Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate
Sham V. Nair, Heather Del Valle, Paul S. Gross, David P. Terwilliger and L. Courtney Smith

You might find this additional information useful...

Supplemental material for this article can be found at:
http://physiolgenomics.physiology.org/cgi/content/full/00052.2005/DC1

This article cites 128 articles, 51 of which you can access free at:
http://physiolgenomics.physiology.org/cgi/content/full/22/1/33#BIBL

Updated information and services including high-resolution figures, can be found at:
http://physiolgenomics.physiology.org/cgi/content/full/22/1/33

Additional material and information about Physiological Genomics can be found at:
http://www.the-aps.org/publications/pg

This information is current as of July 1, 2005.
Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate

Sham V. Nair,1 Heather Del Valle,1 Paul S. Gross,2 David P. Terwilliger, and L. Courtney Smith1

1Department of Biological Sciences, George Washington University, Washington, District of Columbia; and 2Department of Biochemistry, Medical University of South Carolina, Charleston, South Carolina

Submitted 1 March 2005; accepted in final form 6 April 2005

Nair, Sham V., Heather Del Valle, Paul S. Gross, David P. Terwilliger, and L. Courtney Smith. Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate. Physiol Genomics 22: 33–47, 2005. First published April 12, 2005; 10.1152/physiolgenomics.00052.2005.—The purple sea urchin, Strongylocentrotus purpuratus, is a member of the phylum Echinodermata, which is basal to the phylum Chordata within the deuterostome lineage of the animal kingdom. This relationship makes the analysis of the sea urchin immune system relevant to understanding the evolution of the deuterostome immune system leading to the Vertebrata. Subtractive suppression hybridization was employed to generate cDNA probes for screening high-density arrays, conventional cDNA libraries to identify genes that were upregulated in coelomocytes responding to lipopolysaccharide. Results from 1,247 expressed sequence tags (ESTs) were used to infer that coelomocytes upregulated genes involved in RNA splicing, protein processing and targeting, secretion, endosomal activities, cell signaling, and alterations to the cytoskeletal architecture including interactions with the extracellular matrix. Of particular note was a set of transcripts represented by 60% of the ESTs analyzed, which encoded a previously uncharacterized family of closely related proteins, provisionally designated as 185/333. These transcripts exhibited a significant level of variation in their nucleotide sequence and evidence of putative alternative splicing that could yield up to 15 translatable elements. On the basis of the striking increase in gene expression in response to lipopolysaccharide and the unexpected level of diversity of the 185/333 messages, we propose that this set of transcripts encodes a family of putative immune response proteins that may represent a major component of an immunological response to bacterial challenge.

The immune system of the purple sea urchin, Strongylocentrotus purpuratus, lacks the hallmarks of adaptive immunity (103). Yet, as with other invertebrates studied thus far, there are several commonalities among the innate immune reactions of all animals. Current paradigms suggest that invertebrates recognize conserved molecular structures, or pathogen-associated molecular patterns (PAMPs), displayed by invading pathogens. Recognition is mediated by pattern recognition receptors (PRRs), which lack the discriminative capacity of the immunoglobulin-based receptors of the vertebrate adaptive immune system (52, 68). This paradigm is now being challenged by newly emerging data from a number of laboratories, including ours, which may necessitate a reappraisal of immune diversity in invertebrates (15, 16, 24, 63 and reviewed in Ref. 28).

Studies assessing the immune response in sea urchins have shown that the coelomic cavity is rapidly cleared of microbes, foreign cells, and other materials by the activities of phagocytic coelomocytes (reviewed in Ref. 104). Although there are several morphologically distinct classes of coelomocytes, including phagocytes, red spherule cells, colorless spherule cells, and vibratile cells (27, 54, 103), the phagocytes appear to have a central role in host defense. Phagocytes have extensive cytoskeletons and are amoeboid-macrophase-like cells that undergo significant changes in shape during chemotaxis, phagocytosis, encapsulation, clot formation, secretion, and degranulation, all of which are associated with immune responses (32, 34, 99). Analysis of expressed sequence tags (ESTs) prepared from immune-activated coelomocytes identified a simple complement system operating in sea urchins (101). The sea urchin homolog of vertebrate complement component C3 (SpC3) was the first member of the thioester-containing family of proteins identified in an invertebrate (3). It appears that SpC3 is functionally similar to its vertebrate homologs (97) in that it mediates opsonization of foreign matter and pathogens, which leads to augmented uptake by phagocytic coelomocytes (20, 96). A homolog of factor B (Bf), called SpBf, has also been identified in sea urchins (98, 113) along with two cDNAs that encode the mosaic proteins Sp5 and Sp5013 that share a number of domains [including a factor I membrane attack complex (FIMAC) domain] with complement regulatory proteins and members of the terminal complement pathway (70).

The use of large-scale screening technologies employing microarrays, cDNA libraries, and proteomics (118) has provided insights into transcriptional programs that are initiated in animal cells responding to pathogenic threat. These approaches have significantly accelerated the rate of gene discovery in invertebrates, including two genera of shrimp, Penaeus (108) and Litopenaeus (33); two species of oyster, Crassostrea gigas (35) and C. virginica (53); and the sea urchin, S. purpuratus (101). Such studies have been fruitful even for the more familiar and well-characterized invertebrates such as the fruit fly, Drosophila melanogaster (18, 48a), and two mosquito species, Anopheles gambiae (17) and Aedes aegypti (94). Comparative analysis of the emerging data indicates that there are general classes of genes that may be the common denominators of invertebrate immune responses. The molecules encoded by these genes include peptidoglycan recognition proteins, thioester-containing proteins, gram-negative binding proteins, multidomain scavenger receptor family, ficolins, C-
type lectins, and galectins. Other commonly expressed groups of proteins comprise proteases (serine proteases, cysteine proteases, metalloproteases), protease inhibitors (kazal-type serine protease inhibitor, serpin, α2-macroglobulin), antimicrobial peptides, superoxide dismutase, and metal-binding proteins. Proteolytic cascades (such as the prophenoloxidase and the clotting/coagulation systems) and certain signal transduction systems, especially those leading to the activation of rel/NF-κB (17, 35, 108, 33), are also common components of immune activation in invertebrates.

By employing EST studies or differential display, a number of genes in sea urchin coelomocytes have been shown to be induced by immune challenge, including Sp064, which encodes SpC3 (19), the transcription factors SpNFκB and SpRUNT (78), as well as other uncharacterized transcripts (83). The expression of a C-type lectin, Sp056 (GenBank access no. AY663300), is only detectible after LPS challenge and has been used as a reliable marker of coelomocyte activation (70, 113). On the other hand, there are some immune-related genes expressed in coelomocytes that are not induced by immune challenge, including Sp152 (SpBf; Ref. 113), Sp5 and Sp5013 (70), and a large set of scavenger receptors with cysteine-rich (SRCR) domains (76, 77). Furthermore, the transcription factor SpGATAc has been shown to be downregulated in response to immune challenge (78).

At present, only a limited number of immune effector cascades (e.g., complement components, a set of inducible lectins, and SRCR proteins) have been identified in sea urchins, but given the effectiveness of the sea urchin to repel infection, it is expected that the sea urchin immune system will have many more as yet unidentified components and cascades. It is hypothesized that large-scale genomics-based screening approaches should reveal additional genes with central importance in sea urchin immunity. To test this hypothesis, normalized subtracted probes were used to identify expressed genes that were induced by LPS through screening two arrayed, conventional, coelomocyte cDNA libraries, one of which was constructed from bacterially activated cells. Reported here are the detailed results from an analysis of ESTs, which revealed that a number of genes were induced, including many novel ones. Matches to known sequences suggested that coelomocytes respond to gram-negative bacteria by increased expression of genes encoding proteins involved in RNA splicing, protein processing and secretion, signaling pathways, modification of the cytoskeletal architecture, and altered interactions with the extracellular matrix. Of particular note was the observation that the vast majority of the ESTs represented a set of similar transcripts that encoded a family of putative immune response proteins. It is hypothesized that these proteins may be a major immune effector system in the sea urchin based on a striking increase in gene expression in response to LPS plus significant sequence diversity due to variations in the primary nucleotide sequence and putative alternative splicing.

**MATERIALS AND METHODS**

**Sea urchins.** Animals were obtained and housed as previously described (34, 95). Immunoquiescent animals were generated by long-term housing (≥6–8 mo) in a closed aquarium without significant disturbance (19, 32, 95).

**Animal treatments and RNA isolation.** Coelomocytes were withdrawn from sea urchins as described (32, 70), and total RNA was isolated using an RNeasy Mini kit according to the manufacturer’s protocol (Qiagen, Valencia, CA). Sea urchins were injected once on day 1 and twice on day 2 (~9-h interval) with 0.5 μg LPS/ml coelomic fluid, according to the methods of Smith et al. (99). Coelomocytes collected for RNA isolation and cDNA subtractions were obtained before injection (immunoquiescent) and on day 3 (LPS activated).

**RT-PCR and construction of suppressive subtractive hybridization cDNA.** Reverse transcription reactions with random hexamer primers were carried out on 0.5–3.0 μg of total RNA, 200 U of Superscript II RT (Life Technologies, Carlsbad, CA), and 20 μRNAsin (Promega, Madison, WI). The resultant cDNA (1 μl) was amplified by PCR using 1 U of Taq polymerase (Life Technologies) and 1 μM each primer (Supplemental Table S1; available at the Physiological Genomics web site). Construction of suppressive subtractive hybridization (SSH) cDNA was carried out using the PCR-select cDNA subtraction kit (Clontech, Palo Alto, CA) according to the manufacturer’s instructions. Forward cDNA subtraction employed coelomocyte RNA from an LPS-activated sea urchin (urchin) subtracted from RNA prepared from the same animal before challenge (driver), while reverse cDNA subtraction used immunoquiescent RNA (tester) subtracted from LPS-activated RNA from the same sea urchin (driver). The quality of the SSH reaction was assessed using a virtual Northern blot. Samples of forward- and reverse-subtracted cDNAs, in addition to unsubtracted cDNA controls, were amplified by PCR using Clontech primers, electrophoresed, and blotted using standard methods (87). Filters were probed with inserts from specific clones (Supplemental Table S1) that were amplified by PCR and labeled with alkaline phosphatase according to the manufacturer’s instructions (AlkPhos Direct; Amersham Pharmacia Biotech, Piscataway, NJ).

**Riboprobe preparation.** Forward- and reverse-subtracted cDNAs that served as riboprobe templates were amplified by PCR using the P1 forward primer and the nested P2 reverse primer (PCR-select cDNA subtraction kit, Clontech). The amplified cDNAs were labeled with the use of T7 RNA polymerase (Promega) and [32P]UTP (ICN, Costa Mesa, CA). The eluates from G50 Sephadex spin columns (Amersham Pharmacia Biotech) were quantified in an LS6500 scintillation counter (Beckman Instruments, Fullerton, CA). Probes were diluted to a final volume of 500 μl in hybridization solution [50% deionized formamide, 0.25 M phosphate buffer, 0.1% (wt/vol) BSA, 1 mM EDTA, 7% (wt/vol) SDS] and heated to 65°C for 4 min before addition to the filters.

Riboprobes were also generated by run-off polymerization using three 18S/33S clones (Sp0323, Sp0324, Sp0325). Each clone was linearized with XhoI, and [32P]UTP (ICN) was incorporated into a riboprobe with T3, T7, or Sp6 RNA polymerase, depending on the template (see Ref. 70). Unincorporated [32P]UTP was removed with the use of a G50 Sephadex spin column (Amersham Pharmacia Biotech). Probes were heated to 60°C for 4 min before hybridization.

**Macroarray filter screening and clone selection.** Two arrayed cDNA libraries (5 filters per library of 92,160 clones) were used for analysis, one constructed from bacterially activated coelomocytes (14, 78, 82) and a second from nonactivated coelomocytes (3). For both libraries, coelomocytes were pooled from five sea urchins. Each library was screened with [32P]-labeled riboprobes derived from the SSH cDNAs and specific cloned templates that were also used for virtual Northern blot analysis (Supplemental Table S1). Briefly, macroarray filters were prehybridized for 1 h at 42°C followed by hybridization with the riboprobe at 42°C overnight in a rotating oven. Filters were washed twice for 30 min at 65°C in each of the following wash buffers: 4× SSC with 1% SDS, 2× SSC with 1% SDS, and 1× SSC with 1% SDS. Wet filters were sealed within plastic bags and exposed to film.

---

1 The Supplemental Material for this article (Supplemental Tables S1 and S2) is available online at http://physiolgenomics.physiology.org/cgi/content/full/00052.2005/DC1.
at \(-70^\circ C\) using intensifying screens for varying periods. After autoradiography, filters were stripped of the probes by denaturation in 0.4 M NaOH for 30 min at 45°C followed by incubation in stripping buffer (0.1× SSC, 0.1% SDS, 0.2 M Tris–HCl, pH 7.5) for 30 min at 65°C. Final wash was in stripping buffer with 0.2 M EDTA for 10 min at room temperature. After stripping, filters were exposed to film to ensure that there was no residual radioactivity associated with the filters and stored at \(-20^\circ C\).

Clones identified during screening were manually picked, isolated, robotically arrayed into 96-well plates (as described in Ref. 33), and sequenced (Amplicon Express, Pullman, WA). Sequence similarities were identified in the nonredundant nucleic acid database and the EST database (GenBank) using basic local alignment search tool (BLASTn) and BLASTx algorithms accessed from the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/). Only matches with \(e\)-values of \(<10^{-6}\) were considered significant. The ESTs generated in this study have been submitted to GenBank (accession nos. CK828301–CK829214 and Cv652690–Cv652795).

Nonsynonymous-synonymous (\(dN/dS\)) ratio analysis. The 5′-ends of 293 of the best quality 185/333 ESTs were analyzed for nucleotide diversity. The region was 195 bases long and was manually aligned in BioEdit (38). Analysis of the alignment using WinClada (75) indicated that 250 of the 293 sequences were duplicates, which were removed to simplify further analysis. Two clones, Sp0267 and Sp0291, that had stop codons in element 2 were also removed, leaving 42 unique EST clones for phylogenetic analysis by maximum likelihood (PAML) (126). A neighbor-joining tree was generated in molecular evolutionary genetics analysis (MEGA) v.3.0 (56) and used for the PAML analysis.

RESULTS AND DISCUSSION

The availability of immunoquiescent sea urchins has been an important tool for the investigation of specific responses to injury and pathogenic challenge. This is a limiting factor when attempting to assess immune responses in animals collected within a few days or weeks from the wild. Contact with microbes and pathogens in a natural environment activates host defenses continuously in these animals and therefore precludes their use as baseline or downregulated controls. Stress from collection and shipping exacerbates immune activation. Hence, in this study we have used only immunoquiescent animals to generate a baseline of immunological activation. Immunoquiescence has been defined as the condition in which the abundance of certain immune-specific transcripts becomes diminished in sea urchin coelomocytes compared with animals living in their natural environment (32). Immunoquiescence has been demonstrated in S. purpuratus after animals have been housed in closed-circulation aquaria in the laboratory for 6–8 mo (19, 20, 32, 34). Long-term housing does not induce immunosuppression, only immunoquiescence, as immune reactivity can be reversed with injections of bacteria or LPS or from injury. Such challenges result in increases in the relative abundance of the immune-specific transcripts within individual animals (19, 20). In a modified approach to that reported previously (101), immunoquiescent sea urchins were used to obtain RNA from coelomocytes pre- and post-LPS challenge for normalized subtractions. Subtracted cDNAs generated from coelomocytes collected from the same animal were used to produce riboprobes for screening filters from arrayed coelomocyte cDNA libraries. By employing the same animal for these purposes, we avoided inherent problems associated with the genetic diversity present among out-bred sea urchins (99, 100).

**Immunquiescence and RT-PCR analysis of LPS activation.** The immunoquiescent status of four randomly chosen sea urchins was assessed by RT-PCR for expression of several coelomocyte-specific genes. Two animals were found to be immunoquiescent, as evidenced by the lack of initial expression of the gene encoding a C-type lectin (Sp056; Fig. 1A) and its subsequent induction after challenge (Fig. 1B). This result was in agreement with previous work wherein Sp056 proved to be a useful marker for assessing immune activation to LPS challenge (70, 113). A second gene, whose pre- and postchallenge induction is more variable among individual animals under laboratory-rearing conditions, is Sp064, which encodes SpC3 (19). It was found to have low expression or was undetectable before challenge (Fig. 1A, animals 2 and 3) but was inducible in all animals after LPS injection, with one individual exhibiting a large change in expression (Fig. 1B, animal 2). The expression of three additional genes, SpNFkB (78), Sp152 (113) and cytoplasmic actin (CyI; Ref. 58), was analyzed by RT-PCR but showed no change in response to LPS (data not shown). Animal 2, which showed the largest detectable increase in Sp064 expression, was chosen for further analysis.

**cDNA subtraction and virtual Northern blot analysis.** Forward- and reverse-subtracted cDNAs were generated by SSH from isolated RNA samples obtained from animal 2. Forward subtraction selectively enriches the pool of transcripts that increase upon LPS challenge and injury, whereas the reverse subtraction enriches for transcripts that diminish upon LPS challenge and injury. The efficiency of both the forward and reverse subtractions was evaluated by comparing subtracted and unsubtracted cDNAs using virtual Northern blot analysis. Reverse subtraction effectively removed CyI and Sp056 transcripts from the immunoquiescent cDNA pool (Fig. 2, lanes 3 and 11), whereas these cDNAs were enriched in the LPS-activated or forward-subtracted cDNA preparation (Fig. 2, lanes 4 and 12). This indicated that successful subtraction had occurred in both directions. Both Sp056 and CyI were identified in the unsubtracted cDNA from both activated and immunoquiescent coelomocytes, although greater hybridization was observed in the LPS-activated cDNA pools (Fig. 2, lanes 2 and 10). The change in CyI expression may actually reflect a real
increase in actin gene expression in LPS-activated coelomocytes and would correspond with increased profilin expression (99, 100), which is involved in modifications to cytoskeletal shape (81).

Macroarray filter screens. Macroarray filters from two conventionally constructed and arrayed coelomocyte cDNA libraries were screened twice each with reverse-subtracted riboprobes and then stripped and rescreened with forward-subtracted riboprobes. Spots that were positive either for the reverse-subtracted probe or for both probes were not considered further. Filters were also screened for Cyt (actin), Sp056 (lectin), and Sp064 (SpC3), and positive clones were excluded from analysis (data not shown). From ~6,000 clones that were repeatedly identified from the bacterially activated library as genes that were differentially expressed during LPS challenge and/or injury, 1,025 randomly chosen clones were sequenced initially. Approximately 73% (746) showed significant matches to hitherto uncharacterized cDNAs known as DD185 (Ref. 78, GenBank accession no. AF228877) and EST333 (Ref. 101, GenBank accession no. R62081), hereafter provisionally called 185/333. Because there was such a high preponderance of 185/333 sequences among the clones analyzed, filters were rescreened using 185/333 riboprobes, and positives were excluded from the analysis of a subsequent 222 positive clones. Of the 1,247 clones that were analyzed in total, 71 sequences (5.7%) contained too many indeterminate base calls for successful matching, while 328 sequences (26%) were categorized as unknowns either because there was no significant similarity to any sequence (11%) or poor (e-value >10^-5) similarity (15%) by BLAST searches of the nonredundant database. Only 61 ESTs matched known sequences from Strongylocentrotus species, and these represented 23 distinct gene products. Thus this project has deposited >1,100 ESTs into the public domain. These represent novel sea urchin transcripts that are synthesized by immune-activated coelomocytes.

Sequences that could be assigned a putative identification could be further clustered by associated function (Fig. 3). The vast majority of ESTs sequenced were 185/333 transcripts, but a number of immune-related genes were also identified. In addition, a significant number of genes involved in cytoskeletal modulation (cytoskeletal architecture, motility, and cell adhesion), nuclear activities (RNA splicing and transcription), signal transduction, protease activity, and metabolism/energy generation were identified (Table 1; for a complete list see Supplemental Table S2).

185/333 encodes a putative defensive protein. Previous studies identified 185/333 as a transcript that was expressed in sea urchin coelomocytes in response to LPS (101) or to microbial challenge and physical injury (83). Evaluation of 185/333 transcript accumulation in coelomocytes by differential display and Northern blots indicated that expression was maximal at 6 h after bacterial challenge and at 12 h after injury. In agreement with these results, filter screens from both arrayed libraries indicated that the 185/333 transcripts constituted an estimate of 5,929 of the 91,920 clones (6.45%) in the bacterially activated coelomocyte library, whereas there were only 79 positive clones (0.086%) in the nonactivated coelomocyte library of the same size. On the basis of the extraordinary change in expression level in response to LPS challenge, 185/333 was investigated in detail.

Because 60% of the 1,247 ESTs in our database matched to 185/333, the entire coding sequence of the 185/333 message could be reconstructed from these partial sequences. Optimal alignment of 185/333 ESTs could only be accomplished by inserting large gaps between matching stretches of nucleotides that delineated possible exons. This analysis indicated that the complete 185/333 translated sequence was composed of 15 elements (Fig. 4). Because not all elements were present in every 185/333 sequence, this suggested that many of the transcripts might be alternatively spliced. Nucleotide variations in the form of base substitutions were evident throughout the length of the 185/333 messages. Figure 5 illustrates these variations and shows a partial sequence that includes the 3'-end of the leader and the first element. An inspection of the nucleotide substitutions indicated that they were not randomly distributed but occurred at specific positions along the sequences. While many substitutions were conservative, a substantial number of changes resulted in amino acid changes when translated. To determine the ratio of nonsynonymous to synonymous nucleotide changes (dn/ds), 42 unique ESTs (of 293) with high-quality 5'-ends were analyzed with PAML (126). The analysis was restricted to 193 nt, which encoded the leader and element 1, because it maximized the number of ESTs that could be employed in the analysis and also avoided large gaps in the alignment. Results indicated that the dn/ds
ratio was 1.12. Any score above 1.0 implies that there are more nonsynonymous nucleotide changes than synonymous changes, and therefore, nucleotide diversity for the 185/333 genes is being maintained within the population to putatively increase fitness. However, about one-half of the sequence that was employed in the analysis encoded the leader, which appears to be less diverse than the rest of the sequence (see Fig. 4A). Consequently, this suggests that element 1 was the major contributor to this result and that elevated diversity may be present in all downstream elements.

The NH$_2$ termini of the translated sequences contained a hydrophobic leader terminating with a Ser at position 18. The remaining protein could be divided into three regions: 1) an NH$_2$-terminal glycine-rich region (Fig. 4, A and C; gly-rich, elements 1–5) and 2) a histidine-rich central region (his-rich, elements 6–14) and a COOH-terminal region (element 15), which exhibited stop codons at two different positions, depending on the 185/333 sequence examined. Two of the 618 185/333 sequences that were analyzed revealed stop codons in element 2 (Fig. 4A; Sp0269 and Sp0289) that would encode an extremely short protein. Nucleotide insertions/deletions were identified in the leader region resulting in two translated variants (MEVK vs. M-VK) as well as in elements 9 and 15 (Fig. 4A). While no signature sequences for O-linked glycosylation were present, conserved NH$_2$-linked glycosylation sites (up to 12) were noted in the his-rich region, as well as one in the gly-rich region in some sequences (Fig. 4, A and C). Because no cysteines were found in any of the 185/333 coding regions, the structure of the encoded proteins would not be stabilized by disulfide bonds. In addition, a cell adhesion motif (RGD) was identified in element 5.

Short, repeated stretches of amino acids were identified within both the gly-rich and the his-rich regions, using Megalign (DNASTAR, Madison, WI), and by visual inspection [Fig. 4, A (colored brackets) and C, and Table 2]. These included imperfect repeats, which showed variations in their amino acid sequences (Table 2, types 1, 3, 5). Repeat type 1 was only found in the gly-rich region and consisted of four full-length repeats and one short repeat, which was missing five residues from the NH$_2$-terminal end [Table 2 and Fig. 4, A (red brackets) and C]. Repeats 2–5 were only distributed throughout the his-rich region, while the COOH-terminal
region only had repeat type 5. It was interesting that the positioning of the elements and the positioning of the repeats in the gly-rich region did not correspond. Although repeats 2–5 were not found in tandem arrays like type 1, a series of several different repeats were positioned in a tandem arrangement of 5-2-3-4 in elements 14 and 15 (Fig. 4, A and C). Overall, the alignment indicated that the translated 185/333 sequences showed possible alternative splicing, nucleotide polymorphisms, internal repeats, potential NH2-linked glycosylation sites limited mostly to the his-rich region, a variable positioning of the termination codon, and a lack of disulfide bonds.
Unexpected results were revealed from a nucleotide alignment of 290 of the 185/333 ESTs that had the least number of ambiguous bases and included the 5′-UTRs plus 190–196 nucleotides of open reading frame (ORF). This was essentially the same set that was analyzed for the dn/ds ratio above. On the basis of the sequence of the 5′-UTRs, the ESTs selected for this analysis could be allocated into 25 groups, of which 19 groups had between 2 and 101 members and 6 occurred as singletons (Table 3). Thirteen groups, each composed of five or more ESTs with identical 5′-UTRs, were analyzed using WinClada (75) to identify the linkage of an ORF sequence with a particular 5′-UTR. Results indicated that between one and eight ORFs were linked to a given 5′-UTR. If each 5′-UTR sequence was representative of expression from an individual gene, this result suggested that multiple genes with identical 5′-untranslated regions were present in the genome and probably encoded different proteins. The total number of different ORF sequences identified using this approach of analyzing 13 groups (with 5 or more sequences) was 40 (Table 3). On the other hand, when the total number of distinct ORFs was analyzed from all 290 ESTs without grouping the sequences by 5′-UTR identity, there were only 33 different ORFs. This indicated that identical ORFs were associated with different 5′-UTRs.

These data suggested that the 185/333 ESTs represented expression from a large family of closely related genes that may perhaps share sequences and are induced by immune challenge with LPS. Message heterogeneity appears to be a result of putative alternative splicing within and possibly between genes, as well as nucleotide polymorphisms within elements. The messages encode a family of similar but distinct proteins that show an unexpected level of sequence diversity, perhaps under positive selection for diversity as a result of pathogen pressure. On the basis of their expression patterns in response to LPS and the presence of a leader, these proteins may be secreted from the coelomocytes into the coelomic fluid and have an important immune function.

Defense-related proteins. A number of defense-related genes, in addition to the 185/333 family, were identified in this study. Amassin (Sp1200), a 75-kDa sea urchin protein that contains an olfactomedin domain, may be secreted by coelomocytes in response to injury or infection (46). The olfactomedin domain may represent an intercellular adhesion domain involved in coelomocyte-mediated clotting through amassin polymerization (reviewed in Ref. 46). Amassin may also bind to cell surface receptors, such as integrins on coelomocytes. It is noteworthy that a match to integrin-β (Sp0188) was identified that has been reported previously from sea urchin embryos (13, 71). Both the βC- and βL-integrins can be detected serologically on the surface of coelomocytes (R. D. Burke, personal communication).

A galectin-like sequence (Sp0014) was identified that matched with both galectin-8 and galectin-9. Galectins are members of the S-type lectin subfamily that bind β-galactoside structures (36). Galectin-9 functions in the induction of apoptosis (55), acts as an eosiophil chemotactant (67), and is possibly involved in inflammatory reactions as well as cell-cell adhesion (4). The expression of a galectose-binding lectin in coelomocytes may be involved in responses to microbial challenge, perhaps by mediating chemoattraction and cellular accumulations at points of contact with microbes displaying galactoside structures. The induction of the C-type lectin Sp056 in coelomocytes (Fig. 1) is another example of a galactose-binding lectin that is induced by LPS (70, 113).

Two sequences (Sp0157, Sp1107) matched domains that are found in a family of proteins that include thrombospondin (TSP), properdin, semaphorin, angiogenesis inhibitor, hemi-centrin, and fibulin-6 (1). Mosaic proteins with TSP domains function in cell-cell and cell-matrix interactions, cytoskeletal modifications, cell motility and aggregation, inflammatory responses, and wound healing (1). Although these activities are likely to occur in coelomocytes responding to LPS, the identification of a possible properdin homolog that might function in the ancestral sea urchin complement system is of great interest (102). In higher vertebrates, properdin functions to stabilize the formation of C3-convertase by the alternative pathway, and its expression in coelomocytes suggests that the sea urchin complement system may initiate the generation of an echinoderm C3-convertase (for review, see Refs. 98, 102). Sequence analysis of both Sp0157 and Sp1107 indicated that these ESTs represented two tandem TSP1 domains, and it is notable that properdin monomers also have TSP1 domains arranged in a tandem fashion (92).

A single EST (Sp1186) matched to the serine protease inhibitor phosphatidyethanolamine-binding protein (PEBP), which is a phospholipid-binding protein with inhibitory activity toward serine proteases (42). PEBP is highly conserved, and homologs have been identified in a number of species from the animal, plant, and archael kingdoms. The interaction of PEBPs with Raf-1 kinase leads to the inhibition of the mitogen-activated protein (MAP) kinase signaling pathway, which suggests that these proteins may inhibit mitogenic signaling (42, 129). A number of cysteine proteases, such as cathepsin L (Sp0434) and cathepsin B (Sp1108), were identified. Cathepsins are known to function in lysosomes of mammalian macrophages and act to degrade basement membranes and the extracellular matrix of phagocytosed microbes (26, 80). In addition, cathepsins may be released from cells by secretory lysosomes to act on invading microbes in the extracellular milieu (106). Matches to these types of enzymes, as well as to arylsulfatase (EST003, EST004, EST072, EST401; Ref. 101), suggested that coelomocytes may respond to LPS challenge by synthesizing a significant number of proteases that are required for optimal functioning of the endosomal system.

RNA splicing and protein synthesis, processing, and degradation. Many sequence matches identified in this study, inferred from large-scale increases in the expression of genes that encode proteins involved in protein synthesis, processing and targeting, occur in coelomocytes after immunological challenge. There were several matches to proteins involved in the regulation of transcription, such as steroid hormone receptor (Sp1157), LPS-induced tumor necrosis factor-α (Sp1034), immediate-early-response protein (Sp1188), and DNA methyl transferase-associated protein-1 (Sp1093). Other matches included proteins involved in RNA splicing, such as the ET putative translation product (Sp0576), nuclear RNA- and DNA-binding protein (Sp0874), heterogeneous nuclear ribonucleoprotein R (Sp1122), splicing factor 30 (Sp1214), paraspeckle protein (Sp1176), and a conserved nonhistone nucleic acid-binding protein (Sp0896). Increased expression of proteins associated with RNA splicing in coelomocytes suggested an overall increase in the expression of genes that
Fig. 4.

Immune Diversity in a Sea Urchin
require postsynthetic RNA processing. Alternative splicing has been characterized in the expression of Sp152, which encodes the complement homolog SpBf, where one or two of the five short consensus repeats may be spliced out (113). Two transcripts encoding mosaic proteins with FIMAC domains may also undergo alternative splicing (70). It is possible that splicing may also be involved in 185/333 message production.

Matches to proteins involved in protein synthesis included ribosomal proteins (Sp1141, others), translation initiation factor (Sp1244), translation elongation factor (Sp1074), and peptide chain release factor subunit 1 (Sp1059). In addition, matches were identified to Sec61 (Sp9096) and the γ-subunit of the signal sequence receptor or the translocon-associated protein (TRAP; Sp1118). TRAP-γ functions in a complex that binds the signal sequence of proteins that are subsequently produced by ribosomes associated with the endoplasmic reticulum (ER) and fed through the translocation apparatus into the ER lumen (39). Sec61 is a component of the translocation channel through which nascent peptides pass (117). These matches suggested that coelomocytes responding to immune challenge may upregulate protein production, including translation on both free ribosomes and those associated with the ER.

Two protein chaperones were identified: presenilin (Sp1228) and the ER-localized calcistorin/protein disulfide isomerase (ERcalcistorin/PDI; Sp1136). Presenilin is a member of a family of aspartyl proteases that are multipass transmembrane proteins found in rough ER that are involved in processing membrane proteins (e.g., signal peptide removal, amino acid proteolysis of loops extending out from the membrane, and cleaving transmembrane regions) (125, 123). ERcalcistorin/PDI is a putative calcium storage protein in sea urchin eggs that is also found in the ER (65). Its capacity to bind Ca2+ is also found in the ER (65). Its capacity to bind Ca2+ with consequent changes in the conformation of the protein, may be important for its role as a Ca2+-storage protein as well as a chaperone for proper folding of newly synthesized proteins in the ER (57). The identification of protein chaperones in coelomocytes was further evidence of the upregulation of the general cellular machinery for producing and processing proteins in response to immune challenge.

There were a number of matches to proteins involved in protein sorting and trafficking of transport vesicles, including Sec22 (Sp0057), p24A (Sp0565), p24B (Sp1246), transport protein (Sp1209), vacuolar protein sorting-4b (VPS4b; Sp0694), and Rab-7 (Sp0744). The Rab family of proteins...
consists of small GTPases that function as major regulators of intracellular vesicular traffic (121, 130). Sec22, a v-N-ethylmaleimide-sensitive factor attachment protein receptor (vSNARE) protein and a member of the p24 protein family, regulates the direction and targeting of transport vesicles traveling between the ER and the Golgi apparatus (11, 22, 41, 51, 62, 64, 90, 91). VPS4b and Rab-7, on the other hand, direct vesicular traffic between the cell surface and the endosomal system (12, 10, 89, 79, 112). Several matches were identified to proteins involved in the endosomal system, including Rab5-interacting protein (Sp1210), which functions in the fusion and maturation of endosomal elements (47, 86, 124); the mannose-

Fig. 5. Nucleotide alignment of 185/333 ESTs shows significant sequence variation. The highest quality and longest 185/333 ESTs (n = 290) were used in an alignment of sequences that included the 5'-UTR and the open reading frame encoding the leader and first element of the protein. The region shown includes the 3'-end of the leader (arrow) and the entire first element (see Fig. 4 for element locations). Duplicate sequences were reduced to two each for illustration purposes to show variations in nucleotide sequence. Periods (.) indicate conserved nucleotides, and sequence changes (for ESTs Sp0047 to Sp0405, relative to Sp0189) are shown as nucleotides. For example, all ESTs contain an adenine at position 1, except for Sp0562, which has a thymidine at that position.

Table 2. 185/333 Repeats

| Type 1 (red brackets) | MQMGSSGQDGPMGGRRFDGPDSGA |
| Type 2 (blue brackets) | NHTEGHQGH |
| Type 3 (green brackets) | NETGDHPH |
| Type 4 (purple brackets) | RHHSKTVGDDQD |

See brackets in Fig. 4A and numbered symbols in Fig. 4C.

Table 3. Comparison of 185/333 coding* sequences†

<table>
<thead>
<tr>
<th>Group No.‡</th>
<th>No. of Clones</th>
<th>Length, nt</th>
<th>No. of Different Sequences§</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>38</td>
<td>193</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>190</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>196</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>196</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>196</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>196</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>193</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>196</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>196</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>196</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>196</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>196</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>196</td>
<td>3</td>
</tr>
</tbody>
</table>

13 groups 267 clones 40 different sequences

*The leader and the first element (see Fig. 4). †Data excluded from this analysis were 6 groups with 4 or fewer members and 6 unique sequences. ‡Groups were defined based on clustering of the 5'-UTR sequences using PAUP (109). The 5'-UTR was not used in the analysis of the coding region. §The no. of different sequences in each group was calculated in WinClada (75) after subtracting the number of sequences that were different based on ambiguous bases.
6-phosphate receptor (Sp1197), which targets proteins for delivery to the endosomal system (31); and a subunit of the vacuolar H⁺-ATPase (Sp1094), which acidifies the contents of late endosomes and lysosomes (40, 60, 74, 107). Overall, the matches to proteins involved in regulating vesicle transport, trafficking, and targeting and in endosomal function suggested that coelomocytes responding to LPS increased the production of proteins that were processed through the ER, the Golgi apparatus and the endosomal system, thereby requiring increased activities in vesicular traffic.

Matches that suggested significant protein production and targeting to the endosomal system were noteworthy in light of expression of cathepsin B (Sp1108), cathepsin L (Sp0434 and others; EST052; Ref. 101), cathepsin S (EST118; Ref. 101), and arylsulfatase (EST003, EST004, EST072, EST401; Ref. 101), which are proteases that are typically directed toward and function in the endosomal system. Furthermore, the most prevalent ESTs identified in this study. 185/333, all have leader regions suggesting that they may also be processed through the ER/Golgi system and targeted to either the endosomal system or the plasma membrane. Increased secretory activity in response to LPS has been inferred from immunolocalization studies showing that two subsets of phagocytes synthesize and store the complement homolog, SpC3, in small cytoplasmic vesicles (34). After LPS challenge, significant increases in SpC3 occur in the coelomic fluid within 15 min postinjection (19, 102), indicating significant secretory activity. Overall, these data indicate that secretion and endosomal activity are significantly upregulated in responses to LPS.

**Cell motility and the cytoskeleton.** The majority of coelomocytes are macrophage-like, amoeboid, phagocytic cells that appear to be actively engaged in secretion and have extensive and malleable cytoskeletons (43, 44). Many of the coelocyte ESTs identified in this study matched to structural components of the cytoskeleton or to proteins involved in modifying cytoskeletal shape with consequent changes in cell behavior. These included actin (Sp1170 and others) and actin-binding proteins such as profilin (99, 100), cofilin (Sp1152), thymosin-β (Sp0042), and gelsolin (Sp1027). Protein tyrosine kinase-9 (Sp1044), which has a conserved ATP-binding region and a coflin-like domain, may function in depolymerizing actin microfilaments (9). Matches to the mouse mena protein (Sp0029) and the chicken avena protein (Sp1189) were also identified. These proteins are both proline-rich cytoplasmic proteins, and mena is known to bind profilin, perhaps promoting actin polymerization (30). The presence of mena and avena homologs in coelomocytes implies active changes in cytoskeletal shape and cellular motility, which has been suggested previously based on increased expression of sea urchin profilin (99, 100).

Matches to cell surface receptors included proteins that modulate the cytoskeleton and regulate links to the extracellular matrix. For example, liprin-a2 (Sp0350) is believed to be involved in cell-cell and cell-matrix interactions, based on the similarity of its extracellular domain to adhesion proteins (93), while integrins (integrin-βC, Sp0188; Refs. 13, 71) are localized to focal adhesions, where they are known to interact with the extracellular matrix and are important in cell migration (49, 50). Furthermore, integrins interact with β-catenin and G-cadherin, which have been identified in sea urchin embryos (69). Together, expression of this set of genes, which encodes cell surface and cytoskeletal proteins, suggested interactions between the cell and the extracellular environment, perhaps inducing active modulation of the cell shape and behavior in coelomocytes.

Signal transduction, which links cell surface events to intracellular pathways, underpins the responsiveness of coelomocytes to immune challenge. Several components that function in a range of signaling systems were identified in this study, including the receptor for activated protein kinase C (RACK1; Sp0165), which is a member of the G protein (β-subunit) superfamily of proteins and functions at signaling nodes (128). RACK1-mediated signaling is known to affect cytoskeletal rearrangements that manifest in cell adhesion, formation of lamellipodia, and cell migration (23). In coelomocytes, the RACK1 homolog may be a key regulator of chemotactic responses to LPS. Rho GTPase (Sp1096), a member of the Ras family of proteins, interacts with a multitude of proteins and is another key signal transducer involved in modulating the cytoskeleton (21, 37, 66, 72, 110, 111) in addition to mediating changes in transcriptional regulation, protein translation, proliferation, motility, apoptosis, and membrane trafficking (21, 116). The deactivation of Rho proteins is mediated by GTPase-activating proteins (GAPs) and GDP dissociation inhibitors (84), which has been identified previously (EST229; Ref. 101). The expression of Rho GTPase in activated coelomocytes may be involved in coordinating a variety of signaling pathways and cellular responses to immune challenge.

**Cell proliferation.** A finely tuned balance exists between cell proliferation and apoptosis in the vertebrate immune response, and this may hold true for sea urchins as well. In terms of proliferation, Sp1117 matched to polo-like kinase, a serine/threonine protein kinase known to participate in a number of activities associated with mitosis, including mitotic spindle organization and centrosome maturation. Furthermore, pololike kinase is known to interact with β-tubulin (Sp1089), microtubule-associated protein (Sp1195), and microtubule-stabilizing proteins including the translationally controlled tumor protein (Sp0139; Refs. 61, 127). In addition, Sp1188 matched to immediate early response protein-5, a nuclear protein that regulates cellular responses to mitogenic signals (122), while Sp1086 matched to allograft inflammatory factor-1, a cytoplasmic Ca²⁺-binding protein expressed in response to injury (7) whose overexpression has been linked to increased cell proliferation (5, 6). Sp1184 matched to a family of DNA-binding proteins with known helicase activity that includes a DNA-binding protein characterized from *Leishmania major* (120), a DEAD/box transcription factor (85) identified in several species of tunicates (29), and the cellular nucleic acid-binding protein, a conserved protein found in a variety of vertebrates (119). The presence and increased expression of helicases could be interpreted as an indicator of DNA replication leading to mitosis. On the other hand, proliferative responses may actually be due to decreased apoptosis. A match to Bax inhibitor-1 (Sp1158), which is conserved in animals, fungi and plants, suggested that the apoptotic pathway in coelomocytes may be inhibited (48). Overall, the increased expression of genes encoding proteins involved in proliferative and apoptotic pathways is in general agreement with prior studies indicating that numbers of coelomocytes in the coelomic fluid increase in response to injections of LPS (19).
In summary, the change from an immunoquiescent to an immunologically activated state significantly alters the transcriptional program in coelomocytes and invokes a wide range of cellular activities. Sea urchins appear to respond to immunological challenges by active cellular remodeling of their circulating coelomocytes, with commensurate transcriptional upregulation of genes functioning in RNA processing, protein synthesis, protein processing, and trafficking. Among the genes identified in this study that appear to be involved in trafficking, vesicle fusion, and protein sorting, many of the encoded proteins may be targeted to or function within the endosomes and lysosomes, indicating increased activity of these cellular compartments.

The most striking finding among the genes inferred to be upregulated in response to LPS was the identification of a set of closely related transcripts, designated 185/333, which were the most abundant single transcript species in immune-activated coelomocytes. The 185/333 transcripts exhibited high levels of sequence variability and putative alternative splicing of the mRNAs, resulting in encoded proteins with a wide range of size and sequence diversity. On the basis of the observed diversity and the striking increase in expression, as illustrated by the vast difference in the number of 185/333 clones in the two cDNA libraries and from Northern blots (78), a putative function of the encoded proteins is likely to include some aspect of immunity. The implication from the dn/ds ratio for the sequence at the 5'-end of the messages suggests that the genes may be under positive selection to diversify, and we speculate that it may be pathogen pressure that underlies this observation. It is not currently known how many 185/333 genes are present in the genome of an individual purple sea urchin. The data presented here give the impression that there may be hundreds; however, this may have been a result of the library structure, which used pooled coelomocytes from five sea urchins (see MATERIALS AND METHODS). Preliminary and ongoing efforts to quantitate the gene numbers have been based on a variety of approaches. Southern blots indicate that there are not hundreds of genes in a single genome; however, if the 185/333 genes are linked, interpretation of high-molecular-weight bands is difficult (unpublished observation). Comparisons among full-length cDNA sequences from four individual sea urchins responding to LPS have indicated that each animal generates between 11 and 30 different messages based on a combination of both element usage patterns and sequence diversity within elements (unpublished observation). Searches of the trace archives of genome sequences maintained at Caltech (http://issola.caltech.edu/~t/su-traces/blast.cgi) and a direct analysis of the sea urchin genome (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/) have indicated that there are 17 different sequences encoding part of the 5'-UTR plus the leader (unpublished observation). Together, these analyses have indicated that there may be as few as 8 and perhaps as many as 15 loci, and that the number of loci may vary among individual animals. Although our current estimates of gene numbers may be low, as implied from “mixing” between the 5'-UTR and the coding region sequences (Table 3), our preliminary results indicate that there are too few genes to explain the observed nucleotide variability in the ESTs. This suggests that there may be mechanisms for generating sequence diversity in the 185/333 transcripts that have not been previously characterized. On the basis of the similarities among the 185/333 cDNA sequences (and presumably among the genes), speculations on the mechanisms to generate diversity might include gene conversion, gene duplication, and RNA or DNA editing in addition to possible alternative splicing.

The unexpected level of diversity among the 185/333 sequences demonstrates a significantly greater level of complexity in the sea urchin immune system than had been imagined previously. Yet, there are other examples of nucleotide polymorphisms and significant sequence diversity in immune genes from other invertebrates. The sperm receptor located on egg surfaces of the tunicate Halocynthia roretzi, which mediates self sterility in these hermaphroditic animals, is highly variable and is speculated to function in allorecognition (88). Fibrinogen-related peptides (FREPs), which show significant diversity, are expressed by freshwater snail hemocytes responding specifically to the presence of parasitizing trematode worms in their hemocoel (45, 59, 131, 132). Current data suggest that FREP sequence diversity may be a result of both alternative splicing and gene conversion (63). In shrimp, the predominant transcript (6.5–21%) in hemocyte cDNA libraries was from a single family of antimicrobial peptides, the penaeidins, which exhibit significant sequence diversity, giving rise to multiple classes and isoforms within each class (25, 33). Different classes of penaeidins have differing effectiveness against specific microbial and fungal pathogens (24). Hyperdiversity has also been identified in an immunoglobulin-type variable region of an innate immune receptor in the cephalochordate Branchiostoma floridae, which may have chitin-binding activity (15, 16). The diversity shown in the innate immune responses of the sea urchin, snail, shrimp, and Branchiostoma responding to bacterial, parasitic, fungal, and viral challenges suggests that these animals, and perhaps most animals, may have hitherto unrecognized mechanisms to diversify their responses to foreignness. These mechanisms may either result in broad protection against pathogens or in directed expression of specific peptides to combat specific infecting microbes. The analysis of the sea urchin system promises to uncover mechanisms that generate diversity in the immune response, the results of which will contribute to a paradigm shift in our understanding of invertebrate immunity, as suggested by Flajnik and Du Pasquier (28). Identification of novel mechanisms for generating diversity in invertebrates, which has implications for innate immune capabilities in all animals, may result in a better understanding of innate immunity in higher vertebrates.

ACKNOWLEDGMENTS

We acknowledge and thank Dr. Eleanor Shepard, who assisted with the subtractions, and Cynthia Tallent, Sohum Mehta, and students in Cell Biology at George Washington University, who assisted with BLAST searches. Dr. Lakshmi Vasudevan helped to modify the Microsoft Access database program to serve our purposes. Drs. Gary Litman, David Rafos, Virginia Brockton, and Katherine Buckley provided constructive improvements to the manuscript. Present address of S. V. Nair: Dept. of Biological Sciences, Macquarie Univ., North Ryde, NSW, 2109, Australia.

GRANTS

This research was supported by funding from the National Science Foundation (MCB-0077970 to L. C. Smith and EPS-0083102 to P. S. Gross).


