

A family of human receptors structurally related to *Drosophila* Toll

Fernando L. Rock, Gary Hardiman, Jackie C. Timans, Robert A. Kastelein, and J. Fernando Bazan

PNAS 1998;95:588-593
doi:10.1073/pnas.95.2.588

This information is current as of February 2007.

Online Information & Services	High-resolution figures, a citation map, links to PubMed and Google Scholar, etc., can be found at: www.pnas.org/cgi/content/full/95/2/588
References	This article cites 52 articles, 22 of which you can access for free at: www.pnas.org/cgi/content/full/95/2/588#BIBL This article has been cited by other articles: www.pnas.org/cgi/content/full/95/2/588#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

A family of human receptors structurally related to *Drosophila* Toll

FERNANDO L. ROCK, GARY HARDIMAN, JACKIE C. TIMANS, ROBERT A. KASTELEIN, AND J. FERNANDO BAZAN*

Protein Machine Group, Department of Molecular Biology, DNAX Research Institute, 901 California Avenue, Palo Alto, CA 94304-1104

Communicated by Corey S. Goodman, University of California, Berkeley, CA, November 18, 1997 (received for review May 16, 1997)

ABSTRACT The discovery of sequence homology between the cytoplasmic domains of *Drosophila* Toll and human interleukin 1 receptors has sown the conviction that both molecules trigger related signaling pathways tied to the nuclear translocation of Rel-type transcription factors. This conserved signaling scheme governs an evolutionarily ancient immune response in both insects and vertebrates. We report the molecular cloning of a class of putative human receptors with a protein architecture that is similar to *Drosophila* Toll in both intra- and extracellular segments. Five human Toll-like receptors—named TLRs 1–5—are probably the direct homologs of the fly molecule and, as such, could constitute an important and unrecognized component of innate immunity in humans. Intriguingly, the evolutionary retention of TLRs in vertebrates may indicate another role—akin to Toll in the dorsoventralization of the *Drosophila* embryo—as regulators of early morphogenetic patterning. Multiple tissue mRNA blots indicate markedly different patterns of expression for the human TLRs. By using fluorescence *in situ* hybridization and sequence-tagged site database analyses, we also show that the cognate *Tlr* genes reside on chromosomes 4 (TLRs 1, 2, and 3), 9 (TLR4), and 1 (TLR5). Structure prediction of the aligned Toll-homology domains from varied insect and human TLRs, vertebrate interleukin 1 receptors and MyD88 factors, and plant disease-resistance proteins recognizes a parallel β/α fold with an acidic active site; a similar structure notably recurs in a class of response regulators broadly involved in transducing sensory information in bacteria.

The seeds of the morphogenetic gulf that so dramatically separates flies from humans are planted in familiar embryonic shapes and patterns but give rise to very different cell complexities (1, 2). This divergence of developmental plans between insects and vertebrates is choreographed by remarkably similar signaling pathways, underscoring a greater conservation of protein networks and biochemical mechanisms from unequal gene repertoires (3, 4). A powerful way to chart the evolutionary design of these regulatory pathways is by inferring their likely molecular components (and biological functions) through interspecies comparisons of protein sequences and structures (3–5).

A universally critical step in embryonic development is the specification of body axes, either born from innate asymmetries or triggered by external cues (1, 2). As a model system, particular attention has been focused on the phylogenetic basis and cellular mechanisms of dorsoventral polarization (1, 2). A prototype molecular strategy for this transformation has emerged from the *Drosophila* embryo, where the sequential action of a small number of genes results in a ventralizing gradient of the transcription factor Dorsal (6, 7). This signaling pathway centers on Toll, a transmembrane receptor that

transduces the binding of a maternally secreted ventral factor, Spätzle, into the cytoplasmic engagement of Tube, an accessory molecule, and the activation of Pelle, a Ser/Thr kinase that catalyzes the dissociation of Dorsal from the inhibitor Cactus and allows migration of Dorsal to ventral nuclei (7, 8). The Toll pathway also controls the induction of potent antimicrobial factors in the adult fly (9); this role in *Drosophila* immune defense strengthens mechanistic parallels to interleukin 1 (IL-1) pathways that govern a host of immune and inflammatory responses in vertebrates (8, 10). A Toll-related cytoplasmic domain in IL-1 receptors (IL-1Rs) directs the binding of a Pelle-like kinase, IRAK, and the activation of a latent NF- κ B/I- κ B complex that mirrors the embrace of Dorsal and Cactus (8, 10).

We describe the cloning and molecular characterization of five Toll-like molecules in humans—named TLRs 1–5 [as in Chiang and Beachy (11)]—that reveal a receptor family more closely tied to *Drosophila* Toll homologs than to vertebrate IL-1Rs. The TLR sequences are derived from human expressed sequence tags (ESTs); these partial cDNAs were used to draw complete expression profiles in human tissues for the five TLRs, map the chromosomal locations of cognate genes, and narrow the choice of cDNA libraries for full-length cDNA retrievals. Spurred by other efforts (5, 12), we are assembling, by structural conservation and molecular parsimony, a biological system in humans that is the counterpart of a compelling regulatory scheme in *Drosophila*. In addition, a biochemical mechanism driving Toll signaling is suggested by the proposed tertiary fold of the Toll-homology (TH) domain, a core module shared by TLRs, a broad family of IL-1Rs, mammalian MyD88 factors, and plant disease-resistance proteins (13, 14). We propose that a signaling route coupling morphogenesis and primitive immunity in insects, plants, and animals (8, 15) may have roots in bacterial two-component pathways.[†]

MATERIALS AND METHODS

Computational Analysis. Human sequences related to insect TLRs were harvested from the EST database (dbEST) at the National Center for Biotechnology Information by using the BLAST server with the Gonnet protein comparison matrix (16). More sensitive pattern- and profile-based methods (17) were used to isolate the signaling domains of the TLR family that are shared with vertebrate and plant proteins present in nonredundant databases. The progressive alignment of TLR

Abbreviations: TLR, Toll-like receptor; IL-1, interleukin 1; IL-1R, IL-1 receptor; TH, Toll homology; LRR, leucine-rich repeat; EST, expressed sequence tag; STS, sequence-tagged site; FISH, fluorescence *in situ* hybridization; 18w, 18 Wheeler.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U88540 and U88878–81 for human TLRs 1–5).

*To whom reprint requests should be addressed. e-mail: bazan@dnax.org.

[†]This work was presented earlier at the May 1997 Cold Spring Harbor Symposium on "Pattern Formation During Development" [Bazan, J. F., *et al.* (1997) in *Symposium LXII*, ed. Stillman, B. (Cold Spring Harbor Lab. Press, Plainview, NY), p. 23 (abstr).]

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/95588-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

intra- or extracellular domain sequences was carried out by CLUSTALW (18); this program also calculated the branching order of aligned sequences by the neighbor-joining algorithm (10,000 bootstrap replications provided confidence values for the tree groupings).

Conserved alignment patterns were drawn by the CONSENSUS program (internet URL <http://www.bork.embl-heidelberg.de/Alignment/consensus.html>). The PRINTS library of protein fingerprints (<http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html>) (19) reliably identified the myriad leucine-rich repeats (LRRs) present in the extracellular segments of TLRs with a compound "Leurichrpt" motif that flexibly matches N- and C-terminal features of divergent LRRs. Two algorithms whose three-state accuracy is greater than 72%—the neural network program PHD (20) and the statistical prediction method DSC (21)—were used to derive a consensus secondary structure for the TH domain alignment. Fold recognition efforts (for reviews, see refs. 20, 22, and 23) used programs such as PROFIT, I23D and RDP within TOPLIGN (<http://cartan.gmd.de/ToPLign.html>) to "thread" sequences through protein-fold databases with empirically derived potentials; secondary structure information is assimilated by H3P2 (<http://foldserver.mbi.ucla.edu>), PSCAN (<http://www.biokemi.su.se/~arne/pscan>), and TOPITS to best "map" sequence to structure; domain-fold databases encompassed 300 (PROFIT), 347 (PSCAN), 1302 (TOPLIGN), and 1,634 (H3P2) structures.

Cloning of Full-Length Human TLR cDNAs. PCR primers derived from the Toll-like Humrsc786 sequence (GenBank accession no. D13637) (24) were used to probe a human erythroleukemic TF-1 cell line-derived cDNA library (25) to yield the TLR1 cDNA sequence. The remaining TLR sequences were flagged from dbEST, and the relevant EST clones were obtained from the I.M.A.G.E. consortium (26) via Research Genetics (Huntsville, AL); clone identification nos. are as follows: 80633 and 117262, TLR2; 144675, TLR3; 202057, TLR4; 277229, TLR5. Full-length cDNAs for human TLRs 2–4 were cloned by DNA hybridization screening of λ gt10 phage and human adult lung, placenta, and fetal liver 5' Stretch Plus cDNA libraries (CLONTECH), respectively; the TLR5 sequence is derived from a human multiple sclerosis plaque EST. All positive clones were sequenced and aligned to identify individual TLR ORFs as follows: TLR1 (2,366-bp clone, 786-aa ORF), TLR2 (2,600-bp clone, 784-aa ORF), TLR3 (3,029-bp clone, 904-aa ORF), TLR4 (3,811-bp clone, 879-aa ORF), and TLR5 (1,275-bp clone, 370-aa ORF). Probes for TLR3 and TLR4 hybridizations were generated by PCR using human placenta (Stratagene) and adult liver (CLONTECH) cDNA libraries as templates, respectively; primer pairs were derived from the respective EST sequences. PCRs were conducted with *Thermus aquaticus* Taqplus DNA polymerase (Stratagene) under the following conditions: 1 \times (94°C, 2 min), 30 \times (55°C, 20 sec; 72°C, 30 sec; 94°C, 20 sec), and 1 \times (72°C, 8 min). For TLR2 full-length cDNA screening, a 900-bp fragment generated by *EcoRI/XbaI* digestion of the first EST clone (identification no. 80633) was used as a probe.

mRNA Blots and Chromosomal Localization. Human multiple tissue (catalog nos. 7759 and 7760) and cancer cell line blots (catalog numbers 7757–1), containing approximately 2 μ g of poly(A)⁺ RNA per lane, were purchased from CLONTECH. For TLRs 1–4, the isolated full-length cDNAs served as probes; for TLR5, the EST clone (identification no. 277229) plasmid insert was used. Briefly, the probes were radiolabeled with [α -³²P]dCTP by using the Amersham Rediprime random primer labeling kit (product RPN1633). Prehybridization and hybridizations were performed at 65°C in 0.5 M Na₂HPO₄/7% SDS/0.5 M EDTA, pH 8.0. All stringency washes were conducted at 65°C with two initial washes in 2 \times SSC/0.1% SDS for 40 min, followed by a subsequent wash in 0.1 \times SSC/0.1% SDS for 20 min. Membranes were then exposed at –70°C to x-ray film (Kodak) in the presence of

intensifying screens. More detailed studies with cDNA library Southern blots (14) were performed with selected human TLR clones to examine their expression in hemopoietic cell subsets (data not shown).

Human chromosomal mapping was conducted by the method of fluorescence *in situ* hybridization (FISH) as described (27) using the various full-length (TLRs 2–4) or partial (TLR5) cDNA clones as probes. These analyses were performed as a service by SeeDNA Biotech (Ontario, Canada). A search for human syndromes (or mouse defects in syntenic loci) associated with the mapped TLR genes was conducted in the Dysmorphic Human–Mouse Homology Database by an internet server (http://www.hgmp.mrc.ac.uk/DHMH/hum_chrome1.html).

RESULTS

Conserved Architecture of Insect and Human TLR Ectodomains. The Toll family in *Drosophila* comprises at least four distinct gene products: Toll, the prototype receptor involved in dorsoventral patterning of the fly embryo (7), a second receptor named 18 Wheeler (18w) that also may be involved in early embryonic development (11, 28), and two additional receptors are predicted by incomplete Toll-like ORFs downstream of the male-specific-transcript (Mst) locus (GenBank accession no. X67703) or encoded by the sequence-tagged site (STS) Dm2245 (GenBank accession no. G01378) (13). The extracellular segments of Toll and 18w are distinctively composed of imperfect ~24-amino acid LRR motifs (11, 28). Similar tandem arrays of LRRs commonly form the adhesive antennae of varied cell surface molecules and their generic tertiary structure is presumed to mimic the horseshoe-shaped cradle of a ribonuclease inhibitor fold, where 17 LRRs show a repeating β/α -hairpin 28-residue motif (29). The specific recognition of Spätzle by Toll may follow a model proposed for the binding of cystine-knot fold glycoprotein hormones by the multi-LRR ectodomains of serpentine receptors, by using the concave side of the curved β -sheet (30); intriguingly, the pattern of cysteines in Spätzle and an orphan *Drosophila* ligand, Trunk, predict a similar cystine-knot tertiary structure (8, 31).

The 22- and 31-LRR ectodomains of Toll and 18w, respectively (the Mst ORF fragment displays 16 LRRs), are most closely related to the comparable 18-, 19-, 24-, and 22-LRR arrays of TLRs 1–4 (the incomplete TLR5 chain presently includes four membrane-proximal LRRs) by sequence and pattern analysis (16, 17) (Fig. 1). However, a striking difference in the human TLR chains is the common loss of a ~90-residue cysteine-rich region that is variably embedded in the ectodomains of Toll, 18w, and the Mst ORF (distanced 4, 6, and 2 LRRs, respectively, from the membrane boundary). These cysteine clusters are bipartite, with distinct top (ending an LRR) and bottom (stacked atop an LRR) halves (11, 28, 29); the top module recurs in both *Drosophila* and human TLRs as a conserved juxtamembrane spacer (Fig. 1). We suggest that the flexibly located cysteine clusters in *Drosophila* receptors (and other LRR proteins)—when mated top to bottom—form a compact module with paired termini that can be inserted between any pair of LRRs without altering the overall fold of TLR ectodomains; analogous extruded domains decorate the structures of other proteins (32).

Molecular Design of the TH Signaling Domain. Sequence comparison of Toll and IL-1R type I (IL-1R1) has disclosed a distant resemblance of an ~200-amino acid cytoplasmic domain that presumably mediates signaling by similar Rel-type transcription factors (8, 10). More recent additions to this functional paradigm include four plant disease-resistance proteins from tobacco, thale cress, and flax that feature an N-terminal TH module followed by nucleotide-binding and LRR segments (15); by contrast, a death domain precedes the TH chain of MyD88, an intracellular myeloid differentiation marker (13, 14) (Fig. 1). Additional IL-1R-type receptors

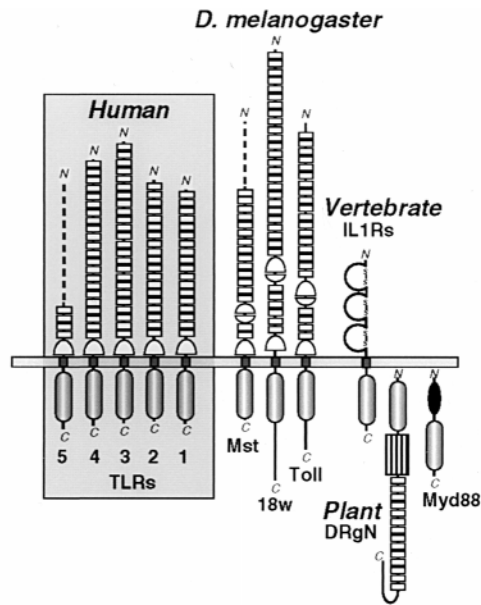


FIG. 1. Schematic comparison of the protein architectures of *Drosophila* and human TLRs and their relationship to vertebrate IL-1Rs and plant disease-resistance proteins. Three *Drosophila* (Dm) TLRs (Toll, 18w, and the Mst ORF fragment) (7, 11, 13, 28) are arrayed beside the four complete (TLRs 1–4) and one partial (TLR5) human (Hu) receptors. Individual LRRs in the receptor ectodomains that are flagged by PRINTS (19) are explicitly noted by boxes; top and bottom Cys-rich clusters that flank the C- or N-terminal ends of LRR arrays are respectively drawn by apposed half circles. The loss of the internal Cys-rich region in TLRs 1–5 largely accounts for their smaller ectodomains (558, 570, 690 and 652 aa, respectively) when compared with the 784- and 977-aa extensions of Toll and 18w. The incomplete chains of DmMst and HuTLR5 (519- and 153-aa ectodomains, respectively) are represented by dashed lines. The intracellular signaling module common to TLRs, IL-1Rs, the intracellular protein Myd88, and the tobacco disease-resistance gene N product (DRgN) is indicated below the membrane. Additional domains include the trio of Ig-like modules in IL-1Rs (disulfide-linked loops); the DRgN protein features a nucleotide-binding domain (box) and Myd88 has a death domain (solid oval).

include IL-1R3, an accessory signaling molecule, and orphan receptors IL-1R4 (also called ST2/Fit-1/T1), IL-1R5 (IL-1R-related protein), and IL-1R6 (IL-1R-related protein 2) (13, 14). With the new human TLR sequences, we have sought a structural definition of this evolutionary thread by analyzing the conformation of the common TH module: 10 blocks of conserved sequence containing 128 amino acids form the minimal TH domain fold; gaps in the alignment mark the likely location of sequence- and length-variable loops (Fig. 2A).

Two prediction algorithms that take advantage of the patterns of conservation and variation in multiply aligned sequences, PHD (20) and DSC (21), produced strong concordant results for the TH signaling module (Fig. 2A). Each block contains a discrete secondary structural element: the imprint of alternating β -strands (labeled A–E) and α -helices (numbered 1–5) is diagnostic of an β/α -class fold with α -helices on both faces of a parallel β -sheet. Hydrophobic β -strands A, C, and D are predicted to form interior staves in the β -sheet, whereas the shorter amphipathic β -strands B and E resemble typical edge units (Fig. 2A). This assignment is consistent with a strand order of B–A–C–D–E in the core β -sheet (Fig. 2B); fold comparison (mapping) and recognition (threading) programs (20, 22, 23) strongly return this doubly wound β/α topology (see *Discussion*). A surprising functional prediction of this outline structure for the TH domain is that many of the conserved charged residues in the multiple alignment map to the C-terminal end of the β -sheet: residue Asp-16 (Fig. 2A, block numbering scheme) at the end of β A, Arg-39 and Asp-40

after β B, Glu-75 in the first turn of α 3, and the more loosely conserved Glu/Asp residues in the β D– α 4 loop or after β E (Fig. 2A). Four other conserved residues (Asp-7, Glu-28, and the Arg-57–Arg/Lys-58 pair) form an epitope at the opposite N-terminal end of the β -sheet (Fig. 2A).

Signaling function depends on the structural integrity of the TH domain. Inactivating mutations or deletions within the module boundaries (Fig. 2A) have been catalogued for IL-1R1 and Toll (33–38). The human TLR1–5 chains extending past the minimal TH domain (8-, 0-, 6-, 22-, and 18-residue lengths, respectively) are most closely similar to the stubby, 4-aa tail of the Mst ORF. Toll and 18w display unrelated 102- and 207-residue tails (Fig. 2A) that may negatively regulate the signaling of their TH domains (37, 38).

The evolutionary relationship between the disparate proteins that carry the TH domain can best be discerned by a phylogenetic tree derived from the multiple alignment (18) (Fig. 3). Four principal branches segregate the plant proteins, the MyD88 factors, IL-1Rs, and Toll-like molecules; the latter branch decidedly clusters the *Drosophila* and human TLRs.

Chromosomal Dispersal of Human TLR Genes. To investigate the genetic linkage of the nascent human *Tlr* gene family, we mapped the chromosomal loci of four of the five genes by FISH (Fig. 4). The *Tlr1* gene has been charted by the human genome project: an STS database locus (dbSTS accession no. G06709, corresponding to STS WI-7804 or SHGC-12827) exists for the Humrsc786 cDNA (24) and fixes the gene to chromosome 4 marker interval D4S1587–D42405 (50–56 centimorgans) about 4p14. This assignment has recently been corroborated by FISH analysis (39). In the present work, we reliably assign the remaining *Tlr* genes to loci on chromosomes 4q32 (TLR2), 4q35 (TLR3), 9q32–33 (TLR4), and 1q33.3 (TLR5). During the course of this work, an STS for the parent TLR2 EST (clone identification no. 80633) has been generated (dbSTS accession no. T57791 for STS SHGC-33147) and maps to the chromosome 4 marker interval D4S424–D4S1548 (143–153 centimorgans) at 4q32—in accord with our findings. There is a \sim 50-centimorgan gap between *Tlr2* and *Tlr3* genes on the long arm of chromosome 4.

Tlr Genes Are Differentially Expressed. Both Toll and 18w have complex spatial and temporal patterns of expression in *Drosophila* that may point to functions beyond embryonic patterning (6–9, 11, 28). We have examined the spatial distribution of TLR transcripts by mRNA blot analysis with varied human tissue and cancer cell lines by using radiolabeled TLR cDNAs (Fig. 5). TLR1 is found to be expressed ubiquitously and at higher levels than the other receptors. Short 3.0-kb and long 8.0-kb TLR1 transcript forms are present in ovary and spleen, respectively (Fig. 5A and B), presumably reflecting alternative splicing. A cancer-cell mRNA panel also shows the prominent overexpression of TLR1 in a Burkitt's lymphoma Raji cell line (Fig. 5C). TLR2 mRNA is less widely expressed than TLR1, with a 4.0-kb species detected in lung and a 4.4-kb transcript evident in heart, brain, and muscle. The tissue distribution pattern of TLR3 echoes that of TLR2 (Fig. 5E). TLR3 is also present as two major transcripts of approximately 4.0 and 6.0 kb in size, and the highest levels of expression are observed in placenta and pancreas. By contrast, TLR4 and TLR5 messages appear to be extremely tissue-specific. TLR4 was detected only in placenta as a single transcript of \sim 7.0 kb. A faint 4.0-kb signal was observed for TLR5 in ovary and peripheral blood monocytes (data not shown).

DISCUSSION

Components of an Evolutionarily Ancient Regulatory System.

The original molecular blueprints and divergent fates of signaling pathways can be reconstructed by comparative genomic approaches (3–5, 12). We have used this logic to identify an emergent gene family in humans—encoding five receptor paralogs at present, TLRs 1–5—that are the direct evolutionary counterparts

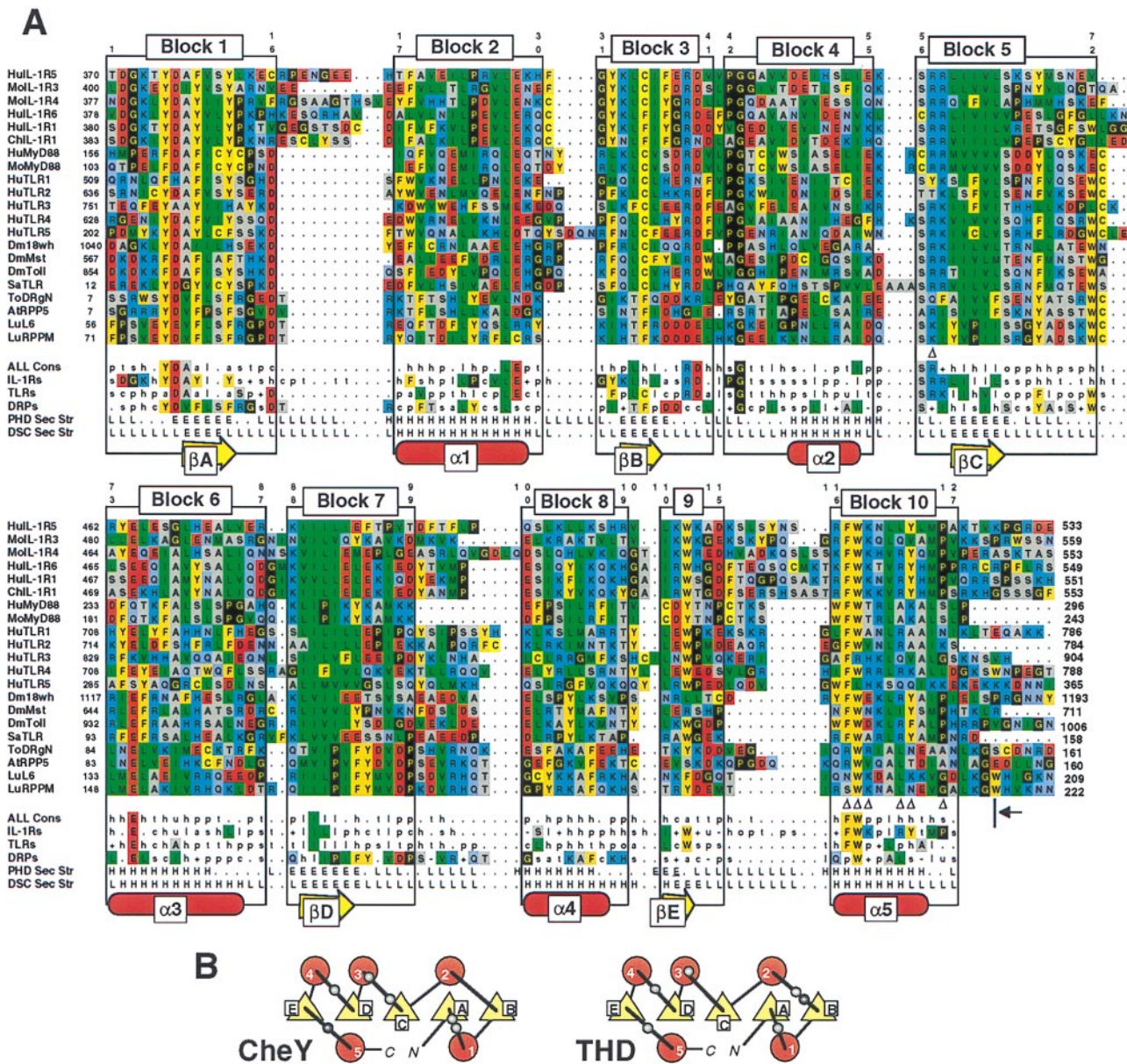


FIG. 2. Conserved structural patterns in the signaling domains of Toll- and IL-1-like cytokine receptors and two divergent modular proteins. (A) Sequence alignment of the common TH domain. TLRs are labeled as in Fig. 1; the human (Hu) or mouse (Mo) IL-1 family receptors (IL-1R1–6) are sequentially numbered as proposed (14); Myd88 and the disease-resistance protein (DRP) sequences from tobacco (To), *Arabidopsis thaliana* (At), and *Linum usitatissimum* (Lu) represent C- and N-terminal domains, respectively, of larger multidomain molecules. Ungapped blocks of sequence (numbered 1–10) are boxed. Triangles indicate deleterious mutations, and truncations N-terminal of the arrow eliminate bioactivity in human IL-1R1 (33). PHD (20) and DSC (21) secondary structure predictions of α -helix (H), β -strand (E), or coil (L) are marked. The amino acid coloring scheme depicts chemically similar residues: green (hydrophobic), red (acidic), blue (basic), yellow (Cys), orange (aromatic), black (structure breaking), and grey (tiny). Diagnostic sequence patterns for IL-1Rs, TLRs, DRPs, and full alignment (ALL) were derived by CONSENSUS at a stringency of 75%. Symbols for amino acid subsets are as follows (see CONSENSUS site for detail): o, alcohol; l, aliphatic; ., any amino acid; a, aromatic; c, charged; h, hydrophobic; -, negative; p, polar; +, positive; s, small; u, tiny; t, turnlike. (B) Topology diagram of the proposed CheY-like TH β/α domain fold. Parallel β -sheets (with β -strands A–E as yellow triangles) are seen at their C-terminal ends; α -helices (red circles labeled 1–5) link the β -strands; chain connections are to the front (visible) or back (hidden). Conserved charged residues at the C-terminal end of the β -sheets are noted in a shaded (Asp) or black (Arg) circle (see text).

of a *Drosophila* gene family headed by *Toll* (Figs. 1–3). The conserved architecture of human and fly TLRs—conserved LRR ectodomains and intracellular TH modules (Fig. 1)—intimates that the robust pathway coupled to Toll in *Drosophila* (6, 7) survives in vertebrates. The best evidence borrows from a reiterated pathway, the manifold IL-1 system and its repertoire of receptor-fused TH domains, IRAK, NF- κ B, and I- κ B homologs (8, 10, 14, 40). It is not known whether TLRs can productively couple to the IL-1R signaling machinery or whether a parallel set of proteins is used. Differently from IL-1Rs, the LRR cradle of human TLRs is predicted to retain an affinity for Spätzle/Trunk-

related cystine-knot factors; candidate TLR ligands (called PENs) that fit this mold have been isolated (R.A.K., unpublished results).

Biochemical mechanisms of signal transduction can be gauged by the conservation of interacting protein folds in a pathway (3, 4). At present, the Toll signaling paradigm involves some molecules whose roles are narrowly defined by their structures, actions, or fates: Pelle is a Ser/Thr kinase (phosphorylation), Dorsal is an NF- κ B-like transcription factor (DNA binding), and Cactus is an ankyrin-repeat inhibitor (Dorsal binding and degradation) (8). By contrast, the functions of the Toll TH domain and Tube remain enigmatic. Like other cytokine receptors (41), ligand-mediated

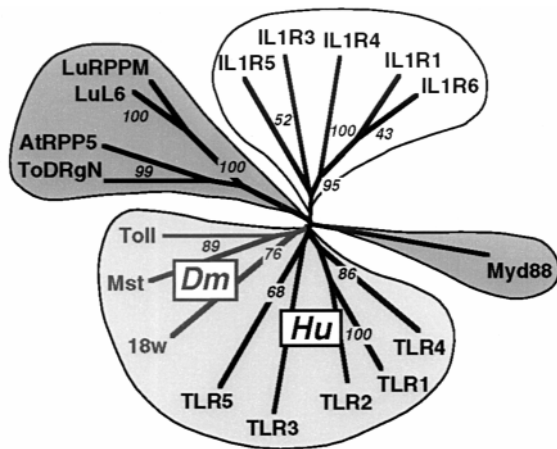


FIG. 3. Evolution of a signaling domain superfamily. The multiple TH module alignment of Fig. 2 was used to derive a phylogenetic tree by the neighbor-joining method (18); 10,000 bootstrapping replications were conducted to assess the reliability of the branching patterns (noted at selected nodes as percents). Proteins are labeled as in the alignment; the tree was rendered with TREEVIEW.

dimerization of Toll appears to be the triggering event: free cysteines in the juxtamembrane region of Toll create constitutively active receptor pairs (35), and chimeric Torso-Toll receptors signal as dimers (42); yet, severe truncations or wholesale loss of the Toll ectodomain results in promiscuous intracellular signaling (37, 43)—reminiscent of oncogenic receptors with catalytic domains (41). Tube is membrane-localized, engages the N-terminal (death) domain of Pelle, and is phosphorylated, but neither Toll–Tube or Toll–Pelle interactions are registered by two-hybrid analysis (42, 44); this latter result suggests that the conformational state of the Toll TH domain somehow affects factor recruitment (38, 42).

At the heart of these vexing issues is the structural nature of the Toll TH module. To address this question, we have taken advantage of the evolutionary diversity of TH sequences from insects, plants, and vertebrates—incorporating the human TLR chains—and extracted the minimal conserved protein core for structure prediction and fold recognition (Fig. 2). The strongly predicted (β/α)₅ TH domain fold with its asymmetric cluster of acidic residues is topologically identical to the structures of response regulators in bacterial two-component signaling pathways (45, 46) (Fig. 2). The prototype chemotaxis regulator CheY transiently binds a divalent cation in an aspartate pocket at the C-terminal end of the core β -sheet; this cation provides electrostatic stability and facilitates the activating phosphorylation of an invariant Asp (45). Likewise, the TH domain may capture cations in its acidic nest, but activation—and downstream signaling—could depend on the specific binding of a negatively charged moiety: anionic ligands can overcome intensely negative binding-site potentials by locking into precise hydrogen-bond networks (47). Intriguingly, the TH domain may not simply act as a passive scaffold for the assembly of a Tube–Pelle complex for Toll—or homologous systems in plants and vertebrates—but, instead, actively participate as a true conformational trigger in the signal transducing machinery. Toll dimerization could promote unmasking by regulatory receptor tails (37, 38), or binding of the TH pocket by small molecule activators, perhaps explaining the conditional binding of a Tube–Pelle complex. However, free TH modules inside the cell (37, 43) could act as catalytic CheY-like triggers by activating and docking with errant Tube–Pelle complexes.

Morphogenetic Receptors and Immune Defense. The evolutionary link between insect and vertebrate immune systems is stamped in DNA: genes encoding antimicrobial factors in insects display upstream motifs similar to acute-phase response elements known to bind NF- κ B transcription factors in mam-

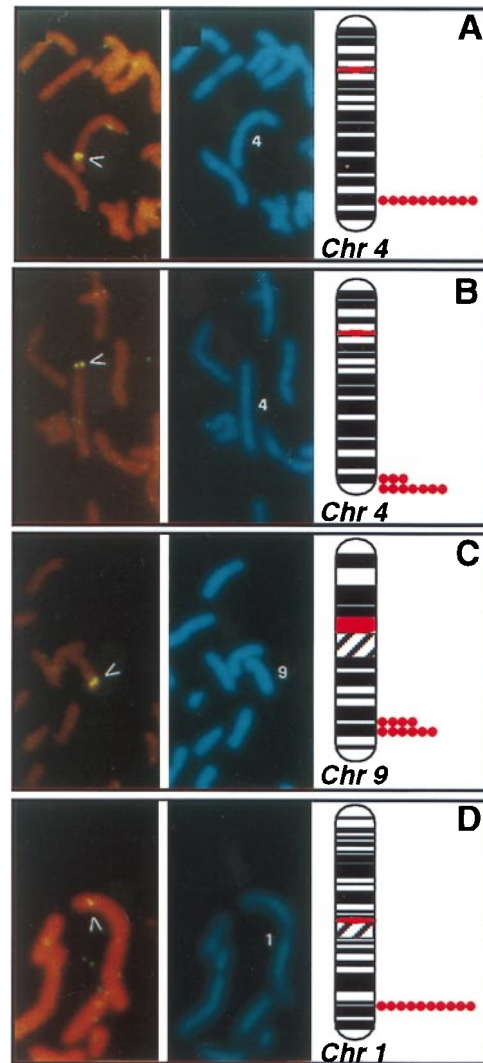


FIG. 4. FISH chromosomal mapping of human TLR genes. Denatured chromosomes from synchronous cultures of human lymphocytes were hybridized to biotinylated TLR cDNA probes for localization. (A) TLR2. (B) TLR3. (C) TLR4. (D) TLR5. Assignment of the FISH mapping data (*Left*) with chromosomal bands was achieved by superimposing FISH signals with 4,6-diamidino-2-phenylindole-banded chromosomes (*Center*) (27). Analyses are summarized in the form of human chromosome ideograms (*Right*).

mals (48). Dorsal and two Dorsal-related factors, Dif and Relish, help induce these defense proteins after bacterial challenge (49–51); Toll or other TLRs probably modulate these rapid immune responses in adult *Drosophila* (9, 52). These mechanistic parallels to the IL-1 inflammatory response in vertebrates are evidence of the functional versatility of the Toll signaling pathway and suggest an ancient synergy between embryonic patterning and innate immunity (8–10, 15, 48–53). The closer homology of insect and human TLR proteins invites an even stronger overlap of biological functions that supersedes the purely immune parallels to IL-1 systems and lends potential molecular regulators to dorsoventral and other transformations of vertebrate embryos (1, 2).

The present description of an emergent robust receptor family in humans mirrors the recent discovery of the vertebrate Frizzled receptors for Wnt patterning factors (12). As numerous other cytokine receptor systems have roles in early development (54), perhaps the distinct cellular contexts of compact embryos and gangly adults simply result in familiar signaling pathways and their diffusible triggers having different biological outcomes at different times—e.g., morphogenesis versus immune defense for

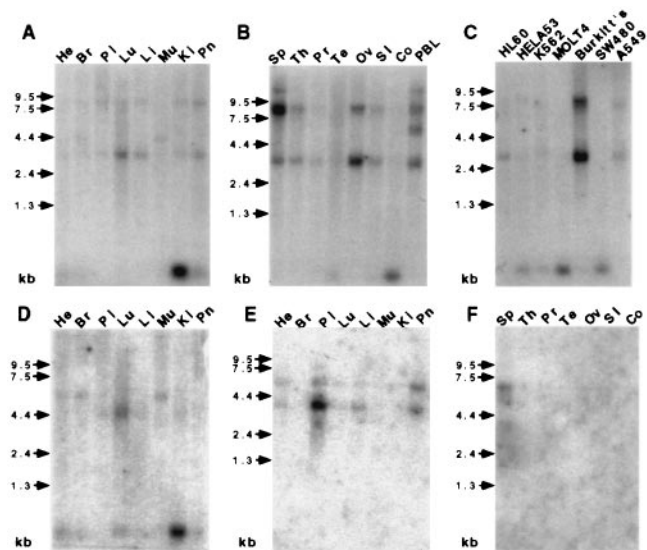


FIG. 5. mRNA blot analyses of human TLRs. Human multiple tissue blots (He, heart; Br, brain; Pl, placenta; Lu, lung; Li, liver; Mu, muscle; Ki, kidney; Pn, Pancreas; Sp, spleen; Th, thymus; Pr, prostate; Te, testis; Ov, ovary; SI, small intestine; Co, colon; PBL, peripheral blood lymphocytes) and cancer cell line (promyelocytic leukemia, HL60; cervical cancer, HELAS3; chronic myelogenous leukemia, K562; lymphoblastic leukemia, Molt4; colorectal adenocarcinoma, SW480; melanoma, G361; Burkitt lymphoma Raji, Burkitt's; colorectal adenocarcinoma, SW480; lung carcinoma, A549) containing approximately 2 μ g of poly(A)⁺ RNA per lane were probed with radiolabeled cDNAs encoding TLR1 (A-C), TLR2 (D), TLR3 (E), and TLR4 (F). Blots were exposed to x-ray film for 2 days (A-C) or 1 week (D-F) at -70°C with intensifying screens. An anomalous 0.3-kb species appears in some lanes; hybridization experiments exclude a message encoding a TLR cytoplasmic fragment.

TLRs. For insect, plant, and human Toll-related systems (14, 15), these signals course through a regulatory TH domain that intriguingly resembles a bacterial transducing engine (46).

Note. After this manuscript was submitted, the cloning of a human Toll homolog identical to our TLR4 sequence was reported (55). As suspected, this molecule activates NF- κ B and triggers the production of several inflammatory cytokines, hallmarks of an innate immune response.

We thank T. McClanahan, Dan Gorman, and their groups for support; D. Liggett for DNA synthesis; K. Anderson for the gift of a *Drosophila* Toll clone; N. Murgolo for threader help; D. Lundell for his experimental lead; and G. Zurawski, R. Debets, and T. Sana for critical comments. DNAX is supported by Schering-Plough.

1. DeRobertis, E. M. & Sasai, Y. (1996) *Nature (London)* **380**, 37-40.
2. Arendt, D. & Nübler-Jung, K. (1997) *Mech. Dev.* **61**, 7-21.
3. Miklos, G. L. G. & Rubin, G. M. (1996) *Cell* **86**, 521-529.
4. Chothia, C. (1994) *Development Suppl.* 27-33.
5. Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitelli, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O. & Ballabio, A. (1996) *Nat. Genet.* **13**, 167-174.
6. St. Johnston, D. & Nüsslein-Volhard, C. (1992) *Cell* **68**, 201-219.
7. Morisato, D. & Anderson, K. V. (1995) *Annu. Rev. Genet.* **29**, 371-399.
8. Belvin, M. P. & Anderson, K. V. (1996) *Annu. Rev. Cell Dev. Biol.* **12**, 393-416.
9. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. (1996) *Cell* **86**, 973-983.
10. Wasserman, S. A. (1993) *Mol. Biol. Cell* **4**, 767-771.
11. Chiang, C. & Beachy, P. A. (1994) *Mech. Dev.* **47**, 225-239.
12. Wang, Y., Macke, J. P., Abella, B. S., Andreasson, K., Worley, P., Gilbert, D. J., Copeland, N. G., Jenkins, N. A. & Nathans, J. (1996) *J. Biol. Chem.* **271**, 4468-4476.
13. Mitcham, J. L., Parnet, P., Bonnert, T. P., Garka, K. E., Gerhart, M. J., Slack, J. L., Gayle, M. A., Dower, S. K. & Sims, J. E. (1996) *J. Biol. Chem.* **271**, 5777-5783.

14. Hardiman, G., Rock, F. L., Balasubramanian, S., Kastelein, R. A. & Bazan, J. F. (1996) *Oncogene* **13**, 2467-2475.
15. Wilson, I., Vogel, J. & Somerville, S. (1997) *Curr. Biol.* **7**, 175-178.
16. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6**, 119-129.
17. Bork, P. & Gibson, T. J. (1996) *Methods Enzymol.* **266**, 162-184.
18. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680.
19. Attwood, T. K., Beck, M. E., Bleasby, A. J., Degtyarenko, K., Michie, A. D. & Parry-Smith, D. J. (1997) *Nucleic Acids Res.* **25**, 212-217.
20. Rost, B. & Sander, C. (1996) *Annu. Rev. Biophys. Biomol. Struct.* **25**, 113-136.
21. King, R. D. & Sternberg, M. J. E. (1996) *Protein Sci.* **5**, 2298-2310.
22. Fischer, D., Rice, D., Bowie, J. U. & Eisenberg, D. (1996) *FASEB J.* **10**, 126-136.
23. Jones, D. T. (1997) *Curr. Opin. Struct. Biol.* **7**, 377-387.
24. Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. & Tabata, S. (1994) *DNA Res.* **1**, 27-35.
25. Kitamura, T., Tojo, A., Kuwaki, T., Chiba, S., Miyazono, K., Urabe, A. & Takaku, F. (1989) *Blood* **73**, 375-380.
26. Lennox, G., Auffray, C., Polymeropoulos, M. & Soares, M. B. (1996) *Genomics* **33**, 151-152.
27. Heng, H. H. & Tsui, L. C. (1994) *Methods Mol. Biol.* **33**, 109-122.
28. Eldon, E., Kooyer, S., D'Evelyn, D., Duman, M., Lawinger, P., Botas, J. & Bellen, H. (1994) *Development (Cambridge, U.K.)* **120**, 885-899.
29. Buchanan, S. G. S. & Gay, N. J. (1996) *Prog. Biophys. Mol. Biol.* **65**, 1-44.
30. Kajava, A. V., Vassart, G. & Wodak, S. J. (1995) *Structure* **3**, 867-877.
31. Casanova, J., Furriols, M., McCormick, C. A. & Struhl, G. (1995) *Genes Dev.* **9**, 2539-2544.
32. Russell, R. B. (1994) *Protein Eng.* **7**, 1407-1410.
33. Heguy, A., Baldari, C. T., Macchia, G., Telford, J. L. & Melli, M. (1992) *J. Biol. Chem.* **267**, 2605-2609.
34. Croston, G. E., Cao, Z. & Goeddel, D. V. (1995) *J. Biol. Chem.* **270**, 16514-16517.
35. Schneider, D. S., Hudson, K. L., Lin, T. Y. & Anderson, K. V. (1991) *Genes Dev.* **5**, 797-807.
36. Norris, J. L. & Manley, J. L. (1992) *Genes Dev.* **6**, 1654-1667.
37. Norris, J. L. & Manley, J. L. (1995) *Genes Dev.* **9**, 358-369.
38. Norris, J. L. & Manley, J. L. (1996) *Genes Dev.* **10**, 862-872.
39. Taguchi, T., Mitcham, J. L., Dower, S. K., Sims, J. E. & Testa, J. (1996) *Genomics* **32**, 486-488.
40. Cao, Z., Henzel, W. J. & Gao, X. (1996) *Science* **271**, 1128-1131.
41. Heldin, C. H. (1995) *Cell* **80**, 213-223.
42. Galindo, R. L., Edwards, D. N., Gillespie, S. K. & Wasserman, S. A. (1995) *Development (Cambridge, U.K.)* **121**, 2209-2218.
43. Winans, K. A. & Hashimoto, C. (1995) *Mol. Biol. Cell* **6**, 587-596.
44. Großhans, J., Bergmann, A., Haffter, P. & Nüsslein-Volhard, C. (1994) *Nature (London)* **372**, 563-566.
45. Volz, K. (1993) *Biochemistry* **32**, 11741-11753.
46. Parkinson, J. S. (1993) *Cell* **73**, 857-871.
47. Ledvina, P. S., Yao, N., Choudhary, A. & Quioco, F. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6786-6791.
48. Hultmark, D. (1993) *Trends Genet.* **9**, 178-183.
49. Reichhart, J. M., Georgel, P., Meister, M., Lemaitre, B., Kappler, C. & Hoffmann, J. A. (1993) *C. R. Acad. Sci. (Paris)* **316**, 1218-1224.
50. Ip, Y. T., Reach, M., Engstrom, Y., Kadalayil, L., Cai, H., Gonzalez-Crespo, S., Tatei, K. & Levine, M. (1993) *Cell* **75**, 753-763.
51. Dushay, M. S., Asling, B. & Hultmark, D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10343-10347.
52. Rosetto, M., Engstrom, Y., Baldari, C. T., Telford, J. L. & Hultmark, D. (1995) *Biochem. Biophys. Res. Commun.* **209**, 111-116.
53. Medzhitov, R. & Janeway, C. A., Jr. (1997) *Curr. Opin. Immunol.* **9**, 4-9.
54. Lemaire, P. & Kodjabachian, L. (1996) *Trends Genet.* **12**, 525-531.
55. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. (1997) *Nature (London)* **388**, 394-397.