Immunogenetics (1999) 49:773–786

ORIGINAL PAPER

Zeev Pancer · Jonathan P. Rast · Eric H. Davidson

Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes

Received: 1 September 1998 / Revised: 20 January 1999

Abstract Echinoderms share common ancestry with the chordates within the deuterostome clade. Molecular features that are shared between their immune systems and that of mammals thus illuminate the basal genetic framework on which these immune systems have been constructed during evolution. The immune effector cells of sea urchins are the coelomocytes, whose primary function is protection against invasive marine pathogens; here we identify six genes expressed in coelomocytes, homologues of which are also expressed in cells of the mammalian immune system. Three coelomocyte genes reported here encode transcription factors. These are an *NFKB* homologue (*SpNFKB*); a GATA-2/3 homologue (*SpGATAc*); and a runt domain factor (*SpRunt-1*). All three of these coelomocyte genes respond sharply to bacterial challenge: *SpNFKB* and *SpRunt-1* genes are rapidly up-regulated, while transcripts of *SpGATAc* factor disappear within hours of injection of bacteria. Sham injection also activates *SpNFKB* and *SpRunt*, though with slower kinetics, but does not affect *SpGATAc* levels. Another gene, *SpHS*, encodes a protein related to the signal transduction intermediate HS1 of lymphoid cells. Two other newly discovered genes, *SpSRCR1* and *SpSRCR5*, encode proteins featuring SRCR repeats. These genes are members of a complex family of SRCR genes all expressed specifically in coelomocytes. The SRCR repeats most closely resemble those of mammalian macrophage scavenger receptors. Remarkably, each individual sea urchin expresses a specific pattern of SRCR genes. Our results imply some shared immune functions and more generally, a shared regulatory architecture which underlies immune system gene expression in all deuterostomes. We conclude that the vertebrate immune system has evolved by inserting new genes into old gene regulatory networks dedicated to immunity.

Key words *NFKB* · GATA · Runt · HS1 · SRCR

Z. Pancer · J.P. Rast · E.H. Davidson (✉)
Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: davidson@mirsky.caltech.edu, Tel: +1-626-3954937, Fax: +1-626-7933047

Introduction

The immune system of sea urchins functions in many ways in a manner similar to the non-adaptive or innate immune system of mammals (Smith and Davidson 1992, 1994; Smith et al. 1996). Sea urchins are echinoderms, and virtually all modern molecular phylogenies place the echinoderms together with vertebrate and invertebrate chordates, and the hemichordates (a form of marine worm with gill slits and a dorsal neural tube) in a monophyletic clade within the animal kingdom, the deuterostomes (see Fig. 1). Deuterostome monophyly means that these phyla have a common ancestory and a common genetic heritage, and therefore share characters not present in other animal groups, for example arthropods or molluscs. It follows that if elements of genetic architecture can be identified which are utilized in both sea urchin and mammalian immune systems, this may reveal the nature of the basal genetic apparatus from which all deuterostome immune systems evolved.

The immune effector cells of sea urchins are the coelomocytes, free-wandering cells which populate the coelomic cavity. In *Strongylocentrotus purpuratus* there are $1\text{-}5 \times 10^6$ coelomocytes per ml of coelomic fluid, about two-thirds of which are phagocytic (Smith and Davidson 1992; Smith et al. 1995). Other types of coelomocytes are vibratile cells, and colorless and red spherule cells. Experimental observations demonstrate that coelomocytes carry out many different functions: these include formation of cellular clots, chemotactic accumulation at sites of injury, and allograft rejection; coelomocytes also clear bacteria and other foreign substances from the coelomic cavity with great effectiveness. Among injected microorganisms and substances

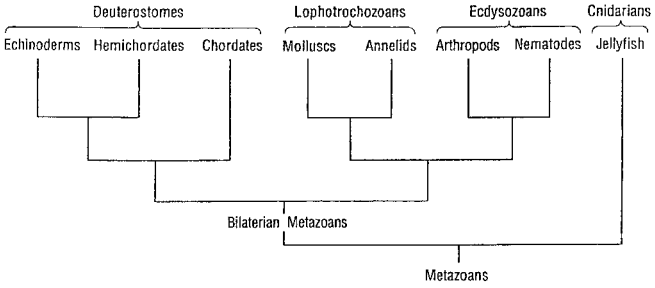

Fig. 1 A very simplified metazoan phylogeny, based on molecular phylogenies of rRNA and *Hox* genes. For references see Aguinaldo and co-workers (1997); Balavoine and Adoutte (1998); Balavoine (1997); Wada and Satoh (1994). Ecdysozoans (Aguinaldo et al. 1997) plus lophotrochozoans (Halanych 1995) equal the traditional protostomes, but the association of these groups within a monophyletic protostome clade does not at present appear to be supported as strongly as are all other features shown in this diagram

which coelomocytes have been reported to phagocytose or otherwise dispose of are bacteriophage, bacteria, yeast, mammalian red blood cells, carmine and carbon particles, and sephadex and latex beads (reviewed by Smith and Davidson 1994).

Coelomocytes respond sharply to injection of LPS, as well as of bacteria (Smith et al. 1995). A molecular response, identified earlier, is the enhanced transcription of profilin, a gene encoding a mediator of changes in cytoskeletal form (Smith et al. 1992, 1995). Coelomocytes also produce at least two proteins of the complement system, one homologous to Factor B (Smith et al. 1998) and the other a component which seems unambiguously a homologue of the vertebrate C3/C4/C5 complement proteins (Smith et al. 1996; Al-Sharif et al. 1998). Most importantly, sea urchin coelomocytes constitute an extremely effective immune system. Sea urchins obviously require such, for they are long lived animals that inhabit environments which sometimes bear heavy concentrations of fungal and bacterial pathogens. Though the maximum life span of *S. purpuratus*, the species we work with, is not known, we have maintained large numbers of these animals for many years in our long-term culture system (Leahy et al. 1978). In this context we have often observed their remarkable abilities to recover from severe and apparently global infections, as well as from infected wounds.

Here we report six different genes that are expressed specifically in coelomocytes of *S. purpuratus*. Each is homologous with a gene or genes also utilized in the immune systems of mammals, and each illuminates not only the functional character of the sea urchin immune system *per se*, but also the evolutionary basis of our own. Three of these genes encode transcription factors. We show that these genes constitute part of the bacterial response system of the coelomocytes. Another coelomocyte gene encodes a protein homologous with signal transduction intermediates of mammalian B and T cells, and two others encode proteins of the macrophage scavenger receptor family. The coelomocytes of each individual sea urchin express a different set of genes belonging to this family, so that each animal would appear to be generating a specific set of putative immune effector molecules.

Materials and methods

Sea urchins

Live specimens of *S. purpuratus* were maintained in the Sea Urchin Maintenance System (Leahy et al. 1978) at the Caltech Kerckhoff Marine Laboratory (Corona del Mar, Calif.).

Recombinant cDNA Libraries

Coelomocyte cDNA libraries representing coelomocyte mRNAs 24 h after bacterial challenge, and also resting coelomocytes, were constructed in the phage cloning vector Lambda ZAP Express (Stratagene, La Jolla, Calif.) according to the manufacturer's instructions. A plasmid bacterial challenged coelomocyte library built in pBK-CMV was also constructed and arrayed in 384-well plates using a Genetix Q-Bot robot. The colonies were printed in a high density 4×4 format, 18432 colonies/22×22 cm filter (Maier et al. 1994).

Bacterial challenge

Bacteria from the epidermis of adult *S. purpuratus* were isolated on Marine plates (Difco 2116). Four such bacterial colonies were grown individually in Marine Broth. For library construction logarithmic phase cultures were rinsed in 0.22 μm (Schleicher & Schuell, Keene, N.H.) filtered sea water (FSW), combined each at $OD_{600} = 0.1$ in filtered sea water and immersed in boiling water for 5 min. One milliliter per animal of the heat-killed bacteria was injected into the coelomic cavity, and coelomocytes were collected for RNA extraction 24 h thereafter. For bacterial induction live bacteria were combined at $OD_{600} = 0.5$ as above, injected alive, and coelomocytes were collected 6, 12, 18 and 24 h thereafter. Sham-injected animals were challenged with 1 ml FSW.

PCR cloning

Standard polymerase chain reaction (PCR) reaction mixtures of 50 μl included 10 pmoles of each primer, 200 μM of each nucleotide, buffer, and 2.5 units of Ampli*Taq* (Perkin Elmer). PCR amplifications were run on a GeneAmp 9600 thermal cycler (Perkin Elmer). Amplification products were purified through a Qiaquick Spin column (Qiagen, Chatsworth, Calif.), cloned into pGEM-T (Promega, Madison, Wis.) and sequenced.

RNA gel blot hybridizations

RNA was extracted from freshly collected coelomocytes, or from liquid nitrogen-pulverized sea urchin tissues, with RNAzol B (Leedo Medical Laboratories, Houston, Tex.). Analysis of transcription was performed in the following tissues and cells of the sea urchin: eggs, 9 h embryos, 15 h embryos, 23 h embryos, 48 h embryos, 72 h embryos, ♂ and ♀ gonads, gut, lantern muscle, and coelomocytes. Ten micrograms of total RNA was electrophoresed through a 1% formaldehyde/agarose gel and blotted onto Nytran membrane, following the instructions of the manufacturer (Schleicher & Schuell). Hybridization to ^{32}P-labeled probes was at 65 °C overnight in: 0.25 M NaH_2PO_4 (pH 7.2), 1 mM ethylenediaminetetraacetate (EDTA), 5% sodium dodecyl sul-

fate (SDS), and 0.5% bovine serum albumin. Washes were done at 65 °C as follows: twice in 0.125 M NaH_2PO_4 (pH 7.2), 2.5% SDS, 0.05 mM EDTA, and then twice in 0.025 M NaH_2PO_4 (pH 7.2), 0.5% SDS, 0.01 mM EDTA. Films were exposed for 1–7 days at –80 °C with an intensifying screen.

Sequence analysis

Cycle sequencing was performed with the ABI Prism Dye Terminator kit and a fluorescence DNA sequencer (ABI PRISM 373). Analysis of sequence composition, structure, and features was performed by the computer program DNASIS for Windows (Hitachi). Database searches and sequence retrieval were via the Internet servers at the European Bioinformatics Institute (BLITZ, FASTA): www.ebi.ac.uk, and the National Institutes of Health (BLAST): www.ncbi.nlm.nih.gov. Multiple alignment was performed with CLUSTAL W version 1.7 (Thompson et al. 1994), and the graphic presentation was composed with GeneDoc (Nicholas 1997). Minimum evolution phylogenetic trees were generated by the method of Rhetsky and Nei (1994) using the METREE program.

Cloning SpGATA-2/3

S. purpuratus GATA transcription factor sequences were initially isolated using a degenerate primer-PCR strategy that targeted conserved regions within the GATA Zn fingers. The primer sequences used in this analysis are as follows: 5' Gata1, CAGCGGCCGCTGYGTNAAYTG; Gata2, CAGCGGCCGCACNGG NCAYTA; Gata-YHKMN, CAGCGGCCGCTAYCAYAARATGA; and 3' Gata3, CAGTCGACRTANGCRTTRCA; Gata-CANCG, CAGTCGACCRTANGCR TTRTA [Y = C/T, R = A/G]. PCR was performed for 30 cycles at 94 °C for 1 min; 40 °C for 1 min; 72 °C for 1 min. A mixture of random primed and poly-T primed first-strand cDNA synthesized from polyA$^+$ RNA was used as template. The product resulting from the reverse transcription of 10 ng of mRNA by both methods was combined and employed in each 50 μl PCR. Both coelomocyte and 15 h embryo RNA was used as template. Bands of expected size were excised from the gel and eluted in 100 μl dH_2O. A second PCR was then performed using 1 μl of eluted template and the original primers. When possible nested or hemi-nested primer sets were also used. After separating on an agarose gel, bands of expected size were electroeluted and cloned into *Sal* I/*Not* I-restriction digested pBSII KS$^+$ and sequenced.

After confirmation of sequence, PCR fragments were random labeled and used to screen both arrayed plasmid cDNA libraries (15 h embryo, 20 h embryo, and unchallenged coelomocyte; 92160 clones each), λZAP 48 h embryo library (1×10^6 pfu) and the lambda ZAP Express coelomocyte library (4×10^5 pfu).

Cloning of SpNFKB

Degenerate PCR primers were used for PCR amplification from the ZAP Express bacteria-challenged coelomocyte cDNA library. Primer Rel.R4 [5′-TCCTTYTGIACCTTRTCRCA, (I = inosine), corresponds to the peptide CDKVQK, aa 273–278 in human SpNFκB P105 (P19838)]. This was used first as a single primer for 5 cycles (20 pmoles Rel.R4; 94 °C for 3 min; 94 °C for 30 s, 40 °C for 2 min, 72 °C for 3 min; 1 μl of the cDNA library – approximately 10^9 pfu), then the ZAP Express 5′-end vector-specific primer BK was added (10 pmoles). Amplification was continued for 30 additional cycles (94 °C for 1 min; 94 °C for 30 s, 55 °C for 2 min, 72 °C for 3 min). The purified products were then reamplified with Rel.R4 and Rel.F2 [5′-TTYMGITAYGARTGYGARGG (M = A/C)], corresponding to the peptide FRYECE, aa 62–68 in *D. melanogaster Dorsal* protein (A30350), 10 pmole each for 30 cycles, using the following parameters: 94 °C for 1 min; 94 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min. The expected 0.6 kilobase (kb) band was picked from the gel, reamplified, and cloned in pGEM-T [SpRel4:635 base pairs (bp), 211 aa; aligns with *SpNFKB* at nt 619–1255]. The degenerate PCR-insert SpRel4 (excised from pGEM-T by digest with *Spe* I and *Sph* I) was used to screen the cDNA library for full-length clones. Six positive phage clones were excised *in vivo* and sequenced; the lengths of these inserts were 4.5 and 6.5 kb. The short and long clones all had the same coding region, but the shorter ones had no polyadenylation consensus site. The 6.5 kb clone included a long 3' trailer sequence possessing a polyadenylation consensus site.

Cloning of coelomocyte SpHS

While attempting to clone sea urchin homologues of vertebrate colony-stimulating factor receptors by degenerate PCR, we observed that one cloned product recovered in pGEM-T, 541 bp in length, instead revealed similarity predominately to the vertebrate Src substrate Cortactin (SRC8) and the hematopoietic cell-specific Lyn substrate 1 (HS1). This insert (nt 904–1444 in *SpHS*) was used to probe the coelomocyte cDNA library. Five positive clones ranging from 3.5–4.5 kb were sequenced, two of them representing the complete cDNA.

Cloning of SpSRCR1 and SpSRCR5

Degenerate PCR primers were designed based upon invertebrate and vertebrate SRCR sequences, and used to amplify single SRCR domains from sea urchin coelomocyte first-strand cDNA. Products of amplification were obtained with the degenerate primers (SRCR.F1:5′-TGGGGIACIRTNTGYGA; SRCR.R2:5′-CAIACIACNCCNGCRTC) were cloned in pGEM-T (237 bp). Of six sequenced clones only two were identical. Equal amounts of PCR-amplified products from five of the unique PCR clones were used to screen coelomocyte cDNA libraries. We recovered a number of different types of *SRCR* genes, of which we report here only two, termed *SpSRCR1* and *SpSRCR5*. Six positive *SpSRCR5* clones were recovered from the arrayed libraries, of which the largest was 3.8 kb. This clone was sequenced. Of the approximately 3.8 kb cDNA we sequenced 2306 bp, including an open reading frame of 1583 bp. To clone the 5' end of *SpSRCR1*, PCR-amplification was performed from the cDNA library: Thus an *SpSRCR*–1 cDNA clone 5675 bp long was extended in the 5' direction as above, by PCR amplification from the cDNA library using a primer VWF.R1 5′-CGGTCTCCGCCATCTTGATC [positions 485–505 in the *SpSRCR1* sequence (114–133 in the cDNA clone)], and the Zap Express specific BK primer, followed by primer SR17.R1 5′-AAGCACCCAGCACGACAGA [positions 445–462 in the *SpSCRC1* sequence (73–91 in the cDNA clone)] and the T3 primer. The combined sequence is 6074 bp in length.

Results

Experimental approach

To activate coelomocytes, sea urchins were injected with mixtures of four different isolates of bacteria that had been cultured from their external surfaces. The bacteria were pooled at an OD_{600} of 0.5, and 1 ml of the mixture was injected per animal (see Experimental Procedures for details). cDNA libraries were constructed from coelomocytes activated by injection of 0.1 OD_{600} of the same mixture of bacteria which had been killed by a brief exposure to 100 °C, and from resting coelomocytes. Most of the clones considered below were initially identified by use of probes generated by means of

degenerate PCR, using conserved features of the respective mRNAs for primer design (see Experimental Procedures).

A coelomocyte GATA-2/3

Vertebrate genomes contain six genes encoding Zn finger transcription factors of the *GATA* family (for review see Weiss and Orkin 1995). The *GATA-4*, *-5*, and *-6* genes are expressed during endodermal and cardiac development. The *GATA–1, –2*, and *–3* genes constitute a subclass, all members of which are utilized for various aspects of hematopoiesis. Thus, *GATA-1* is essential for erythroid differentiation, and *GATA-2* is required for normal development of all hematopoietic lineages (Weiss and Orkin 1995). *GATA-3* is specifically required for T-cell development (Ting et al. 1996). *GATA* target sites occur widely in the *cis*–regulatory elements of erythroid-specific genes (Evans and Felsenfeld 1989; Tsai et al. 1989); in the *cis*–regulatory elements of several lymphoid genes, including the *TCRA*, *TCRB*, *TCRD*, and *CD8A* genes of T cells (reviewed in Arnone and Davidson 1997; Weiss and Orkin 1995), as well as in various mast cell and megakaryocyte genes (Weiss and Orkin 1995).

PCR fragments representing *GATA* mRNAs were amplified in reactions of degenerate primers (see Experimental Procedures) with 15 h (late cleavage stage) sea urchin embryo and coelomocyte cDNAs, and two different classes of *GATA* cDNA clones were thereupon isolated from cDNA libraries. *SpGATAe* clones encode a factor apparently utilized in embryogenesis, as these clones were found in 15 h, 20 h (blastula), and 48 h (late gastrula) embryo cDNA libraries, but not in any of the coelomocyte libraries. Clones representing a coelomocyte *GATA* factor, which we have termed *SpGATAc*, were found in both coelomocyte and 48 h embryo libraries. The sequence of the 431 amino acid protein encoded by *SpGATAc* cDNA is shown in Fig. 2A. When searched against GenBank using BLAST, highest matches for the SpGATAc amino acid sequence are confined to vertebrate *GATA-2* and *-3* and the related *Drosophila* sequence, *dGATAc* (Lin et al. 1995; this *Drosophila* gene is expressed in many regions of the developing embryo, and mutations in it are embryonic lethals, but its specific functions are unknown). Matches between *SpGATAc* and the other members of the family are mainly restricted to the Zn finger and the flanking basic region. The similarity of *SpGATAc* to vertebrate *GATA-2* and *-3* sequences is corroborated in the phylogenetic analysis shown in Fig. 2B. Here it can be seen that, in contrast, *SpGATAe* falls within a group of *GATA* sequences that also includes those of the vertebrate *GATA-4*, *-5*, and *-6* factors; the *Drosophila pannier* sequence (this gene is utilized in neurogenesis as well as for other developmental functions; Ramain et al. 1993). The *Drosophila serpent* sequence also falls into the *GATA-4, -5,* and *-6*

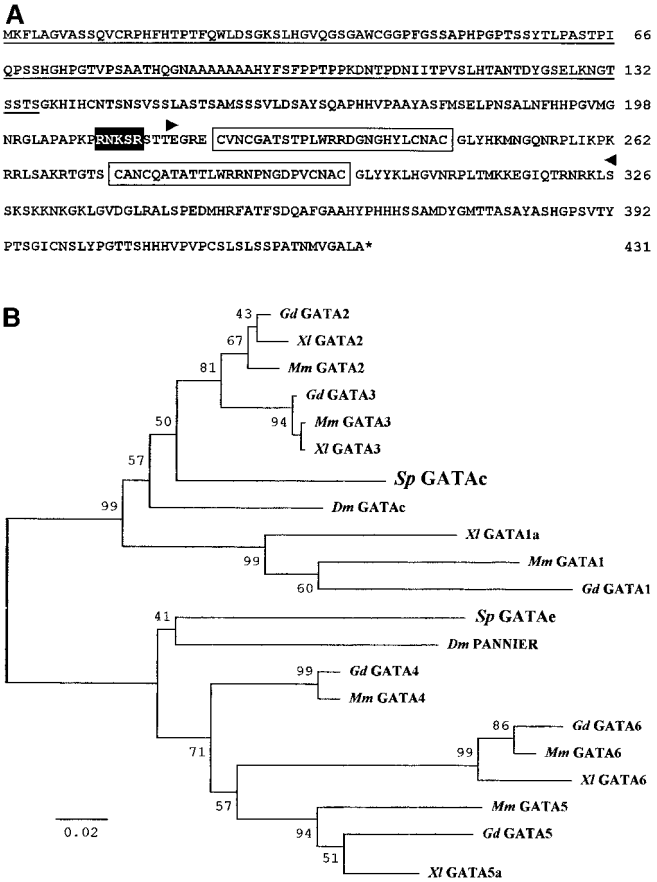

Fig. 2A,B Sequence and relationship of *SpGATAc* (GenBank accession No. AF077674). **A** Protein sequence inferred from DNA sequence of *SpGATAc* cDNA clones. Zinc finger regions are *boxed*. A basic region unique to the vertebrate GATA2/3 subgroup is shown in *reverse type*. This sequence mediates recognition of a particular target site sequence by the N-terminal Zn finger (see text). The probe used for northern blotting corresponds to the region *underlined*. The region that was used in the phylogenetic analysis shown in **B** is designated by *arrows*. **B** Minimum evolution tree (Rzhetsky and Nei 1994) for *Drosophila* and deuterostome GATA transcription factor sequences that contain two Zn fingers. *Sp*GATAe (AF077675); *Mm*GATA1 (X15763); *Mm*GATA2 (AB000096); *Mm*GATA3 (X55123); *Mm*GATA4 (M98339); *Mm*GATA5 (U84725); *Mm*GATA6 (S82462); *Gd*GATA1 (M26209); *Gd*GATA2 (X56930); *Gd*GATA3 (X56931); *Gd*GATA4 (U11887); *Gd*GATA5 (U11888); *Gd*GATA6 (U11889); *Xl*GATA1a (M76566); *Xl*GATA2 (M76564); *Xl*GATA3 (M76565), *Xl*GATA5 A (L13701); *Xl*GATA6 (Y08865); *Dm*GATAc (D50542); *Dm*PANNIER (S68698). Distances are calculated as a proportion of amino acid differences. Positions containing gaps were excluded from all pairwise distance estimates. Two statistically similar minimum evolution trees were found, both of which are identical with respect to the placement of the sea urchin *GATA* genes. *Numbers* on the *left* of the figure indicate the probability that the branch defining the cluster is greater than 0 in length. Species designations are: *Mm*, mouse; *Gd*, chicken; *Xl*, *Xenopus laevis*; *Dm, Drosophila melanogaster*, *Sp*, *Strongylocentrotus purpuratus*. *Scale bar* indicates amino acid substitutions per site

class (this gene was excluded from this phylogenetic analysis because it bears only one Zn finger). *Drosophila serpent* is required for gut, fat body, and hemocyte development (Rehorn et al. 1996). A key feature linking *SpGATAc* to vertebrate GATA-2 and -3 lies in the basic region which flanks the N-terminal Zn finger (see Fig. 2A). This region specifies recognition of *cis*-regulatory target sites which include the sequence AGATCT, an exclusive feature of the vertebrate GATA-2 and -3 proteins (Pedone et al. 1997). The C-terminal Zn finger in all GATA factors is thought to mediate binding to the classic consensus target sequence for GATA factors, WGATAR.

In summary, we discovered two *S. purpuratus GATA* transcription factors. One of these, *SpGATAe*, is expressed in the embryo but not in coelomocytes, and this sequence exhibits sequence affinity with a canonical and perhaps pan-metazoan subclass of *GATA* regulators of gut and cardiac development (Fig. 2B). The other, *SpGATAc*, is expressed in coelomocytes (as well as late embryos) and its sequence falls within the subfamily of *GATA* sequences which in vertebrates serve as essential regulators of hematopoietic gene expression, in erythroid, myeloid, and lymphoid lineages.

An S. purpuratus NFKB homologue expressed in coelomocytes

NFKB/Rel transcription factors regulate immune functions and developmental processes in both insects and vertebrates (Baldwin 1996; Ghosh et al. 1998; Hoffman et al. 1996; Ip et al. 1993; Kanegae et al. 1998). Like the role of GATA factors in endoderm specification, the role of *NFKB/Rel* factors in promoting transcription of protective proteins in response to immune challenge may be a function that is universal in bilaterian organisms (though there is no evidence from any lophotrochozoan; see Fig. 1). In vertebrates, *NFKB/Rel* factors regulate transcription of inflammatory acute phase response genes that are expressed in the liver following bacterial or viral infection, and also mediate expression of cytokine genes in macrophages, and of other myeloid-specific genes. They also serve as key regulators of gene expression in the adaptive immune system. Thus they are required for transcription of both immunoglobulin light chain genes and *TCR* genes, as well as of other genes such as those encoding T-cell cytokines (Baldwin 1996; Ghosh et al. 1998; Verma et al. 1995). Factors of this class provide a canonical example of immediate-early response mechanisms by which the genetic apparatus responds to change in environmental circumstances.

All members of the *NFKB/Rel* family possess a 300 residue amino-terminal Rel homology domain that includes the DNA-binding and dimerization elements of the protein and the nuclear localization site. Based on conserved features of this domain that have been found in both insect and vertebrate sequences, degenerate PCR primers were constructed, and a Rel domain sequence of 635 bp was amplified from a cDNA library made from coelomocytes extracted after bacterial challenge. The PCR fragment was used to screen the cDNA library and a 4552 bp long sea urchin *NFKB* cDNA (termed *SpNFKB*) was isolated. This clone has a 3375 bp open reading frame encoding an 1125 amino acid polypeptide, with a predicted MW of 124 000 M_r. The sequence is compared with those of human *NFK*B P105 (Ghosh et al. 1990) and *Drosophila* Relish (Dushay et al. 1996) in the multiple alignment shown in Fig. 3A. The *S. purpuratus* sequence shares with these a well-conserved Rel homology domain, the nuclear localization site, a glycine-rich domain possibly required for protein processing (Ghosh et al. 1990), a C-terminal region featuring six ankyrin repeats, and a PEST domain (Verma et al. 1995). *SpNFKB* also includes a glutamine-rich domain similar to that found in *Drosophila* dorsal protein.

A minimum evolution phylogenetic tree that compares the Rel domains of both arthropod and vertebrate factors is shown in Fig. 3B. The deuterostome (i.e., sea urchin and vertebrate) sequences cluster together in this tree. Structurally the Relish sequence is more closely related to P105 and *SpNFKB* outside of the Rel domain, i.e., in the ankyrin repeat domain that it shares with these proteins, as shown in Fig. 3A. Within the deuterostome cluster the analysis shows that *SpNFKB* shares a common ancestor with, i.e., is homologous with, the amniote P105 and P65 transcription factors.

RNA gel blot data (not shown) indicate qualitatively that *SpNFKB* is expressed strongly in coelomocytes, but at a level below detection or not at all, in gut, testes, ovary, and lantern muscle. Transcripts appear in 20 h embryo mRNA, though not earlier, but disappear later in embryogenesis.

SpRunt-1 in coelomocytes

Runt domain transcription factors also perform numerous functions in the mammalian immune system (for reviews see Kagoshima et al. 1993; Tenen et al. 1997). These factors consist of an α subunit that contains the runt DNA binding domain complexed with a non-DNA-binding β subunit. Mammalian runt domain proteins include human factors AML-1, AML-2, and AML-3, and their mouse counterparts PEBP2αB, CBFA3, and PEBP2αA. These proteins regulate transcription of differentiation products in both lymphoid and myeloid cells. For example, their target sites occur in the *cis*–regulatory regions of genes such as those encoding granzyme B in cytotoxic T cells (Wargnier et al. 1995), myeloperoxidase in immature myeloid cells (Britos-Bray et al. 1997), the macrophage CSF receptor (Zhang et al. 1996), and granulocyte-macrophage-colony stimulating factor in T cells (Cockerill et al. 1996).

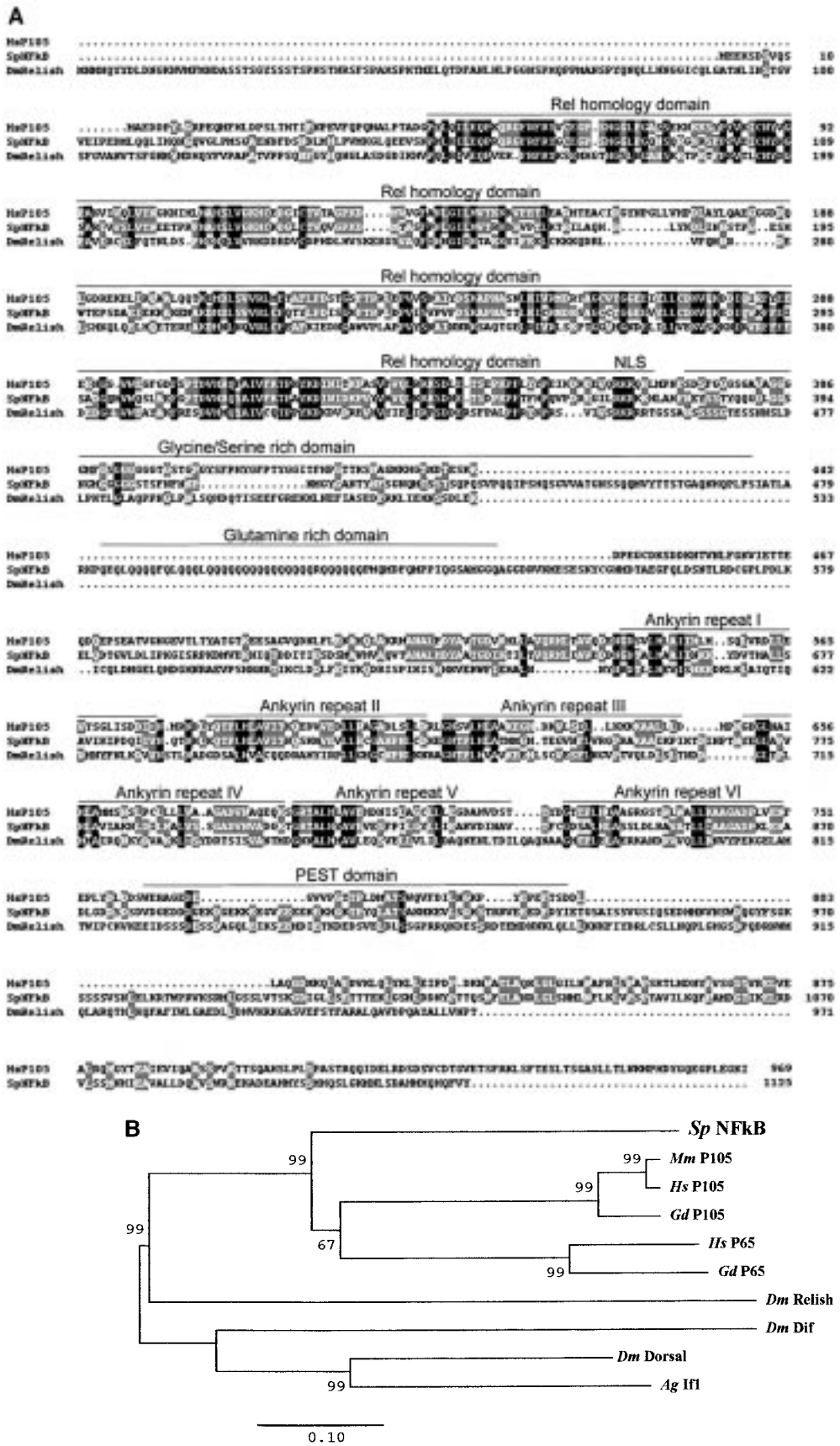

Fig 3A,B Relationship of *SpNFKB* (AF064258) sequence with other Rel domain sequences. **A** Multiple alignment of *SpNFKB*, with human nuclear factor P105 (*HsP105*, P19838) and *Drosophila* Relish (*DmRelish*, gi:1621609). Residues that are identical in all three proteins are shown in *inverted type*, and residues identical among two of the proteins are in *gray shaded type*. Protein domains noted in text are indicated *above* the sequence of HsP105. These are the Rel homology domain, the nuclear localization signal [NLS], a glycine/serine-rich domain, a unique *SpNFKB* glutamine-rich domain, six ankyrin repeats, and the PEST domain. **B** Minimum evolution tree relating Rel domains of *SpNFKB* with those of mouse P105 (*MmP105*, A35697), human P105 (*HsP105*), chicken P65 (GdP65, P98152), *Drosophila* Relish (*DmRelish*), *Drosophila* Dif (Dorsal-Related Immunity Factor) (*DmDif*, P98149), *Drosophila* embryonic polarity Dorsal protein (DmDorsal, A30350), and mosquito (*Anopheles gambiae*) *Ifl* (immune factor 1) (*AgIfl*, X95912). Distances were calculated as proportion of difference. A single shortest tree was obtained. *Scale bar* indicates amino acid substitutions per site. *Numbers* to the *left* indicate the probability that the branch defining clusters is greater than 0 in length

They are also involved in the transcriptional control of expression of TCRs (Giese et al. 1993; Leiden 1993).

An *S. purpuratus* runt domain transcription factor (*SpRunt-1*) has been isolated (Coffman et al. 1996) as a regulator of gene expression during embryogenesis. When applied to a cDNA library made from resting coelomocytes, the *SpRunt-1* probe revealed several clones, the sequences of which turned out to be identical to that of *SpRunt-1*. As we now show, *SpRunt-1*, like *SpNFKB* and *SpGATAc*, is expressed prominently in coelomocytes, in a way that clearly demonstrates a function in the immune responses mounted by these cells.

Responses of SpGATAc, SpNFKB and SpRunt-1 genes to bacterial challenge

The level of expression of all three of the coelomocyte regulatory genes described above change dramatically in response to immune challenge, and this is illustrated in Fig. 4. Groups of several *S. purpuratus* were either injected with bacteria as above, or sham injected (i.e., with sea water), and at the indicated intervals the animals were killed and their coelomocytes were recovered as completely as possible and pooled within each group. RNA was extracted and equal amounts loaded within each lane, as indicated by the hybridization of a probe for the general coelomocyte marker, *SpThymosin*. The control RNA shown in the first lane was extracted from coelomocytes of undisturbed animals. At the earliest time point shown, 6 h postinjection, transcripts of both *SpNFKB* and *SpRunt-1* are very sharply elevated in response to the bacteria, relative to the control. The kinetics according to which the *SpRunt-1* response decays may be slower than those of *SpNFKB*. More detailed kinetic experiments (to be presented elsewhere) show that the *SpNFKB* response actually peaks within about 2 h of bacterial injection, thus differing to a greater extent from the result of sham injection than is evident in Fig. 4. The sham injection response is much slower for both mRNAs, and is clearly less intense as well for *SpRunt-1*, compared with the response on injection of bacteria. We do not yet know whether the sham injection response is due to the inadvertent introduction of a few surface bacteria or to the injury caused by the needle prick. Nor is it clear to what extent injury responses, or a non-specific response to foreign particles, accounts for the transcriptional activation seen on injection of bacteria.

Though the data in Fig. 4 are strictly qualitative, they clearly show that whatever the actual signal(s) affecting the cells, immune challenge causes a large increase in *SpNFKB* and *SpRunt-1* transcript representation per average coelomocyte. That is, the amount of *SpNFKB* mRNA increases at least 10-fold within 6 h of introduction of the bacteria and, compared with the control, that of *SpRunt-1* increases to a much greater extent; in fact the transcript is not detectable at all in resting coelomocyte RNA under these conditions. These effects are probably due to a sharp increase in the rate of transcription of these two genes.

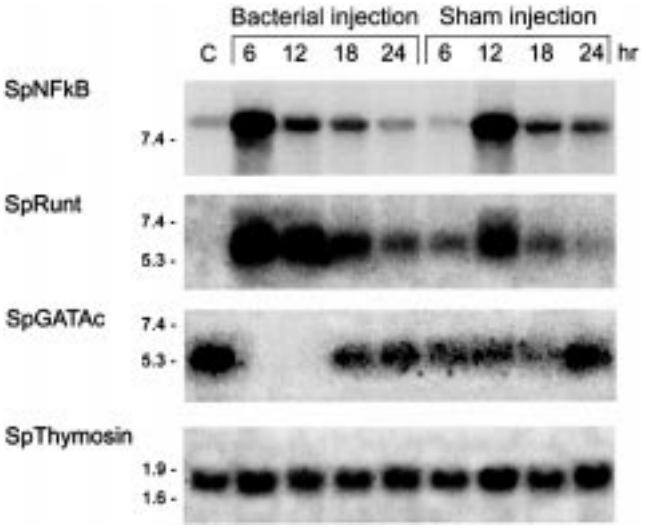

Fig. 4 Inductive change in coelomocyte transcription factor mRNA levels following bacterial challenge. One ml of either sea water alone (sham injection) or of sea water containing $OD_{600} = 0.5$ of live bacteria consisting of a pool of four different isolates that had been cultured from the external skin of *S. purpuratus* were injected into the coelom of each animal, and RNA was extracted (see Experimental Procedures). Coelomocytes from three animals were pooled for RNA extraction for control (C) and 24 h sham-injected samples, and from five animals for each of the other samples. Approximately 10 μg of total RNA was loaded in each lane of a gel, and after electrophoresis blotted to a membrane filter and hybridized. The same blot was reacted in turn with each of the four probes indicated on the *left*. The results shown are completely reproducible and are representative of a number of different experiments. *Numerals* indicate the positions of size markers in kb. Thymosin mRNA serves as a loading control, as this mRNA is present at about the same relatively high prevalence in all coelomocytes, irrespective of their state of stimulation (*SpThymosin* AF076515). Thymosin is an actin-binding peptide which inhibits actin polymerization (Safer and Nachmias 1994). Probes were the PCR insert SpRel4 (see Experimental Procedures) for *SpNFKB*, a 408 bp region beginning at the translation start site for *SpGATA2/1*, and for *SpRunt-1*, a 657 bp fragment beginning at the start of translation and including the runt domain (Coffman et al. 1996). Thymosin, PCR fragment from the 5' end, bp 1-408 in *SpThymosin*

The effect of immune challenge on *SpGATAc* mRNA levels is exactly the converse. SpGATAc message is moderately prevalent in resting or control coelomocytes, unlike *SpNFKB* or *SpRunt-1*, but when bacteria are introduced it essentially disappears. By 18 h the system has nearly recovered, and *SpGATAc* mRNA is again observed. This effect is unlikely to be due to wholesale changes in coelomocyte cell populations, since dividing coelomocytes are rarely observed in adult animals. In contrast to *SpNFKB* and *SpRunt-1* mRNA levels, *SpGATAc* mRNA levels are barely affected by sham injection.

The results shown in Fig. 4 provide a very strong argument that all three of these transcription factors serve as components of the coelomocyte immune re-

sponse system. *SpNFKB* and *SpRunt-1* are likely to function as gene activators, and SpGATAc may act as a repressor of some response genes. The removal of this factor could constitute an element of the mechanism by which the response is triggered at the transcriptional level.

An S. purpuratus coelomocyte homologue of HS-1, a mammalian lymphoid signal transduction intermediate

Figure 5 displays a multiple alignment of a sequence, named *SpHS*, translated from another clone recovered from our coelomocyte cDNA libraries. This sequence clearly belongs to the family defined by mammalian cortactin and *HS-1* (*LckBP-1*), which are intracellular signal transduction intermediates characterized by an src homology (SH3) domain and a particular N-terminal repeat, as indicated in Fig. 5. Both *cortactin* and *HS-1* participate in signal transduction downstream of extracellular ligand-receptor interactions. *HS-1* mediates signaling from IgM B-cell antigen receptors (Yamanashi et al. 1993). In T cells, *Lck* tyrosine kinase interacts specifically with a proline-rich domain of *HS1*, and within 5 min of TCR stimulation, *HS1* becomes tyrosine phosphorylated (Takemoto et al. 1995). *Cortactin*, which is a potent F-actin (filamentous) binding protein, mediates receptor signaling in response to growth factors, and also in response to cell adhesion mediated by integrin, or in response to bacteria (reviewed by

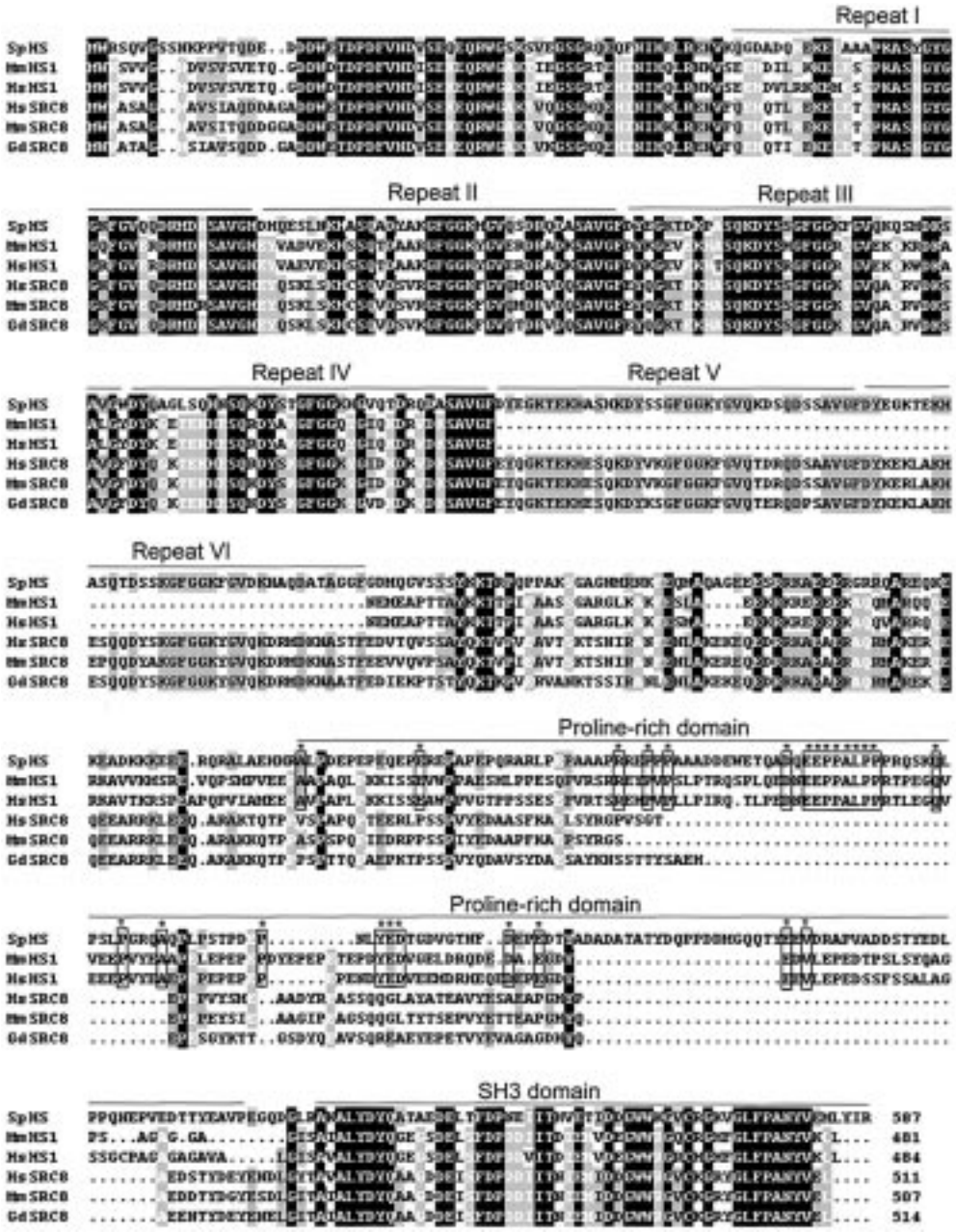

Fig. 5 Multiple alignment of *SpHS* with amniote members of the HS1/cortactin family. The *SpHS* sequence (AF064260) derives from cDNA clones representing a 4.5 kb coelomocyte mRNA, of which 2214 bp were sequenced. The open reading frame predicts a 587 amino acid protein, of MW 65000 M_r. The other sequences are human hematopoietic cell-specific Lyn substrate 1 (*HsHS1*, P14317) and mouse HS1 (*MmHS1*, P49710); human Src substrate cortactin or SRC8 (*HsSRC8*, Q14247); mouse cortactin (*MmSRC8*, Q60598); and chicken cortactin (*GdSRC8*, Q01406). Residues conserved among all sequences are shown in *inverted type*, residues conserved in five of the sequences are in *inverted type* on *light gray*, and conserved residues among four sequences are *shaded*. Domains are as have been delineated for the mammalian proteins (see text for references). Sequence motifs in the proline-rich region uniquely shared between *SpHS* and the mammalian *HS1* sequences are *boxed* and marked by an *asterisk*

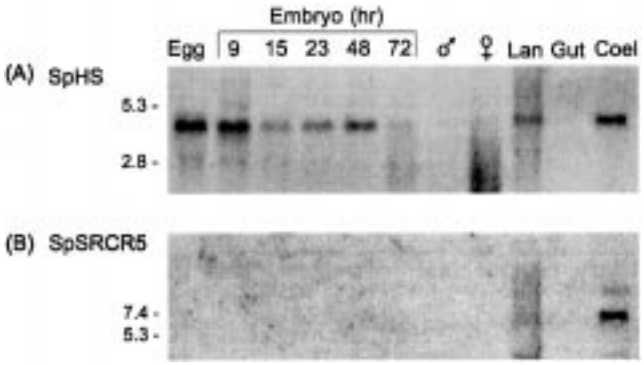

Fig. 6 Embryonic and coelomocyte expression of SpHS, and coelomocyte-specific expression of SpSRCR5 (AF076514). Gel blot hybridizations of probes representing these cDNAs are shown to RNA extracted from coelomocytes, gut, testes, ovary, and lantern muscle, as well as embryos (see Experimental Procedures for details). Probes were: SpSRCR5, a 459 bp *Cla* I/*Pst* I fragment (nt 927–1386); SpHS, a 461 bp PCR fragment corresponding to nucleotides 1154–1615

Zhan et al. 1997). *HS-1* is expressed in all hematopoietic stem cells as an early differentiation marker, as well as in mature lymphoid cells (Kitamura et al. 1989). In contrast, cortactin is expressed in many different organs and tissues in mouse (Miglarese et al. 1994), but not in any hematopoietic cells except megakaryocytes (Zhan et al. 1997). We believe the sequence shown in Fig. 5 is functionally related to *HS-1* rather than to cortactin because it shares with human and mouse *HS-1* a proline-rich domain, which is entirely lacking in the cortactins (25 amino acid identities between the sea urchin protein and mammalian *HS-1*, which are absent in the cortactins are boxed in Fig. 5, including the perfectly conserved motif EEPPALPP). An additional argument that functionally relates the sea urchin protein to *HS-1* rather than to the cortactins is its tissue distribution. Figure 6 A shows that *HS-1* is expressed in embryos and lantern tissues but not in gut, in addition to its strong expression in coelomocytes. We conclude that *SpHS* probably shares a common ancestor with mammalian *HS-1*, and like *HS-1* may function in signal transduction processes specific to the immune response, as well as in some other tissues.

S. purpuratus coelomocyte homologues of the scavenger receptor cysteine-rich repeat family

Scavenger receptor cysteine-rich (SRCR) domains define a superfamily of cell-surface or secreted proteins that presently include two invertebrate and several vertebrate members (reviewed by Pancer et al. 1997; Resnick et al. 1994). The SRCR domain consists of a 110 aa residue motif with a conserved spacing of either six ("Group A" SRCRs; Resnick et al. 1994) or eight cysteines ("Group B" SRCRs) that are thought to form intradomain disulfide bonds. SRCR domains of the sea urchin proteins that we describe in this section are of the Group A type, displaying six diagnostic cysteines in register with the canonical Group A SRCR domains (see Fig. 7A). Furthermore, the sea urchin SRCR1 repeats possess an intron within each SRCR domain (not shown), which is a character of Group A SRCR repeats, while Group B SRCR repeats lack domain introns (Aruffo et al. 1997). The identification of *SpSRCR1* and *SpSRCR5* as Group A homologues rests on these particular features, since in terms of overall sequence similarity it is impossible to class the sea urchin SRCR repeats in one as opposed to the other of these groups. Mammalian Group A SRCR proteins are expressed in myeloid cells. These proteins include a macrophage bacteria binding receptor (MARCO; Eloman et al. 1995); a macrophage scavenger receptor which binds diverse ligands including modified LDL, bacterial endotoxins, and which functions during endocytosis (MRSE; Freeman et al. 1990); the macrophage lectin binding molecules cyclophilin-C-binding protein (Friedman et al. 1993) and the MAC2 binding protein (Koths et al. 1993); and Complement Factor I (Goldberger et al. 1987). Group A SRCR domains are evidently of remote evolutionary origin; cell surface and secreted forms were recently recovered from a marine sponge (Pancer et al. 1997), and they are also utilized in the sea urchin sperm receptor for the egg jelly peptide, SPERACT (Dangott et al. 1989). In Fig. 7A SRCR domains of all of these proteins are aligned with the SRCR domains from *S. purpuratus* SRCR proteins that we are reporting in this work. The sea urchin SRCR sequences align more closely with the mammalian myeloid SRCRs than with the sponge SRCR or the speract receptor SRCR.

Coelomocyte SRCR cDNA clones were isolated using degenerate PCR probes. A number of different genes encoding SRCR domains are expressed in coelomocytes, and are the subject of ongoing investigations in our laboratory. Here we report on two such genes, encoding the proteins SpSRCR1 and SpSRCR5. A schematic diagram of the various elements of which these proteins are composed is shown in Fig. 7B. Except for the SRCR domains the two proteins are dissimilar. Both proteins begin with signal peptides. In addition, SpSRCR1 includes four SRCR domains, a domain similar to an extracellular matrix protein of sea urchins, and four von Willebrand factor domains. SpSRCR5 has two SRCR domains, following a domain which displays no strong similarity to any other known protein.

The RNA gel blot shown in Fig. 6B indicates that the *SpSRCR5* genes are indeed coelomocyte-specific genes, and the same is true of the *SpSRCR1* genes (not shown). Probes for *SRCR1* and *SRCR5* each cross react with additional *SRCR* mRNAs, some of which have been cloned. The SRCR proteins are expressed constitutively and at fairly high levels in coelomocytes whether they are resting or stimulated. However, it is striking that coelomocytes from different individual sea urchins express them to very different extents. Thus for exam-

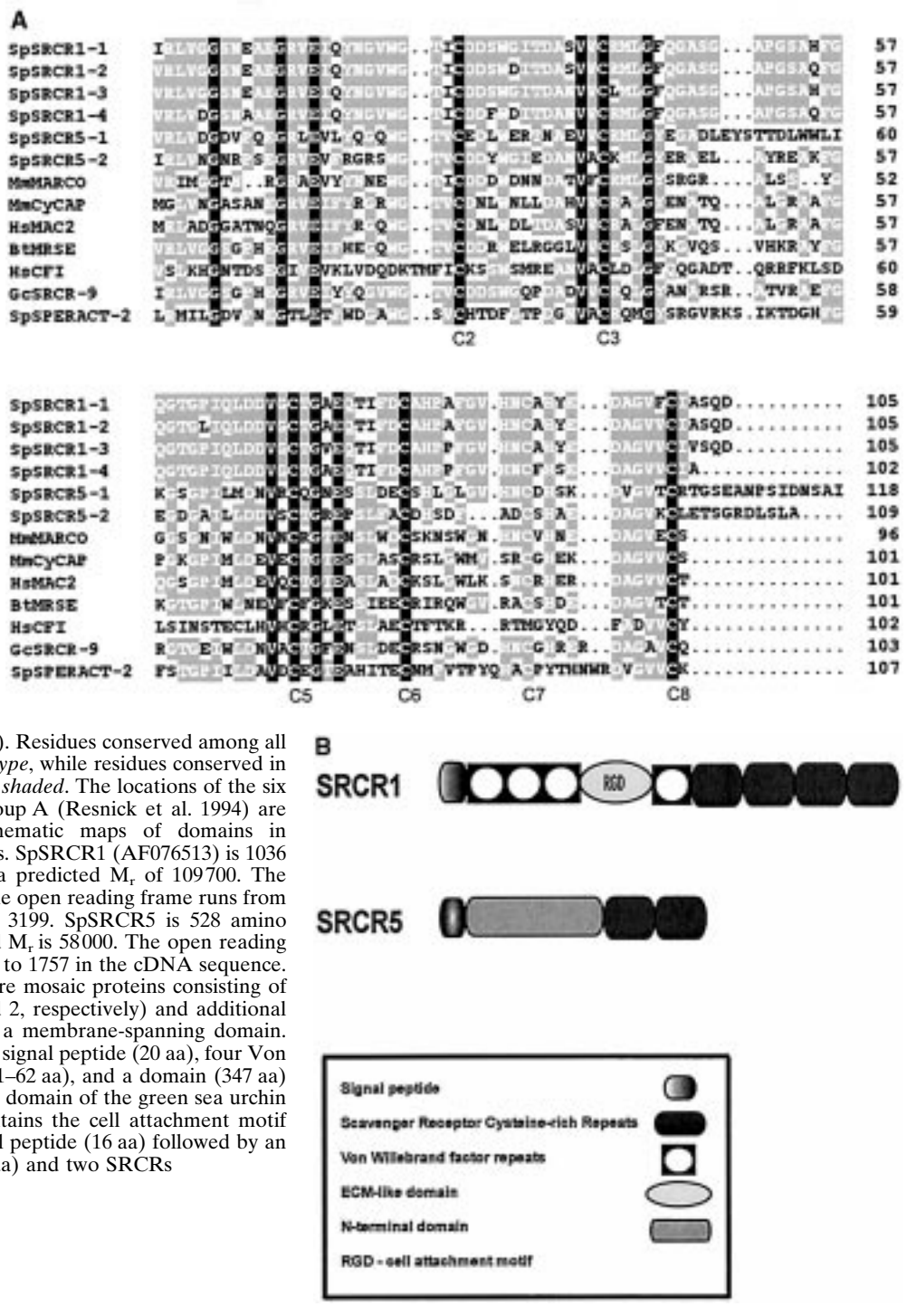

Fig. 7A,B Structure and relationships of *S. purpuratus* SRCR proteins. **A** Multiple alignment of SRCR domains from SpSRCR1 and SpSRCR5 with "Group A" myeloid SRCR domains. Repeats 1–4 of SpSRC1 and repeats 1, 2 of SpSRCR5 are included in the comparison. The four SpSRCR1 repeats are very similar to one another, while the two SRCR domains of SpSRCR5 are about 41% similar. Other proteins in the multiple alignment are: MmMARCO (murine macrophage bacteria-binding receptor, single SRCR, aa 423–1036; X99321); BtMRSE (bovine macrophage type I scavenger receptor, single SRCR, aa 351–452; P21758); MmCyCAP (murine cyclophilin–C-binding protein, single SRCR, aa 24–125; A48231); HsMAC2 (human MAC2-binding protein, single SRCR, aa 24–124; A47161); SpSPERACT (sea urchin sperm egg-peptide receptor, 2nd repeat, aa 153–257; P16264); HsCFI (human complement factor I, single SRCR, aa 114–215; P05156). Residues conserved among all sequences are shown in *inverted type*, while residues conserved in at least four of the sequences are *shaded*. The locations of the six Cys residues characteristic of group A (Resnick et al. 1994) are *marked* and *numbered*. **B** Schematic maps of domains in SpSRCR1 and SpSRCR5 proteins. SpSRCR1 (AF076513) is 1036 amino acids in length, and has a predicted M_r of 109700. The cDNA extends for 6074 bp and the open reading frame runs from positions 89 in this sequence to 3199. SpSRCR5 is 528 amino acids in length, and the predicted M_r is 58000. The open reading frame extends from positions 174 to 1757 in the cDNA sequence. Both SpSRCR1 and SpSRCR5 are mosaic proteins consisting of C-terminal SRCR repeats (4 and 2, respectively) and additional unique segments, and both lack a membrane-spanning domain. SpSRCR1 features an N-terminal signal peptide (20 aa), four Von Willebrand factor-like repeats (61–62 aa), and a domain (347 aa) similar to the extracellular matrix domain of the green sea urchin *Lytechinus variegatus*, which contains the cell attachment motif RGD. SpSRCR5 features a signal peptide (16 aa) followed by an as yet undescribed domain (285 aa) and two SRCRs

ple, as shown in Fig. 8, individuals 2 and 7 express RNAs detected by the *SpSRCR1* probe at relatively high levels, while individuals 4, 12, and 15 express these genes hardly at all. However, the pattern of expression of the *SpSRCR5* gene is different from that of the *SpSRCR1* genes. Such individual-specific distinctions in the expression of genes that may be involved in immune functions are of course of immediate interest. This phenomenon could reflect differences in the number of cells of the subpopulations expressing these proteins; differences in the activity of the transcriptional regulators for the respective genes; differences in mRNA turnover rate; or polymorphic differences in the genomic structures of the genes that affect transcription or RNA turnover. Other experiments (not shown) confirm that given *SRCR* genes are present in the genomes of *S. purpuratus* regardless of whether they are represented in coelomocyte mRNA. Whether differences in level of expression represent genetic or physiological distinctions among these individuals remains to be

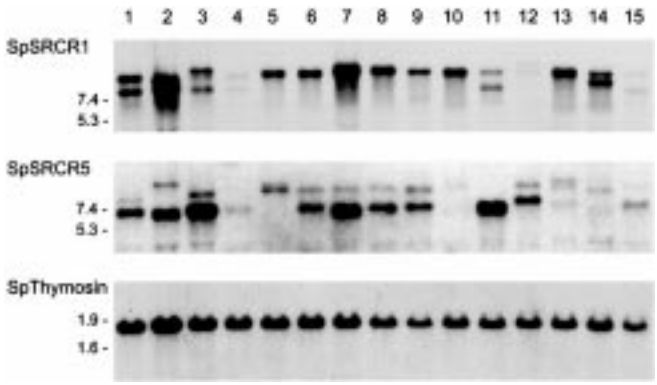

Fig. 8 Individual-specific patterns of transcription of *S. purpuratus* SRCR messages *SpSRCR1* and *SpSRCR5*. Northern blot of coelomocyte total RNA from 15 individual sea urchins (10 μg per lane) was probed either with *SpSRCR1* (*top*) or with *SpSRCR5* (*middle*). In the *bottom panel* the same blot was probed with *SpThymosin* (probes are those described in the legend of Figs. 4 and 6)

seen, though we note that all the animals in this experiment were drawn from apparently healthy populations, maintained in stable long-term culture conditions. Preliminary data indicate individual differences in expression of other *SCRC* genes in addition to *SpSRCR1* and *SpSRCR5*, as well as some SRCR proteins that seem to be expressed with little variation, from animal to animal.

Discussion

Expression of the six genes considered in this paper is directly linked to coelomocyte immune function, either by the inductive change in the levels of their transcripts which follows introduction of bacteria, or the expression by coelomocytes of these genes (i.e., as far as has been determined). Taken together with earlier findings from a coelomocyte EST project (Smith et al. 1996) one can begin to identify some of the specific molecular functions that coelomocytes perform. For example, they appear to utilize a complement system as do mammalian myeloid cells, in that they express genes encoding proteins homologous with certain complement proteins of mammals (Al-Sharif et al. 1998; Smith et al. 1996). Our current findings greatly extend the presently slim body of molecular level information about echinoderm coelomocyte functions and mechanisms.

Molecular insights into the functional character of the echinoderm immune system

We show here that *S. purpuratus* coelomocytes mount an extensive transcription-level response to immune challenge. Introduction of bacteria, or even the small injury caused by the prick of a fine-gauge needle and injection of 1 ml of filtered sea water (with perhaps a concomitant local infection that follows), sharply affects the levels of three different mRNAs encoding transcription factors. Obviously this implies the existence of batteries of facultatively expressed genes that must be activated downstream of those transcription factors for which message levels rise and fall after challenge. Bacterial response of one of these genes, that encoding *SpNFKB*, is indeed expected, since genes encoding homologous factors are up-regulated on bacterial infection in both insects and mammals (reviewed by Baldwin 1996; Hoffman et al. 1996); these animal clades are so distant that this response system is likely to be pan-bilaterian (see Fig. 1). However, the even sharper response of *SpRunt-1* mRNA, and the down regulation of *SpGATAc* mRNA occurring in coelomocytes after bacterial injection (Fig. 4) were surprising. It will be interesting to determine whether the *SpRunt-1* response occurs downstream of the same signal transduction pathway as they utilize to activate their *NFKB* genes, or whether this is an additional, separate signal/response pathway also activated by bacterial products and in consequence of sham injection. The kinetics of the *SpRunt-1* and *SpNFKB* responses to large amounts of injected bacteria appear similar in Fig. 4, and both responses seem directly reciprocal to that of *SpGATAc*. Thus this message virtually disappears as the levels of the others dramatically increase, and then as the latter fade back to a normal condition *SpGATAc* mRNA reappears. The linkage could be causal, i.e., *SpGATAc* could repress *SpNFKB* and *SpRunt-1* except on bacterial challenge. However, we note that *SpNFKB* and *SpRunt-1* are both expressed at enhanced levels, though less dramatically, after sham injection, but *SpGATAc* downregulation is much less pronounced in sham injected animals. This is inconsistent with an obligatory reciprocal relation between SpGATAc expression and that of the other two genes. This result also implies that the gene battery or batteries responding to bacterial introduction include but extend beyond those activated by sham injection alone. The *SpGATAc* response occurs only on exposure to bacteria, while the *SpNFKB* and *SpRunt-1* responses occur in both contexts.

The observations in Fig. 4 highlight one simple and obvious point: a major function of coelomocytes is to respond transcriptionally to bacteria. We have shown previously that injection of LPS, a component of bacterial coats, activates the profilin gene of coelomocytes (Smith et al. 1995). Therefore coelomocytes possess at least one mechanism that would allow them to respond specifically to bacteria, in addition to whatever response is provoked by introduction of particles or by minor injury. Though there is as yet no direct evidence, the positive *SpNFκB* and *SpRunt* responses, and the negative *SpGATAc* response to bacteria, are likely also to represent specific reactions to potential pathogens. This is scarcely surprising; the immune systems of invertebrates exist fundamentally to protect the organism against pathogens (Medzhitov and Janeway 1998). Such

is of course also a major function of our innate immune systems, and to the extent that the transcriptional apparatus which carries out these functions is similar in sea urchins and mammals, it represents a shared basal character, a specific genetic system inherited from ancestral deuterostome founder stocks.

The most unexpected discovery in regard to sea urchin immune system function that we present here is that illustrated in Fig. 8. Here we see that the coelomocyte-specific scavenger receptor homologues *SpSRCR1* and *SpSRCR5* are expressed differently in different individual *S. purpuratus*. Some of those animals that do not express the *SpSRCR1* genes nonetheless express the *SpSRCR5* genes. *SpSRCR1* and *SpSRCR5* contain distinct SRCR repeat units, which have different functional attributes. In addition, while *SpSRCR5* may have a simpler structure (Fig. 7B), *SpSRCR1* is evidently polyfunctional, as it includes two other functional domains as well as SRCR repeats, i.e., the von Willebrand factor domains and a domain similar to a known extracellular matrix protein of sea urchins. Though not shown here, we have discovered other SRCR-containing proteins expressed in *S. purpuratus* coelomocytes which have yet other characteristics, and which are also expressed to very different extents, individual to individual. Differential representation of the SRCR mRNAs in the 15 individual coelomocyte RNA preparations is most likely due to differential transcription of the same genes, animal to animal, though as noted above other explanations cannot yet be excluded. These SRCR proteins are likely to be involved in the coelomocyte immune response, because of their similarities with the Group A SRCR domains of the myeloid scavenger receptor and lectin-binding proteins (Fig. 7A), and by virtue of their relatively intense coelomocyte-specific expression. However, the levels of SRCR mRNAs are not affected, at least on a short time scale either by bacterial challenge or by sham injection. Figure 8 could be taken to imply a more or less constitutive, individual-specific heterogeneity in the nature of the immune capabilities of each animal. If this heterogeneity in SRCR mRNA levels is of transcriptional origin, it could depend on differences in the immunological, i.e., "medical" history of each animal, which might have affected either the constitution of its coelomocyte population, or the level of expression of specific genes in the same coelomocyte population (or both). In either case, Fig. 8 suggests that we may be seeing the tip of an iceberg with respect to the molecular definition of immune individuality in our distant deuterostome cousins.

Process of immune system evolution in the deuterostomes

Sea urchins have no B or T cells or rearranging immunoglobulins, so far as anyone has been able to discover, and it is extremely unlikely that they have an adaptive immune system anything like that which vertebrates display (reviewed by Smith and Davidson 1992, 1994). For example, they have no immunological memory in respect to allograft rejection. Yet homologues of three transcription factors which appear to display immune responses to bacterial injection in sea urchin coelomocytes are also utilized in the lymphoid genes that mediate the evolutionarily "new" adaptive immune responses of vertebrates. The implication is that evolution of the adaptive immune system has proceeded by insertion of new effector genes into an ancient regulatory network. This network consists of gene batteries which function in the immune response, together with their upstream regulatory controls. In the sense of evolutionary descent from a common ancestor, gene batteries encoding immune functions defined by the use of homologous transcription factors to control their expression can be considered homologous regulatory structures. Thus the immune induction biochemistry upstream of the activation of the genes encoding these transcription factors may also be homologous. In different evolutionary descendants these batteries may indeed retain some of the same effector genes, e.g., the *Toll/IL-1* intracellular signaling pathway (Gonzales-Crespo and Levine 1994) or the *SpHS* and SRCR genes studied here. New or different genes have been added, as well as other induction systems, the termini of which have been inserted into otherwise homologous *cis*-regulatory elements. But if the developmental inputs are different because the organism is structurally very different, what is homologous is only the genetic regulatory apparatus, not the cells in which it runs, nor its particular immune functions. Sea urchin coelomocytes have no evident embryological or developmental similarity to vertebrate hematopoietic cells. Echinoderm coelomocytes cannot arise from equivalent regions of the ventral embryonic mesoderm or from developing blood vessels as do hematopoietic precursors in vertebrates, because no such structures exist in any obvious sense in the echinoderm body plan or its embryonic forms. Thus, the evolutionarily significant homology in distantly related deuterostome immune systems resides in their basal genetic regulatory organization, not in the particular cells or body parts in which they are expressed. Different genes have been inserted in an old regulatory apparatus in the course of immune system evolution in deuterostomes, leading to some shared, and many particular immune functions in each group. Fundamentally, sea urchins and vertebrates share a genetic regulatory network dedicated to the control of whatever protective immune functions have proved advantageous in each group.

Acknowledgments We are grateful to our Caltech colleagues, Prof. Ellen Rothenberg, Dr. Kevin Peterson, and Dr. Michele Anderson, for reviews and discussion of this manuscript. Miki Yun provided invaluable assistance in the analysis of the many new clones described in this work, and Patrick Leahy was of enormous assistance in handling the sea urchins and for the immune response induction experiments. We thank Michael Rhoades for

assistance in cloning the GATA factors. The research was supported by HFSP Grant RG-333/96; Z.P. was supported by NIH Training Grant HD-07257; and J.P.R. by NIH Individual NRSA GM-18478.

References

Aguinaldo AMA, Tubeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493

Al-Sharif WZ, Sunyer JO, Lambris JD, Smith LC (1998) Sea urchin coelomocytes specifically express a homologue of the complement component C3. J Immunol 160:2983–2997

Arnone M, Davidson EH (1997) The hardwiring of development: Organization and function of genomic regulatory systems. Development 124:1851–1864

Aruffo A, Bowen MA, Patel DD, Hynes BF, Starling GC, Gebe JA, Bajorath J (1997) CD6-ligand interactions: a paradigm for SRCR domain function? Immunol Today 18:498–504

Balavoine G (1997) The early emergence of platyhelminths is contradicted by the agreement between 18 S rRNA and *Hox* gene data. Life Sci 320:83–94

Balavoine G, Adoutte A (1998) One or three Cambrian radiations? Science 280:397–398

Baldwin AS Jr (1996) The NFκB and IκB proteins: New discoveries and insights. Annu Rev Immunol 14:649–81

Britos-Bray M, Friedman AD (1997) Core binding factor cannot synergistically activate the myeloperoxidase proximal enhancer in immature myeloid cells without c-Myb. Mol Cell Biol 17:5127–5135

Cockerill PN, Osborne CS, Bert AG, Grotto RJ (1996) Regulation of GM-CSF gene transcription by core-binding factor. Cell Growth Differ 7:917–922

Coffman JA, Kirchhamer CV, Harrington MG, Davidson EH (1996) SpRunt–1, a new member of the runt domain family of transcription factors, is a positive regulator of the aboral ectoderm-specific *CyIIIa* gene in sea urchin embryos. Dev Biol 174:43–54

Dangott LJ, Jordan JE, Bellet RA, Garbers DL (1989) Cloning of the mRNA for the protein that crosslinks to the egg peptide speract. Proc Natl Acad Sci USA 86:2128–2132

Dushay MS, Asling B, Hultmark D (1996) Origins of immunity: relish, a compound Rel-like gene in the antibacterial defense of *Drosophila*. Proc Natl Acad Sci USA 93:10343–10347

Elomaa O, Kangas M, Sahlberg C, Tuukkanen J, Sormunen R, Liakka A, Thesleff I, Kraal G, Tryggvason K (1995) Cloning of a novel bacteria-binding receptor structurally related to scavenger receptors and expressed in a subset of macrophages. Cell 80:603–609

Evans T, Felsenfeld G (1989) The erythroid-specific transcription factor Eryf1:A new finger protein. Cell 58:877–885

Freeman M, Ashkenas J, Rees DJ, Kingsley DM, Copeland NG, Jenkins NA, Krieger M (1990) An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. Proc Natl Acad Sci USA 87:8810–8814

Friedman J, Trahey M, Weissman I (1993) Cloning and characterization of cyclophilin C-associated protein: a candidate natural cellular ligand for cyclophilin C. Proc Natl Acad Sci USA 90:6815–6819

Ghosh S, Gifford AM, Riviere LR, Tempst P, Nolan GP, Baltimore D (1990) Cloning of the P50 DNA binding subunit of NFκB: homology to *rel* and *dorsal*. Cell 62:1019–29

Ghosh S, May MJ, Kopp EB (1998) NFκB and rel proteins: evolutionarily conserved mediators of immune responses. Annu Rev Immunol 16:225–260

Giese K, Kingsley C, Kirshner JR, Grosschedl R (1995) Assembly and function of a TCRα enhancer complex is dependent on LEF-1 induced DNA bending and multiple protein-protein interactions. Genes Dev 9:995–1008

Goldberger G, Bruns GA, Rits M, Edge MD, Kwiatkowski DJ (1987) Human complement factor I: Analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. J Biol Chem 262:10065–10071

Gonzales-Crespo S, Levine M (1994) Related target enhancers for dorsal and NFκB signaling pathways. Science 264:255–258

Halanych KM, Bacheller JD, Aguinaldo AMA, Liva SM, Hillis DM, Lake JA (1995) Evidence from 18 S ribosomal DNA that the lophophorates are protostome animals. Science 267:1641–1643

Hoffmann JA, Reichhart J-M, Hertru C (1996) Innate immunity in higher insects. Curr Opin Immunol 8:8–13

Ip TY, Reach M, Engstrom Y, Kadalayil L, Cai H, Gonzales-Crespo S, Tatei K, Levine M (1993) *Dif*, a *dorsal*–related gene that mediates an immune response in *Drosophila*. Cell 75:753–763

Kagoshima H, Shigesada K, Satake M, Ito Y, Miyoshi H, Ohki M, Pepling M, Gergen P (1993) The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet 9:338–341

Kanegae Y, Tavares AT, Belmonte JCI, Verma IM (1998) Role of Rel/NFκB transcription factors during the outgrowth of the vertebrate limb. Nature 392:611–614

Kitamura D, Kaneko H, Miyagoe Y, Ariyasu T, Watanabe T (1989) Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage. Nucl Acids Res 17:9367–9379

Koths K, Taylor E, Halenbeck R, Casipit C, Wang A (1993) Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. J Biol Chem 268:14225–14249

Leahy PS, Tutschulte TC, Britten RJ, Davidson EH (1978) A large-scale laboratory maintenance system for gravid purple sea urchins (*Strongylocentrotus purpuratus*). J Exp Zool 204:369–380

Leiden JM (1993) Transcriptional regulation of T-cell receptor genes. Annu Rev Immunol 11:539–570

Lin W-H, Huang L-H, Yeh J-Y, Hoheisel J, Lehrach H, Sun YH, Tsai S-F (1995) Expression of a *Drosophila* GATA transcription factor in multiple tissues in the developing embryo. J Biol Chem 270:25150–25158

Maier E, Meier-Ewert S, Ahmadi AR, Curtis J, Lehrach H (1994) Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridization. J Biotech 35:91–203

Medzhitov R, Janeway CA Jr (1998) An ancient system of host defense. Curr Opin Immunol 10:12–15

Miglarese MR, Mannion-Henderson J, Wu H, Parsons JT, Bender TP (1994) The protein tyrosine kinase substrate cortactin is differentially expressed in murine B lymphoid tumors. Oncogene 9:1989–1997

Nicholas KB, Nicholas HB Jr (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments (version 2.4.000)

Pancer Z, Münkner J. Müller I, Müller WEG (1997) A novel member of an ancient superfamily: sponge (*Geodia cydonium*, Porifera) putative protein that features scavenger receptor cysteine-rich repeats. Gene 193:211–218

Pedone PV, Omichinski JG, Nony P, Trainor C, Gronenborn AM, Clore GM, Felsenfeld G (1997) The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. EMBO J 16:2874–2882

Ramain P, Heitzler P, Haenlin M, Simpson P (1993) *Pannier*, a negative regulator of *achaete* and *scute* in *Drosophila* encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. Development 119:1277–1291

Rehorn KP, Thelen H, Michelson AM, Reuter R (1996) A molecular aspect of hematopoiesis and endoderm development common to vertebrates and *Drosophila*. Development 122:4023–4031

Resnick D, Pearson A, Krieger M (1994) The SRCR superfamily: A family reminiscent of the Ig superfamily. Trends Biochem Sci 19:5–8

Rzhetsky A, Nei M (1994) METREE: A program package for inferring and testing minimum-evolution trees. Comput Appl Biosci 10:409–412

Safer D, Nachmias VT (1994) Beta thymosins as actin binding peptides. BioEssays 16:473–479

Smith LC, Davidson EH (1992) The echinoid immune system and the phylogenetic occurrence of immune mechanisms in deuterostomes. Immunol Today 13:356–361

Smith LC. Davidson EH (1994) The echinoderm immune system. Ann NY Acad Sci 712:213–226

Smith LC, Britten RJ, Davidson EH (1992) *SpCoel1*: a sea urchin profilin gene expressed specifically in coelomocytes in response to injury. Mol Biol Cell 3:403–414

Smith LC, Britten R.J, Davidson EH (1995) Lipopolysaccharide activates the sea urchin immune system. Dev Comp Immunol 19:217–224

Smith LC, Chang L, Britten RJ, Davidson EH (1996) Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Immunology 156:593–602

Smith LC, Shih C-S, Dachenhausen SG (1998) Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system. J Immunol 161:6784–6793

Takemoto Y, Furuta M, Li X-K, Strong-Sparks WJ, Hashimoto Y (1995) LckBP1, a proline-rich protein expressed in haematopoietic lineage cells, directly associates with the SH3 domain of protein tyrosine kinase $p56^{lck}$. EMBO J 14:3403–14

Tenen DG, Hromas R, Licht JD, Zhang D-E (1997) Transcription factors, normal myeloid development, and leukemia. Blood 90:489–519

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Ting CN, Olson MC, Barton KP, Leiden JM (1996) Transcription factor GATA-3 is required for development of T-cell lineage. Nature 384:474–478

Tsai SF, Martin DI, Zon LI, D'Andrea AD, Wong GG, Orkin SH (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature 339:446–451

Verma IM, Stevenson JK, Schwarz EM, Van Antwerp D, Miyamoto S (1995) Rel/NFκB/IκB family: intimate tales of association and dissociation. Genes Dev 9:2723–2735

Wada H, Satoh N (1994) Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18 S rDNA. Proc Natl Acad Sci USA 91:1801–1804

Wargnier A, Legros-Maida S, Bosselut R, Bourge JF, Lafaurie C, Ghysdael CJ, Sasportes M, Paul P (1995) Identification of human granzyme B promoter regulatory elements interacting with activated T-cell-specific proteins: Implication of Ikaros and CBF binding sites in promoter activation. Proc Natl Acad Sci USA 18:6930–6934

Weiss MJ, Orkin SH (1995) GATA transcription factors: key regulators of hematopoiesis. Exp Hematol 23:99–107

Wijngaard PL, Metzelaar MJ, MacHugh ND, Morrison WI, Clevers HC (1992) Molecular characterization of the WC1 antigen expressed specifically on bovine CD4-CD8-gamma delta T lymphocytes. J Immunol 149:3273–3277

Wu H, Parsons JT (1993) Cortactin, an 80/85-kilodalton pp60src substrate, is a filamentous actin-binding protein enriched in the cell cortex. J Cell Biol 120:1417–1426

Yamanashi Y, Okada M, Semba T, Yamori T, Umemori H, Tsunasawa S, Toyoshima K, Kitamura D, Watanabe T, Yamamoto T (1993) Identification of HS1 protein as a major substrate of protein-tyrosine kinase(s) upon B-cell antigen receptor-mediated signaling. Proc Natl Acad Sci USA 15:3631–3635

Zhan X, Haudenschild CC, Ni Y, Smith E, Huang C (1997) Upregulation of cortactin expression during the maturation of megakaryocytes. Blood 15:457–464

Zhang DE, Hohaus S, Voso MT, Chen HM, Smith LT, Hetherington CJ, Tenen DG (1996) Function of PU.1 (Spi-1), C/EBP, and AML1 in early myelopoiesis: regulation of multiple myeloid CSF receptor promoters. Curr Top Microbiol Immunol 211:137–147