A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity

Ruslan Medzhitov[*]**, Paula Preston-Hurlburt**
& Charles A. Janeway Jr[*]

*Section of Immunobiology, Yale University School of Medicine, and * Howard Hughes Medical Institute, New Haven, Connecticut 06520-8011, USA*

Induction of the adaptive immune response depends on the expression of co-stimulatory molecules and cytokines by antigen-presenting cells. The mechanisms that control the initial induction of these signals upon infection are poorly understood. It has been proposed that their expression is controlled by the non-clonal, or innate, component of immunity that preceded in evolution the development of an adaptive immune system in vertebrates[1]. We report here the cloning and characterization of a human homologue of the *Drosophila toll* protein (Toll) which has been shown to induce the innate immune response in adult *Drosophila*[2–4]. Like *Drosophila* Toll, human Toll is a type I transmembrane protein with an extracellular domain consisting of a leucine-rich repeat (LRR) domain, and a cytoplasmic domain homologous to the cytoplasmic domain of the human interleukin (IL)-1 receptor. Both *Drosophila* Toll and the IL-1 receptor are known to signal through the NF-κB pathway[5–7]. We show that a constitutively active mutant of human Toll transfected into human cell lines can induce the activation of NF-κB and the expression of NF-κB-controlled genes for the inflammatory cytokines IL-1, IL-6 and IL-8, as well as the expression of the co-stimulatory molecule B7.1, which is required for the activation of naive T cells.

The Toll protein controls dorsal–ventral patterning in *Drosophila* embryos and activates the transcription factor Dorsal upon binding to its ligand Spatzle[8]. In adult *Drosophila*, the Toll/Dorsal signalling pathway participates in an anti-fungal immune response[2]. Signalling through Toll parallels the signalling pathway induced by the IL-1 receptor (IL-1R) in mammalian cells: IL-1R signals through the NF-κB pathway, and Dorsal and its inhibitor Cactus are homologous to NF-κB and I-κB proteins, respectively[5,6]. Moreover, the cytoplasmic domain of *Drosophila* Toll is homologous to the cytoplasmic domain of IL-1R (ref. 9). Remarkably, the tobacco-virus-resistance gene that encodes N-protein is also similar to Toll in that it contains both a Toll signalling domain and an LRR domain[10]. It thus appears that the immune-response system mediated by Toll represents an ancient host defence mechanism[6] (Fig. 1). To investigate the possibility that this pathway has been retained in the immune system of vertebrates, we used sequence and pattern searches[11] of the expressed-sequence tag (EST) database at the National Center for Biotechnology Information (NCBI). A search with a sequence profile of the Toll/IL-1R signalling domain identified a matching sequence in the EST database derived from human fetal liver/spleen library (Genebank accession number H48602, corresponding to clone 202057 from the IMAGE consortium[12]). Sequencing of this clone showed that it corresponds to the 3′ untranslated region (UTR) and the C terminus of the coding region of the messenger RNA. A fragment of the clone amplified by using the polymerase chain reaction (PCR) was used to screen a human spleen complementary DNA library by hybridization, and the 5′ end of the cDNA was cloned using the 5′-RACE technique as described[13,14]. The full-length 4,874-base-pair (bp) cDNA clone contained an open reading frame of 2,523 bp which encoded an 841-amino-acid protein chain (Fig. 2a), as well as a LINE-1 reverse transcriptase pseudogene in the 3′-UTR. Alignment of the sequences of the human and *Drosophila* Toll proteins shows that there is homology over the entire length of the protein chains (Fig. 2b). Notably, the similarity between the cytoplasmic domains is higher than between the human proteins Toll and IL-1R (not shown). The extracellular domain of human Toll (hToll) contains 21 tandemly repeated leucine-rich motifs separated by a non LRR region, similar to *Drosophila* Toll (dToll). At the N-terminal end of the LRR domain there is a 31-amino-acid long N-flanking region that is also present in several other LRR-containing proteins, for example RP105, Decorin and Biglycan[15]. The C-terminal end of the LRR domain is flanked by a cysteine-rich domain which is also present in dToll and some other transmembrane proteins[16,17].

To examine the expression pattern of hToll, a 720-bp cDNA fragment was used to probe northern blots containing poly(A)+ RNA from several organs. Most organs expressed two mRNA species: one of ~5 kilobases (kb) was predominant in most tissues except peripheral blood leukocytes (PBL), and corresponded to the length of the cDNA that we cloned. The lower band was ~4 kb long and this band was predominant in the PBL. The 4-kb band was not detectable in kidney, and liver did not contain any mRNA at all (Fig. 3). We also tested different mouse and human cell lines for expression of hToll mRNA by using PCR with reverse transcription (RT-PCR). We found mRNA for hToll in monocytes, macrophages, dendritic cells, γ/δ T cells, Th1 and Th2 α/β T cells, a small intestinal epithelial cell line, and a B-cell line (data not shown). The hToll gene is expressed most strongly in spleen and PBL (Fig. 3); its expression in other tissues may be due to the presence of macrophages and dendritic cells, in which it could act as an early-warning system for infection. Alternatively, hToll may be widely expressed because hToll signals through the conserved NF-κB pathway (see below) and NF-κB is a ubiquitous transcription factor.

To characterize hToll functions and see whether it can induce transcription of immune response genes like dToll, we generated a dominant-positive mutant of hToll because the natural ligand of hToll is unknown. To produce a constitutively active mutant of hToll, we made use of genetic information from dToll: analysis of ventralizing mutants in *Drosophila* embryos had identified the function of the ectodomain C-flanking cysteine-rich region in dToll[16] as controlling the activity of dToll in signal transduction. In three dominant
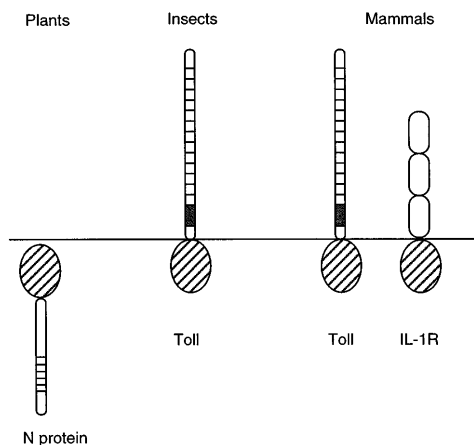
Figure 1 Ancient immune defence systems of plants, insects and vertebrates. A homologous immune response system based on the Toll signalling domain is used in plants, insects and vertebrates. In mammals, Toll induces signals required for the activation of both an innate and an adaptive immune response (see text). The figure is modified from ref. 6. Diagonal hatching represents the Toll signalling domain; striped rods, leucine-rich repeat domains; black rectangles, C-terminal cysteine-rich domains; white ellipsoids, immunoglobulin domains.

mutations, one of four conserved cysteine residues was changed to a tyrosine, rendering dToll constitutively active[16]. In another dominant mutant allele of dToll, this region is deleted altogether, together with the rest of the extracellular part of dToll[18], indicating that a constitutively active dominant mutant of dToll can be generated by either disrupting the disulphide bonds in the C-terminal region flanking the ectodomain or deleting this region and the LRR region of the ectodomain[16,18]. As this region is conserved in hToll (Fig. 2b), we constructed a recombinant hToll receptor that mimics features of the dominant-positive mutants of dToll in which the ectodomain of the mouse CD4 (mCD4) molecule was spliced with the transmembrane and cytoplasmic domains of hToll so that almost all of the ectodomain of hToll (Fig. 4a) and most of the C-terminal flanking region, including three of the four conserved cysteine residues, were removed. Mouse CD4 acted as an epitope tag to monitor the expression of the construct in human cells by using monoclonal antibodies. This chimaeric construct was cloned into an expression vector and transfected into the human monocytic cell line THP-I which endogenously expresses hToll and should thus contain the downstream signalling components. Seven stably transfected clones were selected for further analysis. Expression of the CD4/hToll chimaera in these clones was identified by surface mCD4 expression detected by FACS analysis (Fig. 4b). We used RT-

PCR on seven independent clones to test whether the dominant-positive mutant of hToll can induce expression of NF-κB-controlled immune-response genes; untransfected or mock-transfected THP-I cells acted as controls. Each experiment was repeated 8–12 times and the results from two representative CD4/hToll transfected clones and untransfected TH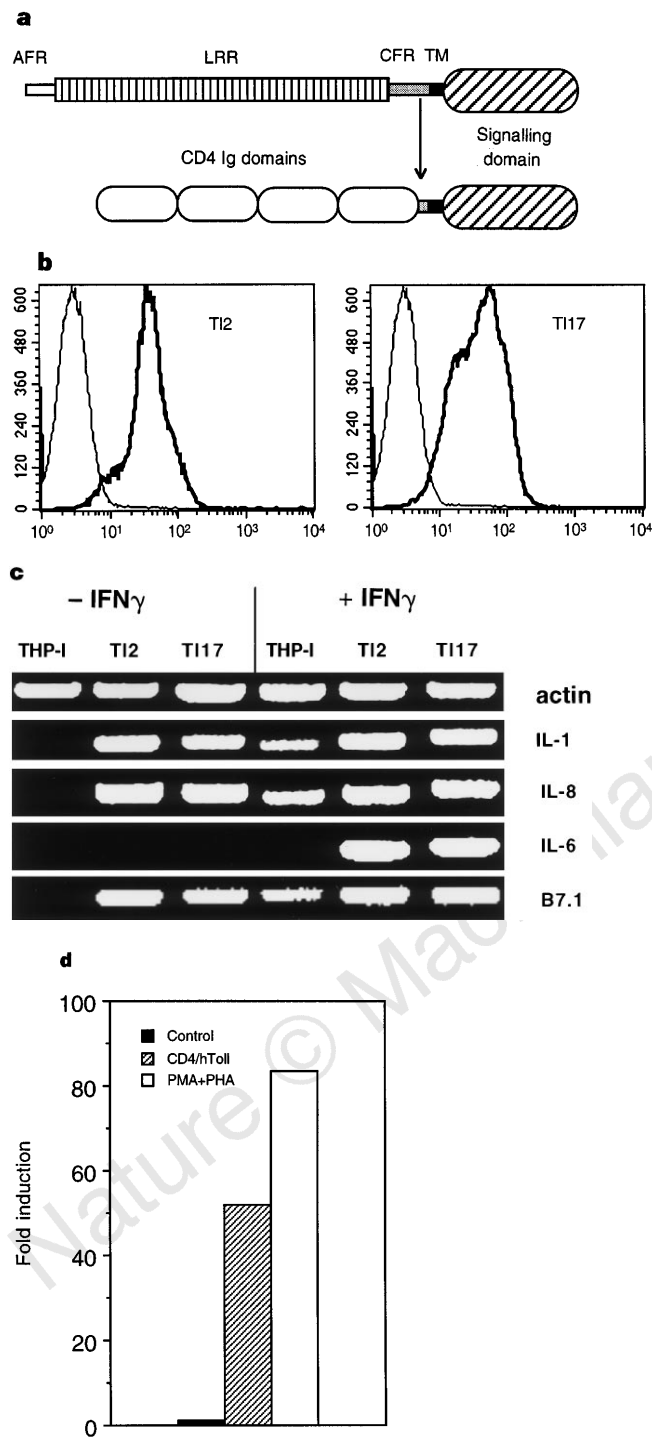P-I cells are shown in Fig. 4c. Expression of a constitutively active mutant of hToll caused induction of the IL-1, IL-8 and B7.1 genes. In the presence of interferon (IFN)-γ, IL-6 mRNA was also strongly induced, which was not due to induction of IL-1 by IFN-γ because during the time of the experiments (8–20 hours) IL-1 was induced in control cells but IL-6 was not. Other genes that might be controlled by signalling through hToll, for example inducible nitric oxide synthase (iNOS) and B7.2, although expressed in the transfectants, could not be tested in this system because of constitutive expression (B7.2) or clonal heterogeneity in the bulk population of control cells (iNOS).

As dToll signals through the NF-κB pathway, we tested whether hToll could induce activation of NF-κB. The CD4/hToll chimaeric construct was transiently transfected into Jurkat cells together with an NF-κB-promoter-driven luciferase reporter gene. As shown in Fig. 4d, expression of CD4/hToll causes induction of NF-κB, confirming that Toll/NF-κB signalling is conserved in *Drosophila* and humans.

a

```
MISASPLAGT LIPAMAFLSC VKPKSWEPCV EVVPNITYQC MELNFYKIPD NLPFSTKNLH    60
LSFNPLRHLG SYSFFSFPKL QVLDLSRCEI QTIEDGAYQS LSHLSTLILT GNPIQSLALG   120
AFSGLSSLQK LVAVETNLAS LENFPIGHLK TLKELNVAHN LIQSFKLPEY FSNLTNLEHL   180
DLSSNKIQSI YCTDLRVLHQ MPLLNLSLDL SLNPMNFIQP GAFKEIRLHK LTLRNNFDSL   240
NVMKTCIQGS GWFRSPFVWV SGENLEMKET WKSLTNLLRG LCNLTIEEFR LAYLDYYLDD   300
IIDLFNCLNQ MFLHFFPLESV TIERVKDFSY NFGWQHLELV NCKFGQFPTL KLKSLKRLTF   360
TSNKGGNAFS EVDLPSLEFL DLSRNGLSFK GCCSQSDFGT TSLKYLDLSF NGVITMSSNF   420
LGLEQLEHLD FQHSNLKQMS EFSVFLSLRN LMYLDISHTH TRVAFNGIFN GLSSLEVLKM   480
AGNSFQENFL PDIFTELRNL TFLDLSQCQL EQLSPTAFNS LSSLQVLNMS HNNFFSLDTF   540
PYKCLNSLQV LDYSLNHIMT SKKQELQHFP SSLAFLNLTQ NDFACTCEHQ SFLQWIKDQR   600
QLLVEVERME CATPSDKQGM PVLSLNITCQ MNK**TIIGVSV LSVLVVSVVA VLV**YKFYFHL   660
MLLAGCIKYG RGENIYDAFV IYSSQDEDWV RNELVKNLEE GVPPFQLCLH YRDFIPGVAI   720
AANIIHEGFH KSRKVIVVVS QHFIQSRWCI FEYEIAQTWQ FLSSRAGIIF IVLQKVEKTL   780
LRQQVELYRL LSRNTYLEWE DSVLGRHIFW RRLRKALLDG KSWNPEGTVG TGCNWQEATSI  841
```

b

```
 30  VEVVPNITYQCMEL-NFYK-IPDNLPFSTKNLHLSFNPLRHLGSYSFFSFPKLQVLDLSRCEIQTIEDGAYQSLSHLSTLILTGNPIQSLALGAFSGLSS  127
     :|:::...:|.:.:|:.:|::| .:|.|:::|::|.|::::|.||:.|:|::.|::|:::::| :. |.::: |::.|:.::|
196  LENLESIEFGSNKLRQMPRGIFGKMP-KLKQLNLWSNQLHNLTKHDFEGATSVLGIDIHDNGIEQLPHDVFAHLTNVTDINLSANLFRSLPQGLFDHNKH  294

128  LQKLVAVETNLASLENFPIGHLKTLKELNVAHNLIQSFK-LP-EYFSNLTNLEHLDLDLSSNKIQSIYCTDLRVLHQMPLLNLSLDLSLNPMNFIQPGAF-K  224
     |:::...|::.|::|.::| :.  |||.:| .::.|::|:.:.| |::|.||.||:::||:|:: |||||||:: |::: |::  |:|:| |:||.:
295  LNEVRLMNNRVP-LATLPSRLFANQPELQILR-LRAELQSLPGDLFEHSTQITNISLGDNLLKTL-PATL-LEHQVNLL--SLDLSNNRLTHLPDSLFAH  388

225  EIRLHKLTLRNNFDSLNVMKTCIQGS-GWFRSPFVWVSGENLEMKETWKSL-TNLLRGLCNLTIEEFRL--AYLDYYLD-DIIDLFNCLNQMFLHFFPLES  319
     ...|.:|.|::|::| :.|::  |::|::: |:.:| : ||:::.:|.::|: |:.:::|::| | :::.||:  | .:::|||:| : |::.: |:|||:
389  TTNLTDLRLEDNL--LTGISGDIFSNLGNLVT-LV-MSRNRLRTIDSRAFVSTNGLRHL-HLDHNDIDLQQPLLDIMLQTQINSPFGYMHGL-LTLNLRN  482

320  VTIERVKDFSYNFGWQH--LEL--VNCKFGQFPTL---KLKSLKRLTFTSNKGGNAFSEVDLPSLEFLDLSRNGLSFKGCCSQSDFGTTSL--KYLDLSF  410
     :.| :| || :| :|::.:|: |::.  .:.::::   .|  :|..   : |::|  |::::||    | :.:.||    |   |:|...| |::|.|:
483  NSI--I--FVYN-DWKNTMLQLRELDLSYNNISSLGYEDLAFLSQNRLHVNMTHNKIRRIALPE-D-VHL---GEGYNNNLVHVDLNDNPLVCDCTILWF  572

411  NGVI--TMSSNF---LGLEQLEHLDFQHSNLKQMSEFSVFLSLRNLMY-LDIS-HTHTRVAFNGIFNGLSSLE---VLK-MAGNSFQENFLPDIFTELRN  499
     :: .::.::|. : |:.::|  ::||. |:|.::|.|:. :.|:::|.|:| ::.|: .| :::: | ::| :||:| ||:|:.:|::::.|
573  IQLVRGVHKPQYSRQFKL-RTDRLVCSQPNVLEGTPVRQIEP-QTLICPLDFSDDPRERKCPRGCNCHVRTYDKALVINCHSGNLTHVPRLPNLHKNMQL  670

500  LTFLDLSQCQLEQLSPTAFNSLSSLQVLNMSHNNFFSLDTFPYKCLNSLQVLDYSLNHIMTSKKQELQHFPSSLAF--LNLTQNDFACTCEHQSFLQWIK  597
     :.|:|:.|  |:|:::. .:|  |:: |::.::| | |::|:|: .|  .:|| |:  |.::|:| .: |  |. ...:|.:. ..: |
671  ME-LHLENNTLLRLPSANTPGYESVTSLHLAGNNLTSIDVDQLP--TNLTHLDISWNHLQMTANTVLGFLNRTMKWRSVK**LSGNPWMCDCTAKPLLLFTQ**  767

598  DQRQLLVEVERMECATPSDKQGMPVLSLNITCQMNKTI---IGVSV-LSVLVVSVSVVAVLVYKF--------YFH-LML-LAGCIKYGRGENIYDAFVIYS  683
     |:::::.|.|::..|...  .|  | ||.|:.. :::.   :|:|:  |:.  :.|. ||.:||| |||  | || |.|:| |.|:.:
768  <u>DNFERIGDRNEMMCVNAEMPTRMVELSTND</u>ICPAEKGVFIALAVVIALTGLLAGFTAALVYKFFQTEIKIWLYAHNLLLWFVTEEDLDKDKK-FDAFISYS  866

684  SQDEDWVRNELVKNLEEGVPPFQLCLHYRDFIPGVAIAANIIHEGFHKSRKVIVVVSQHFIQSRWCIFEYEIAQTWQFLSSRAGIIFIVLQKV-EKTLLR  782
     :||:::.|:|.:.. |::::.|.||||:.|.| ||:|.:.|:..:.|   .|::|::::.| |::.|||:.|:.|.|||:|.|||..:::|:|: .|.||
867  HKDQSFIEDYLVPQLEHGPQKFQLCVHERDWLVGGHIPENIMR-SVADSRRTIIVLSQNFIKSEWARLEFRAAHRSALNEGRSRIIVIIYSDIGDVEKLD  965

783  QQVELYRLLSRNTYLEWEDSVLGRHIFWRRLRKALLDGKSWNPEGTVGTG  832
     :::. |..||:||.|: ||.|.|:|| |::.|.:.: :.|.|:|:|
966  EELKAY--LKMNTYLKWGDP-----WFWDKLRFALPH---RRPVGNIGNG 1005
```

Above residue position 598/768 in sequence b, an asterisk (*) is marked.

Figure 2 a, Amino-acid sequence of human Toll. The predicted signal peptide (residues 1–23) is underlined. The predicted transmembrane segment (residues 634–653) is in bold and underlined. The cDNA sequence (Genbank accession U93091) contains two coding regions: one has an open reading frame (position 104-2,627) and encodes the 841-amino-acid human Toll. The second coding region (position 3,970-4,489) contains a pseudogene, *LINE-1* reverse-transcriptase homologue (not shown). **b**, Alignment of the protein sequences of hToll (top) and dToll (bottom). A vertical line indicates identity and a colon represents similarity between the corresponding pair of amino-acid residues. The location of the cysteine-rich domain is indicated by underlining in the *Drosophila* sequence, and the site used to prepare chimaeric hToll is indicated by an asterisk above the hToll sequence. Parameters of the alignment were as follows: Ktuple: 2, gap penalty: 3, gap length penalty: 8.

Figure 3 Northern blots of poly(A)+ RNA demonstrate predominant expression of hToll in spleen (Sp) and peripheral blood leukocytes (PBL). The upper band corresponds to the length of the hToll cDNA we have cloned (~5 kb); the lower band is ~4 kb. The difference in the sizes of the two bands is probably due to the differential processing of the 3'UTR, as detected by RT-PCR analysis of this region (not shown) and might be due to the presence of the *LINE-1* reverse-transcriptase pseudogene in this region. Tissues examined: Hr: heart; Br: brain; Pl: placenta; Lu: lung; Li: liver; SM; skeletal muscle; Ki: kidney; Pa: pancreas; Th: thymus; Pr: prostate; Te: testis; Ov: ovary; SI: small intestine; Co: colon.

395

Figure 4 Signalling function of hToll. **a**, Representation of the mouse CD4/hToll chimaera that renders the signalling domain of hToll constitutively active. hToll cut at the site of the asterisk in Fig. 2b. AFR and CFR represent the amino-terminal flanking and carboxy-terminal cysteine-rich regions. TM, transmembrane region. **b**, FACS analysis of mCD4/hToll transfected clones Tl2 (left) and Tl17 (right, dark lines) stained with a monoclonal antibody for mouse CD4. Untransfected THP-1 cells (fine line) stained with the same antibody were used as a control. **c**, RT-PCR analysis of the gene expression induced by signalling through hToll. IFN-γ treatment was 8–20 h in various experiments which gave the same results using 20 ng human recombinant IFN-γ (Gibco BRL). **d**, Activation of NF-κB by the dominantly active CD4/hToll chimaera in Jurkat cells, as measured by reporter gene activity. Controls were transfected with empty vector (negative controls) and activated with phorbol myristyl acetate (PMA; 20 ng ml⁻¹) and phytohaemagglutinin (PHA; 2 μg ml⁻¹) for 3–4 h. All transfectants contained a pB2XLuc reporter construct.

These findings indicate that Toll functions in vertebrates as a non-clonal receptor of the immune system. Pro-inflammatory cytokines such as IL-1, IL-6 and IL-8 are induced early during infection but, apart from CD14, the cellular receptors that are activated during this initial phase have not yet been identified. It has been argued that the innate immune response has been preserved in vertebrates not only to resist infection before adaptive immunity is induced, but also to induce signals that inform the adaptive system of the presence of pathogens within the internal milieu[1,19,20]. We have shown that the Toll/NF-κB host defence pathway is conserved from *Drosophila* to humans and that the human homologue of the *Drosophila* Toll protein can induce signals for activating both an innate and an adaptive immune response in vertebrates. □

Methods

Cloning and sequence analysis. BLAST searches of the NCBI sequence database were made using sequence profiles of the dToll/IL-1R signalling domains. Matching sequence (Genbank accession number H48602) corresponded to the IMAGE consortium clone number 202057. PCR primers were designed based on the EST sequence and used to amplify a 220-bp fragment from a human spleen cDNA library (Clontech). The PCR-amplified fragment was used to screen a human spleen cDNA library as described[13]. The 5′ end of the cDNA was cloned by 5′-RACE protocol as described[14]. Cloned products were subjected to automated sequencing (ABI).

RNA expression. Tissue northern blots were purchased from Clontech and analysed according to the manufacturer's instructions. The presence of equal amounts of RNA on the blots was confirmed by hybridization with a β-actin probe.

RT-PCR analysis. Total RNA was isolated using TRIzol Reagent (Gibco BRL) and 0.5 μg was used for reverse transcription using AMV reverse transcriptase (Seikagaku America) or Superscript reverse transcriptase (Gibco BRL). One-tenth to one-twentieth of this reaction was used as a template for PCR amplification with *Taq* polymerase (Gibco) for 25–35 cycles at 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min. Sequences of the primers were taken from ref. 21.

Construction of mCD4/hToll chimaera. An *Sph*I site was introduced between V365 and N366 in the mouse CD4 gene by site-directed mutagenesis. An internal *Sph*I site in hToll at the position between M620 and P621 was used for the construction of the chimaera.

Monoclonal antibodies and FACS analysis. Anti-mouse CD4 antibody (L3T4) (Pharmingen) was used for staining. Flow cytometry analysis was done on a FACScan (Becton–Dickinson).

NF-κB assay. Jurkat cells were transiently transfected using Lipofectamine reagent (Gibco BRL) according to the manufacturer's instructions. 1 μg pB2XLuc plasmid, containing NF-κB-driven luciferase gene was cotransfected with 1 μg pSRαN expression vector with or without (control sample) the insert encoding CD4/Toll chimaera. For a positive control, cells were treated with PMA (phorbol myristyl acetate; 20 ng ml⁻¹) and PHA (phytohaemagglutinin; 2 μg ml⁻¹) for 3–4 h. Cells were lysed 2 or 3 days later for measurement of luciferase activity using reagents from Promega.

1. Janeway, C. A. Jr Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symp. Quant. Biol.* **54,** 1–13 (1989).
2. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette *spatzle/toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86,** 973–983 (1996).
3. Rosetto, M., Engström, Y., Baldari, C. T., Telford, J. L. & Hultmark, D. Signals from the IL-1 receptor homolog, Toll, can activate an immune response in a *Drosophila* hemocyte cell line. *Biochem. Biophys. Res. Commun.* **209,** 111–116 (1995).
4. Ip, Y. T. & Levine, M. Molecular genetics of *Drosophila* immunity. *Curr. Opin. Genet. Dev.* **4,** 672–677 (1994).
5. Wasserman, S. A. A conserved signal transduction pathway regulating the activity of rel-like proteins dorsal and NF-κB. *Mol. Biol. Cell* **4,** 767–771 (1993).
6. Hultmark, D. Insect immunology: Ancient relationships. *Nature* **367,** 116–117 (1994).
7. Belvin, M. P. & Anderson, K. V. A conserved signaling pathway: the *Drosophila* Toll–dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12,** 393–416 (1996).
8. Morisato, D. & Anderson, K. V. The *spatzle* gene encodes a component of the extracellular signaling pathway establishing the dorsal–ventral pattern of the *Drosophila* embryo. *Cell* **76,** 677–688 (1994).
9. Gay, N. J. & Keith, F. J. *Drosophila* Toll and IL-1 receptor. *Nature* **351,** 355–356 (1991).
10. Whitham, S. *et al.* The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* **78,** 1101–1115 (1994).
11. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. Issues in searching molecular sequence databases. *Nature Genet.* **6,** 119–129 (1994).

12. Lennon, G., Auffray, C., Polymeropoulos, M. & Soares, M. B. The IMAGE Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152 (1996).
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
14. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.(eds) *PCR Protocols: A Guide to Methods and Applications* (Academic, Orlando, 1992).
15. Miyake, K., Yamashita, Y., Ogata, M., Sudo, T. & Kimoto, M. RP105, a novel B cell surface molecule implicated in B cell activation, is a member of the leucine-rich repeat protein family. *J. Immunol.* **154**, 3333–3340 (1995).
16. Schneider, D. S., Hudson, K. L., Lin, T. Y. & Anderson, K. V. Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal–ventral polarity in the *Drosophila* embryo. *Genes Dev.* **5**, 797–807 (1991).
17. Takahashi, N., Takahashi, Y. & Putnam, F. W. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich alpha 2-glycoprotein of human serum. *Proc. Natl Acad. Sci. USA* **82**, 1906–1910 (1985).
18. Winans, K. A. & Hashimoto, C. Ventralization of the *Drosophila* embryo by deletion of extracellular leucine-rich repeats in the Toll protein. *Mol. Biol. Cell.* **6**, 587–596 (1995).
19. Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* **272**, 50–53 (1996).
20. Medzhitov, R. & Janeway, C. A. Jr On the semantics of immune recognition. *Res. Immunol.* **147**, 208–214 (1996).
21. Yamamura, M. *et al.* Defining protective responses to pathogens: cytokine profiles in leprosy lesions. *Science* **254**, 277–279 (1991).

A protein related to splicing factor U2AF³⁵ that interacts with U2AF⁶⁵ and SR proteins in splicing of pre-mRNA

Hélène Tronchère*, Jiwu Wang* & Xiang-Dong Fu

Division of Cellular and Molecular Medicine, Department and School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA

* These authors contributed equally to this work.

Recognition of a functional 3′ splice site in pre-mRNA splicing requires a heterodimer of the proteins U2AF⁶⁵/U2AF³⁵. U2AF⁶⁵ binds to RNA at the polypyrimidine tract[1,2], whereas U2AF³⁵ is thought to interact through its arginine/serine-rich (RS) domain with other RS-domain-containing factors bound at the 5′ splice site, assembled in splicing enhancer complexes, or associated with the U4/U6.U5 small nuclear ribonucleoprotein complex[3–7]. It is unclear, however, how such network interactions can all be established through the small RS domain in U2AF³⁵. Here we describe the function of a U2AF³⁵-related protein (Urp), which is the human homologue of a mouse imprinted gene. Nuclear extracts depleted of Urp are defective in splicing, but activity can be restored by addition of recombinant Urp. U2AF³⁵ could not replace Urp in complementation, indicating that their functions do not overlap. Co-immunodepletion showed that Urp is associated with the U2AF⁶⁵/U2AF³⁵ heterodimer. Binding studies revealed that Urp specifically interacts with U2AF⁶⁵ through a U2AF³⁵-homologous region and with SR proteins (a large family of RS-domain-containing proteins) through its RS domain. Therefore, Urp and U2AF³⁵ may independently position RS-domain-containing factors within spliceosomes.

We previously purified and cloned a mammalian serine kinase, SRPK1, which is highly specific for the RS-rich superfamily of splicing factors[8–10]. This stringent substrate specificity enabled us to do a yeast two-hybrid screen to search for new RS-domain-containing proteins. We isolated a positive clone from a HeLa complementary cDNA library that contains a typical RS domain at the carboxy terminus (Fig. 1a). The recombinant protein was efficiently phosphorylated in the RS domain, like other RS-domain proteins, by

SRPK1 *in vitro* (data not shown). This protein of 483 amino acids does not contain any RNA-recognition motif characteristic of the SR family of splicing factors[11,12]; instead, it contains two stretches of sequences, H1 and H2 (boxed in Fig. 1b), homologous to the human splicing factor U2AF³⁵. We therefore named it Urp, for U2AF³⁵-related protein. Northern blotting analysis indicates that both U2AF³⁵ and Urp are ubiquitously expressed in all human tissues and cell lines examined (data not shown). Homology searches revealed that Urp had been independently cloned from humans as U2AF1-RS2 (ref. 13), which is nearly identical (94% amino-acid identity) to another human gene, U2AF1-RS1 (ref. 13) (Fig. 1b). These two human genes are equally homologous (~80% identical) to two mouse genes[14], one of which is imprinted[15], and the other has a high transmission distortion in interspecific backcross progeny[16]. Thus, several U2AF³⁵-related proteins are expressed in mammalian cells and may be members of the superfamily of RS-domain-containing splicing factors[11].

To determine the function of Urp, we raised rabbit polyclonal antibodies against recombinant Urp expressed in bacteria. Immunoblotting of HeLa cell nuclear extracts demonstrated that anti-Urp antibodies specifically detect a single band of relative molecular mass 66K (Fig. 2a), which corresponds to endogenous Urp as it matches the size of Urp translated in reticulocyte lysate (data not shown). To see whether Urp is essential for splicing, we constructed affinity columns with anti-Urp antibodies or preimmune serum. Control depletion with preimmune serum had no effect, but passing nuclear extract through an anti-Urp affinity column efficiently depleted it of Urp, as evidenced by western blotting and complete inhibition of splicing (data not shown). Although this result is consistent with Urp being a splicing factor, this Urp-depleted extract could not be complemented by recombinant Urp, probably because of co-depletion of other essential splicing factors (see below). We therefore did the immunodepletion in the presence of high salt (0.5 M KCl) to minimize co-depletion. As shown in Fig. 2c, human β-globin pre-mRNA splicing in Urp-depleted nuclear extract was significantly decreased (compare lanes 1 and 2). The second step of the splicing reaction, which generates released lariat intermediate and
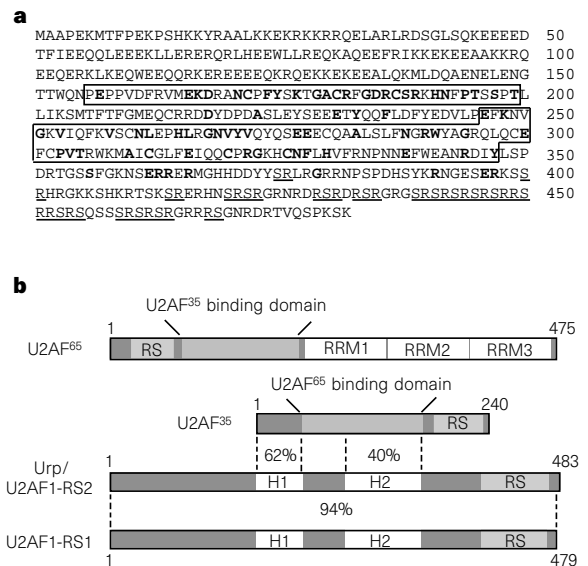
a

```
MAAPEKMTFPEKPSHKKYRAALKKEKRKKRRQELARLRDSGLSQKEEED    50
TFIEEQQLEEEKLLERERQRLHEEWLLREQKAQEEFRIKKEKEEAAKKRQ   100
EEQERKLKEQWEEQQRKEREEEEQKRQEKKEKEEALQKMLDQAENELENG   150
TTWQNPEPPVDFRVMEKDRANCPFYSKTGACRFGDRCSRKHNFPTSSPTL   200
LIKSMTFTTFGMEQCRRDDYDPDASLEYSEEETYQQFLDFYEDVLPEFKNV   250
GKVIQFKVSCNLEPHLRGNVYVQYQSEEECQAALSLFNGRWYAGRQLQCE   300
FCPVTRWKMAICGLFEIQQCPRGKHCNFLHVFRNPNNEFWEANRDIYLSP   350
DRTGSSFGKNSERRERMGHHDDYYSRLRGRRNPSPDHSYKRNGESERKSS   400
RHRGKKSHKRTSKSRERHNSRSRGRNRDRSRDRSRGRGSRSRSRSRSRRS   450
RRSRSQSSSRSRSRGRRRSGNRDRTVQSPKSK                    483
```

b

Figure 1 Urp structure and sequence comparison with U2AF³⁵. **a**, Amino-acid sequence of Urp. It is identical to U2AF1-RS2 (GenBank D49677; ref. 13). The two U2AF³⁵ homologous domains, H1 and H2, are boxed and amino-acid residues identical to those in U2AF³⁵ are shown in bold. Repeats of SR or RS dipeptides are underlined. **b**, Domain structure of the large and small subunits of U2AF and comparison of human U2AF³⁵, Urp/U2AF1-RS2 and U2AF1-RS1. Percentage identity between homologous domains is indicated.