed two arginine residues at R$_{101}$ and R$_{199}$ for directing donor glutathione substrate towards the catalytic center. A potential N-glycosylation site $^{87}$NCTN$_{90}$ was identified using the PROSITE program, which has been observed in other GPX amino acid sequences.

ClustalW pairwise amino acid sequence alignment was performed to determine the identity percentage (%) of Se-GPX with other species Se-GPXs. BLAST analysis showed that the amino acid sequence of Se-GPX deduced in the present study has extremely high identity with that of *Pinctada fucata* reported from China. Multiple alignment of Se-GPX sequence of *Pinctada fucata* with that of other species are shown in Fig.4.5. The results showed that characteristic selenocysteine residue, $^{75}$LGFCNQF$^{82}$ signature motif2, $^{163}$WNFEKF$^{168}$ active site, functional residues of Glutamine-Gln (Q$_{81}$,Q$_{85}$), tryptophan-Trp (W$_{153}$,W$_{163}$) residues were aligned across all the selected species. The Se-GPx sequence resolved in this study revealed the main characteristic motifs and functional amino acid residues of the Se-GPx protein family. As there is a general lack of sequence information on these genes from other bivalves, further phylogenetic analysis were not possible.
Fig. 4.4 Nucleotide sequence of GPX cDNA from *P. fucata* and its deduced amino acid sequence. GPX signature sequence motif 2, consisting of 8 amino acids at \textsuperscript{75}LGFPCNQF\textsuperscript{82} and active site motif located at \textsuperscript{163}WNFEKF\textsuperscript{168} are shaded gray with black lettering. Glutamine-Gln (Q\textsubscript{81}, Q\textsubscript{85}), tryptophan-Trp (W\textsubscript{153}, W\textsubscript{163}) residues, which are responsible for fixation of selenium are shaded black with white lettering. Arginine residues, R\textsubscript{101} and R\textsubscript{199} for directing donor glutathione substrate towards the catalytic center. Selenocysteine encoded by a TGA were boxed. The start codon is in bold and termination codon is indicated with asterisk (*).
Molecular Genetic Profiling of Functional Genes of Pearl Oyster, Pinctada fucata

Fig. 4.5 Alignment of the deduced amino acid sequence of GPX of different species. The characteristic conserved regions are shaded.

4.1.1.3 Glutathione-s-transferase (GST)

In case of GST, the initial PCR of the cDNA using GST specific primers resulted in the amplification of a partial segment of 596 bp which was sequenced. RACE-PCR carried out revealed the complete ORF region and the UTR comprising of 1652 bp (Fig. 4.6). Deduced amino acid sequence encoded by it on analysis with signalP program revealed no signal peptides. NCBI BLAST analysis revealed that the gene sequence identified in this study belong to omega GST. Conserved domain search at NCBI revealed that the omega GST has spatially distinct domain: a small thioredoxin-like N-terminal domain and a larger helices rich C-terminal domain. The N and C-terminal domains in the position between 18-96 and 101-237 were identified by Prosite software. The GST active site was composed of a specific glutathione binding site (G-site) common to all GSTs, and a nonspecific substrate binding site (H-site) which varied between different classes and isotypes. Residues from the N-terminal thioredoxin domain form the G-site while the H-site was
comprised mainly of residues from the C-terminal alpha helical domain. The residues of G-site (Cys-28, Pro-29, Phen-30, Ala-31, Leu-52, Lys-55, Leu-67, Val-68, Glu-80 and Ser-81) were conserved. The composition of cysteine-containing tetramer (C-P-F-A) is marked on figure 4.6. The H-site (Gly-116, Thr-119, Gly-120, Tyr-123, Glu-124, Trp-174, Arg-177 and His-221) were also identified.

BLAST analysis showed that the deduced amino acid sequence of omega GST has extremely high identity with the omega GST of *Pinctada fucata* reported from China. Multiple alignment of omega GST sequence with other species (Fig.4.7) showed that the characteristic cysteine-containing tetramer (C-P-F-A) and G-site (Cys-28, Pro-29, Phen-30, Ala-31, Leu-52, Lys-55, Leu-67, Val-68, Glu-80 and Ser-81) were aligned in all the species used for alignment.

Phylogenetic analysis using omega GST from pearl oyster, other invertebrates and vertebrates was carried out with cyanobacteria as the out-group (fig.4.7). The GSTs of *P. fucata* and abalone *H. discus discus* only formed a single cluster while other species did not form part of it.
Fig. 4.6 Nucleotide sequence of GST cDNA from *P. fucata* and its deduced amino acid sequence. Cysteine-containing tetramer (C-P-F-A) is underlined. The amino acids required for G-site (Cys-28, Pro-29, Phen-30, Ala-31, Leu-52, Lys-55, Leu-67, Val-68, Glu-80 and Ser-81) are shaded gray with white lettering. The residues for H-site (Gly-116, Thr-119, Gly-120, Tyr-123, Glu-124, Trp-174, Arg-177 and His-221) are shaded black with white lettering. The start codon is in bold and termination codon is indicated with an asterisk (*).
4.1.2 Pattern Recognition Receptor Protein Genes

The three pattern recognition receptor protein genes namely F-type lectin, galectin and LGBP were taken up for molecular identification and characterization. Full length characterization with sequence information of complete coding region, deduced amino acid sequence, identification of binding sites of putative proteins were worked out for F-type lectin. In case of galectin and LGBP genes, characterization of partial segment was only achieved. Detailed account of the results is presented below:
4.1.2.1 F-type lectin

Polymerase Chain Reaction (PCR) carried out to amplify the F-type lectin gene using the primers designed from the conserved region gave positive amplicon. A single PCR product of the expected size (<600 bp) was obtained. Purification followed by sequencing of this product revealed the nucleotide sequence of 588 bp length. This partial sequence provided the necessary information to obtain an additional 257 bp sequence by 3’RACE, and 46 bp sequence by 5’RACE. Finally, the full-length sequence information of the F-type lectin cDNA was obtained by overlapping these three sequences (588+257+46). The nucleotide sequence is shown in Fig.4.9. The full-length F-type lectin cDNA is of 891 bp, consisting of a 46 bp 5’-terminal untranslated region (UTR), 588 bp open reading frame (ORF), 257 bp 3’-terminal UTR with a poly(A) tail of 21 bp, and a putative polyadenylation consensus signal (AATAAA). The ORF encodes a polypeptide of 196 amino acids (Fig.4.9). A signal peptide was identified at the N-terminus of the deduced polypeptide by signal P program and its cleavage site was located between the positions of Gly\textsuperscript{19} and Tyr\textsuperscript{20}. The cDNA sequence of F-type lectin and its deduced amino acid sequence were submitted to the NCBI GenBank under accession no. JX103557.

In order to determine the variation and similarity, BLAST searches were conducted in NCBI. The BLAST analysis showed that the deduced amino acid sequence of F-type lectin has 100% identity with F-type lectin of *Pinctada martensii*, whereas, it has 71% identity with *Xenopus laevis*, *Morone saxatilis*, and *Anolis carolinensis*. F-type lectin of *P. fucata* was found to possess only one domain, which is similar to the F-lectin of *Pinctada martensii*. 
Multiple alignment of deduced amino acid sequences (Fig.4.10) with other closely related F-type lecin sequences showed that seven cysteines (Cys 12, Cys 92, Cys 124, Cys 125, Cys 147, Cys 164 and Cys 188) are present in the mature F-type lectin and six of them (Cys 92, Cys 124, Cys 125, Cys 147, Cys 164 and Cys 188) were conserved. Conserved domain search at NCBI revealed that the F-type lectin domain extends from Lys 55 to Val 192. The first carbohydrate binding residue was serine.

Phylogenetic relationships of F-type lectin from pearl oysters and other invertebrates and vertebrates were estimated. F-lectin of *Drosophila melanogaster* was used as the out-group. As shown in Fig.4.11, F-type lectin of congeneric species of pearl oysters *Pinctada fucata* (present study) and *Pinctada martensii* formed one distinct cluster, supported by high bootstrap values (NJ-100%) indicating their close genetic relationship, whereas, the F-lectins from *Xenopus laevis*, *Morone saxatilis* and *Anguilla japonica* formed another distinct clade. Interestingly, the F-lectin from the edible oyster *Crassostrea gigas* formed a third cluster along with other vertebrate species *Lateolabrax japonicas*, and *Anolis carolinensis* though with low bootstrap values.
Chapter 4                  Results

Molecular Genetic Profiling of Functional Genes of Pearl Oyster, Pinctada fucata

Fig. 4.9 Nucleotide sequence and deduced amino acid sequences of F-type lectin of *Pinctada fucata*. Nucleotide numbers are shown on the left. The predicted signal peptide sequence is underlined. Six conserved cysteine residues are highlighted in grey, and three residues for sugar binding are highlighted in black. Polyadenylation consensus signal sequences are shaded. The start codon is in bold and the termination codon is indicated with asterisk (*).

Fig. 4.10 Alignment of the deduced amino acid sequence of F-type lectin of different species. The characteristic conserved regions are shaded.
**Fig. 4.11** Phylogenetic analysis using amino acid sequence of F-type lectin from pearl oyster *Pinctada fucata* and other species. Note: Numbers represent the bootstrap values. GenBank accession numbers are shown next to each species.

### 4.1.2.2 Galectin

Amplification of the galectin gene segment of *Pinctada fucata* through PCR using cDNA of the target gene and primer designed from the conserved regions successfully produced amplicon (<950 bp) as visualized through agarose gel electrophoresis. Optimization of PCR parameters helped to get specific amplicons free from nonspecific PCR product. For the identification and characterization of gene, the PCR products were sequenced. Sequencing of the purified amplicon in both directions gave the sequence information of 925 bp length. The sequence were aligned by using BioEdit software. The nucleotide sequence obtained were subjected to similarity search in NCBI-BLAST and confirmed to be the galectin gene fragment.

RACE-PCR using the primers designed from the above partial sequence information revealed sequence of 1750 bp portion of the galectin gene, comprising of the ORF region and part of UTR. The ORF sequence was used for prediction of the
amino acids being coded, using Transq (EMBL). The ORF encodes a polypeptide of 551 amino acids. The nucleotide sequences of galectin and the amino acid sequences encoded by them are presented in Fig.4.12. No signal peptide was present in the galectin sequence. Conserved domain search at NCBI revealed that it possess four galectin specific domains, each of which are indicated in Fig.4.12. The alignment of galectin sequence of \textit{P. fucata} with that of various vertebrate and invertebrate species (Fig.4.13) revealed that 7 amino acid residues were observed to be conserved in all the four galectin specific domains viz. His44, Asn46, Arg48, Asn61, Trp68, Glu71, and Arg73; the numbers corresponding to the bovine galectin-1 and each of these four galectin domains are connected by a unique 11 amino acid residues, underlined in Fig.4.12.

BLAST analysis showed that the deduced amino acid sequence of galectin deduced in the present study has high identity with the galectin of \textit{Pinctada fucata} reported from China. Multiple alignment of amino acid sequences of galectin of \textit{P. fucata} with other selected species are presented in Fig.4.13. The results indicated that the seven characteristic residues in all the four galectin domains were conserved across these species.

Phylogenetic analysis of galectin (Fig.4.14) of congeneric species of pearl oysters \textit{Pinctada fucata} (present study) and \textit{Pinctada fucata} (AC036044) formed one distinct cluster, supported by high bootstrap values (NJ-100%) indicated their close genetic relationship, whereas, the galectin from \textit{Crassostrea virginica}, \textit{Agropecten irradians} and \textit{Xenopus laevis} formed distinct clades. \textit{Drosophila melanogaster} was used as the out-group.
Fig. 4.12 Nucleotide and deduced amino acid sequences of galectin of *Pinctada fucata*. Nucleotide numbers are shown on the left. Seven conserved residues in four galectin domains are highlighted in grey. Four galectin domains are in brackets. The start codon is in bold and the termination codon is indicated with an asterisk (*).
Fig. 4.13 Alignment of the deduced amino acid sequence of galectin of different species. The characteristic conserved residues within four domains are shaded.
4.1.2.3 Lipopolysaccharide and β-1, 3-Glucan Binding Protein (LGBP)

In case of LGBP, the initial PCR of the cDNA using LGBP specific primers resulted in the amplification of a partial segment of 628 bp, which was sequenced. This partial sequence was used to design new primers and carry out 3' and 5' RACE. RACE-PCR product of LGBP on cloning and sequencing revealed additional sequence data of 1112 bp making up the total 1740 bp. This is a partial segment of the ORF, and it encodes 563 amino acids of the LGBP. The partial coding sequences, the amino acid sequences being encoded by it and its characteristics are presented in Fig.4.15. Glycoside hydrolase family 16 were identified in the sequence (277-480). Furthermore, *P. fucata* LGBP contained a β-1,3-glucanase site with active residues of W (Trp), E (Glu), I (Ile) and D (Asp) at positions of 365, 370, 371 and 372, respectively. Threonine-rich region (T-96, T-132) and a glycine-rich region (G-141, G-199) were also found. The LGBP sequence contained a protein kinase C phosphorylation site (321SAK323) and a modified cell adhesive site (345KGD347). Moreover, a polysaccharide binding motif (340-357), a LPS-binding site (380-396) and a β-1,3-linkage recognition motif of polysaccharides (428-446) were also found.
in the LGBP sequence. Two potential “N-linked glycosylation” sites consisting of 3 amino acids, each were identified at positions N-255, R-256, S257 and N-441, I-442, S-443.

BLAST analysis showed that the deduced amino acid sequence of LGBP has high identity with the LGBP of *Pinctada fucata* reported from China. Multiple alignment of LGBP sequence with other species (Fig.4.16) showed that the potential N-linked glycosylation sites, protein kinase C phosphorylation site (321SAK323), modified cell adhesive site (345KGD347), LPS-binding site, polysaccharide binding motif and β-1,3-linkage recognition motif of polysaccharides are aligned in all the species used for alignment.

Phylogenetic tree using LGBP sequence of the pearl oyster and other species was constructed (Fig.4.17), in which the congenic species of pearl oysters *Pinctada fucata* (present study) and *Pinctada fucata* (ACN76701) formed one distinct cluster, supported by high bootstrap values (NJ-100%) indicated their close genetic relationship, whereas LGBP from other species formed separate clades. *Thermotoga neapolitana* was used as the out-group.
Fig. 4.15 Nucleotide and deduced amino acid sequences of LGBP of *Pinctada fucata*. Nucleotide numbers are shown on the left. Thrreonine-rich region (T-96, T-132) and a glycine-rich region (G-141, G-199) are shaded black with white lettering and grey with black lettering. Two potential N-linked glycosylation sites are boxed. The amino acid residues of protein kinase C phosphorylation site (S-321SAK323) and a modified cell adhesive site (S-345KGD347) are underlined. The termination codon is indicated with asterisk (*).
Fig. 4.16 Alignment of the deduced amino acid sequence of LGBP of different species. The characteristic conserved residues such as polysaccharide binding motif (PsBM), glucanase motif (GM) and β-1,3-linkage recognition motif (βGRM) are shaded.

Fig. 4.17 Phylogenetic analysis using amino acid sequence of LGBP from pearl oyster *Pinctada fucata* and other species. Note: Numbers represent the bootstrap values. GenBank accession numbers are shown next to each species.
4.1.3 Pearl Forming Genes

The pearl formation genes of *P. fucata* taken up for molecular identification and characterization in this study were nacrein, prismalin-14 and N-19. PCR amplification of a segment of each of these gene was initially carried out using the cDNA synthesized from the total RNA. These gene segments were sequenced and the sequence data were used for RACE to develop the nucleotide sequence on its 3’ and 5’ regions. The amino acid sequence of the proteins being encoded by these gene segments were deduced and their characteristic features worked out.

4.1.3.1 Nacrein

PCR of the nacrein gene of *Pinctada fucata* using the cDNA and primers designed from the conserved region successfully amplified a segment of the gene (<1000 bp) as visualized through agarose gel electrophoresis. Optimization of the PCR parameters helped to get specific amplicon free from nonspecific PCR product. For the identification and characterization of gene, the PCR product was purified and sequenced. Sequencing of the amplicon in both directions gave the sequence information of 984 bp length. Similarity search of the sequence was carried out in the NCBI data base. The sequences were aligned by using Bio Edit software. The nucleotide sequence obtained on similarity search in NCBI using BLAST confirmed it to be a segment of the nacrein gene. This partial nucleotide sequence showed 100% homology with the nacrein gene sequence of *P. fucata* available in the NCBI GenBank.

RACE-PCR using the primers designed from the above partial sequence information revealed the sequence of a 1444 bp portion of the nacrein gene, comprising of part of the ORF region and part of UTR. The ORF sequence was used
for prediction of the amino acids being coded by it, using Transq (EMBL). The ORF was found to encode a polypeptide of 438 amino acids.

BLAST analysis showed that the deduced amino acid of nacrein deduced in the present study has high identity with the nacrein of *Pinctada fucata* reported from Japan. Alignment of the amino acid sequences of nacrein of *P. fucata* (present study) with that from Japan revealed that the characteristic features of both of them were observed to be similar.

Nucleotide and deduced amino acid sequences of nacrein are presented in Fig. 4.18. Thirty six residues of the active site (R-46, S-72, Q-107, N-108, R-109, A-110, P-111, E-112, E-114, H-136, N-137, H-139, H-141, E-151, H-152, E-162, H-164, V-166, L-182, F-192, V-194, G-196, Y-370, Y-372, L-376, T-377, T-378, P-379, P-380, T-382, S-384, V-385, W-387, V-389, T-421, R-423) and three zinc binding histidine residues (H-141, H-152, H-164) were also identified in the sequence.
Fig. 4.18  Nucleotide and deduced amino acid sequences of nacrein of *Pinctada fucata*. Nucleotide numbers are shown on the left. Thirty six residues of the active site (R-46, S-72, Q-107, N-108, R-109, A-110, P-111, E-112, E-114, H-136, N-137, H-139, H-141, E-151, H-152, E-162, H-164, V-192, V-194, G-196, Y-370, Y-372, L-376, T-377, T-378, P-379, P-380, T-382, V-385, W-387, V-389, T-421, R-423) are shaded. Three zinc binding histidine residues are boxed. Termination codon is indicated with asterisk (*).
4.1.3.2 Prismalin-14

PCR amplification of the prismalin-14 gene segment of *Pinctada fucata* using cDNA of the target gene and the primer designed from the conserved regions successfully produced amplicon (<200 bp) as visualized through agarose gel electrophoresis. Specific amplicon free from nonspecific products could be amplified through optimization of the PCR parameters. For the identification and characterization of the gene, the PCR product was sequenced. Sequencing of the purified amplicon in both directions gave the sequence information of 195 bp length. The sequences were aligned by using Bio Edit software. The nucleotide sequence obtained were subjected to similarity search in NCBI-BLAST and confirmed to be the prismalin-14 gene fragment.

Sequence of 400 bp portion of the prismalin-14 gene was further resolved through RACE-PCR using the primers designed from the sequence of the initial PCR product, and this segment was found to be comprising of the ORF region and part of UTR of the prismalin-14 gene. The ORF sequence was used for prediction of the amino acids being coded, using Transq (EMBL). The ORF encoded a polypeptide of 121 amino acids. The coding sequences of prismalin-14 and the amino acid sequences encoded by them are presented in Fig.4.19. BLAST analysis showed that the deduced amino acid sequence prismalin-14 deduced in the present study has high identity with the prismalin-14 of *Pinctada fucata* reported from Japan. A signal peptide was identified at the N-terminus of the deduced polypeptide by signal P program and its cleavage site is located between the positions of Gly<sup>17</sup> and Tyr<sup>18</sup>. 
Fig.4.19 Nucleotide and deduced amino acid sequences of prismalin-14 of *Pinctada fucata*. Nucleotide numbers are shown on the left. The predicted signal peptide sequence is underlined. The start codon is in bold and the termination codon is indicated with asterisk (*).

4.1.3.3 N19

Amplification of the N19 gene segment of *Pinctada fucata* through PCR using cDNA of the target gene and primer designed from the conserved regions successfully produced an amplicon (<600 bp) as visualized through agarose gel electrophoresis. Optimization of PCR parameters helped to get specific amplicons free from nonspecific PCR product. Purification of this PCR product followed by sequencing revealed that it consists of 557 nucleotides (Fig.4.20). The sequence was aligned by using Bio Edit software. The nucleotide sequence obtained on similarity search in NCBI-BLAST confirmed to be N19 gene. Deduction of amino acid sequence using online software Transeq of European Molecular Biology Laboratory (EMBL) revealed that 185 amino acids were encoded by the ORF. The nucleotide and amino acid sequence of N19 has shown to be matching with only *P. fucata*. RACE-PCR did not successfully work in case of N19 gene.
4.2 Gene Expression Studies

Expression levels of the defense related genes and pearl forming genes of *Pinctada fucata* following gene induction through exposure to lipopolysaccharide or Mabe implantation, were estimated through semi-quantitative RT-PCR, along with that of the untreated controls. The relative gene expression levels are presented below.

4.2.1 Expression of Defense Related Genes

Live adult individuals of *P. fucata* acclimatized for two weeks in the laboratory at 25°C in tanks containing static aerated seawater (0.5 L/oyster, 25 ppt salinity) were used as controls as well as for exposure to lipopolysaccharide. Test animals which received LPS dissolved in PBS is showed enhanced expression levels of antioxidant genes and PRP genes. The expression levels at different time points
following LPS treatments were tested for significance the differences. Expression in control animals administrated with PBS alone showed slight variations attributable to the effect of PBS/physiological variations. Since variations among the controls were statistically non-significant, it could be ignored. Statistical analysis to evaluate significance of the enhanced gene expression levels over that of the control at different time intervals were also carried out before arriving at conclusion. Expression levels of the different genes under study are presented below.

4.2.1.1 Antioxidant Gene Expression

Cu/Zn-Superoxide dismutase (Cu/Zn SOD)

Semi-quantitative PCR estimation of the mRNA levels revealed that while a certain baseline level of expression of SOD gene was noticed in the control animals, LPS stimulation significantly increased SOD mRNA levels in the haemocytes in the treated animals in a time-dependent manner. These differences in the expression levels at all the time points were statistically significant. Quantification of expression was made by measuring the intensity of the amplified PCR product in the agarose gel, using the imageJ software. After 4 hours of exposure the level showed a significant increase, which further increased and reached the maximum at 8 h post treatment and then dropped to basal levels by 36 h (Fig.4.21). Comparison of the gene expression in LPS treated and the control at different time intervals were made, and the statistical analysis showed that the expression level at 4 and 8 hours after LPS treatment are significantly different (P<0.05) from that of the control (Fig.4.21). At the maximum, the expression of SOD showed a 48% increase over the control.
Fig. 4.21 Expression level of Cu/Zn SOD mRNA in haemolymph of the control and LPS challenged pearl oysters. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) between control and LPS treated are indicated with the asterisk (*).

Glutathione peroxidase and glutathione-S-transferase (GPX and GST)

Both the GPX and GST involved in degrading the peroxides that was produced as a result of the breakdown of reactive oxygen species by the SOD. Following LPS stimulation, the pattern of expression of GPX and GST genes were similar to that of SOD - mRNA levels of both genes increased. The difference in expression levels at all the time points were compared, and found to be statistically significant. Expression levels of both these genes in the LPS challenged specimens showed significant difference from that of the control, reaching the maximum at 8 h, and thereafter, decreasing slowly to control levels. At the maximum (8 h), the relative mRNA expression of both GPX and GST was 34% more than the control. The time dependent variation in the expression of GPX and GST following LPS stimulation were represented in graph (Fig.4.22 and Fig.4.23). The statistical analysis showed that the gene expression level at 8 h after LPS are significantly different (P<0.05) from control.
4.2.1.2 Pattern Recognition Receptor Protein Genes (PRP)

Pattern recognition receptor protein (PRP) genes are involved in the recognition of different immune modulators, and they are capable of binding specifically to the conserved portion of microbial cell wall component.
F-type lectin

In the present study, LPS stimulation substantially increased the expression of F-type lectin in the haemocytes (Fig.4.24). The mRNA level of F-type lectin in the challenged animals reached the maximum at 4 h and then gradually decreased over the time. However, the level of expression at each time points were significantly different from each other. Though, the level of F-type lectin in the challenged animals were higher than the control at all the time points, it was significantly higher at 4 hr and 8 hr only. At the maximum (4 hr) the relative mRNA expression of F-type lectin showed a 50% increase over the control. The graphical representation of the expression level quantified using the imageJ software is presented in graph (Fig.4.24). Result of the statistical analysis of the expression level at different time intervals following exposure to LPS, using SPSS software is also presented in the graph. The statistical analysis showed that the gene expression level at 4 and 8 hours after LPS are significantly different (P<0.05) from control.

![Graph showing expression level of F-type lectin mRNA in haemolymph of the control and LPS challenged pearl oyster.](image)

**Fig.4.24** Expression level of F-type lectin mRNA in haemolymph of the control and LPS challenged pearl oyster. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) between control and LPS treated are indicated with the asterisk (*).
Galectin

Galectin is important for recognizing β-galactoside ligand of the pathogen by their conserved carbohydrate recognition domains (CRD) and play crucial role in innate immunity. Expression of the galectin gene on exposure of the animal to LPS were higher than the control, from 4 to 36 h post challenge. Graphical representation of the expression level quantified using the imageJ software (Fig.4.25) shows that it reached the maximum at 8 h post treatment, and then dropped to basal level at 36 h. Statistical analysis of the gene expression levels in LPS treated and control animas at different time intervals, using SPSS software showed that the expression level at 8 hours post exposure was only significantly different (P<0.05) from the control. However, the expression levels at different time points, when compared without considering the control were significantly different.

![Expression level of galectin mRNA in haemolymph of the control and LPS challenged pearl oyster. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) between control and LPS treated are indicated with the asterisk (*).](image)

**Fig.4.25** Expression level of galectin mRNA in haemolymph of the control and LPS challenged pearl oyster. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) between control and LPS treated are indicated with the asterisk (*).
Lipopolysaccharide and β-1, 3-glucan binding protein (LGBP)

In the present study, LPS stimulation significantly increased the mRNA expression of LGBP in the Haemocytes in a time-dependent manner (Fig.4.26). The level of LGBP mRNA in the challenged animals increased significantly compared to the control, and reached maximal levels at 8 h and then gradually decreased over the time. At the maximum the mRNA level of LGBP was 24% more over the control. The graphical representation of the expression level at different time intervals following exposure to LPS, quantified using the imageJ software is presented in Fig.4.26 along with the result of the statistical analysis using SPSS software. The statistical analysis showed that the gene expression level at 8 hours after LPS were significantly different (P<0.05) from the control. However the LPS treated samples at different time intervals were showing statistically significant differences when compared without considering the control.

![Graph showing expression level of LGBP mRNA in haemolymph of the control and LPS challenged pearl oyster.](image1)

**Fig.4.26** Expression level of LGBP mRNA in haemolymph of the control and LPS challenged pearl oyster. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) between control and LPS treated are indicated with the asterisk (*).
4.2.2 Expression of Pearl Forming Genes

Expression levels of the pearl forming genes in *Pinctada fucata* following Mabe implantation estimated through semi-quantitative RT-PCR along with that of the controls are presented below.

Acclimation of oyster at the optimum salinity under the controlled condition was carried out to understand the normal gene expression, which would serve as the baseline to compare the gene expressions following implantation. At each predetermined day (10, 30 & 40) following implantation, five oysters were harvested. Mantle surrounding the Mabe were carefully isolated and subjected to transcriptome analysis. The gene coding for GAPDH RNA was used as housekeeping gene, as its expression is not effected by environmental stressors, and therefore, could serve as an internal control in the relative gene expression studies following implantation. The expression levels of the four pearl forming genes at different time intervals are presented below.

**Nacrein**

Expression of the nacrein gene after Mabe implantation have shown initial up regulation. As shown in fig.4.27 endogenous expression of nacrein gene increased in the implanted animals. The increased level of nacrein mRNA in the experimental oysters was significantly higher than the control up to the 30th day. The increase was 53% more than the control at day 10. On the 30th day the gene expression was marginally reduced than the 10th day level. However, there was a steep decrease on day 40 compared to the earlier days, but was still higher than the control. Fluctuations in the control animals during the period was only very marginal. Graphical representation of the expression level quantified using the imageJ
software, which detects the intensity of the amplified gene product in the agarose gel are presented in Fig.4.27. Statistical analysis of the gene expression levels at different days following Mabe implantation, using SPSS software showed that the gene expression level at day 10 and 30 were significantly higher (P<0.05) than the control (Fig.4.27).

![Graph showing gene expression levels](image)

**Fig.4.27** Expression level of Nacrein gene in mantle surrounding the Mabe quantified by semi-quantitative PCR. Vertical bars represent the mean ±S.E (N=5). Significant differences (P<0.05) are indicated with the asterisk (*).

**Prismalin-14**

As shown in Fig.4.28 prismalin-14 gene expression increased in the implanted animals. The level of prismalin-14 mRNA in the experimental oysters was significantly higher in the initial period itself than the control, and it was sustained up to 30 days. The increase was 35% more than the control on day 10. On the 30th day the gene expression was marginally less than the 10th day level. This level was sustained on the 40th day also. Fluctuations in the control during the period was only very marginal. The graphical representation of the expression level quantified using the imageJ software which detected the intensity of the amplified gene product in the
agarose gel is presented in fig.4.28. Statistical analysis of the gene expression levels at different days following Mabe implantation using SPSS software showed that the gene expression level at day 10 post exposure was significantly higher (P<0.05) than the control (Fig.4.28).

![Expression level of Prismalin-14 gene in mantle surrounding the Mabe measured by semi-quantitative PCR. Vertical bars represent the mean ±S.E (N=5). Significant differences (P<0.05) are indicated with the asterisk (*).](image)

Fig.4.28 Expression level of Prismalin-14 gene in mantle surrounding the Mabe measured by semi-quantitative PCR. Vertical bars represent the mean ±S.E (N=5). Significant differences (P<0.05) are indicated with the asterisk (*).

N19

Marginal increase in the expression of N19 gene was observed at day 10 compared to the control as shown in fig.4.29. However, at day 30 gene expression of N19 increased further, and was significantly higher than the control. At 40th day, though there was a slight decrease, compared to 30th day, the level was higher than that on the 10th day. Fluctuations in the control animals during different time intervals was very marginal. The gene expression at different days estimated through imageJ software are graphically represented in fig.4.29. The statistical analysis using SPSS software showed that the gene expression level at 30th day after Mabe implantation was significantly different (P<0.05) from control.
Fig. 4.29 Expression level of N19 gene in mantle surrounding the Mabe measured by semi-quantitative PCR. Vertical bars represent the mean ± S.E (N=5). Significant differences (P<0.05) are indicated with the asterisk (*).