

Research

Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (*TLR4*)

Irina Smirnova*, Alexander Poltorak*, Edward KL Chan[†],
Colleen McBride* and Bruce Beutler*

Addresses: *The University of Texas Southwestern Medical Center and the Howard Hughes Medical Institute, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9050, USA. [†]The Scripps Research Institute, Department of Molecular and Experimental Medicine, 10550 N Torrey Pines Road, La Jolla, California 92037, USA.

Correspondence: Bruce Beutler. E-mail: beutler@howie.swmed.edu

Published: 27 April 2000

Genome Biology 2000, **1**(1):research002.1–002.10

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2000/1/1/research/002>

© GenomeBiology.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 29 December 1999

Revised: 10 February 2000

Accepted: 15 February 2000

Abstract

Background: Differences in responses to bacterial surface lipopolysaccharides (LPSs) are apparent between and within mammalian species. It has been shown in mice that resistance to LPS is caused by defects in the Toll-like receptor 4 gene (*Tlr4*), the product of which is thought to bind LPS and mediate LPS signal transduction in immune system cells.

Results: We have sequenced the Toll-like receptor 4 gene of humans (*TLR4*; 19.0 kilobases, kb) and mice (*Tlr4*; 91.7 kb), as well as the coding region and splice junctions of *Tlr4* from 35 mouse (*Mus musculus*) strains, from the chimpanzee and from the baboon. No other discernible genes or regions of interspecies conservation lie close to *Tlr4* and, in both humans and mice, flanking sequences and introns are rich in repeats of retroviral origin. Interstrain analyses reveal that Tlr4 is a polymorphic protein and that the extracellular domain is far more variable than the cytoplasmic domain, both among strains and among species. The cytoplasmic domain of the Tlr4 protein is highly variable at the carboxy-terminal end.

Conclusions: We suggest that selective evolutionary pressure exerted by microbes expressing structurally distinguishable LPS molecules has produced the high level of variability in the Tlr4 extracellular domain. The highly variable carboxy-terminal region of the cytoplasmic domain is likely to determine the magnitude of the response to LPS within a species.

Background

Cells of the innate immune system sense host invasion by detecting structural determinants that are broadly conserved among pathogens of a given phylogenetic group [1]. The lipopolysaccharides (LPS or endotoxin) that decorate the outer membrane of Gram-negative bacteria are excellent examples of such determinants. In response to minute concentrations of LPS derived from certain Gram-negative bacteria, macrophages secrete cytokines such as tumor necrosis

factor (TNF), interleukin-1 (IL-1) and IL-6, which contribute to the containment of infection and help to initiate a specific immune response. On the other hand, overstimulation of the innate immune system through this channel can lead to acute systemic inflammation and shock [2,3].

Dramatic differences in LPS responses are apparent between closely related species [4], and there are substantial differences even among genetically heterogeneous members of the

comment

reviews

reports

deposited research

refereed research

interactions

information

same species [5]. For example, whereas humans and chimpanzees are generally considered to be very sensitive to LPS [4], baboons and most other primates are highly resistant. It is likely that part of the difference in sensitivity may be explained at a very proximal level, although differences in responses to cytokines (for example TNF) may also have a role. Neither interspecific differences nor interindividual differences in LPS responses have, until recently, been accessible to systematic genetic analysis. Advances in understanding LPS signal transduction now permit these issues to be examined.

Although LPS was once thought to exert its effects through intercalation into biological membranes, or to bind to many different receptors on the cell surface, it has been clear for nearly three decades that there is, in fact, a single biochemical pathway for LPS detection. This was indicated by the observation that allelic [6] mutations of a single gene (*Lps*) in mice could entirely abrogate the response to LPS, and did so with great specificity. Mice of the strains C3H/HeJ [7] and C57BL/10ScCr [8] are highly resistant to LPS, showing none of the usual biological effects, yet respond normally to other bacterial products and to most cytokines induced by LPS [9].

We recently identified the *Lps* locus through positional cloning [10] and showed that the LPS-resistance phenotype was caused by defects in the Toll-like receptor 4 gene (*Tlr4*) [11]. In C3H/HeJ mice, a point mutation (P712H; single-letter amino-acid notation) modifies the protein within the cytoplasmic domain, creating a codominant inhibitory effect on LPS signal transduction. In C57BL/10ScCr mice the gene is deleted, yielding a recessive abolition of the LPS response. Subsequently, overexpression of the wild-type Tlr4 protein was found to enhance LPS signal transduction in wild-type macrophages, lowering the effective concentration (EC₅₀) for LPS by a factor of 30, whereas overexpression of the Tlr4^{Lps-d} isoform represented in C3H/HeJ mice almost completely suppresses signaling [12]. Furthermore, genetic complementation studies have demonstrated that LPS and Tlr4 enter into close physical proximity in the course of signal transduction - Tlr4 appears to bind directly to LPS. Hence, the species origin of Tlr4 is the sole determinant of species preference for a given LPS structure [13].

Mice of the C3H/HeJ and C57BL/10ScCr strains are abnormally susceptible to infection by certain Gram-negative bacteria, suggesting that timely recognition of LPS is essential for successful containment of infection [14,15]. Because deleterious mutations of *Tlr4* have become fixed spontaneously in two strains of mice, we considered it possible that other functionally important mutations might be identified in mice and humans. Moreover, information on the degree of *Tlr4* polymorphism in these and other species might allow inferences about the importance of different Tlr4 domains. Accordingly, we decided to sequence the entire *Tlr4* gene of both humans and mice, and to survey genetic variation at the locus in each species. We also examined the *TLR4* sequence

from two species of subhuman primates that have dramatically different responses to LPS.

Results

Overall structure of *TLR4* and *Tlr4*

The mouse *Tlr4* gene is somewhat longer than its human counterpart *TLR4*, owing to the greater length of intronic sequence - 15,337 base pairs (bp) from beginning to end of the transcribed sequence in the mouse, compared with 11,467 bp in the human. There are three exons in *Tlr4*, and each corresponds to a homologous sequence in the human gene. Rock *et al.* [16] reported a human cDNA sequence (GenBank accession number U88880) that includes a fourth exon, positioned between the 'normal' first and second introns. When included in the processed transcript, however, this exon specifies early termination of the polypeptide chain. Although it is possible that translation is initiated distal to the added stop codon, and that a shorter product results in the human than in the mouse, this would be unusual, given the length of the 5' untranslated region (UTR) that would then exist and the presence of multiple upstream initiation codons. Moreover, there is no murine sequence homologous to the alternative second exon of the human gene. The biological significance of this exon is therefore unclear and, in all likelihood, its inclusion in the mRNA leads to the formation of a nonfunctional protein.

Neither the human nor the mouse gene has a TATA element or CAAT box in the proximal promoter region. A number of conserved promoter and enhancer motifs are apparent on alignment of the murine and human 5' flanking sequences, and are described in detail elsewhere [17]. Both *Tlr4* and *TLR4* lie amid repetitive sequences of retroviral origin, and no other genes are detected close to either of them using homology searches or the gene prediction algorithm GRAIL. In Figure 1, the grayscale images of the human and mouse genes call attention to the repetitive elements in the region and illustrate the relationship between exons and spacing in the two species.

Genetic variation at the mouse *Tlr4* locus

Among 35 strains of *Mus musculus*, 10 different alleles were identified on the basis of mutations at 22 sites with respect to the reference sequence. Of these, 13 create amino-acid substitutions (Table 1, Figure 2). The most common murine allele was represented at a frequency of only 69%. The ancestry of different *Tlr4* alleles can be traced by haplotype analysis, as many deviations from the reference allele occur in conjunction with one another. A plausible arrangement of strain relationships is presented in Figure 3. Some strains have accumulated many more mutations than others. For example, the P/J strain *Tlr4* gene exhibits 11 mutations that distinguish it from the most common haplotype, six of them specifying changes in the Tlr4 amino-acid sequence; the SEA/GnJ strain differs by nine mutations, and the strains NZW/J and VM/Dk, which are identical to one another,

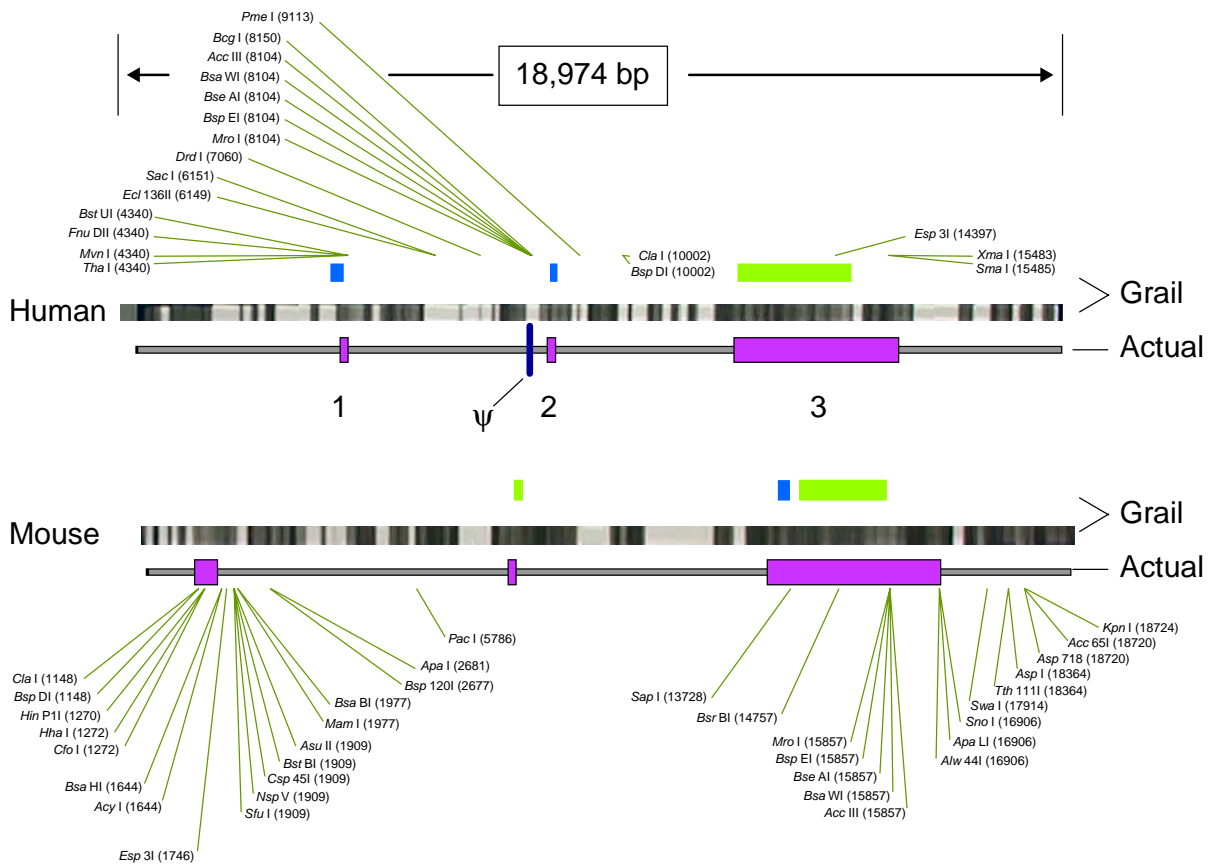


Figure 1

The landscape of genomic DNA in the region of human and mouse *Tlr4* genes. Approximately 19 kb of sequence is shown from each species. Exons are numbered 1 to 3. In the human gene model, an added exon (ψ) is also portrayed, as found in the alternative sequence of Rock *et al.* [16], in which early truncation of the protein is predicted. The grayscale image was generated using X-GRAIL, version 1.3c, and depicts GC content as well as repetitive elements (both complex and simple) identified by RepeatMasker (which appear as unbroken stretches of white). GC-rich areas appear darker than AT-rich areas. Grail exons are shown in green (highest quality) and blue (intermediate quality) above each grayscale image. Restriction sites indicate enzymes that cut at single sites within the interval.

differ from the most common haplotype at six sites. Shared mutations suggest that introgression took place after mutational separation had occurred, leading to the introduction of groups of mutations by genetic recombination. Hence, mice of the P/J, NZW/J, and VM/Dk strains have several mutations that are observed in the A/J and BALB/c strains, but also lack some of the mutations of the latter strains and have unique mutations of their own.

Most of the murine mutations reside within exon 3, and only two substitutions are noted that modify the cytoplasmic domain (Figure 2). Of these, however, one mutation (R761H) is fairly common among the strains surveyed, and the corresponding residue has been reported as an H in the hamster. A single conservative substitution (V637I) was noted within the transmembrane domain of the P/J strain.

Anthropoid ape and lower primate *TLR4* sequences, and their relationship to the human and rodent sequences

The human and chimpanzee amino-acid sequences are nearly identical over the interval studied, distinguished by only three substitutions. The baboon sequence is 93.5% identical to the human in the extracellular domain, differs in the transmembrane domain by one substitution out of 30 residues, and differs in the proximal cytoplasmic domain by only one residue in 155. At the carboxyl terminus, however, homology is badly disrupted, so that 16 of the last 21 human residues are not replicated in the baboon protein, which is 13 amino acids shorter than the human protein. Similarly, among rodents, the carboxyl terminus of the protein is the least conserved. Overall, the order of conservation with respect to domain is: proximal cytoplasmic domain > transmembrane domain > extracellular domain > distal cytoplasmic domain (Table 2, Figure 4).

Table 1

Polymorphisms of <i>Tlr4</i> in mice					
Mouse strain*	Genomic nucleotide affected	Exon/intron affected	Altered amino acid	Affected receptor domain [†]	Conserved? [‡]
10	26,400 A → G	Exon 1	-	-	
4, 8, 20, 21, 22	37,685: (T) ₁₀	Intron 2	-	-	
34, 35	37,685: (T) ₁₂	Intron 2	-	-	
23	37,754 G → A	Exon 3	D94N	Ecto	Yes
4, 8, 10, 20, 21, 34, 35	38,101 G → A	Exon 3	M209I	Ecto	Yes
21	38,130 A → G	Exon 3	D219G	Ecto	No
4, 8, 21	38,234 G → A	Exon 3	V254I	Ecto	Yes
10	38,584 A → G	Exon 3	-	-	
21	38,742 A → T	Exon 3	Q423L	Ecto	Yes
10	38,794 G → A	Exon 3	-	-	
10	38,903 G → T	Exon 3	A477S	Ecto	Yes
18	39,020 A → G	Exon 3	T516A	Ecto	Yes
4, 8, 10, 20, 21, 22, 34, 35	39,199 C → T	Exon 3	-	-	
4, 8, 10, 20, 21, 34, 35	39,253 A → C	Exon 3	E593D	Ecto	Yes
23	39,273 A → T	Exon 3	N600I	Ecto	No
10	39,294 C → T	Exon 3	A607V	Ecto	Yes
10	39,383 G → A	Exon 3	V637I	TM	No
10	39,544 G → A	Exon 3	-	-	
19	39,604 T → C	Exon 3	-	-	
4, 8, 20, 21, 34, 35	39,631 C → T	Exon 3	-	-	
4, 8, 20, 21, 34, 35	39,756 G → A	Exon 3	R761H	Cyto	Yes
10	39,907 T → G	Exon 3	N811K	Cyto	Yes

*Strains are as follows: 1, NZO/HILt; 2, SI/Col; 3, DBA/J; 4, A/J; 5, EL/Suz; 6, CBA/J; 7, AKR/J; 8, BALB/c; 9, DDY/Jc; 10, P/J; 11, MRL/Mp; 12, SJL/J; 13, NOD/Lt; 14, 129/J; 15, FL/IRe; 16, MA/MyJ; 17, SWR/J; 18, LP/J; 19, PRO/IRel; 20, SOD/IEi; 21, SEA/Gn; 22, SM/J; 23, KK/HIJ; 24, ST/b; 25, WB/Re; 26, YBR/Ei; 27, FVB/NJ; 28, PL/J; 29, LT/ChReSv; 30, RIIS/J; 31, RF/J; 32, NZB/BINJ; 33, AU/Ss; 34, NZW/Lac; 35, VM/Dk.

[†]Exon 1, 26041-26424; Exon 2, 32397-32563; Exon 3, 37732-41297. ATG at 26335; TGA at 39982. Ecto, extracellular domain; Cyto, cytoplasmic domain; TM, transmembrane domain.

[‡]Yes implies one or two forms; No implies three to five forms among the six mammalian species examined.

Discussion

The pathway by which LPS activates host innate immune defenses has been illuminated by the positional cloning of *Lps* and establishment of its identity with *Tlr4* [11], a mammalian representative of an ancient family of receptors [16] that serve both developmental and defensive functions. The function of *Tlr4* as the LPS signal transducer became clear when mapping [10] and sequencing [11] data revealed *Tlr4* as the only gene in a critical 2.6 megabase (Mb) region and, furthermore, showed that two strains of LPS-resistant mice are homozygous for mutations of *Tlr4* that are absent in closely related LPS-sensitive strains.

In *Drosophila*, the prototypic homolog of the mammalian Toll-like receptors (Toll) defends against fungal infection [18],

whereas the protein 18-wheeler defends against bacterial infection [19]. In the case of Toll, there is no evidence of direct contact between the receptor and the microbial pathogen or its components. Rather, infection activates a proteolytic cascade that leads to the production of an endogenous ligand (Spätzle), which in turn engages Toll. The situation is apparently different in mammals, in which *Tlr4* is clearly a direct interface with the microbial world. As such, the primary structure of a *Tlr4* molecule determines ligand specificity [13], accounting for the well-known observation that mouse cells can recognize tetra-acyl lipid A as an agonist [20] whereas human cells recognize it as an antagonist [20-24].

As mice of the C3H/HeJ strain are highly susceptible to infection by Gram-negative bacteria, it would seem plausible

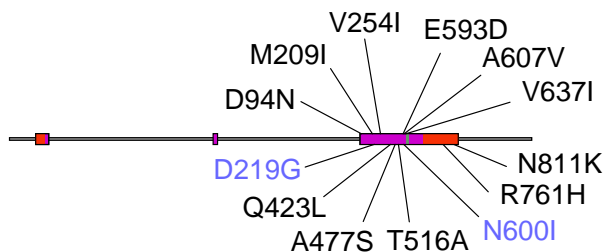


Figure 2
 Distribution of coding mutations found in *Tlr4* of 35 *Mus musculus* strains. All coding mutations reside within exon 3. Most occupy portions of the gene corresponding to the extracellular domain. The transmembrane domain is denoted by a blue-green bar. Mutations occurring at sites that are relatively conserved among species (only one or two forms among six species) are shown in blue; those occurring at sites that are less conserved (three to five forms among six species) are shown in black.

that among human patients with Gram-negative sepsis, some individuals may have been at risk by virtue of mutations in *TLR4*. The first step in determining whether different isoforms of Tlr4 confer added risk of, or protection

against, sepsis is the assessment of genetic variability at this locus in the normal population.

Whereas ablative mutations of *Tlr4* abolish the capacity to detect LPS, the question arises as to whether less severe mutation might alter either the specificity of LPS detection, as discussed above, or the magnitude of the LPS signal. Our present knowledge of Tlr4 structural variation in mice may permit an answer to this question, insofar as the mutations might easily be recreated, and measurements of signal-transducing activity through the modified receptors carried out in immortalized C3H/HeJ macrophages [13]. Animals of the A/J and P/J strains have defective tumoricidal capacity [25-29], although in neither case is the *Lps* locus implicated in the defect.

Deleterious mutations of *TLR4* might reasonably be sought in individuals who have developed serious Gram-negative infections, on the premise that mice with deleterious mutations of *Tlr4* are rendered susceptible to such infections. Similarly, in birds, a polymorphism at the *tenascin* locus (which lies a few megabases proximal to *Tlr4* in mice) predicts susceptibility to Gram-negative infection [30], suggesting that it may lie in linkage disequilibrium with a particular form of the avian *Tlr4* gene. Beyond this, it may be assumed, given the powerful pro-inflammatory activity of the receptor, that germline or

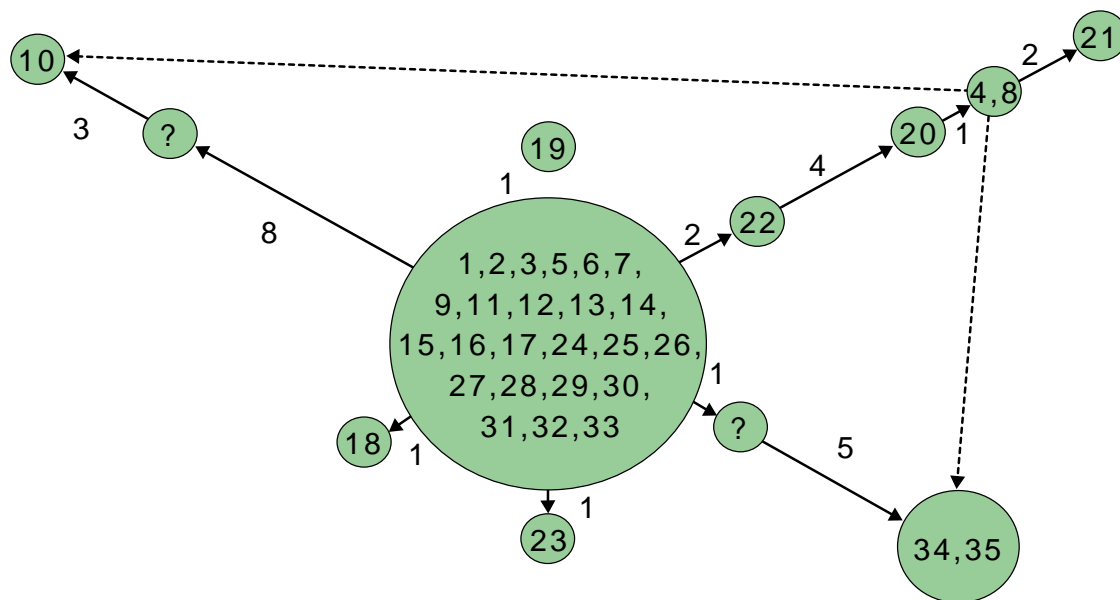


Figure 3
 Genetic distance and probable ancestral relationships among *Tlr4* genes of 35 *Mus musculus* strains. Numbers within circles denote strains, in accordance with the legend to Table 1. Numbers adjacent to arrows indicate the mutational distance (number of mutations separating each strain from its presumed ancestor), with reference to both coding and noncoding substitutions listed in Table 1. Arrows point in the direction of descent, and their lengths are proportional to distance. Dotted lines suggest introgression, given the similarity of the haplotypes observed. The symbol '?' denotes the likelihood of an intermediate form before interbreeding of strains.

Table 2

Conservation of Tlr4 among six mammalian species, calculated according to region

	Human	Chimp	Baboon	Rat	Mouse	Hamster
Extracellular domain						
Human	100%					
Chimp	99.6%	100%				
Baboon	91.5%	91.5%	100%			
Rat	61.3%	60.7%	59.4%	100%		
Mouse	61.9%	62.0%	60.1%	82.9%	100%	
Hamster	64.3%	64.2%	62.5%	73.8%	74.8%	100%
Proximal cytoplasmic domain						
Human	100%					
Chimp	99.4%	100%				
Baboon	99.4%	98.7%	100%			
Rat	91.7%	91.0%	91.0%	100%		
Mouse	90.4%	89.7%	89.7%	98.1%	100%	
Hamster	91.7%	91.0%	91.0%	97.4%	95.5%	100%
Transmembrane domain						
Human	100%					
Chimp	100%	100%				
Baboon	97.1%	97.1%	100%			
Rat	67.7%	67.7%	67.7%	100%		
Mouse	70.6%	70.6%	70.6%	91.2%	100%	
Hamster	73.5%	73.5%	73.5%	79.4%	79.4%	100%
Distal cytoplasmic domain						
Human	100%					
Chimp	100%	100%				
Baboon	50%	50%	100%			
Rat	38.1%	38.1%	ns	100%		
Mouse	26.3%	26.3%	ns	63.2%	100%	
Hamster	40.9%	40.9%	ns	40.9%	45.5%	100%

With respect to the human sequence, the extracellular domain is amino acids 1-624; transmembrane domain, amino acids 625-658; proximal cytoplasmic domain, amino acids 659-618; distal cytoplasmic domain, amino acids 619-638. Homology estimates were generated by multiplex FASTA comparison; ns, not significant.

somatic mutations of *TLR4* could in some instances cause constitutive signal transduction (as observed with the *Drosophila Toll* locus [31]). Such mutations might, in principle, account for certain inflammatory diseases. In fact, in the *Lpr* model of systemic lupus erythematosus, the *Lps* locus was believed a likely modulator of phenotype [32]. This would seem still more likely at present, given the paucity of other candidate genes in the immediate vicinity of *Tlr4* [10].

The evolutionary conservation of *Tlr4* is of particular interest, in that different species show preferential responses to some LPS forms and not others [20-24], and have particular 'set points' for responses to toxic LPS molecules. An assessment of variability may be made through comparison of different species, but is complemented by the study of a large number of individuals within species. Both approaches reveal that the extracellular domain of Tlr4 is highly variable compared with the transmembrane domain and proximal cytoplasmic domain of the protein. Pooling the number of mutable sites in the extracellular domain and transmembrane domain of humans and mice, 17 coding changes are observed, compared with two in the proximal cytoplasmic

domain. Moreover, variability does not seem confined to any particular region of the extracellular domain, but is spread uniformly across its length.

The extracellular domain of any receptor is concerned principally with ligand recognition. Given that the ligand is an endogenous protein, extracellular domain conservation tends to be strict, insofar as mutations affecting extracellular domain structure are likely to diminish specificity or affinity of binding. Hence, protein-binding receptors tend to be minimally polymorphic. The presumed ligand for Tlr4 is LPS itself, presented alone or in conjunction with another protein (for example CD14) [13], and the relatively high frequency of polymorphism observed in Tlr4 may be viewed as a consequence of the protective effect rendered by LPS recognition and the variability of LPS structure.

The cytoplasmic domain of Tlr4 is far more stringently conserved than the extracellular domain. This is probably due to the fact that the cytoplasmic domain of Tlr4 is not required to cope with a ligand of variable structure. Rather, when called upon to signal, it must do so, utilizing transducing

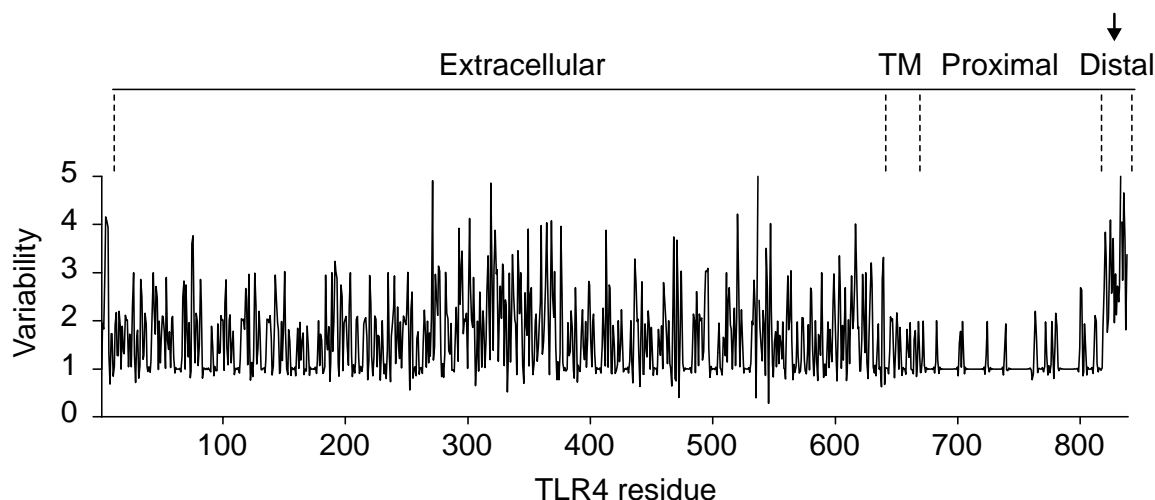


Figure 4

Spline curve illustrating interspecific sequence variation across the *Tlr4* protein. A multiple alignment of *Tlr4* sequences from three rodent species (mouse, rat and hamster) and three primate species (human, chimpanzee and baboon) was generated using the GCG program Pileup. The number of amino acids observed at each residue was plotted using the program Prism 3.0 (a value of 1 was assigned if a single amino acid was observed in the six species; a value of 5 was assigned if five different amino acids were observed among the six species, and so on). The points were then connected using a cubic spline curve. TM, transmembrane domain. Numbering refers to the human sequence. Where a deletion was introduced by Pileup, a single mismatch was assumed. Where the sequence was truncated, each missing residue was tabulated as a separate mismatch.

molecules with conserved structures. On the other hand, the intensity of the LPS signal has apparently been optimized for each species [33], perhaps ensuring that the response to LPS is appropriately integrated into the immune response as a whole. Humans, ungulates and rabbits, for example, exhibit an intense response to low concentrations of LPS, whereas the lower primates and most rodents are relatively more resistant [4]. The highly variable carboxy-terminal end of the *Tlr4* cytoplasmic domain may be seen as the embodiment of interspecific differences in LPS sensitivity, although the poor conservation of this portion of the molecule might alternatively be taken to indicate a neutral effect of mutation. It is also possible that this region of the molecule is subject to a higher rate of mutation than that applying to the rest of the protein. Although most such mutations might be removed by selection, some might be discovered in populations defined by the occurrence of Gram-negative sepsis.

Materials and methods

Determination of the complete mouse (*Tlr4*) and human (*TLR4*) genomic sequences

The mouse bacterial artificial chromosome (BAC) 152C16 (from the 129/J strain, Research Genetics) was previously shown by us to contain the *Tlr4* gene in its entirety [10], and a small fraction of *Tlr4* was also found in the overlapping BAC 309I17 [10]. Human *TLR4* was identified in BAC 110P15 (Genome Systems) by hybridization screening using a PCR-amplified human *TLR4* cDNA sequence as a probe.

All three BACs were fragmented by ultrasound, shotgun cloned into the vector pBluescript-KS, and extensively sequenced using ABI model 373 and 377 sequencers, using Big Dye terminators; 959 reads were obtained from 390I17, 1503 reads from 110P15, and 2731 reads from 152C16. The average read length was approximately 700 nucleotides. To concentrate data acquisition efforts on the *Tlr4* and *TLR4* genes themselves, PCR primers were fashioned to match regions flanking each gene. A 16 kb fragment was amplified from the mouse BAC 152C16, and a 12 kb fragment was amplified from the human BAC 110P15, each containing all exons of the respective gene. These fragments were also shotgun cloned and extensively sequenced, so that the depth of sequence reached an average of 12 reads over the area of greatest interest. Sequence assembly used the programs phred and Phrap (obtained from Brent Ewing and Phil Green, University of Washington Genome Center). Interpretation of repetitive elements was achieved with the program RepeatMasker (obtained from Arian Smit, University of Washington Genome Center). A contiguous high-quality sequence 18,974 bp in length, containing *TLR4*, was obtained from the human BAC, and a contig 91,748 bp in length, containing *Tlr4*, was obtained from the mouse BAC. Over these intervals, the error rate was estimated at <1 per 10^4 bp. The sequences have been posted to GenBank in annotated form (accession number AF177767 for the murine sequence and AF177765 for the human sequence). All data related to mutations are presented with reference to these sequences.

Table 3

Oligonucleotide primers used to amplify and sequence mouse and human *Tlr4* genes

Mouse	Primate
Amplification	
Exon 1	
↑ CAGTCGGTCAGCAAACGCCTTCTTC	↑ GCTCGGTAAACGGTGATAG
↓ CAAGGCAGGCTAGCAGGAAAGGGTG	↓ TGAGAAGTTCTGGGCAGAAG
Exon 2	
↑ TTATTTCATCTTTGGAGAGGAGTGG	↑ TCTCTGGTCTAGGAGAGG
↓ AAGGAAGTTTAGTTAGAACCACCTTG	↓ CCAGTCCAATAATGAAATG
Exon 3	
↑ TCTCCTGCTCACACCATCATCACCTG	↑ CCATCACATCTGTATGAAGAGCTGGATGAC
↓ CATGTGTTCCATGGGCTCTCGGTC	↓ TGACTTTCCTTTGTCATGGGTTCTTGACTG
Sequencing	
Exon 1 same as above	
Exon 2 same as above	
Exon 3 primary	
1 ↓ ATGCCATGCCTTGTCTTC	3.1 ↑ GAGCTGGATGACTAGGATTAATATTC
2 ↓ TTAAATTCFCCAAG	3.1 ↓ TCAAATTGCACAGGCCCTCTAG
3 ↓ CAGCTCTCTAGACC	3.2 ↑ CAATCTCTCTTTAGACCTGTCC
4 ↑ TGTGAACATCAGAAATTCCT	3.2 ↓ AATACTTTAGGCTGGTTGTCCC
5 ↑ TGAGATTGCTCAAACATGG	3.3 ↑ GAAGTTGATCTACCAAGCCTTG
6 ↑ TTGAAACAATTGAAGACAAGGC	3.3 ↓ GGAAGTCATTATGTGATTGAGAC
7 ↑ CCTGGCTGGTTTACACGTC	3.4 ↑ CTTCTGGACCTCTCTCAGTGTCAAC
8 ↑ TTTCATGGGTCTAGAAGAGCTG	3.4 ↓ GAAGCAGAGCTGAAATGGAGG
9 ↓ AAGAAGTCTCTCTGTTCC	3.5 ↑ TCAGATGAATAAGACCATCATTGGTG
10 ↓ TCAGAAACTGCCATGTTTG	3.5 ↓ AACAAAGTGTGGACCCAG
Exon 3 secondary	
5' ↑ TGAGCTGGTAAAGAATTTAG	1 ↑ GTAAATTTGGACAGTTTCC
7' ↑ CTGACGAACCTAGTACATGTG	2 ↑ TTCAGTATTCCTATCACTCAG
9' ↓ ATGTC AAGTTTGTGTGTT	3 ↑ TTATAAGTGTCTGAACTCCC
	4 ↑ TCGGTCCTCAGTGTGCTTG
	5 ↑ GTGTCCAGCACTTCATC
	6 ↑ AACCTCCTGAGGCATTTTC
	7 ↓ GTTTCAAATGGAAATGCTG
	8 ↓ AAGGAAACGTATCCAATG
	9 ↓ AAGCACACTGAGGACCGAC
	10 ↓ GATGAAGTGTGGGACAC
	11 ↓ TCCTCTTCAGATAGATGTTG
	12 ↓ TTTCTTTGTCATGGGTTTC
	0 ↑ TTTAGGTTCTTATTTCAGCAG
	0 ↓ GCTCTAGATTGGTCAGATTAG

↑ Primer matches + strand; ↓ primer matches - strand.

Sequencing DNA from mouse, chimpanzee and baboon samples

Mouse DNA, obtained from animals of 35 *Mus musculus* strains, was ordered from the Jackson Laboratories. Chimpanzee and baboon DNA were obtained from Kurt Benirschke (University of California, San Diego) and Gregory Delzoppo (Scripps Research Institute), respectively.

The three principal exons of *Tlr4* were amplified independently from mouse genomic DNA samples, leaving a margin of approximately 50 bp to each side of the exons so as to identify intronic mutations that might alter splicing. All exons of the chimpanzee were amplified and sequenced using the same primers used to amplify and sequence the human exons. For the baboon, the first two exons were also amplified using these same primers; however, the third exon of the baboon was amplified with a substituted primer at the 5' end.

The PCR products were isolated by agarose gel electrophoresis. Exons 1 and 2 were sequenced using the same

primers that were used for amplification. Exon 3 was sequenced using the flanking primers as well as a collection of eight internal primers. In this manner, the entire coding region and all splice junctions of *Tlr4* could be covered with a total of 14 sequencing reads. All primers used for amplification and sequencing are presented in Table 3; separate sets were used to amplify and sequence mouse and the primate samples.

Independent assembly of each sample was required as a condition for further analysis, and if such assembly failed, additional reads were executed using a secondary collection of primers. Thereafter, mutations were identified *en masse*, by pooling all of the reads from 25 to 30 samples at a time and reassembling with the program polyphred, using the phredPhrapPoly script (obtained from Natalie Kolker, University of Washington Genome Center). Consed_alpha (obtained from David Gordon, University of Washington Genome Center) was used to visualize reads and mutations.

The annotated chimpanzee exon sequences have been submitted to GenBank with the accession numbers AF179218, AF179219 and AF179220. The baboon sequences have been submitted with the accession numbers AF180962, AF180963 and AF180964. For genetic comparisons, rat (GenBank AF057025) and hamster (AF153676) *Tlr4* sequences were also used.

Genetic computation

A 500 MHz DEC-alpha system equipped with 256 Mbytes of memory was used for direct analysis of sequence data as described above. In addition to the programs already mentioned, the GCG software (version 9.0) was used for alignment analysis, with the program Pileup used in multiple alignments of protein sequences. The windows-based program Generunner 3.0 (Hastings Software) was used for the design of oligonucleotide primers. A spline curve describing heterogeneity of the *Tlr4* polypeptide sequence from different species was produced using the program Prism 3.0 (Graphpad Software). Sequences were prepared for submission with the use of the program Sequin 2.90 (obtained from the National Center for Biotechnological Information).

Acknowledgments

We thank the Zoologic Society of San Diego, which was the source of the chimpanzee genomic DNA used in these studies.

References

- Janeway CA Jr: **The immune system evolved to discriminate infectious nonself from noninfectious self.** *Immunol Today* 1992, **13**:11-16.
- Beutler B, Milsark IW, Cerami A: **Passive immunization against cachectin/tumor necrosis factor (TNF) protects mice from the lethal effect of endotoxin.** *Science* 1985, **229**:869-871.
- Tracey KJ, Beutler B, Lowry SF, Merryweather J, Wolpe S, Milsark IW, Hariri RJ, Fahey TJJ, Zentella A, Albert JD, et al.: **Shock and tissue injury induced by recombinant human cachectin.** *Science* 1986, **234**:470-474.
- Redl H, Bahrami S, Schlag G, Traber DL: **Clinical detection of LPS and animal models of endotoxemia.** *Immunobiology* 1993, **187**:330-345.
- Molvig J, Baek L, Christensen P, Manogue KR, Vlassara H, Platz P, Nielsen LS, Svejgaard A, Nerup J: **Endotoxin-stimulated human monocyte secretion of interleukin 1, tumour necrosis factor alpha, and prostaglandin E₂ shows stable interindividual differences.** *Scand J Immunol* 1988, **27**:705-716.
- Coutinho A, Meo T: **Genetic basis for unresponsiveness to lipopolysaccharide in C57BL/10Cr mice.** *Immunogenetics* 1978, **7**:17-24.
- Sultzter BM: **Genetic control of leucocyte responses to endotoxin.** *Nature* 1968, **219**:1253-1254.
- Coutinho A, Forni L, Melchers F, Watanabe T: **Genetic defect in responsiveness to the B cell mitogen lipopolysaccharide.** *Eur J Immunol* 1977, **7**:325-328.
- Vogel SN: **The Lps gene: Insights into the genetic and molecular basis of LPS responsiveness and macrophage differentiation.** In *Tumor Necrosis Factors: The Molecules and their Emerging Role in Medicine*, edited by Beutler B. New York: Raven Press; 1992:485-513.
- Poltorak A, Smirnova I, He XL, Liu MY, Van Huffel C, McNally O, Birdwell D, Alejos E, Silva M, Du X, et al.: **Genetic and physical mapping of the Lps locus - identification of the toll-4 receptor as a candidate gene in the critical region.** *Blood Cells Mol Dis* 1998, **24**:340-355.
- Poltorak A, He X, Smirnova I, Liu M-Y, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, et al.: **Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene.** *Science* 1998, **282**:2085-2088.
- Du X, Poltorak A, Silva M, Beutler B: **Analysis of *Tlr4*-mediated LPS signal transduction in macrophages by mutational modification of the receptor.** *Blood Cells Mol Dis* 1999, **25**:328-338.
- Poltorak A, Ricciardi-Castagnoli P, Citterio A, Beutler B: **Physical contact between LPS and *Tlr4* revealed by genetic complementation.** *Proc Natl Acad Sci USA* 2000, **97**:2163-2167.
- O'Brien AD, Rosenstreich DL, Scher I, Campbell GH, Macdermott RP, Forma LSB: **Genetic control of susceptibility to *Salmonella typhimurium* in mice: role of the LPS gene.** *J Immunol* 1980, **124**:20-24.
- Rosenstreich DL, Weinblatt AC, O'Brien AD: **Genetic control of resistance to infection in mice.** *CRC Crit Rev Immunol* 1982, **3**:263-330.
- Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF: **A family of human receptors structurally related to *Drosophila* Toll.** *Proc Natl Acad Sci USA* 1998, **95**:588-593.
- Rehli M, Poltorak A, Schwarzfischer L, Krause SW, Andreesen R, Beutler B: **PU.1 and interferon consensus sequence binding protein (ICSBP) regulate the myeloid expression of the human Toll-like receptor 4 gene.** *J Biol Chem* 2000, in press.
- Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA: **The dorsoventral regulatory gene cassette *spatzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults.** *Cell* 1996, **86**:973-983.
- Williams MJ, Rodriguez A, Kimbrell DA, Eldon ED: **The 18-wheeler mutation reveals complex antibacterial gene regulation in *Drosophila* host defense.** *EMBO J* 1997, **16**:6120-6130.
- Golenbock DT, Hampton RY, Qureshi N, Takayama K, Raetz CR: **Lipid A-like molecules that antagonize the effects of endotoxins on human monocytes.** *J Biol Chem* 1991, **266**:19490-19498.
- Kovach NL, Yee E, Munford RS, Raetz CR, Harlan JM: **Lipid IVA inhibits synthesis and release of tumor necrosis factor induced by lipopolysaccharide in human whole blood ex vivo.** *J Exp Med* 1990, **172**:77-84.
- Wang MH, Flad HD, Feist W, Brade H, Kusumoto S, Rietschel ET, Ulmer, AJ: **Inhibition of endotoxin-induced interleukin-6 production by synthetic lipid A partial structures in human peripheral blood mononuclear cells.** *Infect Immun* 1991, **59**:4655-4664.
- Kitchens RL, Ulevitch RJ, Munford RS: **Lipopolysaccharide (LPS) partial structures inhibit responses to LPS in a human macrophage cell line without inhibiting LPS uptake by a CD14-mediated pathway.** *J Exp Med* 1992, **176**:485-494.
- Delude RL, Savedra R, Zhao HL, Thieringer R, Yamamoto S, Fenton MJ, Golenbock DT: **CD14 enhances cellular responses to endotoxin without imparting ligand-specific recognition.** *Proc Natl Acad Sci USA* 1995, **92**:9288-9292.
- Boraschi D, Meltzer MS: **Defective tumoricidal capacity of macrophages from A/J mice. I. Characterization of the macrophage cytotoxic defect after in vivo and in vitro activation stimuli.** *J Immunol* 1979, **122**:1587-1591.
- Boraschi D, Meltzer MS: **Defective tumoricidal capacity of macrophages from A/J mice. II. Comparison of the macrophage cytotoxic defect of A/J mice with that of lipid A-unresponsive C3H/HeJ mice.** *J Immunol* 1979, **122**:1592-1597.
- Boraschi D, Meltzer MS: **Defective tumoricidal capacity of macrophages from A/J mice. III. Genetic analysis of the macrophage defect.** *J Immunol* 1980, **124**:1050-1053.
- Boraschi D, Meltzer MS: **Defective tumoricidal capacity of macrophages from P/J mice: tumoricidal defect involves abnormalities in lymphokine-derived activation stimuli and in mononuclear phagocyte responsiveness.** *J Immunol* 1980, **125**:777-782.
- Boraschi D, Meltzer MS: **Defective tumoricidal capacity of macrophages from P/J mice: characterization of the macrophage cytotoxic defect after in vivo and in vitro activation stimuli.** *J Immunol* 1980, **125**:771-776.
- Hu JX, Bumstead N, Barrow P, Sebastiani G, Olien L, Morgan K, Malo D: **Resistance to salmonellosis in the chicken is linked to NRAMP1 and TNC.** *Genome Res* 1997, **7**:693-704.
- Schneider DS, Hudson KL, Lin TY, Anderson KV: **Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo.** *Genes Dev* 1991, **5**:797-807.
- Wang Y, Nose M, Kamoto T, Nishimura M, Hiai H: **Host modifier genes affect mouse autoimmunity induced by the *Ipr* gene.** *Am J Pathol* 1997, **151**:1791-1798.

33. Berczi I, Bertok L, Bereznai T: **Comparative studies on the toxicity of *Escherichia coli* lipopolysaccharide endotoxin in various animal species.** *Can J Microbiol* 1966, **12**: 1070-1071.