HOW MAMMALS SENSE INFECTION: FROM ENDOTOXIN TO THE TOLL-LIKE RECEPTORS

Nobel Lecture, December 7, 2011

by

BRUCE BEUTLER

Center for Genetics of Host Defense, the University of Texas Southwestern Medical Center, Dallas, TX, USA.

HOW I CAME TO WORK ON THE QUESTION OF INNATE IMMUNE SENSING

About 30 years after the fact, I remembered a walk my father (Figure 1) and I took through a grove of redwoods in Sequoia National Park. I was perhaps 10 or 12 years of age. “Why is it that trees don’t simply rot?” I asked him, aware that plants had none of the lymphoid or myeloid cells that confer immunity to vertebrates. He explained there were tannins and perhaps other molecules in trees that made them resistant to decay. “But they rot after they die, and the tannins are still there,” I countered. The discussion went on, venturing into infections of live plants such as potatoes and wheat, and I tentatively concluded that plants must have some form of immunity that was actively maintained in the sense that it depended on their vitality. But at least to the two of us, not much seemed to be known about it. Of course, I didn’t know then that I would discover a mechanism of disease resistance in mammals that had its counterpart in most multicellular life forms, including insects and plants. But our conversation in Sequoia was to return to my mind almost immediately when I did.

Figure 1. Ernest Beutler, M.D. (1928–2008).
This was one of many thousands of discussions about science I had with my father, who always challenged me, counseled me, and helped to prepare me for whatever I wanted to do. He encouraged my love of science from the time I was a small child. From him, I learned to work in the lab, to isolate and analyze proteins, to think in evolutionary terms, and to evaluate experimental results. It was he who suggested I should go to medical school, to gain broad familiarity with the special processes that make living things what they are. Among the most important pieces of advice I recall was to “know what problems are important.” The message is one I try to pass on to students today. By the time my own interest in the field of innate immunity had become highly resolved, I felt secure in the knowledge that I was working on something of great importance. This inspired me to see a tough project through to its completion, despite all the difficulties my team and I encountered.

AMONG THINGS THAT ARE IMPORTANT, INFECTION IS A COMPLEX PROBLEM

In the depths of prehistory, infection probably killed most of our forebears. In historic times, there is no doubt that it did. Neither famine nor warfare nor cancer nor cardiovascular diseases have caused as many deaths as infection, for as long as humans have kept records of mortality and its causes. Even in the present century, with all our resources, infection claims nearly a quarter of all human lives¹ (Figure 2). Smallpox alone is said to have been the most frequent single cause of death among Homo sapiens during the 20th century² and great plagues of other kinds may have been close behind. Particularly because they strike down so many people before or during reproductive age, microbes constitute the strongest selective pressure with which our species must contend, and we may assume that microbes have shaped the human genome more than any other selective pressure in recent times. The autoimmune and autoinflammatory diseases we experience – by themselves major causes of morbidity and sometimes death – are the legacy of the intense selection our species has endured.

Figure 2. Leading causes of death worldwide. Approximately 15 million of the 57 (~25%) million annual deaths are the result of infectious diseases (listed in the table). Data were obtained from figures published by the World Health Organization (see http://www.who.int/whr/en). The figure above is adapted from Morens, Folkers, and Fauci.³
From almost any point of view, few phenomena are more complex than infections, which represent the clinical manifestations of the battle between host and microbe. Infection is a process in which thousands of biological processes go awry all at once as host and microbe compete with one another.

So many changes occur simultaneously during infection that it was once difficult to take a reductionist, mechanistic approach to the subject. The rhetorical question an investigator might ask was “where to begin?” But within this question was the seed of the right question, because there was reason to think that the host response might be initiated in a comparatively simple way: by a handful of receptors, recognizing the tell-tale molecular signatures of microbes and sounding an alarm. Our work was directed toward finding these receptors. Find them, we thought, and we would find the “eyes” of the immune system. Find them, and we might also understand how sterile inflammation is initiated. These were goals worth struggling for.

Genetics has provided the critical breakthrough in many biological problems, and it did so in the analysis of host responses to infection. By starting with distinguishable biological states (i.e., phenotypes) that are heritable, and identifying the genetic determinant of the difference, one may elucidate the molecules that play a key role in the phenomenon of interest. Using genetics, my colleagues and I determined one of the principal means by which mammals become aware of infection when it occurs, and deliver a response.

FRAMING THE QUESTION: INFECTIONS AND HOW THEY HARM US

Some of the most basic questions about how we fight infection remain unanswered to this day. The question as to how we sense infection was once in this category. To address it, we asked a more focused question: how do we sense endotoxin, a structurally conserved component of Gram negative bacteria? We did so in the hope that the answer would shed light on the more global picture of microbe sensing. We wanted to identify the first molecular events that initiate the immune response and all that goes with it.

The question about endotoxin was one that had endured for more than 100 years. Microbes had been discovered in the 17th century. But only in the late 19th century was their relationship to infection established, principally by Pasteur and Koch. It immediately occurred to many scientists of the time to ask how microbes actually do harm, and the possibility that toxins emanate from microbes was entertained.

Endotoxin was discovered by a German army surgeon, Richard Pfeiffer, who joined Koch and his team in 1887. Encouraged by Koch to study cholera, Pfeiffer noted that guinea pigs died when injected with a large inoculum of V. cholerae, even if passively or actively immunized against the microbe. Yet adaptive immunity had done its work: no living vibrio could be retrieved from the host. Pfeiffer’s name became attached to this phenomenon, and he extended his observation, noting that heat-killed vibrio were also lethal to guinea pigs. He called the toxic principle “endotoxin,” and endotoxin became a paradigmatic molecule in microbial pathogenesis (Figure 3).
WHAT WAS ENDOTOXIN, AND HOW DID IT WORK?

With the passage of decades it was understood that endotoxin was, chemically speaking, a lipopolysaccharide (LPS). It was the major glycolipid of the outer leaflet of the outer membrane of nearly all Gram-negative bacteria. Mesrobeanu and Boivin extracted LPS in fairly pure form in the 1930s; Lüderitz and Westphal introduced a phenol extraction protocol for LPS isolation subsequently. Ultimately, LPS from diverse microbial sources was structurally characterized. The lipid moiety of LPS, named “Lipid A,” was seen to constitute the toxic center of LPS, and by 1984, Lipid A had been synthesized artificially by Imoto et al., and shown to possess all of the activity of natural preparations. The pathway by which LPS was naturally synthesized was also deciphered, principally by Christian Raetz and his colleagues. LPS partial structures were isolated, and a number of structural rules for endotoxicity were established. Of importance to our work later on, it was noted that some LPS partial structures are agonistic when applied to mouse cells, but antagonize LPS when applied to human cells. The best example of this was Lipid IVa, which lacked acyl-oxyacyl side chains, and had only four lipid chains.

At nanomolar concentrations LPS was capable of activating leukocytes in vitro. If administered to living animals, it was strongly pyrogenic. It caused an immediate fall in the peripheral leukocyte count in animals as a result of the margination of circulating cells, and was also known to induce both the local and generalized Shwartzman reactions. If administered in sufficient quantities, just as Pfeiffer had observed, it could have a lethal effect. As a
class, mammals are more sensitive to LPS than other vertebrates, and among mammals, humans, rabbits, ungulates, and certain other taxa were exquisitely sensitive to LPS. It was clear that LPS must contribute to the often dramatic shock and tissue injury observed in Gram-negative infections. For this reason more than any other, LPS became a central interest in biomedicine, and efforts were made to interdict and neutralize it, often with antibodies.

Intriguing “beneficial” effects of LPS were also noted. By the 1950s, Johnson et al. had shown that purified LPS was endowed with adjuvant activity, greatly augmenting the antibody response to ovalbumin. Over the next decades, organized efforts to use LPS and LPS derivatives with diminished toxicity in vaccines were pursued by Ribi and others. LPS could also induce non-specific resistance to infections for a period of time after its administration. It was able to induce the necrosis of tumors in mice. And it was known to have a protective effect against otherwise lethal doses of gamma irradiation.

EVIDENCE FOR THE EXISTENCE OF AN LPS RECEPTOR, DEPENDENT UPON A SINGLE GENE.

There was no clear consensus as to what the LPS receptor might be. But as to the existence of an LPS receptor, there was a high degree of confidence from the 1960s onward. And much was known about the general characteristics of the receptor: that it could detect many structural variants of LPS, for example, but was not involved in the perception of other inflammatory molecules made by microbes. The evidence came from mouse genetics.

In 1965, Heppner and Weiss reported that mice of the C3H/HeJ strain were highly resistant to the toxic effects of LPS. Sultzer later documented the absence of leukocyte responses to LPS in these mice, in that they failed to form a peritoneal exudate when injected with LPS. The C3H/HeJ substrain had been separated from other C3H substrains only a few years earlier, and evidently, a recessive or semi-dominant mutation had become fixed in the population, forbidding responses to LPS. C3H/HeN mice, and C3H/OuJ mice stood as controls for LPS responsiveness, but were nearly identical to C3H/HeJ mice.

In 1977, Coutinho observed that mice of the strain C57BL/10ScCr mimicked mice of the C3H/HeJ strain, in that they had absent B cell responses to LPS. Allelism testing showed that the mutation in C57BL/10ScCr mice affected the same locus that was affected in C3H/HeJ mice.

Both C3H/HeJ and C57BL/10ScCr strains had highly specific defects. They responded normally to all microbial ligands tested, save LPS. This gave reason for confidence that the LPS receptor itself was affected by the mutations, rather than a broadly utilized transducing protein. And it suggested that the LPS receptor was quite specific. For example, certain lipopeptides from Borrelia burgdorferi seemed to utilize a distinct receptor to elicit TNF production.
In 1974, Watson and Riblet determined that a single locus mutation abolished the response to LPS in C3H/HeJ mice. In their study, they utilized B cell division and IgM production as indicators of the LPS response. Then, using classical phenotypic markers and a total of 14 recombinant inbred strains of mice derived from C57BL/6 and C3H/HeJ parents, they established linkage between the newly named Lps locus and the Major Urinary Protein (Mup1) locus on chromosome 4. Using a backcross strategy, the Lps locus was further confined to the interval flanked by Mup1 and Polysyndactyly (Ps) loci. This critical region was of unknown size, but immense (occupying about 1/8 of the chromosome), and could not be narrowed until much later.

LPS-resistant mice revealed profound facts about endotoxicity. The lethal effect of LPS was shown to be conferred by cells of hematopoietic origin (although LPS undoubtedly triggers responses in other cells as well). Mice that could not sense LPS were markedly compromised in their ability to survive infection by Gram-negative bacteria. Therefore, whatever the harmful effects of LPS, detecting it operates to the benefit of the host under conditions in which a small inoculum of bacteria has been introduced. Moreover, all effects of LPS were apparently mediated by the Lps locus; hence adjuvant effects, B cell mitogenesis, IgM production, and lethality all depended on a single gene.

One of the most important conclusions of work with LPS-resistant mice concerned the affirmative link between LPS sensing and host resistance. If mice were unable to sense LPS, they were vulnerable to infection by Gram-negative microbes, despite the fact that they were spared damage caused by LPS itself. This was first observed in animals infected with Salmonella typhimurium, then in E. coli, and later F. tularensis and Rickettsia akari. One plausible interpretation of these results is that LPS sensing contributes to detection of microbes during the earliest stages of an infection, permitting the host to mount a response that contains or eliminates them. If the host remains ignorant of the infection, containment does not occur; hence the burden of microbes becomes much greater. By the time the microbes are detected because of other molecules they produce (for example, flagellin, lipopeptides, nucleic acids), it is too late to contain the infection, and the host is overwhelmed. The existence of inducers of an inflammatory response other than LPS, and specific receptors for their detection, is implicit in this interpretation. So, too, is the primacy of the LPS detection system where these particular Gram negative microbes are concerned.

MY OWN INTEREST IN LPS AND HOW IT DEVELOPED

Some of the discoveries described above took place before I was interested in LPS, and indeed before I was born. But I began to think about LPS at a young age. In 1975, during a summer term at UCSD where I was a student, I approached Abraham Braude (Figure 4) to ask whether I might work in his laboratory. Braude had been a pioneer in the use of passive immunization against LPS as a means of countering sepsis: an approach that never gained
general broad acceptance by the medical community. He referred me to
Arthur Friedlander, a postdoctoral associate in his group who was then study-
ing the capsular polysaccharide of Cryptococcus, and its ability to induce
chemotaxis. Friedlander, who later made impressive advances in the study of
the lethal toxin of Bacillus anthracis, put me to work studying the responses
of rabbit leukocytes to purified polysaccharide. In this environment, I first
learned of LPS and its ability to activate leukocytes, induce fever, and cause
shock. At that stage, my consciousness of LPS mainly concerned its potential
to cause experimental artifacts, and the need to destroy it by baking glass-
ware at 180°C (LPS is resistant to autoclaving).

As a medical student at the University of Chicago (1977–1981), and as
a house officer at the University of Texas Southwestern Medical Center
at Dallas (1981–1983), I treated patients suffering from Gram negative
sepsis, and began to see the clinical effects of LPS firsthand. This certainly
impressed me as to the importance of LPS as a clinical problem, and as to
the magnitude of the disturbances LPS could cause. But my formal entry into
the field of LPS research began later. As a postdoctoral associate and then an
assistant professor, I worked in the lab of Anthony Cerami at the Rockefeller
University (1983–1986). There I isolated and characterized cachectin, an
LPS-induced macrophage factor.

Cachectin was named before my arrival in the lab, for its postulated role
as mediator of cachexia, the wasting process seen in many chronic diseases.
At the time I arrived, it was a crude factor, defined by its ability to suppress
expression of lipoprotein lipase (LPL) produced by fat cells (or in the usual
case, cultured 3T3-L1 pre-adipocytes). LPL is an enzyme required for the
hydrolysis of triglycerides to generate free fatty acids, permitting the entry
of plasma lipids into energy storage tissues. Cachectin activity was secreted in abundance by LPS-activated macrophages or LPS-activated immortalized macrophage cell lines. But no headway had been made in isolating the factor. Certain candidate mediators had been obtained as crude preparations from other laboratories, and tested for cachectin activity. One of the candidate mediators was tumor necrosis factor, obtained from the laboratory of Lloyd Old. It was found to have no cachectin activity; hence there was considerable surprise when I purified cachectin and determined what it actually was.

In succession I developed two purification protocols to isolate mouse cachectin from the conditioned medium of LPS-activated macrophages (RAW 264.7 cells). The first of these consisted of pressure dialysis, liquid-phase isoelectric focusing, ConA sepharose chromatography, and non-denaturing polyacrylamide gel electrophoresis (PAGE). This strategy allowed me to measure the quantity of cachectin produced by macrophages (it was about 2% of their secretory product during the early hours following LPS activation), to determine its specific activity, and to visualize it (Figure 5) as a 17.5 kDa protein species on a polyacrylamide gel.32 I also raised a strong antiserum against cachectin in rabbits. But it was not possible to obtain the amino acid sequence of the protein, which evidently became N-terminally modified in the course of purification.

![Figure 5](image)

**Figure 5.** Cachectin purification from conditioned medium of LPS-activated macrophages (RAW 264.7 cells). Following pressure dialysis, ConA sepharose chromatography, liquid-phase isoelectric focusing, and non-denaturing polyacrylamide gel electrophoresis, cachectin was identified as a 17.5 kDa protein by polyacrylamide gel electrophoresis (SDS-PAGE). Cachectin was subsequently identified as the mouse orthologue of human tumor necrosis factor (TNF; space-filling representation shown on right).

I therefore devised a second purification method, involving the newly-developed FPLC system from Pharmacia, which included pressure dialysis, anion exchange (Mono Q) chromatography, and gel filtration (Superose 12). This yielded a product from which an N-terminal sequence could be obtained by Edman degradation.33 Cachectin was strongly similar in sequence to human tumor necrosis factor (TNF), which had been isolated only a few
months earlier by workers at Genentech. Moreover, cachectin showed TNF bioactivity equivalent in terms of specific activity to that of purified recombinant human TNF: a fact first pointed out to us by John Mathison, a postdoctoral associate in the laboratory of Richard Ulevitch at The Scripps Research Institute, to whom we had sent some of our purified material.

It was thus suspected that cachectin was the mouse orthologue of human TNF: a conclusion verified by cDNA cloning a short time later. However, its de novo purification from mouse cells, based on a different biological activity, opened a new window on what TNF actually did.

TNF, so-named by Lloyd Old, who worked at Sloan Kettering Cancer Research Institute across the street from Rockefeller University, had a history intertwined with the history of microbes and LPS. As described by Old, the search for an endogenous mediator of tumor necrosis during sepsis was predicated on the observations of William Coley, who had used microbes and their products to induce remissions in patients with inoperable tumors during the early 20th century. Old had discovered TNF as the mediator of this effect, showing that LPS-injected mice produced a serum factor, apparently a protein, that could induce hemorrhagic necrosis of transplantable tumors grown in mice. This factor was also capable of killing tumor cells, but not normal cells, in vitro. It was viewed as a potentially nontoxic chemotherapeutic agent.

The fact that both TNF and cachectin activities emanated from a single molecule suggested to me that many of the effects of LPS might be TNF-dependent, and that TNF might mediate a strong inflammatory response. Indeed, I speculated that the lethal effect of LPS might depend upon TNF. In order to test this hypothesis, I raised an antibody against mouse TNF in rabbits, affinity purified the immunoglobulin, and made Fab’2 fragments from it. I used both intact antibody and Fab’2 fragments to passively immunize mice prior to LPS challenge. Mice that were blocked in their ability to respond to TNF were demonstrably though partially LPS resistant, indicating that TNF was one of the major factors responsible for endotoxicity, though not the sole factor. Moreover, I observed that TNF was remarkably toxic in mice, causing death when as little as 20 ug of active protein was administered intravenously. All in all, the animals resembled mice that had been injected with LPS, developing diarrhea, prostration, and organ injury. Later, more detailed toxicological studies were performed in rats and primates, with more or less the same outcome. In humans toxicity was observed also. While isolated limb perfusion with TNF did lead to remission of tumors such as melanoma, systemic toxicity barred its routine use in chemotherapy.
Figure 6. TNF is a major factor responsible for LPS-induced endotoxicity. (A) TNF blockade attenuates the lethal effect of LPS. Mice were treated with immune (“anti-TNF”, triangles) and pre-immune serum, circles. Figure adapted from Beutler, Milsark, and Cerami. (B) Mice injected with TNF (20 μg, purified from macrophages), became severely ill and often died. Animals injected with heat-inactivated material showed no untoward effects. These studies demonstrated that TNF was a major mediator of LPS toxicity.

The discovery that TNF could mediate the lethal effect of LPS led directly to experiments in many laboratories, in which the inflammatory potential of this cytokine was probed. It was found to be produced and influential in diverse model systems, and in particular, to affect both leukocytes and vascular endothelial cells so as to foster inflammatory responses (Figure 7). TNF blockade did not only prevent inflammation, but rendered animals highly susceptible to certain infections: especially infections with intracellular microbes such as *Listeria monocytogenes* and *Mycobacterium bovis*. TNF thus behaved as a clear executor of innate immunity.

Figure 7. The many biological activities of TNF. Ligand binding to T cells (either T cell receptor specific or non-specific) and macrophages (e.g., by LPS or other microbial ligands) cause intracellular signaling that causes the secretion of TNF. Modulatory influences on signaling can be exerted by IFNγ, glucocorticoids, or TGF-β. Secreted TNF had measurable (usually inflammatory) effects on virtually all receptor-expressing cells.
TNF receptors, first isolated and cloned by David Wallach and by David Goeddel and their colleagues, were found to exist on many cells throughout the body, and to trigger inflammatory responses when exposed to the ligand. A practical consequence of our work was the use of anti-TNF antibodies and soluble versions of the TNF receptors as inhibitors of TNF activity in human inflammatory diseases. One of the inhibitors, a fusion protein in which the ectodomain of the TNF receptor was linked to the hinge and Fc fragment of an IgG heavy chain, was invented in my laboratory in Dallas, patented, and sold to Immunex, which later manufactured an equivalent molecule, Enbrel. In time, TNF blockade was used to effectively treat several diseases, including rheumatoid arthritis, Crohn’s disease, ankylosing spondylitis, and psoriasis (Figure 8).

But a great question remained, relevant to all of these diseases. In sterile inflammation, what elicited production of TNF and other cytokines in the first place? And where Gram-negative infection was concerned, how did LPS trigger a biological response? In short, the cytokine response was clearly what orchestrated inflammation. But there was no understanding as to how the cytokine response began.

TNF was clearly a biologically relevant endpoint to follow in understanding responses to LPS. It was an “apex” cytokine, produced earlier than most other cytokines and capable of inducing many of them. Hence, I began to use TNF as a measurable marker of LPS-induced macrophage activation. And I did so with the conviction that a single molecule – the LPS receptor – must be found if we were to understand the very first events in the response to infection.
GROWING OBSESSION WITH THE C3H/HEJ MOUSE

It was during the course of my work with TNF that I first became aware of the existence of LPS resistant strains of mice: probably in 1983. Masanobu Kawakami, a postdoctoral associate who preceded me in the Cerami lab, had used macrophages from these animals as a control, to show that LPS induced cachectin activity and did not itself possess this activity when applied to adipocytes.48 Despite all that had been learned from these mice, already discussed above, nothing was yet known about the LPS receptor, or how it signaled. Gradually, the genetic lesion of the C3H/HeJ mouse began to occupy center stage in my mind, particularly after I left the Rockefeller University and set up my own laboratory at UT Southwestern Medical Center at Dallas. There I was jointly appointed as a member of the Howard Hughes Medical Institute, and was encouraged to pursue a focused “high-risk, high-impact” project.

I began to reflect on the fact that aside from the N-formyl-methionyl-leucyl-phenylalanine (fMLP) receptor, known to be a plasma membrane GTP binding protein, almost no avenues for the perception of microbes had been established. While it was obvious that host cells perceived molecular signatures indicative of broad microbial taxa (fMLP and LPS being only two examples among many), little was known about how this was accomplished. And the fMLP receptor could not really compare with the LPS receptor in terms of its biological relevance: it had not been shown to be crucial for the events of sepsis or for resistance to infection. Finding the LPS receptor seemed the critical question in innate immunity, inasmuch as LPS itself was the archetypal microbial elicitor molecule. Finding the LPS receptor might tell how most infections sound an alarm.

How might one identify the protein affected by mutations of the Lps locus? Several “easy” methods were potentially at hand. For example, one might simply look for a distinction between C3H/HeJ mice and C3H/HeN mice at the protein level. Or one might attempt to raise an antibody against cells from C3H/HeN proteins in C3H/HeJ recipient. Such an antibody might pinpoint the protein defective in the C3H/HeJ strain. One might try insertional mutagenesis: make an F1 heterozygote by crossing C3H/HeJ to C3H/HeN; make an immortalized macrophage line (expected to be LPS responsive), and then attempt to destroy the one “good” copy of the Lps locus with a retrovirus and isolate an unresponsive clone. A standard cDNA rescue approach might also have worked. We went so far as to make an immortalized C3H/HeJ cell line in collaboration with Paola Ricciardi Castagnoli, hoping to use it in this way. Later, it was put to a different use, as described below.

Each of these approaches was diligently pursued, but each was unsuccessful. We understood that in other laboratories, affinity purification methods were used in an attempt to isolate the LPS receptor. But no substantial publications resulted. The nature of the LPS receptor remained entirely mysterious. There were, however, numerous papers hinting at what the receptor and/or the product of the Lps locus “might” be. Some referred to the
putative involvement of the cell surface molecule CD18 in LPS signaling. Through differential display studies, SLPI, a serine protease inhibitor, was indirectly implicated. A small GTP binding protein called Ran/TC4 was suggested as well. Studies with inhibitors suggested the involvement of a tyrosine kinase. Or perhaps a histidine kinase was encoded by Lps, based on the upstream activator of the p38 equivalent in yeast, HOG1. Or perhaps the Lps locus encoded a member of the protein kinase C family. During the course of our positional cloning work, we needed to ignore such hints, because hypothesis-driven targeted searches would have distracted us from our primary mission: to find all candidate genes within the Lps critical region, and ultimately find the causative mutation responsible for LPS resistance.

One suggestion as to the nature of the LPS receptor could not be ignored, because it was so clearly true and compelling. In the year 1990, Sam Wright, working in Richard Ulevitch’s group, showed that antibodies against the monocyte surface marker CD14 were able to inhibit responses to LPS. The Ulevitch lab subsequently showed that overexpression of CD14 in 70Z/3 pre-B cells would greatly enhance LPS responsiveness in these cells, which were otherwise minimally responsive to LPS. CD14, a leucine-rich repeat protein anchored to the surface of cells, had no obvious means of inducing a transmembrane signal, in that it had no cytoplasmic domain. Moreover, it was encoded by a gene unlinked to the Lps locus (it is now known to be on chromosome 18). It was considered, therefore, that CD14 might be part of the LPS receptor complex. But at least one missing part of the complex, crucial for signaling, would necessarily be encoded by the Lps locus (Figure 9).

Figure 9. An unknown receptor in LPS-mediated TNF secretion. The figure reflects the state of knowledge that existed in the early 1990s. Overexpression of CD14 was observed to enhance LPS responses of 70Z/3 cells, although it had no means of inducing a transmembrane signal. Therefore CD14 was considered to be an essential part of the receptor complex, which must include an unknown membrane-spanning protein, likely defective in C3H/HeJ and C57BL/10ScCr mice. In unstimulated macrophages, TNF mRNA is repressed translationally through its AU-rich elements (AREs). In LPS-stimulated cells, the unidentified receptor relays a signal for the NF-κB subunits p50/p65 to the nucleus and transcribe the TNF gene. The blockade of translation (via the AREs) is also alleviated upon LPS-mediated activation. The mRNA is subsequently translated and the TNF protein is processed and secreted.
During the late 1980s and early 1990s, my colleagues and I analyzed control points that governed biosynthesis of TNF in LPS-activated macrophages, with the general thought that we might work backward toward the receptor if we understood specific molecular events that transpired in the cell. We showed that both transcriptional and translational activation steps occurred in macrophages in response to LPS, and noted that acting in concert with one another, these regulatory mechanisms permitted a several thousand-fold increase in TNF secretion by activated cells as compared to quiescent cells. From the work of Jongeneel and his colleagues, it was known that transcriptional activation depended minimally upon NF-κB translocation to the nucleus. But the translational activation step remained quite enigmatic, and to a large extent still does. Consistent with earlier work of Kruys and Huez, we concluded that translational repression of the TNF mRNA was maintained in quiescent cells by a cis-acting UA-rich sequence which we had identified in the 3'-untranslated region of the molecule. The same sequence motif, independently identified by others, was shown to cause instability of many mRNA molecules.

We studied the effects of various inhibitory drugs (particularly glucocorticoids and pentoxifylline) on the TNF response to LPS, and created reporter mice bearing a transgene in which a chloramphenicol acetyltransferase (CAT) coding sequence was substituted for the TNF coding sequence. These animals permitted us to determine where TNF was produced during in vivo LPS challenge. While useful, these analyses were tangential to the key question of what the receptor might be. Increasingly, I felt we were dancing around the problem rather than attacking it, and in 1993, I resolved to focus exclusively on the \textit{Lps} locus.

**POSITIONAL CLONING OF THE LPS LOCUS.**

An unbiased genetic approach to finding the \textit{Lps} locus became feasible when the density of markers in the mouse genome increased enough to permit at least some narrowing of the critical region established much earlier by Watson et al. At the time we began our positional cloning work in 1993, only 317 microsatellite markers had been published, and only a fraction of these were informative between any two selected strains. We set about to map the \textit{Lps} locus to high-resolution, hopeful that we could confine the mutation to a relatively small part of mouse chromosome 4.

Christophe Van Huffel took the lead in this work, and was soon joined by Alexander Poltorak, Irina (Ira) Smirnova, Xiaolong He, and Mu-Ya Liu, all postdoctoral associates in my lab. Not all of them would see the work through to completion. Alexander and Ira were the exceptions, and deserve most of the credit for our success. They were truly devoted, smart, and innovative in their work, and passionate in their desire to find the mutation. I must also mention the contribution of Ms. Betsy Layton, my administrative assistant then and to the present day. She not only organized our work when it came to determining the position of markers and constructing accurate maps, but al-
so participated in some of the laboratory effort, picking colonies, inoculating broth, and reading and searching DNA sequences. To these people, I owe a great debt of gratitude (Figure 10).

![Figure 10. (clockwise, from top left) Ms. Betsy Layton, Dr. Alexander Poltorak, Dr. Irina (Ira) Smirnova and Dr. Christophe Van Huffel. Photos taken in early to middle 1990s.]

The *Lps* cloning project grew to enormous proportions, although we never imagined it would in the beginning. It was accomplished in three phases. First came the *genetic mapping* phase. To narrow the critical region, we used a total 2,093 meioses, derived from crosses of C3H/HeJ to either SWR/J or C57BL/6J mice, with backcrosses of F1 hybrid animals to the C3H/HeJ parent. The F2 mice were examined for LPS responsiveness, measuring the ability of their peritoneal macrophages to secrete TNF when challenged with LPS, by means of a biological assay. The effort was not entirely straight-
forward, because the mutation was semi-dominant, and some phenotypic assignments were ambiguous. Mindful that a single mistake could put us out of the critical region, we tested and re-tested animals until we were absolutely sure of their \( Lps \) genotypes.

In the final analysis, we were able to confine the mutation to an interval 2.6 million base pairs (Mb) in length – or so we thought. This estimate was based on fluorescence in situ hybridization (FISH) studies using labeled bacterial artificial chromosomes (BACs) isolated from the region, on measurements of BAC sizes established through pulsed-field gel electrophoresis, and on probabilistic estimates of BAC overlap. In reality, all of these methods are rather imprecise, and the minimum size of the region, ultimately revealed by whole genome DNA sequencing, was approximately 5.8 Mb. No other critical region of such size had ever been successfully explored for a point mutation, at least in the mouse, and had we known the authentic size, we might have thought twice about continuing our work.

Try as we might, we could not reduce the critical region by genetic mapping. A large interval of the chromosome, incorporating approximately 5 Mb of DNA, was refractory to meiotic recombination, at least with the strain combinations we had selected.

The next phase, physical mapping, was actually initiated in parallel with the mapping work (which continued almost until the end of the project). The objective was to clone all of the genomic DNA of the critical region. Our purpose was both to estimate the physical size of the region and to explore it for genes, each of which would stand as a candidate until exonerated by DNA sequencing in C3H/HeN and C3H/HeJ strains. In the final gene identification phase, these candidates would be cloned and sequenced at the cDNA level, one by one.

It must be recalled that the draft sequence of the mouse genome was not published until the year 2002, and the gene content of the \( Lps \) critical region was completely unknown. Moreover, exact details of its syntenic relationship with the human genome were also unknown: owing to an ancestral translocation event, the human \( LPS \) locus, if such existed, could either be on chromosome 9p or chromosome 9q (the likelihood was about equal for either possibility). And basic assumptions about the number of genes in the mouse genome were dramatically incorrect: it was typically bantered about that there must be “100,000 genes.” If correct, there could easily have been 100 genes or more in the critical region we finally established. We now know that the total gene number was overestimated four-fold, and that the \( Lps \) critical region is rather poor in genes, though rich in pseudogenes.

Using D4Mit microsatellites as markers (they were listed at genetic intervals of 1 cM), we built a contiguous overlapping collection of yeast artificial chromosomes (YACs) and later BACs that covered the critical region. In common parlance, this is called a “contig.” Ultimately a total of 66 BAC clones and 2 YAC clones were isolated to cover the region in depth. From close analysis of these clones, a minimum tiling path was established that contained a total of 24 overlapping BAC clones and one YAC clone (Figure 11).
Figure 11. Physical map of the chromosomal region surrounding Tlr4. Mapping by Watson et al. confined Lps to an interval between *Map1* and *Ps* loci (top). Mapping was repeated using microsatellite markers extending over approximately 1/4 of the chromosome (D4Mit111 through D4Mit77). On 493 meioses Lps was first confined to an interval bounded by D4Mit218 and D4Mit80, and physical mapping was initiated. On 1600 additional meioses Lps was confined to an interval bounded by new markers B and 83.3 (isolated from the growing BAC and YAC contig). A contig was assembled to span the interval, which could not be genetically reduced although new polymorphic markers (circles) were identified within it. Each bar is a BAC, except the largest, which is a YAC. Yellow bars indicate sequencing to near-completion. Pink bars indicate heavy, but incomplete sequencing. Red bars indicate BACs that contained Tlr4. Genes are shown in green. Including a small gap near the 5’ end of the contig, the entire critical region is known today to be approximately 5.8 Mb in size, rather than the 2.6 Mb that it was formerly believed to be, and the distance from Tlr4 to 83.3 is 4.2 Mb.

I will dispense with most of the details of the physical mapping and fine-mapping in this text. The map was published elsewhere as a prelude to publishing the identity of the *Lps* receptor.69 But I would like it to be clear that physical mapping was a daunting task. The contig had, at first, “islands” of BACs separated from each other until they could be joined by chromosome walking. These islands would often flip in orientation, as we came to understand that one marker must be proximal to another; then the conjoined islands might flip again as they were extended to join with other islands. In the process of chromosome walking, we established new markers: unique sequences based on the ends of BAC clones, and also new microsatellites not seen by others. All microsatellites would be checked for polymorphism in the hope that we might use them in mapping, and a number of them could indeed be used. The final contig was the largest ever built in a mouse positional cloning foray, and no contig of equal size is likely to be built again.
It became unnecessary to build contigs once the genome of the mouse was sequenced and annotated.

One by one, the BAC and YAC clones were fragmented using ultrasound, and the fragments were polished, sub-cloned, and sequenced bidirectionally. When we began our work, there was no such assembly program as Phrap, which we later used extensively to generate long contiguous sequences (sometimes approaching 100 KB in length). In the beginning, we used much simpler programs such as Wordsearch and FastA to align sequences and reconstruct the landscape of each BAC, and to join BAC clones together.

HUNTING FOR GENES

The final phase of the project, as mentioned above, was the concerted search for genes, and for the mutation responsible for the defect in C3H/HeJ mice. In searching for genes within the trackless wilderness of our contig, we progressed from the most primitive methods (exon trapping and hybridization selection), to the most sophisticated (searching genomic sequences against expressed sequence tag [EST] databases using the matching algorithm BLAST [Basic Local Alignment Search Tool]) over the course of our search. We also used a computational method, the program GRAIL, developed by Richard Mural at ORNL, to search for gene candidates in silico based on the base hexamer composition of genomic DNA. It proved remarkably sensitive and specific, although at first we had to use it essentially on faith.

Exon trapping, which soon went out of fashion, depended on cloning BAC DNA into special vectors with donor and acceptor splice sites. If a piece of DNA happened to have an exon in it, the exon would be spliced when the vector was transfected into mammalian cells, yielding a colony color difference based on the expression of beta-galactosidase activity. The cDNA would then be amplified from these cells, and the nature of the “trapped exon” would be determined by DNA sequencing. A total of 169 exons were trapped by Christophe Van Huffel, but many were from degenerate pseudogenes or from a fragment of the gene Pappa (encoding pregnancy associated plasma protein A), which we eventually eliminated as a candidate.

Hybridization selection depended upon the hybridization of cDNA from macrophages, which we knew to express the Lps locus, to BAC DNA of the contig (after appropriate blocking with Cot-1 DNA). The cDNA was then eluted and cloned, and a total of 568 cDNAs were isolated and identified in this manner. Among these were some validated candidate genes, but also a large number of false positive identifications, including conserved pseudogenes within the contig that were similar in sequence to mRNAs expressed from authentic genes elsewhere in the genome.

While we struggled with these difficult gene finding methods, EST databases were being developed, and these excited our interest. Some were proprietary (for example the TIGR database) and inaccessible to us without special subscription fees and institutional agreements; one (dbEST) was public. But there was a sense that one had to have access to all the data, lest the
all-important gene go undetected. And for that matter, it might be important to continue exon trapping and hybridization selection because we couldn’t be absolutely sure that these methods wouldn’t lead to the identification of the gene we were seeking. Gradually, we gained confidence that EST searching was the best approach.

EST databases were simply databases of cDNA sequence, derived from processed mRNA expressed in many different tissues at many different embryonic stages. If one found a strong match between a genomic DNA sequence from the contig and an EST, one could be reasonably confident that this particular piece of genomic DNA must be expressed as mRNA. There were some caveats. EST databases were often contaminated with genomic DNA to some degree. Even foreign sequences (from other species, including microbes) had a way of creeping into the record. One could not therefore be totally confident about a particular match. But EST searching was certainly something we couldn’t dispense with, and it became our preferred approach. Alexander and Ira became consummate artists at fragmenting BAC clones, producing complex libraries, and feeding the sequencing operation.

From the start, we were badly constrained by a shortage of sequencing power. When we began our work we were sequencing by hand. One of us would read the sequencing ladder on an X-ray film to another who would type it into a computer for later BLAST searching. There was a certain pathos to this, because we knew this was not the best way to proceed, but could not afford to purchase an automated sequencer with our existing funds, and HHMI declined to supplement our budget despite plaintive appeals on my part. Semi-automated slab sequencers, which could actually call bases, came into fairly wide use by the mid-1990s. We began to decentralize our operation, sending samples to three separate core sequencing labs. We also appealed to the UT Southwestern Sequencing Center (GESTEK) run by Glen Evans, hoping he might help us. We additionally turned to Bruce Roe, who ran a sequencing center at the University of Oklahoma. In this way we did manage to acquire hundreds and sometimes thousands of reads per week. Some of the sequencing was done by Dale Birdwell, a technician whom I paid out of pocket to work weekends. Still frustrated, I bought a somewhat antiquated ABI 373 sequencer at personal expense. Because it was under my exclusive control, we could run it around the clock to capture sequence. With this machine, as it happened, we made the critical breakthrough that allowed us to find the mutation. Later, when it was totally outmoded, I donated it to the San Diego Zoo. I don’t know where it is today, but must confess I feel rather nostalgic about it.

Sequencing produced data that had to be analyzed in an efficient manner. It wasn’t sufficient to BLAST a sequence once and then forget about it, because the EST databases were always being updated, and again, there was the feeling that one might miss the critical sequence. After a time, Betsy and I began to find that all our time was taken up with the manual submission of flat sequence files for blasting. At the advice of David Gordon, the author of consed (a program we had begun to use extensively to view
sequences), I taught myself to program in Perl in order to automate the file manipulations needed for recursive sequence analysis. I wrote a script called Central_Command, which sent sequences for analysis and flagged those with likely matches for further analysis by a human observer. But every sequence was studied individually by a human observer at some point, because I feared we might otherwise miss a critical match. One day I received a telephone call from the National Center for Biotechnology Information. They complained that I was using an excess of computational resources by BLASTing on their servers and told me I must BLAST less or they would cut me off. Somehow I complied, decreasing the frequency of recursive searches and doing some of the BLAST searches on a Linux computer of our own.

FIRST GLIMPSE OF THE GENE

Psychologically, genetic work is addictive in the same way that gambling is addictive. We felt as gamblers at the slot machines probably do when we waited for crossovers that might help us narrow the critical region, and when reading through BLAST results to see whether we had found a new candidate gene. Most gamblers know how a losing streak feels, and in our own case, many months had elapsed with little progress either in genetic confinement of the mutation or in gene identification to show for the investment of time and money. But like a gambler who has committed a great deal of money without seeing a big payoff, we could hardly bear to give up having invested so much. At the same, we quietly begin to worry: are we in the correct area at all? Have we made a mistake in mapping, or in our contig construction?

These worries were compounded by external pressure. In April of 1998, I learned that my funding at HHMI would be terminated in September of 2000. I rued the decision, because we had indeed worked on an important problem with focus and industry. But we had not succeeded quickly enough. Like a gambler who sees he is down to his last few dollars, I decided to stay the course, whatever perils it might hold.

By our best estimates, approximately 90% of the critical region had been thoroughly explored by late summer of 1998, and only a modest collection of pseudogenes and a single authentic gene had been identified. Seemingly there were far fewer genes in the region than in most parts of the genome. One can invent a story about almost any gene – and even some pseudogenes – convincing oneself that at last the gene has been found. But a disciplined approach was necessary: the presence of a mutation distinguishing the gene in LPS sensitive mice from the gene in LPS resistant mice had to be found, or failing that, at least there needed to be a dramatic difference in expression. Time and again, the candidates failed these tests.

One of the candidates, known at the time as KIAA0029, was particularly fiendish, because it was large and complex, and expressed in an enormous number of variant splice forms. We began to wonder whether the innate immune system might depend upon splicing to generate receptor diversity, and whether one and only one splice variant might serve the recognition
of LPS: a plausible idea at the time. We had, therefore, to look at hundreds of cDNA clones from this gene, to see whether a particular splice variant might be missing in C3H/HeJ. In the end, we decided this wasn’t the case; moreover it became clear that the chromosome 4 version of KIAA0029 was a pseudogene. Irritatingly, it had absorbed much of our sequencing power while the intrigue lasted.

On the evening of September 5, 1998, I got my first look at the last gene we were to find in the contig. I was working at home, and it was about 9:30 pm. I was electrified by what I saw, but not entirely convinced we had reached the end of our search, given past experience.

Having endured a period of months during which I had seen absolutely nothing in terms of credible BLAST results, I was strongly confident that this was a genuine gene rather than a pseudogene. Two clones derived from BAC I17 scored as hits with the EST database, and both ESTs were derived from the same gene. Both matches were virtually flawless, and either end of the transcript was struck. By the next day, sequences from an overlapping BAC, C16, provided a third hit (and later a third overlapping BAC showed fragments of the gene). But as always, the proof would depend upon finding a distinguishing mutation.

The gene we had identified was \textit{Tlr4}, and it was interesting to be sure. I saw immediately that it had cytoplasmic domain homology to the IL-1 receptor. I knew that Drosophila Toll had cytoplasmic domain homology to the IL-1 receptor as well: a fact that had surprised me since I learned of it in a lecture given by Steven Wasserman at UT Southwestern several years earlier. This had originally been noticed by Nick Gay in 1991. Why should a developmental protein in the fly resemble an immunological protein in the mouse? A quirk of evolution, I had thought, that common signals could be co-opted for very different purposes. The IL-1 receptor, of course, mediated inflammatory responses, and that was a good sign where the candidate LPS receptor was concerned. Moreover, the leucine-rich repeats of the TLR4 ectodomain were structurally reminiscent of CD14, which we knew on strong experimental grounds to be involved in LPS sensing.

I had a dim recollection that Toll was necessary for the response to fungal infection in Drosophila, which depended upon NF-κB mediated induction of Drosomycin, an antimicrobial peptide: the work of Jules Hoffmann, published in \textit{Cell} two years earlier. Confirming this recollection within a few minutes, I realized that this situation was highly analogous to the LPS paradigm, since mutations affecting the LPS receptor had long been known to confer susceptibility to Gram-negative infection. There were differences as well: the fact that Toll engaged a protein ligand rather than a molecule of microbial origin. Nonetheless, I thought perhaps the host resistance mechanisms of the mouse and the fly were more similar than anyone had realized before.

That evening I also became aware for the first time of a paper from 1997 by Medzhitov, Preston-Hurlburt, and Janeway in which one member of the human TLR family had been cloned at the cDNA level and dubbed “hToll”\textsuperscript{71}. Unlike our work, theirs was derived from the discoveries of the Hoffmann
group in Drosophila. They had demonstrated that human TLR4 could activate NF-κB if expressed as a fusion protein with extracellular CD4 sequences, designed to cause constitutive activation. It had been well established that both Toll and the IL-1 receptor could activate NF-κB; hence it was no surprise that TLR4 could do so. Transfected into myeloid cells, the modified TLR4 construct would cause upregulation of costimulatory molecules. It was suggested that this, too, might result from NF-κB activation.

The link to NF-κB was, by itself, no proof that TLR4 had an immunological function, since NF-κB has both developmental and immunological roles to play, in both mammals and in insects. In the fly, in fact, among nine members of the Toll family, only Toll itself has anything to do with immunity. Nonetheless, the paper advanced the hypothesis that TLR4 was a “pattern recognition receptor” using the name Janeway had earlier coined to designate innate immune receptors recognizing broadly conserved molecules such as LPS. But it presented no evidence in support of this hypothesis, nor did it name the ligand that TLR4 was supposed to recognize.

All of the afore-mentioned considerations gave grounds for speculation, but only a mutational difference between Tlr4 in LPS sensitive and resistant strains would provide strong evidence that TLR4 was involved in LPS signaling. Speculations notwithstanding, our positional data – and our positional data alone – pointed to TLR4 as the LPS receptor. In a telephone conversation that night, Alexander, Ira, and I all took note of the structural similarity between TLR4 and the IL-1 receptor and CD14, but mostly, we were swayed by the fact that the bulk of the contig had been explored. With only a few hundred thousand nucleotides (and presumably only a few thousand coding nucleotides) left to sequence in what seemed to be a gene desert, this was the only viable candidate in hand.

That night I designed primers to amplify the entire TLR4 cDNA by long-range PCR, and they were ready by the following afternoon. Alexander amplified both C3H/HeN and C3H/HeJ mRNA samples, extracted the bands, fragmented them with ultrasound, and shotgun cloned them into a sequencing vector. Within a week, we had sequenced both libraries, covering the cDNAs with 100 or more reads each: perhaps to a mean depth of ten reads or so. We first saw the mutation using the consed viewer on September 15th: a C→A transversion in the third exon of the gene. The mutation caused the substitution of a histidine for a conserved proline in the cytoplasmic domain of the protein (P712H). It was quickly verified at the genomic level. For some weeks, we kept returning to the computer screen to gaze at the trace file.

This discovery, while exciting, did not formally prove that Tlr4 was the relevant gene in the Lps critical region, required for LPS responses. The mutation might have been an irrelevant and functionally neutral substitution that had occurred independently of the causative mutation, and like it, had become fixed in the C3H/HeJ strain. However, data developed from our analysis of the C57BL/10ScCr and C57BL/10ScSn strain combination were definitive. These strains, maintained for many years by our collaborators Chris Galanos and Marina Freudenberg, who provided us with much valu-
able insight into LPS biology both before and during the Lps cloning work, had been reserved for a final confirmatory experiment. Alexander was, in the case of LPS-unresponsive C57BL/10ScCr strain, unable to amplify the TLR4 cDNA. But he succeeded with amplification of the TLR4 cDNA in the case of the LPS-responsive C57BL/10ScSn strain. Northern and Southern blots suggested complete deletion of the Tlr4 locus had occurred in C57BL/10ScCr: a conclusion substantiated by DNA sequencing soon thereafter. A 74 kb interval of genomic DNA was cleanly excised, removing Tlr4 but sparing all other genes (Figure 12).

Figure 12. The mutations in C3H/HeJ and C57BL/10ScCr strains, as detected for the very first time. Right, top: photograph of computer screen, consed display, showing a C to A transversion in exon 3 at position 2342 of the C3H/HeJ Tlr4 cDNA, causing the substitution P712H in the polypeptide chain. Left: Northern blot and ethidium stain of mRNA from C57BL/10ScSn (LPS responsive) and C57BL/10ScCr (LPS unresponsive) macrophages. Right, bottom: RT-PCR of Tlr4 shows absence of detectable Tlr4 mRNA in C57BL/10ScCr strain mice, but not in closely related C57BL/10ScSn mice, nor in C3H/HeJ or C3H/HeN mice. Transferrin mRNA is expressed by all strains.

We thus knew of two allelic variants of Tlr4, one of them overtly destructive and the other likely to be. These variants were observed in the C3H/HeJ and C57BL/10ScCr strains, but not in closely related control strains that had normal LPS responses. Having published the mapping data on September 14, we submitted our major paper establishing the identity of Lps and Tlr4 to Science on September 30. It was accepted with minimal revisions and published on December 11, 1998. It soon became the most highly cited publication in the innate immunity field because it had revealed the key sensors used by the mammalian innate immune system to detect infection, and had also revealed the conservation of this system for innate immune activation from mammals to insects (Figure 13). As I write this lecture, it has been cited 3,970 times.

Not long after our discovery, two published reports asserted that TLR2 was the LPS receptor. Workers at Genentech and Tularik, unaware of our work, each transfected mammalian cells to overexpress TLR2, and had found that high concentrations of LPS could drive NF-κB activation in these cells. Their papers, published in September and in December of 1998, were not quantitative in the sense that there was no way to compare the magnitude of the LPS response to that in a positive control cell with truly normal LPS signaling potential (for example, a macrophage cell line). Moreover, no genetic test of the TLR2 hypothesis was made. To this day, it is not entirely clear why the observed results were obtained, but it has been suggested that the LPS preparations used were contaminated with lipopeptides that may have triggered a TLR2 response.

We knew immediately that the core conclusions of both studies were erroneous, and that the error occurred as a result of weak methodology: the use of transfection, rather than a true genetic approach, as the basis for inference. Lack of TLR4 signaling, as observed in C3H/HeJ or C57BL/10ScCr mice, completely abolished LPS sensing. This excluded the existence of an alternative pathway in which TLR2 might act as an autonomous LPS receptor. The notion that TLR2 could make any contribution to LPS signaling lost all credibility when Osamu Takeuchi and colleagues from the group of Shizuo Akira targeted Tlr2, and showed that TLR2 deficient mice respond normally to LPS. Nonetheless, for the next few years, numerous publications referred to “two LPS receptors,” until the idea faded away.
The discovery that TLR4 is necessary for LPS responses did not address the question as to whether it was indeed a physical receptor for LPS: a difficult question to answer given the hydrophobic character of the putative ligand, and the resultant difficulty of performing classical binding studies. In Drosophila, the homologous sensor Toll had no contact with any product of microbes, and genetic evidence indicated that the protein Spaetzle, cleaved by upstream proteases that were activated in response to infection, was the proximal ligand for Toll. Considerable doubt remained as to whether direct interaction between LPS and TLR4 actually occurred.

To address this question, we made use of a fact mentioned earlier: certain LPS partial structures, notably Lipid IVa, antagonize LPS when it is applied to human mononuclear cells, but act as agonists in the mouse. Lipid IVa differed from Lipid A only by the absence of two acyl side chains in the former and their presence in the latter. We hypothesized that TLR4 itself would “decide” whether those chains were present or absent, and if indeed it could do so, it must be in very close contact with lipid A or lipid IVa. A C3H/HeJ cell line had been created for us by Paola Ricciardi Castagnoli, and this cell line formed the perfect vehicle within which to test whether TLR4 actually made the decision. Lacking an active TLR4 molecule itself, but endowed with all of the other machinery needed to respond to LPS, this line was transfected to express either human or mouse TLR4 proteins, or neither. A clear outcome was obtained. In cells expressing mouse TLR4, both Lipid A and Lipid IVa could induce TNF production. In cells expressing human TLR4, only Lipid A, but not Lipid IVa, could induce TNF production. We concluded that TLR4 must indeed directly “see” LPS: a conclusion also reached by Lien, et al., who used a conceptually similar approach (Figure 14).
A little known paper published by our group inquired into the copy number of TLR4 on the surface of LPS-responsive macrophages. In the RAW 264.7 cell line, we estimated that only a few hundred receptors exist per cell. Yet these cells respond vigorously to LPS, consistent with strong signal amplification. Knowing that the lethal effect of LPS is delivered by myeloid cells in the mouse, and knowing the approximate number of TLR4 molecules per cell, I calculated that the dramatic shock syndrome and lethal effect of LPS are delivered by only a few nanograms of TLR4 protein in the mouse. It was truly a tiny spark that lit the fire of endotoxic shock.

But not all of the receptor complex had been discovered in 1998. The following year, Miyake and colleagues reported that a small protein, MD-2, was tightly associated with TLR4 and was also important for LPS perception. This report was soon verified by gene targeting. A study analyzing the ability of MD-2 to discriminate between agonistic and non-agonistic LPS partial structures soon indicated that this molecule, too, must have direct contact with the ligand.
Several years elapsed before X-ray crystallography illustrated the exact mode of interaction between LPS and the receptor complex. The acyl chains of Lipid A are mostly contained within a hydrophobic pocket formed by MD-2, but parts of the molecule also have direct contact with TLR4. A reaction mechanism favoring rotational rearrangement of the TLR4/MD-2 homodimer to produce a signal has been proposed, based on the different interactions of LPS as compared with an LPS antagonist (Figure 15).

Figure 15. 3D crystallographic model of the interaction between LPS with the TLR4/MD-2 receptor complex. The acyl chains of the Lipid A component of LPS (yellow [carbons] and red [oxygens]) are mostly contained within a hydrophobic pocket formed by MD-2 (purple), but parts of the molecule also have direct contact with TLR4 (light blue/green). The TIR domain of TLR4 mediates homo- and heterotypic protein interactions during signal transduction. Reproduced from the work of Park, B.S. et al. (96).

OTHER TLRs RECOGNIZE OTHER MICROBIAL LIGANDS

By the time TLR4 was identified as the LPS receptor, four other TLRs were already known to exist, and we and others were soon to identify and clone several others, until a total of 12 TLRs were identified in mice and 10 in humans. But the specificities of the other TLRs remained unknown. The fact that TLR4 is a specific receptor for LPS suggested the possibility that each of the other TLRs recognize other microbial ligands. Shizuo Akira led the way in testing this hypothesis, targeting all of the TLR-encoding genes in the mouse. In due course, it was clear that each TLR did recognize specific molecules of microbial origin. Hence, the qualitative similarity in responses to many different microbial ligands was explained by the similarity of their receptors.

X-ray crystallography has now shown that different ligands bind their respective TLRs in strikingly different ways (Figure 16). Some do so in conjunction with helper proteins, or co-receptors, as discussed below. In all
instances, signaling is mediated by the recruitment of adaptor proteins, with structural similarity to the cytoplasmic domains of the TLRs themselves. In turn, protein kinases are recruited to the activation complex; then ubiquitin ligases modify and recruit still other proteins, some with other kinase activities. These events lead to the transcriptional and translational activation events seen in TLR signaling, and ultimately, to cytokine release and still more events downstream.

**Figure 16.** TLR ligands and interactions with receptors. (A) Three-dimensional structures of the lipopeptide Pam2CSK4 (from 3A79), double-stranded RNA (from 3CIY), LPS (from 3FXI), flagellin (3K8V), tRNA as a model of single-stranded RNA (2L9E), and unmethylated CpG-DNA (from 3QMB). Gray, blue, red, orange, and yellow spheres represent carbon, nitrogen, oxygen, phosphorus, and sulfur atoms, respectively. The chemical structures of resiquimod, imiquimod, and loxoribine are also shown. Possible microbial sources of ligands are indicated. (B) Structures of TLR2–TLR6–Pam2CSK4 lipopeptide (3A79), TLR2–TLR1–Pam3CSK4 lipopeptide (2Z7X), TLR3–dsRNA (3CIY), and TLR4–MD-2–LPS (3FXI). Side view (upper panels) and top view (lower panels) are shown. Protein Databank ID numbers are indicated in parentheses. (All figures were generated with Schroedinger PyMol).

**EVOLUTIONARY IMPLICATIONS: MAMMALS, INSECTS, AND PLANTS**

As noted above, even before the immunological function of Toll was known in Drosophila, it had been noticed that the TIR domain of Toll and the IL-1 receptor were similar, and furthermore, that both shared homology to certain plant pathogen resistance proteins; hence the designation TIR (for Toll/IL-1 receptor/Resistance motif). It was soon noticed that leucine rich repeat motifs were commonly associated with resistance factors in plants, but for a time, the only examples were represented in cytoplasmic proteins, rather similar in overall domain structure to the NOD-like receptors (NLRs)
of mammals. Pamela Ronald’s discovery in 1995 of a cell surface LRR known as XA21, responsible for resistance to *Xanthomonas oryzae pv. Oryzae* in domestic rice (*Oryza sativa*), was exceptionally important, and receives far less attention than it should. This protein does not signal by way of a TIR motif, but way of a non-RD kinase motif. Like TLRs in mammals (but not in flies), it recognizes a conserved molecule of microbial origin: as Ronald later showed, a sulfated peptide, AXYS22, produced by the microbe. These discoveries cement the relationship between cell surface LRR proteins in mammals, insects, and plants as sensors of infection, and reveal how truly ancient and strongly conserved innate immune sensing mechanisms are (Figure 17). A related and strictly personal note: Pamela Ronald and I, whose interests in innate immunity developed entirely independently and then converged, are third cousins, both descendants of Fanny Frank (b. 1834) and Julius Rothstein (b. 1834) (Figure 18). Perhaps our mutual interest in innate immunity was itself innate!

![Figure 17. TIR domains in immunity across the tree of life. Plant disease resistance genes, Drosophila Toll, and mammalian Toll-like receptors all bear the TIR domain, which contributes to immunity. Some plant resistance proteins, such as XA21, have leucine-rich repeats (but no TIR domain) and directly engage microbial activators at the cell surface as Toll-like receptors do.](image-url)
Figure 18. The Beutler and Ronald families, related by descent. XA21 and TLR4 may be as well, and have obvious functional similarity.

In insects, Jules Hoffmann and his colleagues identified a Toll-independent sensory system that recognizes Gram-negative bacteria. Termed the Imd pathway, it mimics at several points the mammalian TNF signaling pathway. One may therefore think of the mammalian TLR signaling pathways, inexorably linked to the production of TNF and then to TNF signaling, as equivalent to both the Toll and Imd pathways in the fruit fly (Figure 19). The “connecting” role of TNF, which joins the two pathways, gives some insight as to why TNF blockade is particularly effective as a therapy. TNF signaling is obviously one of the major mechanisms by which innate immunity and inflammation are implemented, and has been conserved across the evolutionary divide between vertebrates and invertebrates.
Figure 19. Comparison of the Toll and Imd signaling pathways with mammalian TLR and TNF signaling pathways. Each red X in the TLR signaling pathways corresponds to one or more mutations produced using ENU and isolated by its phenotypic effect. Homologues in the mammal and insect are given similar shapes and colors.

PUTTING PHENOTYPE FIRST: ENU MUTAGENESIS IN THE ANALYSIS OF INNATE IMMUNITY

When the positional cloning of Lps lay behind us, we sought to use forward genetics to further dissect innate immunity in mammals. We and others anticipated sweeping advances in mouse genetics, brought about by two developments. First, the sequencing and annotation of the mouse genome, soon to be completed, would make it unnecessary to build contigs or search for genes within critical regions. The complete gene content of the mouse would be known. Second, new sequencing technologies, a bit further off, would make it possible to find mutations with unprecedented speed. To exploit these advances, I foresaw the need to create new immunologically relevant phenotypes, and to do so using a random process.

In the year 2000, I relocated my laboratory to The Scripps Research Institute, where I began to mutagenize mice using the germline mutagen N-ethyl-N-nitrosourea (ENU) (Figure 20). In due course, we created hundreds of phenovariant mice, many with altered immune function. Over the years, we screened more than 150,000 animals for recessive defects of immunity, and identified many new and informative genetic diseases. We developed computational and robotic methods to target our search for mutations to the coding region and splice junctions of the mouse genome. This
allowed us to tackle critical regions vastly larger than we could have in the past, which of course meant much less genetic mapping. In some instances, we were able to build quite elaborate models that shed light on immune reactions, based on the identification of dozens of mutations affecting defined immune phenomena.

![Diagram of mutagenesis strategy and resulting phenotypes](image)

**Figure 20.** Making new phenotypes in mice. The germline N-ethyl-N-nitrosourea (ENU) mutagen was used to generate mutations in C57BL/6J mice. Left, the mating strategy used to generate G3 mice homozygous for a fraction of the ENU induced mutations (red box) from mutagenized (G0) progenitors. G3 mice were subjected to screening. Right, examples of appearance-altering mutations in G3 mice (clockwise, from top left): mutations in *Tmprss6* (mask), *Lepr* (business class), *npr3* (eel), *Sox10* (Dalmatian), and *Hr* (mister clean). Just as mutations cause visible and/or behavior phenotypes, they may also disrupt immune function. Mutations of this type can be detected by phenotypic screens that test immune competence. For more information on these mutations visit [http://mutagenetix.utsouthwestern.edu/home.cfm](http://mutagenetix.utsouthwestern.edu/home.cfm).

For example, we asked: “what genes are essential for survival during infection with mouse cytomegalovirus?” Many mutations were found to affect survival, and a reasonable picture of the events that must occur to allow the host to survive could be assembled. In this particular case, TLR sensing (especially via TLRs 3 and 9) are crucial; so is activation and effector function in the NK cell compartment; and so is the ability to produce NK cells, conventional dendritic cells, and inflammatory monocytes, along with the capacity for cardiovascular adaptation to the cytokine response. Several postdoctoral associates, including Karine Crozat, Ben Croker, Micha Berger, Celine Eidenschenk, Nengming Xiao, and Carrie Arnold participated in finding these mutations and analyzing them (Figure 21).
Figure 21. Genes essential for survival during mouse cytomegalovirus infection. Mutations causing MCMV susceptibility are denoted in red near their respective gene names (black) and/or protein structures. For more information on these mutations visit http://mutagenetix.utsouthwestern.edu/home.cfm. TLR signaling (specifically via TLRs 3 and 9) in conventional dendritic cells (cDCs) and inflammatory monocytes is crucial for survival; so is activation and effector function in the NK cell compartment; the ability to produce NK cells, cDCs, and inflammatory monocytes, and the capacity for cardiovascular adaptation to the cytokine response.

We also asked, “What genes maintain intestinal homeostasis when mice are challenged with oral dextran sodium sulfate (DSS)?” DSS, administered at a low dose, causes mild, reparable damage to the mucosa of the gastrointestinal tract. But exceptional mice develop severe inflammatory bowel disease as a result of mutations that interfere with mucosal proliferation, or with the immune function of hematopoietically derived cells adjacent to the epithelial layer. Again, TLRs were found to be involved in the repair process, and appear to act in mucosal cells, triggering the release of growth factors needed to close the wounds that DSS causes. Other essential events include the secretion of granules from Paneth cells and goblet cells; the degranulation of immune cells; the uptake of water within rapidly dividing cells of the epithelium; and the ability to manage ER stress. Katharina Brandl and Wataru Tomisato have taken the lead with this screen, which implies, overall, that there may be many monogenic causes of IBD, presumably in humans as in mice (Figure 22).
Figure 22. Maintenance of intestinal homeostasis after dextran sodium sulfate (DSS) challenge, as deduced from ENU mutagenesis. Microbe sensing after initial disruption of the epithelial barrier is accomplished by TLRs, which stimulate the release of growth factors needed to close epithelial defects. Other essential events include the secretion of granules from Paneth cells and goblet cells; the degranulation of immune cells; the uptake of water within rapidly dividing cells of the epithelium; and the ability to manage ER stress. For more information on these mutations visit http://mutagenetix.utsouthwestern.edu/home.cfm.

In more focused screens, we looked closely at the signaling pathways utilized by TLRs themselves, and found a number of surprises (Figure 23). TLRs 3, 7, and 9, which sense nucleic acids, must be escorted to endosomal compartments within which they signal. A protein called UNC93B1 is essential for this process, and a mutation identified in our laboratory, called 3d to connote a triple defect of nucleic acid sensing, could abrogate signaling via TLRs 3, 7, and 9.92 Another mutation, called Lps2 because it closely mimicked the phenotype imparted by the classical Lps mutation, was seen to inhibit TLR4 signaling and to completely abolish TLR3 signaling, suggesting a common adaptor protein. It was tracked to the gene encoding a new adaptor, independently identified and called TRIF by the Akira group,93 and TICAM1 by the Seya group.94 The Lps2 allele, created prior to the knockout, first revealed the basis of MyD88-independent signaling by the TLRs.95 A mutation called Oblivious revealed the importance of CD36, a class B scavenger receptor, in signaling via TLR2 heterodimers.96 And a mutation called Feeble showed that a solute channel, SLC15A4, is essential for plasmacytoid dendritic cells to detect nucleic acids via TLRs 7 and 9.97
Figure 23. Overview of Toll-like receptor (TLR) signaling pathways. Shown are the signaling events downstream of TLR activation that ultimately lead to the induction of thousands of genes including TNF and type I IFN, which are critical in activating innate and adaptive immune responses. TLR1,2,4,5 and 6 are located at least largely at the cell surface, while TLR3,7, and 9 are located in endosomes. Once TLR complexes recognize their ligands, they recruit combinations of adaptor proteins (MyD88, TICAM, TRAM, TIRAP) via homotypic TIR domain interactions. Death domains (DD); osteopontin (OPN), vesicular stomatitis virus glycoprotein G (VSV-G); lipoteichoic acid (LTA); diacyl-lipopeptide (LP2). Phosphorylation events are represented by small yellow circles labeled with a “P”. Other proteins are indicated with standard nomenclature. In boxes, ENU mutations, often representing multiple alleles of individual genes, that helped to elucidate pathways are listed (red text) with indication as to which genes they affect (black text). For more information on mutations in the TLRs or TLR-associated pathways visit http://mutagenetix.utsouthwestern.edu/home.cfm.

Another screen pursued by Carrie N. Arnold from my group in collaboration with Gunilla Karlsson Hedestam, Gerry McInerny, and Pia Dosenovic at the Karolinska Institute, has asked simply, “What is needed for an antibody response to an administered antigen?” This screen probes both T-dependent and T-independent immunization. One interesting observation is that while many mutations abolish the antibody response to an administered antigen, none of those identified to date seem to do so solely by virtue of an effect on the innate immune response.

The beauty of forward genetics is that one renounces hypothesis as much as possible. This is an act of humility; an admission that the system one is studying is too mysterious for guesses. Without hypotheses, the search for fundamental causes is far less susceptible to bias, and therefore, mistakes. At the same time, one may hope to discover the unexpected, because one has no strongly held expectations. And once a phenotype exists where none did before, it may spawn new views of how the system operates.
AFTERMATH AND FUTURE PROSPECTS

When a fundamental discovery has been made, there is a tendency to overreach. TLRs explain much of what happens during infection, and that is why they are so important. But not *everything* that happens during infection begins with TLRs. There are other systems for sensing infection, which emerged in the wake of the TLR4 discovery: the NOD-like receptors (NLRs), the RIG-I like helicases (RLHs), and the C-type lections (CTLs), for example. These must not be ignored in seeking to explain infection-related phenomena, and there may be still other sensors of which we are currently unaware.

As previously mentioned, LPS has an adjuvant effect, known since the 1950s.\(^9^8\) It was well documented that this effect, like almost all effects of LPS, is dependent on the *Lps* locus.\(^9^9\) Therefore, from the moment TLR4 was identified as the critically altered protein in C3H/HeJ mice, it was explicitly clear that TLR4 mediates the adjuvant effect of LPS.

However, blanket statements that the TLRs are “necessary” or “required” for adaptive immune responses are frankly incorrect. It is easy to demonstrate that TLRs are not necessary for adaptive immune responses to antigens administered with classical adjuvants, including those that employ microbes or their products.\(^1^0^0\) And from what we know so far, few if any authentic infections depend upon TLR signaling to elicit an adaptive immune response. Nor are TLRs necessary for allograft rejection, although as discussed below, they do play an important part in the development of certain forms of autoimmunity, particularly systemic lupus erythematosus.\(^1^0^1\) It may be inferred that the pathways leading to activation of an adaptive immune response are rather redundant.

What will our understanding of the TLRs do for us in the future? We have begun to realize, as with rheumatoid arthritis and Crohn’s disease in the past, that the systems that evolved to limit infection are the same systems that malfunction to cause inflammatory diseases. Ann Marshak-Rothstein and her colleagues have shown that TLR signaling causes an important forward feedback loop that propels autoimmunity at the B cell level. It appears that a B cell with specificity for nucleic acids will engage DNA or RNA complexes upon encountering them. At that point, internalization of the nucleic acids may drive TLR-mediated expansion of the B cell clone. This general mechanism seems particularly important in systemic lupus erythematosus (SLE) pathogenesis, and may account for the prevalence of antibodies against nuclear components in that disease.

SLE is well modeled in mice of several strains; for example, the MRL\(^Lpr\) strain, with its critical Fas mutation, the NZW x NZB F1 hybrid, and the BXSB strain with its well-known Y-linked accelerator of autoimmunity (*Yaa*) locus. In the latter strain, the critical contribution of TLR signaling to development of autoimmunity was revealed by the observation that the *Tlr7* locus is duplicated within the pseudoautosomal region of the Y chromosome.\(^1^0^2\) Homozygosity for *Tlr7* or *Myd88* knockout alleles, or for the \(^3d\) allele of *Unc93b1*, will each suppress autoimmunity caused by the Fas\(^Lpr\) mutation.\(^1^0^3\)
These observations give reason to think that highly specific therapies for amelioration of SLE may be based on antagonism of individual TLRs or proteins that support their action. Other autoimmune and autoinflammatory diseases may similarly be found to be TLR dependent.

We know, too, that death from infection sometimes results from mutational defects in the TLR signaling pathways. Innate immune deficiency is not a newly discovered phenomenon, but more and more mutations that cause it are being identified. The vulnerabilities these mutations impart may be relatively selective, and one wonders if each of us has his own Achilles’ heel.

Finally, returning to infection, and to sepsis where our story began: might TLR blockade, instituted promptly and in conjunction with appropriate antibiotic therapy, spare some patients the severe injury that sepsis causes? One may hope so, and be reasonably optimistic.


