THE ROYAL SOCIETY

Challenge of investigating biologically relevant functions of virulence factors in bacterial pathogens

Richard Moxon* **and Christoph Tang**

Molecular Infectious Diseases Group, Oxford University Department of Paediatrics, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK

Recent innovations have increased enormously the opportunities for investigating the molecular basis of bacterial pathogenicity, including the availability of whole-genome sequences, techniques for identifying key virulence genes, and the use of microarrays and proteomics. These methods should provide powerful tools for analysing the patterns of gene expression and function required for investigating host–microbe interactions *in vivo*. But, the challenge is exacting. Pathogenicity is a complex phenotype and the reductionist approach does not adequately address the eclectic and variable outcomes of host–microbe interactions, including evolutionary dynamics and ecological factors. There are difficulties in distinguishing bacterial 'virulence' factors from the many determinants that are permissive for pathogenicity, for example those promoting general fitness. A further practical problem for some of the major bacterial pathogens is that there are no satisfactory animal models or experimental assays that adequately reflect the infection under investigation. In this review, we give a personal perspective on the challenge of characterizing how bacterial pathogens behave *in vivo* and discuss some of the methods that might be most relevant for understanding the molecular basis of the diseases for which they are responsible. Despite the powerful genomic, molecular, cellular and structural technologies available to us, we are still struggling to come to grips with the question of 'What is a pathogen?'

Keywords: microbial infections; bacterial virulence; pathogenesis; genomics

> '. . . although microorganisms grown *in vitro* are conveniently studied, in this field of pathogenicity they can be incomplete and even misleading.' (Smith 1972)

1. INTRODUCTION

From the moment we are born, humans are exposed to a myriad of micro-organisms, some of which become resident on skin, mucous membranes and in the gastrointestinal tract—the process of colonization. In most instances, microbes and man coexist in mutually benign or even symbiotic partnership; commensal organisms may stimulate host immunity and provide some essential nutrients and co-factors. Rarely, these resident or newly acquired organisms result in disease in which specific interactions between host and microbe culminate in damage to host tissues and impairment of health. This potential of microbes to injure and even kill their hosts is an essential characteristic of pathogens and represents a constant challenge to human health in every part of the globe.

But, this simple perspective does not do justice to the challenge of defining pathogenicity, a complex phenotype that is dependent on the interactions between a spectrum of different host and microbial molecules. This mutuality of host and microbe, and its implicit co-evolutionary implications, are key concepts. The distinction between commensal and pathogenic behaviour is not a sharp cut-off, but a continuum and must take into account hosts that are compromised (opportunistic infections) and those with competent innate and acquired immune mechanisms. Falkow (1997) defines a pathogen as 'any micro-organism whose survival is dependent upon its capacity to replicate and persist on or within another species by actively breaching or destroying the cellular or humoral host barrier that ordinarily restricts or inhibits other micro-organisms'. Recognizing additional complexity, he emphasizes the importance of (i) ecological factors, for example the capacity to reach a unique host niche free from microbial competition, (ii) variations in host clearance mechanisms, and (iii) evolutionary factors that determine the efficiency with which microbes are transmitted to new, susceptible hosts. But, elusive though it may be, the definition of what is a pathogen is not just an issue of semantics, but a conceptual issue of substance. What we understand by pathogenicity determines and, importantly, confines how we think and therefore has major implications for the theoretical and experimental approaches that we deploy to understand its biological basis (Read *et al.* 1998). (Some use pathogenicity and virulence interchangeably. Others use pathogenicity as a qualitative term, the capacity to damage host tissues or decrease host fitness, and virulence as a quantitative term, for example the challenge dose of organisms that kills a proportion of hosts under defined experimental conditions.)

*Author for correspondence (moxon@paediatrics.ox.ac.uk).

The survival of pathogenic bacteria depends on transmission from one host to another. This requires factors that facilitate dissemination, translocation and survival between hosts, as well as colonization (Lipsitch & Moxon 1997). The relationship between the microbial factors supporting fitness, commensal and virulence behaviour is complex. Early in the last century, Topley (1919) recognized that selection for enhanced transmissibility might drive pathogens towards heightened pathogenicity under circumstances where the virulence of the pathogen was linked to its rate of replication and transmission.

Ewald (1994) has built upon this general thesis and suggested that water and food are 'cultural vectors' and, by analogy to the transmission of parasites by insects, play an important role in the evolution of bacterial virulence. Our behaviour—living in towns, working in institutions, infringing on rainforests, waging wars, travelling in jet planes—are major factors affecting microbial transmission and therefore the evolution of pathogenic bacteria. The evolution of pathogenicity is all about trade-offs (Levin & Svanborg-Eden 1990). Natural selection may favour properties that facilitate successful transmission of a microbe, the major driving force being the number of secondary infections that result from a primary infection (Anderson & May 1982). However, singular explanations of the evolution of virulence are too simplistic. The meningococcus, for example, is one of the most virulent microbes of man. But it is not a simple matter to explain how the devastating virulence of this organism benefits its transmission to other humans, because the population of bacteria *in vivo* (for example in blood or spinal fluid) of a fatal case of meningococcal disease is not transmitted to other persons. What drives the evolution of its virulent behaviour?

Over the past three decades, molecular biology has brought about a revolution in the methodology of experiments on pathogenic bacteria. The classical approach to investigate a pathogen, enshrined in Koch's postulates, was transformed by advances in molecular biology (Falkow 1988). Most conspicuously, the availability of molecular genetic techniques encouraged an approach based on the use of genetically defined strains that resulted in an explosion of information on key microbial characteristics contributing to pathogenicity. However, reductionism may obscure key concepts. Too often, pathogenicity is approached as if it could be captured through identifying and characterizing a subset of the genes of a particular microbe. While appealing, this perspective is naive. The challenge of coming to grips with pathogenicity is not so very different from trying to fathom the genetic basis of, say, a champion athlete. Nobody would seriously consider that any fixed set of genes would be common to those who run 100 m in less than 10 s! Our colleagues in human genetics face similar challenges. Consider single gene disorders such as cystic fibrosis or phenylketonuria and the contrast with polygenic diseases such as hypertension or diabetes mellitus. In the field of pathogenicity, there are parallels. For example, tetanus is a disease of a single gene that can be precisely reproduced by injecting an experimental animal with the purified toxin. Contrast tetanus with the pathogenesis of bacterial meningitis in which multiple genes are implicated in a complex series of interdependent, sequential steps in pathogenesis. Crucial to our understanding of these complexities is the need to translate reductionist methods for investigation of pathogens into experiments that enlighten *in vivo* function.

Over many years, few have championed this issue more tenaciously than Harry Smith. In numerous reviews, he has lamented how investigators failed to grasp this nettle. His message is insistent, unambiguous and compelling: stop procrastinating on what might happen *in vivo* and get your teeth into an effort to find out what does happen (a paraphrase of Smith (1996)). The current symposium seems a judiciously timed opportunity to review and renew our efforts to meet the challenge of a better understanding of what a pathogen is. In this presentation, we will offer a personal perspective on what we know about the behaviour of pathogenic bacteria *in vivo* and our thoughts about how we might make progress in the future. As with all biology, the challenge is as exciting as it is daunting. Although we subscribe to the view that our knowledge, where possible, should seek to delineate functions at the molecular level, we caution that the reductionist's dream may be the holist's nightmare!

2. BACTERIAL PATHOGENS *IN VIVO*

Molecules that interact with host factors in the pathogenic process can be classified as follows (Falkow 1988; Moxon 1997):

(i) tropism for host tissues, or more specifically for particular cells of the host (adhesins);
(ii) invasion of host tissues by dissemination on, within or through host cells (invasins);
(iii) facilitating microbial survival of the host clearance mechanisms (evasins);
(iv) the strategies for acquiring nutrients and co-factors for *in vivo* growth (pabulins);
(v) damaging, directly or indirectly, host tissues (cytotoxins).

In a typical experimental approach, the role of a candidate virulence gene is studied by comparing the behaviour (in a 'biologically relevant' assay) of the wild-type bacterial pathogen with that of a variant lacking this gene. There is a compelling logic and simplicity to this paradigm, but to what extent does it stand up to scrutiny? We need to think carefully about the assumptions underlying the use of isogenic strains. The interpretation of experiments in which a single gene is expressed or not, ignoring for the moment technical genetic issues such as polar effects, may not be straightforward. The loss of one surface molecule may have significant secondary effects on other determinants. For example, capsule-deficient mutants may have altered membrane proteins (Loeb & Smith 1980), and truncated forms of lipopolysaccharide (LPS) may affect the distribution of membrane phospholipids (Zwahlen *et al.* 1985). Despite the obvious merits of using isogenic strains, these caveats and their implications cannot be ignored.

However, by far the most difficult issue is the selection of an experimental system that allows biologically relevant insights into the function of virulence factors. The best system is to investigate well-characterized bacterial pathogens in spontaneous or experimental infections of

Table 1. *Usefulness of infant rats as a model for investigating the pathogenesis of* H. influenzae *bacteraemia and meningitis*

bacteria reach the central nervous system by the haematogenous route, not via the cribrifrom plate, following intranasal challenge (Ostrow *et al*. 1979)
the probability of meningitis is directly related to the concentration of bacteria in the blood and to their exceeding a critical threshold of greater than 10^3 organisms ml^{-1} blood (Moxon & Ostrow 1977)
the population of organisms recovered from the blood or CSF in the acute phase of bacteraemia and meningitis is the progeny of a few founder bacteria, often a single clone (Moxon & Murphy 1978)
opsonophagocytosis, not bactericidal killing, is the major mechanism of *in vivo* clearance (Weller *et al*. 1978)
organisms grown *in vivo* are relatively resistant to antibody-dependent, complement-mediated killing (*in vitro*), when compared with organisms grown *in vitro* (Shaw *et al*. 1976)
expression of phosphorylcholine, a phase-variable component of LPS, facilitates nasopharyngeal colonization, whereas its absence increases intravascular survival (Weiser & Pan 1998)
expression of type b capsule, *per se*, confers heightened virulence compared with other capsular polysaccharides (Zwahlen *et al*. 1989)
mutations of LPS can attenuate systemic infection even when the bacteria are fully encapsulated (Zwahlen *et al*. 1986)
mean generation time of organisms is not less than 50 min in the acute phase of infection (Moxon 1992)
the blood is a major site of replication of organisms in the pathogenesis of bacteraemia (Rubin *et al*. 1985)
prior nasopharyngeal infection, or induction of inflammation, increases susceptibility to bacteraemia following intranasal challenge (Myerowitz 1981)
decreased concentrations of glucose in CSF during acute meningitis do not result from increased consumption by organisms or inflammatory cells (Moxon *et al*. 1979)
seeding the peritoneal cavity of splenectomized rats with fragments of spleen enhances protection against lethal challenge (Moxon & Schwartz 1980)
serum antibodies to type b capsular polysaccharide facilitate clearance of organisms from the nasopharynx (Moxon & Anderson 1979)

their natural hosts. However, for many of the most significant bacterial pathogens, this is not an option. *Haemophilus influenzae* is a good example. This potentially pathogenic bacterium is usually a harmless commensal of the human upper respiratory tract. As far as anyone knows, humans are its only host and therefore appropriate studies on the *in vivo* relevance of its virulence factors are difficult to approach experimentally. Infant rats have been used for many years as an experimental model of *H. influenzae* bacteraemia and meningitis. Table 1 summarizes some of what has been learned about *in vivo* events. However, the limitations of this model must be emphasized. To give a few examples: in almost all of these experiments, the bacterial inoculum was prepared by growing organisms in rich, artificial medium with a surfeit of essential nutrients and co-factors. Yet, it is well known that *H. influenzae* undergoes a phenotypic shift *in vivo*. When organisms were obtained directly from the nasopharynx, or from the blood of bacteraemic animals, they showed relative resistance to complement-mediated killing when compared with organisms grown *in vitro* (Shaw *et al*. 1976; Rubin & Moxon 1985). The relevant small molecular weight host factors have been identified and, at least some of the phenotypic correlates are associated with phase-variable phenotypic changes in LPS (Roche & Moxon 1995; Weiser & Pan 1998). The frequency distribution of LPS variants in a challenge inoculum prepared for experimental assays of infection is therefore critical. For example, LPS variants will be a major factor in determining the efficiency with which *H. influenzae* colonizes the nasopharynx in experimental infections of infant rats or survive intravascular clearance following systemic inoculation. The LPS of *H. influenzae* has evolved digalactoside structures, gal-α-1-4-gal-β, that mimic the glycosylation residues of human cells (Virji *et al*. 1990), but in the rat, the prevalent glycosylation on host cells is gal-α-1-3-gal-β. Whereas rats recognize this epitope, humans do not (Weiser *et al*. 1997). Thus, *in vivo*, molecular mimicry dependent on species-specific differences in host glycosylation patterns may be critical. Many of the adhesins, invasins and evasins of *H. influenzae* that interact with the host use ligand–receptor interactions that are specific to human tissues, but not those of rats. For example, IgAl proteases cleave the relevant human, but not rat, immunoglobulin (Plaut *et al*. 1977). However, some cellular features of humans and rats may be similar. Although not a natural commensal or pathogen of rats, the efficiency with which even a small inoculum of *H. influenzae* can establish colonization in the experimental setting is impressive (Moxon & Murphy 1978).

Given the difficulties in identifying relevant models of infection, it is not surprising that much research that attempts to investigate *in vivo* relevance has used cultured cells to investigate the molecular basis of the host–microbe interactions that underlie attachment and invasion. But these studies can be extremely difficult to interpret. For example, the interactions of *Salmonella typhimurium* with epithelial cells has been studied in depth, but there is a considerable body of evidence that the pathogenesis of *Salmonella typhi* (the pathogenic species of most relevance to humans) involves entry through M cells (reviewed by Falkow 1996). All too often, this aspect of *in vivo* relevance and interpretation is given short shrift. Clear distinctions might need to be made between the scientific merit of using bacteria as probes for investigating cellular functions, such as endocytosis or

trafficking, and the pragmatic issue of their relevance to pathogenesis. For example, the pathogenicity of enteropathogenic *Escherichia coli* is associated with the effacement of microvilli and cytoskeletal rearrangements. The latter have been confirmed through ultrastructural studies on jejunal epithelial cells observed in clinical specimens (Rothbaun *et al.* 1982, 1983). *In vitro* investigations have identified the molecular details of signalling proteins that induce these changes, which include protein phosphorylation and inositol and calcium fluxes, and the role of environmental factors such as the effects of temperature, growth phase of the bacteria, calcium and osmolarity (Kenny *et al.* 1997). These investigations offer a detailed coalition of *in vivo* and *in vitro* research, but how these events lead to diarrhoea is unknown!

Environmental factors, such as osmolarity, pH and available nutrients, influence bacterial pathogenicity by affecting the production of virulence determinants and by controlling the growth rate (Miller *et al.* 1989). For example, when *S. typhimurium* is present in the intestinal lumen, several environmental and regulatory conditions modulate the expression of factors required for bacterial entry into host cells. An excellent example of the complex and coordinate regulation of virulence genes *in vivo* is expression of six different invasion genes, located on the pathogenicity island SPI-1, which are coordinately regulated by oxygen, osmolarity, and pH (Bajaj *et al.* 1996). An important point is that the activation of these virulence genes is obligatorily dependent on a multiplicity of environmental cues. Specific, combinatorial responsiveness provides a mechanism whereby bacteria behave in a very organized manner, subject to subtle variations in the different extracellular or intracellular environments encountered in the host. Another fascinating issue is that, as a general rule, physical contact between bacteria and host cells is required for the expression of virulence genes in animal hosts. Upon contact with a host cell, *Yersinia pseudotuberculosis* increases the rate of transcription of certain virulence genes. This microbe–host interaction triggers export of LcrQ, a component of a type III secretion system and a negative regulator of the expression of a family of secreted proteins (Yops). A decrease in the intracellular concentration of LcrQ mediates increased expression of Yops. Thus, the type III secretion system plays a key role in the coordinate elaboration of virulence factors after physical contact with the target cell (Pettersson *et al.* 1996). Interestingly, in microbial–plant interactions, substantial molecular signalling occurs under circumstances where the infecting bacterium is spatially distanced from the target tissues.

Expression of virulence factors may depend on the microbial population attaining a critical density to trigger the elaboration of bacterial cell-to-cell signalling molecules (e.g. *N*-acyl-L-homoserine lactones in *Pseudomonas aeruginosa*) (Winson *et al.* 1995). Although relatively little work has been done *in vivo* to confirm the role of quorum sensing systems (but see Williams *et al.*, this issue) the importance of cell-to-cell signalling within bacterial populations is undoubtedly both neglected and important. *Staphylococcus aureus* uses a global regulator, *agr*, which is activated by secreted autoinducing peptides, to control the expression of its major virulence genes. The autoinducing peptides show sequence variation that affects their specificity. These peptides are thiolactones that either activate or inhibit depending on the conformation of the ligand–receptor interaction, the fine structural details of which are now understood in detail (Ji *et al.* 1997; Balaban *et al.* 1998). This has made it possible to use synthetic variants of these peptides for *in vitro* and *in vivo* assays to establish their relevance and to open the door to novel strategies of infection control (Mayville *et al.* 1999).

There has been much emphasis on the important contributions of gene regulation and regulatory networks. But, gene regulation provides only one example of the genetic mechanisms that have evolved to facilitate bacterial acclimatization to their host. Pathogenic bacteria face especially demanding tests of their adaptive potential, a consequence of the diversity of host polymorphisms (innate immunity) and the capacity to generate extensive repertoires of T and B cells (acquired immunity). Typically, bacterial infections occur within a matter of hours in which bacteria *in vivo* encounter host landscapes of such diversity that prescriptive strategies, such as two-component sensory–transducer systems or classical gene regulation, may be inadequate to encompass the plethora of potential variables. The challenge is all the more striking in that many infections stem from the clonal expansion of a single bacterial cell (Meynell & Stocker 1957; Moxon & Murphy 1978). How then do the small numbers of bacteria that make up a pathogen population in an individual host generate the necessary diversity to adapt to host polymorphisms and immune clearance mechanisms? One strategy is through increasing the mutation frequency of those genes that are involved in critical interactions with their hosts. In many pathogenic bacteria, this hypermutability occurs because selected genes have evolved sequences containing runs of repetitive DNA (Stern *et al.* 1986; Weiser *et al.* 1989). Nucleotide repeats (microsatellites) are highly mutable, because of their propensity to undergo slippage (Streisinger *et al.* 1966; Levinson & Gutman 1987). The resulting mutations, involving spontaneous loss or gain of nucleotides, result in reversible, high-frequency on–off switching of genes through altered transcription, altered binding of RNA polymerase to promoters, or translation, frame shifts (Moxon & Wills 1999). Given that there are several such genetic loci in a single pathogen genome and that the mutations occur at random and independently of one another, the combinatorial effect on the phenotypic diversity of the pathogen population can be substantial. These genes have been called contingency loci, to emphasize their potential to enable at least a few bacteria in a given population to adapt to unpredictable and precipitous contingencies [sic] within, and in transmission between, different host environments (Moxon *et al.* 1994). Traits encoded by contingency genes include those governing recognition by the immune system, motility, attachment to and invasion of host cells, and acquisition of nutrients (table 2).

A specific example of a contingency gene with a proven role in pathogenesis is *lic1a* in *H. influenzae*. The *lic1* locus is required for the synthesis of choline phosphate (ChoP), a component of *H. influenzae* LPS (Weiser *et al.* 1997). The *lic1a* gene contains multiple copies of a tetranucleotide, CAAT, in the 5′-portion of the gene and, through slippage, *lic1a* switches on and off at high frequency; the result is loss or gain of $ChoP^+$ and $ChoP^-$

Table 2. *Why might we not be identifying relevant* (in vivo) *pathogenicity factors?*

factors	reasons
adhesins, invasins	dependent on appropriate 'model' or assay
pabulins	difficulties in distinguishing 'permissive' or 'fitness' genes from virulence factors
evasins	dependent on host, immune status, route of inoculation, etc.
toxins	some methods do not detect, e.g. signature-tagged mutagenesis

LPS glycoforms. In an infant rat model, $ChoP^+$ variants colonize the nasopharynx more efficiently than $ChoP^-$ variants. However, $ChoP^+$ variants are more susceptible to C-reactive protein (CRP)-dependent, complement-mediated killing and variants found in blood and cerebrospinal fluid (CSF) are almost exclusively $ChoP^-$. Thus, the switching of this contingency gene confers an adaptive strategy that is selectively advantageous to the survival of *H. influenzae* at different stages in its pathogenesis (Weiser & Pan 1998).

Survival and commensal or pathogenic behaviour *in vivo* is dependent on the availability of nutrients, and these requirements may be the basis for tissue tropism. This was shown many years ago in the context of *Proteus mirabilis* and urea concentrations in the kidney (MacLaren 1970), and of *Brucellae* and the abundance of erythritol in ungulate foetal tissues (Smith *et al.* 1962). Although much attention has been focused on the provision of iron and its essentiality for bacterial growth, the same cannot be said for other key nutrients. How, for example do bacteria scavenge zinc to provide the metalloproteins for DNA and RNA polymerases? It is only very recently, based on the availability of whole-genome sequences, that information on these important mechanisms has been uncovered for *H. influenzae* and *H. ducreyi* (Lewis *et al.* 1999). Nicotinamide adenine dinucleotide (NAD^+) is also essential for growth of *H. influenzae* and *H. ducreyi* yet, despite its essentiality, the molecular mechanisms by which NAD^+ is taken up and transported remain uncharacterized. A periplasmic pyrophosphatase that hydrolyses NAD^+ or flavin adenine nucleotide to the mononucleotide has been described and analogues capable of inhibiting this enzymatic activity impair growth of *H. influenzae* (Anderson *et al.* 1985). However, much more remains to be learned about the contribution of these essential mechanisms to commensal and virulence behaviour.

Compared with replication *in vitro*, *in vivo* multiplication is complicated by the uncertain contribution of host immunity since population numbers are the result of replication and clearance (Brock 1971). Meynell devised an elegant method for estimating bacterial growth rates *in vivo* that gives the true division rate based on the dilution of non-replicating plasmids. This was used to estimate the growth rate of *S. typhimurium* following intravenous inoculation of mice and their recovery from the spleen (Maw & Meynell 1968). The division rates in the organisms in the spleen proved to be only 5–10% of the maximum observed *in vitro*. Interestingly, however, the death rate of the bacteria in the spleen was also extremely small. Thus, the bacteria presumably could persist in the spleen for long periods. These and similar methods could yield information of great value on many of the problems discussed.

Putative virulence factors expressed *in vitro* may not be expressed *in vivo*. Alternatively, molecules made *in vivo* may not be made *in vitro*. *S. aureus* synthesizes poly-*N*-succinyl β-1-6 glucosamine (PNSG) as a surface polysaccharide during human and animal infection, but these strains do not produce PNSG *in vitro* (McKenney *et al.* 1999). On the other hand, the *ail* gene of *Y. enterocolitica* facilitates serum resistance and invasion of cultured cells *in vitro*, but when examined in mouse models, had no detectable influence on colonization or LD_{50} following oral challenge (Wachtel & Miller 1995).

Taken together, these observations indicate that although we know many of the questions, there is a considerable shortfall, even neglect, of some of the most fundamental issues concerning the functional relevance of virulence factors. Yet, in addition to the huge body of descriptive, clinical information on bacterial infections—experiments of nature—we have available to us an array of extremely powerful resources and methods to facilitate experiments in the laboratory and the field. However, future studies should take into account the need to consider critically their limitations, as well as their strengths. For example, we can predict that many adhesins and invasins would not be identified by signature-tagged mutagenesis (STM) because the tissues of the animal model may not express the appropriate ligands, or the route of inoculation may bypass relevant tissues. STM does not identify cytotoxins, and redundancy or complementation may obscure the identification of some evasins or pabulins. In the case of *in vivo* expression technology (IVET) and differential fluorescence imaging, we must be cautious in interpreting *in vivo*, but not *in vitro*, gene transcription as a strong indicator of a biologically relevant function. This message has been powerfully affirmed by studies on yeast (Winzeler *et al.* 1999). Furthermore, constraints on export, rather than transcription and translation, are limiting factors of relevance for *in vivo* function of molecules produced by type III secretion systems. The approach to pathogens in the future must secure a greater emphasis on collaboration through research consortia. This makes sense in the context of projects driven by whole-genome sequences, where a range of expertise in genetics, cell biology (including immunology), microscopy and structural biology can be assembled. Importantly, microbiologists, pathologists and infectious diseases physicians have a wealth of experience gathered through years of observation, as well as precious collections of strains and clinical specimens. Too often, communications between the clinic and the molecular biology laboratory are inadequate to bring about the necessary convergence of information and methodologies to execute comprehensive biologically relevant *in vitro* and *in vivo* assays of virulence factors in different hosts and at different stages of an infection.

3. WHOLE-GENOME SEQUENCING HAS REVOLUTIONIZED STUDIES OF PATHOGENIC BACTERIA

The availability of complete bacterial genome sequences has had a major impact on the opportunities for investigating

the biological basis of pathogenicity, although enormous challenges still remain (Strauss & Falkow 1997). The impact of genomics is far reaching because it provides the most economical means of acquiring large amounts of information and because it has forged the creation of new technologies to exploit these data. Furthermore, some pathogens are highly specialized microbes that are difficult to grow in the laboratory. As we move from the era of the gene to the era of the genome, it is essential that we move beyond the functional characterization of isolated genes. To make sure genome analysis fulfils its potential with respect to pathogenesis it is essential to make bioinformatic tools, services, and methodologies accessible to the scientific community (Bork *et al.* 1998). This requires combining traditional reductionist approaches with studies that consider genome-level organization and population biology. Gene context, gene organization and gene acquisition must be considered in the experimental approaches to bacterial pathogens. This will require collaboration between computational biologists and those experimentalists who intimately know their respective organisms at the molecular and whole-organism levels. We should also not forget that the ultimate challenge and opportunity for genomics is the completion of the host genome sequence, when we can evaluate pathogens within the genetic context of their hosts.

Whole-genome sequences make a unique contribution, one that differs and contrasts importantly with the tools of the classical methods and their technological refinements. By providing an inventory of every nucleotide of a pathogen, experimental approaches based on whole-genome sequences can command unprecedented rigour; the great advantage of a 'top-down' approach is that it is comprehensive, a feature that distinguishes it from most other methodologies, such as classical genetics. For example, random mutagenesis, offers a powerful approach; but because mutations in many genes are absolutely or conditionally lethal, many potentially important virulence factors remain obscure. Even worse, we cannot know how many we are missing!

Appropriate informatics is essential to unlock and make available the immense fund of information contained in whole-genome sequences (Field *et al.* 1999). Every genome project has reported the identification of multiple putative virulence-related genes scored by homology. Given the large number of genome sequences now available, it is becoming common to apply 'species-filter' approaches to data, mining for information about genes that control virulence phenotype, a process that entails subtracting all homologues of one genome from another to define common, as opposed to unshared, genes. There are 116 genes in *H. influenzae* that have no homologue in *E. coli* but have a known function or have a similarity to a gene of another genome. Out of these, a search was made for proteins potentially involved in the interaction of *H. influenzae* with its host. Almost half of the open reading frames (ORFs) fulfilled the criteria of being found exclusively in pathogens or having sequence similarity to putative adhesins, invasins or cytotoxins (Huynen *et al.* 1997). An elegant example of the application of microarray methods to virulence concerns bacille Calmette–Guérin (BCG) vaccines that are used as live attenuated vaccines against diseases caused by *Mycobacterium tuberculosis*. Comparative hybridizations between the genomes of *M. tuberculosis*, *M. bovis* and the various daughter BCG strains were undertaken. Twelve novel deletions were found in daughter lineages compared with the progenitor strain, findings that may shed light on the varying effectiveness of different lots of BCG vaccine and their association with attenuated virulence (Behr *et al.* 1999).

It is essential to understand how genes are organized within genomes to produce the complex phenotypic traits underlying commensal or virulence behaviour. One example is the development of databases capable of piecing together metabolic pathways. The concept of the '*in silico*' cell is still in its infancy but will surely be refined so as to give us strong clues as to how bacterial cells can function in particular environments. This hypothesis-generating approach can be coupled to the use of microarrays (see § 4), a method allowing genome-based surveys of genetic variation and differences in expression profiles under different conditions, including infected tissues. Another important contribution of whole-genome sequences is the identification of virulence-related genes acquired by gene transfer. These gene acquisitions are often extremely obvious through an analysis of the nucleotide composition of the complete genome by the detection of sequences that vary significantly from those typical of the species (Censini *et al.* 1996).

If the information from genomes is not accessible to those who have the expertise to exploit it, then progress will be thwarted. It has taken a few years to orchestrate the liaison between genome information and the various experimental methods for its exploitation. One of the lessons learned is how essential it is that those who have a command of the biology are involved in whole-genome sequencing projects from the earliest stage. The absence or deficiency of these appropriate collaborations is an important reason why, to date, there are relatively few examples of genome-driven translational biology in the literature, even though the first complete genome sequence of a pathogenic bacterium was completed in 1995. However, this message has been assimilated and consortia are now being organized to undertake comprehensive, genome-based analyses of pathogens with an emphasis on biologically relevant functions. A consortium approach was taken from the outset to sequence *Bacillus subtilis*. This has proved to have great advantages in the subsequent coherence of exploiting the genome information, but another clear lesson is that specialized units or institutes can do the sequencing immeasurably more efficiently and economically. Thus, sequencing projects should be carried out in a seamless collaboration that also provides appropriate expertise in bioinformatics, systematic tagging of all genes (protein-coding and stable RNA genes), mutational, transcriptional and translational studies in concert with *in vitro* and *in vivo* assays.

The availability of the 1.83 megabase-pair sequence of the *H. influenzae* strain Rd genome facilitated significant progress in investigating the biology of its LPS, a major virulence determinant of this human pathogen. By searching the genomic database with sequences of known LPS biosynthetic genes from other organisms, 25 candidate LPS genes were identified and cloned. Construction of mutant strains and characterization of the LPS by

reactivity with antibodies, polyacrylamide gel electrophoresis (PAGE) fractionation patterns and electrospray mass spectrometry confirmed a potential role in LPS biosynthesis for the majority of these candidate genes. This allowed a series of virulence studies in an infant rat model to estimate the minimal LPS structure required for intravascular dissemination (Hood *et al.* 1996).

Informatics provides a critical 'hypothesis-generating' starting point for biologically relevant experiments. However, its usefulness is dependent on intelligent and expert application. Sequences identify ORFs, i.e. predicted but not proven genes, whose putative functions can be hypothesized but not assumed without critical appraisal of evidence. Database entries vary enormously in their accuracy, appropriateness and the extent to which they have been validated experimentally. An excellent example to illustrate this point is provided by work on calmodulin, a small Ca^{2+}-binding protein that translates Ca^{2+} signals into cellular responses in yeast. An analysis of mutants indicated that deletion of the gene was lethal but reduction or even complete ablation of calcium binding did not interfere with efficient growth. Thus, calmodulin can perform essential functions without the apparent ability to bind Ca^{2+}. Far from diminishing the importance of genetics and biochemistry, genomics has heightened their importance in investigating the assumptions, extrapolations and speculations that habitually surround assignments of function (Geiser *et al.* 1991).

The power of genomics in raising and refining key questions must be emphasized. For example, a striking finding in the completed genome sequences of every bacterium to date has been the number of genes lacking either any homology to existing sequences in the current databases or with homologies to genes of unknown function. Depending on criteria, we are unable to assign functions to about 15–40% of genes, although the genome sequence of *M. tuberculosis* seems to be an exception. It is reasonable to suppose that many of these genes in bacterial pathogens will be important virulence factors *in vivo*, and in a few instances, this has been shown to be so (Heithoff *et al.* 1997; Martindale *et al.* 2000).

Five of the six genomes published in 1998–1999 were human pathogens, all host adapted. Four of these were obligate intracellular pathogens and their study is providing novel insights into host–pathogen interactions and co-evolution. Many have a lifestyle that is characterized by a relative reduction in genome size, the numbers of genes and the complexity of regulatory elements as well as an increased dependence on the host environment for essential nutrients. Several genome projects are very significant because they mark the beginning of a significant trend in the sequencing of closely related genomes (e.g. *Mycobacteria*, *Neisseria*, *Mycoplasma*, *E. coli*, *Helicobacter*). This is important because the availability of a complete sequence from a single strain is only the starting point in understanding the genetic diversity of the natural population and its implications for commensal and virulence behaviour.

The information from one index genome of a pathogen must be extended to embrace the diversity of the species. And one obvious drawback of genomics is that usually only one strain of a pathogen is sequenced, although there are exceptions to this (*M. tuberculosis*, *Helicobacter pylori*, *Neisseria meningitidis*, *E. coli*). Nonetheless, the substantial diversity that exists in the natural populations of the bacterial species must be emphasized. In particular, an appreciation of the extent of horizontal transfer of DNA from the global gene pool is critical. Thus, as well as allelic diversity through shift and drift, some strains within a species may possess genes (or sequences within genes) that are denied to others. An example of this is Opc, an adhesin and invasin, found in only a proportion of *N. meningitidis* strains (Seiler *et al.* 1996). Some genes may be crucial to virulence, or contribute to heightened virulence. The capsule of *H. influenzae* is clearly implicated in the potential to cause invasive infections, and of the six distinct capsular polysaccharides, that of serotype b (polyribosyl-ribitol phosphate) heightens pathogenicity in a fashion denied to other capsular serotypes (Moxon & Kroll 1990).

We need to provide a clear understanding of the differences underlying pathogenic and commensal behaviour, or the potential for different diseases. *E. coli* is responsible for many different clinical syndromes of gastrointestinal disease, and to a great extent these differences can be attributed to the acquisition of specific virulence genes through horizontal transfer. *E. coli* well illustrates the depths of our ignorance about *in vivo* behaviour. Enterotoxigenic *E. coli* strains elaborate a toxin that is similar to that of cholera toxin. A great deal of genetic, structural and biochemical data have been accrued on the toxins of both these pathogens, but why after an initially clinically similar prodromal illness do classic cholera strains go on to cause a much more severe, life threatening disease? Another seeming paradox is the production of toxin (PT), a major virulence factor in *Bordetella pertussis*. *In vivo* evidence provides strong support for its key role in pathogenesis since a subunit vaccine based on PT protects against disease. However, *B. parapertussis* lacks PT, but clinical cases of comparable severity occur (Hoppe 1999; He *et al.* 1998). So what is the explanation?

Whole-genome sequences are also opening the door to understanding the importance of higher-order processes (expression patterns, gene regulation, kinetic properties, localization and concentration effects, environmental influences, and fitness contributions). When we do observe what happens *in vivo*, we need to consider not merely qualitative aspects of function (e.g. the potential to induce inflammation), but the quantitative correlates (e.g. the extent to which the factor induces inflammation). Correlation of genotype and virulence potential is not black and white, but replete with subtleties. These issues are critical in our efforts to evaluate vaccine or drug targets (for example, where absolute conservation of an epitope is sought) as contrasted with, for example, novel diagnostic possibilities that depend on identifying variations. Furthermore, we have not been in a position to appreciate the variability in metabolic pathways and these may critically impact on *in vivo* behaviour at different steps in pathogenesis.

4. SYSTEMATIC, GENE-BY-GENE, ANALYSIS OF BACTERIAL PATHOGENS

There is now great excitement among the community of molecular microbiologists concerning prospects for the

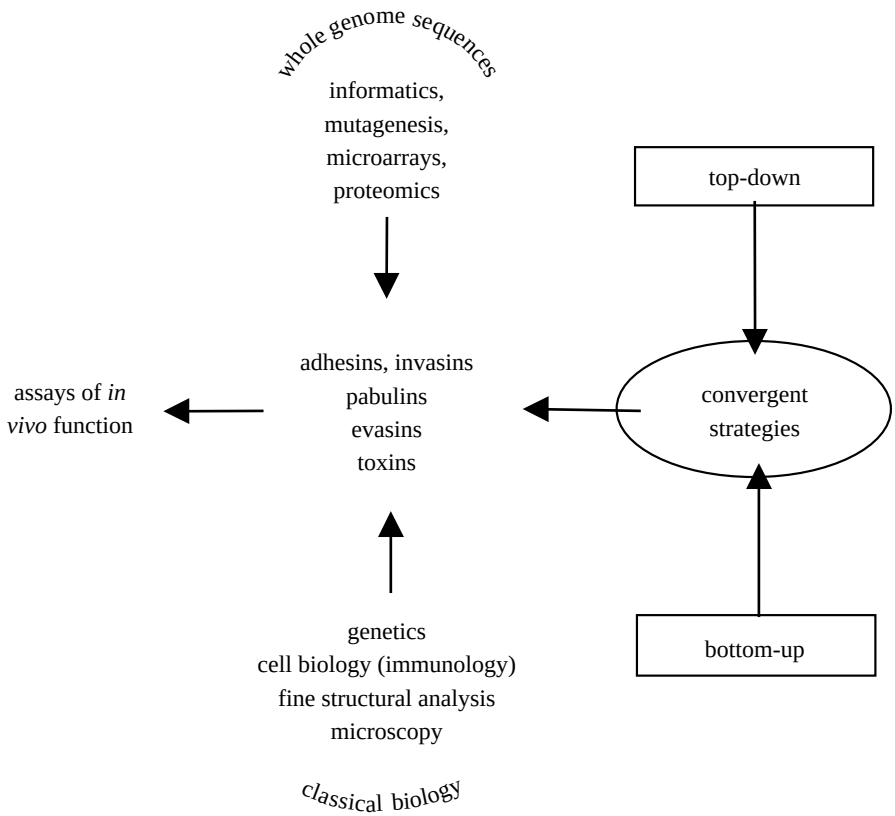

Figure 1. Convergence of strategies for investigation of bacterial pathogen *in vivo*. The goal is to identify molecules of pathogenic bacteria that are crucial to virulence. Adhesins, invasins, pabulins, evasins and toxins (see § 1 for further details of these) can be identified by a combination of genomic and classical genetic approaches. Both genomic and classical genetic approaches have a variety of powerful supporting technologies to facilitate their exploitation. The big challenge is to translate these tools of discovery into an understanding of biological function in pathogenesis.

comprehensive analysis of bacterial pathogens. For many of the major pathogens, there is access to complete genome sequences, appropriate techniques for genetic transfer are in place and importantly, well-validated animal models are available. There is a wealth of accumulated observational data on the biology of infections, as well as many years of accumulated data obtained by microbiologists. Also, there are extensive strain collections, and specimens of infected tissues from biopsies and autopsies—not forgetting the important insights that can be afforded by studies on ancient DNA!

The potential for substantial progress is reflected in recent publications on functional genomics (Pallen 1999; Fields *et al.* 1999; Winzeler *et al.* 1999). The starting point is the identification, through annotation of the pathogen's genome, of the coding sequences for all proteins and stable RNAs. For a host-adapted pathogenic bacterium, this typically amounts to between 1500 and 2000 putative genes. For many of these genes, no functions are known or the sequences lack any homologies to entries in available general databases. Indeed, incisive data on the function of the majority of these genes will be lacking. Precise deletion of all genes, for example using inverse PCR with sequence tagging (bar-coding) and/or fluorescein labelling, can be attempted using a variety of directed strategies, which are likely to realize some 500–700 mutants. These mutants can be investigated using high throughput screening techniques (see § 5(b)), starting with growth in rich and minimal media *in vitro* and thence to tissue or organ culture, embryonated chicken eggs, small animals (rats, mice) and eventually natural hosts including, possibly, humans (Cohen *et al.* 1994). For many of these experiments, pools of deletion mutants can be used in the first instance and aliquots sampled at various times during the growth of cultures or stages of infection. Tags can then be amplified and hybridized to microarrays (see § 5(b)). Although apparently simple in outline, such ambitious, comprehensive analyses represent a formidable task. Ideally, this programme of investigation should use a consortium approach (figure 1).

The pathogenesis of bacterial infections is accompanied by a series of molecular changes in both the bacterium and the host cells. These events result in altered expression of a number of genes. There are now a number of methods for comparing global gene expression profile changes in the bacterium and the host cells. These methods include oligonucleotide arrays (Lockhart *et al.* 1996), serial analysis of gene expression (SAGE) (Velculescu *et al.* 1995), differential cDNA screening (Peitu *et al.* 1996), expressed sequence tag database comparison (Vasmatzis *et al.* 1998), and two-dimensional gel electrophoresis of cellular proteins. Many of these methodologies are labour-intensive and not yet suitable for high-throughput use, for example those dependent on gel-based methods such as SAGE. However, oligonucleotide arrays are a realistic high-throughput methodology, although many technological problems still

need to be refined and validated. Nonetheless, there are already many published examples where microarrays have been used to provide identity and expression levels of selected genes simultaneously, providing that the gene sequences are known; i.e. the complete genome sequence of the relevant pathogen is available. The technological advances in the past decade have lead to miniaturization of the appropriate synthesis and attachment chemistry; thousands of DNA or RNA molecules can be arrayed on a few square centimetres of a solid phase, e.g. a glass slide (Cheng *et al.* 1998). This allows monitoring of the expression level of genes.

As a demonstration of the feasibility of this technology, RNA was extracted from *Streptococcus pneumoniae* cultures, with or without exposure to the competence-stimulating peptide (CSP), and labelled with biotin. This allowed investigation of differences in the expression of genes involved in the development of natural competence for DNA transformation by probing a microarray of *S. pneumoniae* genes. Whereas most of the studied genes showed no differences in expression, whether or not they had been exposed to CSP, genes from the competence operon were induced (de Saizieu *et al.* 1998). There are now several demonstrations of the feasibility of this approach to monitor and compare gene expression of organisms in different environmental conditions (Kononen *et al.* 1998; DeRisi *et al.* 1997). However, convincing demonstrations of its application to pathogenic bacteria *in vivo* are yet to come, although we would expect this to be achieved to the point that, within a few years, such an approach is as routine a methodology as automated sequencing.

5. THE CONVERGENCE OF METHODS

(a) *Gene expression* in vivo

An innovative and powerful method for identifying virulence factors was described and used prior to the availability of whole-genome sequences. IVET was devised to select positively for bacterial genes that are specifically induced when bacteria infect host tissues (Mahan *et al.* 1993). Many of these *in vivo*-induced genes play key roles in the infectious process. Several versions of the IVET principle have been developed, each offering different capabilities including universal (rescue of *purA* auxotrophy (Mahan *et al.* 1993)), intracellular (use of antibiotic selection (Mahan *et al.* 1995)) or stage-specific (use of genetic recombination (Camilli *et al.* 1994)) gene expression in host tissues. A limitation of IVET is that it will not detect many relevant genes, for example, where there are subtle, but biologically significant, *in vitro* versus *in vivo* differences in gene activity. Not surprisingly, many of the genes discovered by IVET are implicated in *in vivo* metabolic or physiological adaptations of pathogenic bacteria (including pabulins), rather than being classical virulence factors (adhesins, invasins, evasins and toxins). The availability of whole genomes is an obvious bonus in that only a few nucleotides of any gene selected by IVET are needed to access its entire coding and flanking sequences. The principles of IVET can also be used in a 'top-down' approach in which the genome sequence can be used to select all or subsets of genes and submit them to microarrays, as one component of a multifaceted high-throughput screening approach.

(b) *STM and complementation of gene libraries*

In STM, transposons carrying unique DNA sequence tags are used to produce a bank of mutants (Hensel *et al.* 1995). The tags from a mixed population of bacterial mutants representing the inoculum (input) and bacteria recovered from infected hosts (output) are detected by amplification, radiolabelling and hybridization analysis. This allows detection of *in vivo* bacterial virulence genes by negative selection; those mutants present in the input pool that cannot be recovered from the output pool are presumably deficient in some function at one or more stages in the infectious process. STM can be adapted to screen site-directed mutants of complete sets, or subsets, of genes based on information from whole-genome sequences. A considerable advantage of STM is that large numbers of mutants can be screened in pools of approximately 100 per animal.

In the absence of well-defined systems for genetic transfer, complementation analysis *in vivo* provides a means for identifying genes associated with virulence. Integrating shuttle cosmid libraries that replicate as multicopy plasmids in *E. coli* and integrate into the mycobacterial chromosome was used to identify key virulence factors using a pair of genetically related *M. tuberculosis* strains, one fully virulent, the other attenuated (Pascopella 1994).

Despite the power of information from genomes and its integration with high-throughput screening of mutants *in vivo*, it is necessary to characterize the role of each of these putative virulence factors using tissue or organ cultures, whole-animal assays, cell biology (including immunology), microscopy and structural biology.

(c) *Immune responses as an indicator of* in vivo *function*

A limiting factor in many functional genomic strategies is the sheer volume of manipulation required to screen for virulence genes and also direct these studies so as to identify genes of *in vivo* relevance. Using host immune responses to indicate genes expressed *in vivo* is both powerful and specific. There are numerous examples of using bacteria obtained from patients or infected animals followed by analysis using sodium dodecyl sulphate PAGE and immunoblotting with convalescent sera (multiple references are cited in Smith (1996)). However, an innovative recent approach is to capitalize upon the surprising finding that amplified PCR fragments can be rendered transcriptionally active when injected into animals, so-called DNA or nucleic-acid immunization (Hoffman *et al.* 1995). For example, given the 4000 genes revealed by the whole-genome sequence of *M. tuberculosis*, each gene could be individually amplified from the genome by PCR. Each PCR product would be annealed to fragments encoding promoter and terminator sequences and pools (approximately 50) of these linear expression elements inoculated into mice (Sykes & Johnston 1999). With minor modifications, we suggest that this approach can be used to carry out a systematic and comprehensive analysis in which the antibodies obtained to *in vivo*-expressed proteins are used to detect the microbial products expressed during infection. Such analysis could use fluorescence for *in situ* work or, more simply, an enzyme-linked immunosorbent assay on *ex vivo* material. In the latter, plates could be coated with extracts of different infected tissues, normal tissues

serving as controls, or the same tissues at different stages of infection to obtain a profile of *in vivo* expression during pathogenesis. This approach would not of course be useful for identifying macromolecules such as capsule or LPS and would suffer from the potential loss of epitopes critical to antibody recognition through *in vivo* degradation or changes in conformation.

A different strategy for selecting bacterial genes expressed only *in vivo* which can take advantage of the availability of whole-genome sequence data is the use of expression libraries screened with two types of antibodies. This approach has some advantages for studying pathogens that lack efficient genetic transfer systems. Suk *et al.* (1995) used this approach to identify genes in *Borrelia burgdorferi*. Reactivities of antibodies obtained from animals immunized with killed organisms and from infected hosts were compared. The genomic clones that reacted with antibody from infected animals, but not from immunized animals, indicated genes expressed only in the host.

(d) *Fluorescence-activated cell sorter (FACS) to select bacteria bearing transcriptionally active* gfp *fusions in vivo*

Valdivia & Falkow (1996) have devised an elegant strategy that lends itself perfectly to a convergent approach using genomics and high-throughput screening, for example using FACS. The essence of this methodology is that, under particular conditions *in vivo*, transcriptionally active genes fused to *gfp* will fluoresce. FACS can then be used to identify either fluorescent bacteria or host cells containing fluorescent bacteria. The only genetic requirements are that the bacterial pathogen is able to maintain an episomal element and express the functional *gfp*. Although, as initially described, this system used random DNA fragments inserted upstream of a promoterless *gfp* gene, the system could be used for a site-specific strategy using a comprehensive approach based on whole-genome sequences. The fluorescence intensity of individual bacteria grown in tissue culture can be compared with the same bacterial clone after release from infected cells. This allows selection of bacteria that are transcriptionally active only within the host cell (for example, a macrophage (Valdivia & Falkow 1997)). In the original study, 14 insertion mutants with intracellular-dependent activities were identified. This study was done prior to the availability of the relevant whole-genome sequence of *S. typhimurium*, so the investigators were obliged to characterize these genes the hard way, i.e. by cloning, DNA sequencing and further characterization. However, with the availability of the whole-genome sequences of a pathogen under study, a high-throughput, efficient and comprehensive approach should be feasible since, once the promoter is identified, the entire gene to which it corresponds can be easily isolated. Additionally, it will be possible to identify those promoter fusions that are unstable. An advantage of using fluorescence detection is that there is no dependence upon growth. Many pathogens enter viable, but non-culturable states, e.g. *Vibrio cholerae*. This is therefore a powerful strategy for studying environmentally triggered gene induction in bacteria obtained from infected animals or host cells that accurately reflect conditions of natural infections. As has been stressed, the activity of virulence genes is sensitive to precise combinations of environmental conditions. It is therefore useful to use such strategies to track differences in gene expression at different times and at different stages of pathogenesis.

(e) *Photonic detection of bacterial pathogens in living hosts*

Real-time, non-invasive analyses of pathogenic events can now be performed *in vivo* through marking bacterial pathogens that have been transfected with a plasmid conferring constitutive expression of bacterial luciferase. The transmission of light through opaque tissues *in vivo* has been used for non-invasive optical imaging and spectroscopy in both animals and humans. Further, localization of photons from bioluminescent molecules has been used quantitatively to monitor gene expression in plants (Millar *et al.* 1995) and promoter activity in mammalian cells (Hooper *et al.* 1990). These observations suggested that an optical method for evaluating disease progression is possible. This was elegantly demonstrated by Contag *et al.* (1995) who showed that bioluminescent *Salmonella* could be observed adhering to and entering epithelial cells and macrophages. The detection of photons transmitted through tissues of animals infected with bioluminescent *Salmonella* allowed localization of the bacteria to specific tissues so that progressive infections were distinguished from those that were persistent or abortive. Through this technique, the patterns of bioluminescence suggest that the caecum may play a pivotal role in the pathogenesis of *Salmonella* infection.

(f) *Confocal microscopy*

The traditional tools for examining infectious processes *in vivo* include dose–responses of mortality (LD_{50}), or some appropriate end-point such as bacteraemia, and infection kinetics of organs, such as the liver, lungs or kidneys. These studies, combined with observations from tissue culture experiments do not allow detailed localization of host–pathogen interactions *in vivo*. An excellent example of how more recent techniques can overcome these deficiencies is afforded by studies on the intracellular pathogen *S. typhimurium* (Richter-Dahlfors *et al.* 1997). To pinpoint which particular cell type harbours *S. typhimurium*, confocal laser scanning electron microscopy and computerized image analysis techniques were used to determine bacteria–cell associations at an early stage in the infection of mouse liver cells. This study was able to derive data using relatively small inocula (approximately 100 organisms), a much more convincing scenario than the artificially high challenge doses used in preceding studies. The key to the sensitivity of these experiments was the use of thick sections and the powerful resolution afforded by confocal microscopy and computer-assisted imaging. Using antibodies directed against different host cell types, it was possible to identify the intracellular location of the bacteria and to determine the specific cell type involved.

(g) *Molecular approaches to histopathology*

For studies of bacteria *in vivo*, advances in the application of molecular histopathological techniques to clinical and experimental specimens offer huge potential (Quirke

& Mapstone 1999). There are already valuable techniques for the detection of infectious agents in paraffin-embedded tissue, including non-culturable bacteria. Newer thermal cyclers combined with Taqman^{TM} fluorescent dyes make amplification and detection of bacterial DNA possible in a few minutes and include quantitative PCR (Belgrader *et al.* 1999). Among the leading-edge technologies, atomic-force microscopy is a type of physical microscopy that allows the shape of a structure under analysis to be visualized down to the atom. It can be used on routine paraffin sections with or without immunochemistry, its power being derived from its ability to image DNA and individual proteins. This opens up the possibility of studying normal and abnormal protein–DNA interactions in real time under physiological conditions. This ability to investigate the surface morphology, at an unprecedented level of resolution, has been used already to study the surface of *E. coli* exposed to antibiotics, but there would seem to be great potential for examining the effects on bacterial morphology of host factors (Braga & Ricci 1998). Through digital storage of information, the image of bacteria can be rotated allowing cross-sections of the bacterium obtained at any point. The initial steps in bacterial adhesion with host cells can be studied to gauge the precise manner in which ligand–receptor, van der Waals and electrostatic interactions are affected by cell surface components (Razatos *et al.* 1998). Interactions between the cantilever tip and confluent monolayers of isogenic bacteria will allow appreciation of subtle effects of cell surface macromolecules, such as LPS and capsular polysaccharide, on host tissues.

6. SUMMARY

The main points are as follows: pathogenicity is a complex entity and its definition is elusive. The strong reductionist culture of today's science, while having great strengths, may obscure the goal of integrating molecular detail into a solid framework of *in vivo* relevance to pathogenesis. For the student of bacterial infections, a critical and convergent approach is needed. Whole-genome sequences of major bacterial pathogens are, or will be, available and can provide an invaluable source of information, and as well as being one of the ways of initiating relevant hypotheses about *in vivo* function. It is crucial that the information from genomes is accessible in 'user friendly' form to provide the basis of a variety of data-mining algorithms and the powerful analyses possible through comparative genomics. A variety of high-throughput screening methods, such as nucleic-acid microarrays and proteomics, are available to make bridges from bacterial genomes to the experimental biology of pathogenic and commensal behaviour. But the crucial studies of *in vivo* function will require not only the relevant biological assays but also a convergence of clinical observations, carefully amassed strain collections and naturally or experimentally infected host tissues. These resources must be integrated with the judicious application of genetics, cell biology, fine structural analyses and high-resolution microscopy. Expertise in physiology, intermediary metabolism and biochemistry has been relatively neglected, but must be resurrected to meet the challenges of the post-genomic era. The task is challenging and may be best met through consortium approaches since many individual research groups are likely to be overwhelmed by the amount of data and resources that emanate from genome projects, as well as the contingent demand for eclectic expertise. Nonetheless, the current era offers unprecedented challenges and excitement to spur our creativity as we try to understand the depths and details of bacterial pathogenicity.

We thank our colleagues in the Molecular Infectious Diseases Group, Department of Paediatrics, for their many useful discussions, Paul Rainey (Department of Plant Sciences, University of Oxford) and Jay Hinton (now at the Institute of Food Research, Norwich) for their insightful comments.

REFERENCES

Anderson, R. M. & May, R. M. (eds) 1982 *Population biology of infectious diseases. Report of the Dahlem workshop on population biology of infectious disease agents*. Berlin: Springer.

Anderson, B. M., Kahn, D. W. & Anderson, C. D. 1985 Studies of the 2′:3′-cyclic nucleotide phosphodiesterase of *Haemophilus influenzae*. *J. Gen. Microbiol.* **131**, 2041–2045.

Bajaj, V., Lucas, R. L., Hwang, C. & Lee, C. A. 1996 Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression. *Mol. Microbiol.* **22**, 703–714.

Balaban, N. (and 10 others) 1998 Autoinducer of virulence as a target for vaccine and therapy against *Staphylococcus aureus*. *Science* **280**, 438–440.

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane., S. & Small, P. M. 1999 Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523.

Belgrader, P., Bennett, W., Hadley, D., Richards, J., Stratton, P., Mariella Jr, R. & Milanovich, F. 1999 PCR detection of bacteria in seven minutes. *Science* **284**, 449–450.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. & Yuan, Y. 1998 Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**, 707–725.

Braga, P. E. & Ricci, D. 1998 Atomic force microscopy: application to investigation of *Escherichia coli* morphology before and after exposure to cefodizime. *Antimicrob. Agents Chemother.* **42**, 18–22.

Brock, T. D. 1971 Microbial growth rates in nature. *Bacteriol. Rev.* **35**, 39–58.

Camilli, A., Beattie, D. T. & Mekalanos, J. J. 1994 Use of genetic recombinations as a reporter of gene expression. *Proc. Natl Acad. Sci. USA* **91**, 2634–2638.

Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. 1996 *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl Acad. Sci. USA* **93**, 14 648–14 653.

Cheng, J., Sheldon, E. L., Wu, L., Uribe, A., Gerrue, L. O., Carrino, J., Heller, M. J. & O'Connell, J. P. 1998 Preparation and hybridization analysis of DNA/RNA from *E. coli* on microfabricated bioelectronic chips. *Nat. Biotech.* **116**, 541–546.

Cohen, M. S., Cannon, J. G., Jerse, A. E., Charniga, L. M., Isbey, S. F. & Whicker, L. G. 1994 Human experimentation with *Neisseria gonorrhoeae*: rationale, method and implications for the biology of infection and vaccine development. *J. Infect. Dis.* **169**, 532–537.

Contag, C. H., Contag, P. R., Mullins, J. I., Spilman, S. D., Stevenson, D. K. & Benaron, D. A. 1995 Photonic detection of bacterial pathogens in living hosts. *Mol. Microbiol.* **18**, 593–603.

DeRisi, J. L., Iyer, V. R. & Brown, P. O. 1997 Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.

de Saizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W. & Mous, J. 1998 Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat. Biotech.* **16**, 45–48.

Ewald, P. 1994 *The evolution of infectious disease*. New York: Oxford University Press.

Falkow, S. 1988 Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10**, S274–S276.

Falkow, S. 1996 The evolution of pathogenicity in *Escherichia*, *Shigella*, and *Salmonella*. In Escherichia coli *and* Salmonella *cellular and molecular biology*, 2nd edn (ed. F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter & H. E. Umbarger), pp. 2723–2728. Washington, DC: American Society for Microbiology Press.

Falkow, S. 1997 What is a pathogen? *Am. Soc. Microbiol. News* **63**, 359–365.

Field, D., Hood, D. & Moxon, R. 1999 Contribution of genomics to bacterial pathogenesis. *Curr. Opin. Genet. Dev.* **9**, 700–703.

Fields, S., Kohara, Y. & Lockhart, D. J. 1999 Functional genomics. *Proc. Natl Acad. Sci. USA* **96**, 8825–8826.

Finlay, B. B. & Falkow, S. 1997 Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**, 136–169.

Geiser, J. R., Van Tuinen, D., Brockerhoff, S. E., Neff, M. M. & Davis, T. N. 1991 Can calmodulin function without binding calcium? *Cell* **65**, 949–959.

He, Q., Viljanen, M. K., Arvilommi, H., Aittanen, B. & Mertsola, J. 1998 Whooping cough caused by *Bordetella pertussis* and *Bordetella parapertussis* in an immunized population. *J. Am. Med. Assoc.* **280**, 635–637.

Heithoff, D. M., Conner, C. P., Hanna, P. C., Julio, S. M., Hentschel, U. & Mahan, M. J. 1997 Bacterial infection as assessed by *in vivo* gene expression. *Proc. Natl Acad. Sci. USA* **94**, 934–939.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. 1995 Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403.

Hoffman, S. L., Doolan, D. L., Sedagah, M., Gramzinski, R., Wang, H., Gowda, K., Hobart, P., Margalith, M., Norman, J. & Hedstrom, R. C. 1995 Nucleic acid malaria vaccines. Current status and potential. *Ann. NY Acad. Sci.* **772**, 88–94.

Hood, D. W., Deadman, M. E., Allen, T., Martin, A., Brisson, J. R., Fleischmann, R., Venter, J. C., Richards, J. C. & Moxon, E. R. 1996 Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis. *Mol. Microbiol.* **22**, 951–965.

Hooper, E. C., Ansorge, R. E., Browne, H. M. & Tomkins, P. 1990 CCD imaging of luciferase gene expression in single mammalian cells. *J. Biolum. Chemilum.* **5**, 123–130.

Hoppe, J. E. 1999 Update on respiratory infection caused by *Bordetella parapertussis*. *Pediatr. Infect. Dis. J.* **18**, 375–381.

Huynen, M. A., Diaz-Lazcoz, Y. & Bork, P. 1997 Differential genome display. *Trends Genet.* **13**, 389–390.

Ji, G., Beavis, R. & Novick, R. P. 1997 Bacterial interference caused by autoinducing peptide variants. *Science* **276**, 2027–2030.

Kenny, B., Abe, A., Stein, M. & Finlay, B. B. 1997 Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect. Immun.* **65**, 2606–2612.

Kononen, J., Bubendorf, L., Kallioniemi, A., Bärlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G. & Kallioniemi, O.-P. 1998 Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **4**, 844–847.

Levin, B. R. & Svanborg-Eden, C. 1990 Selection and evolution of virulence in bacteria: an ecumenical excursion and modest suggestion. *Parasitology* **100**, S103–S115.

Levinson, G. & Gutman, G. A. 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221.

Lewis, D. A., Klesney-Tait, J., Lumbley, S. R., Ward, C. K., Latimer, J. L., Ison, C. A. & Hansen, E. J. 1999 Identification of the *znuA*-encoded periplasmic zinc transport protein of *Haemophilus ducreyi*. *Infect. Immun.* **67**, 5060–5068.

Lipsitch, M. & Moxon, E. R. 1997 Virulence and transmissibility of pathogens: what is the relationship? *Trends Microbiol.* **31**, 31–37.

Lockhart, D. J. (and 10 others) 1996 Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotech.* **14**, 1675–1680.

Loeb, M. R. & Smith, D. H. 1980 Outer membrane protein composition in disease isolates of *Haemophilus influenzae*: pathogenic and epidemiological implications. *Infect. Immun.* **30**, 709–717.

McKenney, D., Pouliot K. L., Wang, Y., Murthy, V., Ulrich, M., Döring, G., Lee, J. C., Goldmann, D. A. & Pier, G. B. 1999 Broadly protective vaccine for *Staphylococcus aureus* based on an *in vivo*-expressed antigen. *Science* **284**, 1523–1527.

MacLaren, D. M. 1970 The influence of urea on the growth of *Proteus mirabilis*. *Guys Hosp. Rep.* **119**, 133–143.

Mahan, M. J., Slauch, J. M., Hanna, P. C., Camilli, A., Tobias, J. W., Waldor, M. K. & Mekalanos, J. J. 1993 Selection for bacterial genes that are specifically induced in host tissues: the hunt for virulence factors. *Infect. Agents Dis.* **2**, 263–268.

Mahan, M. J., Tobias, J. W., Slauch, J. M., Hanna, P. C., Collier, R. J. & Mekalanos, J. J. 1995 Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. *Proc. Natl Acad. Sci. USA* **92**, 669–673.

Martindale, J., Stroud, D., Moxon E. R. & Tang, C. M. 2000 Genetic analysis of *Escherichia coli* K1 gastrointestinal colonization. *Mol. Microbiol.* (Submitted.)

Maw, J. & Meynell, G. G. 1968 The true division and death rates of *Salmonella typhimurium* in the mouse spleen determined with superinfecting phage P22. *Br. J. Exp. Pathol.* **49**, 597–613.

Mayville, P., Ji, G., Beavis, R., Yang, H., Goger, M., Novick, R. P. & Muir, T. M. 1999 Structure–activity analysis of synthetic autoinducing thiolactone peptides from *Staphylococcus aureus* responsible for virulence. *Proc. Natl Acad. Sci. USA* **96**, 1218–1223.

Meynell, G. G. & Stocker, B. A. D. 1957 Some hypotheses on the aetiology of fatal infections in partially resistant hosts and their application to mice challenged with *Salmonella paratyphi*-B or *Salmonella typhimurium* by intraperitoneal injection. *J. Gen. Microbiol.* **16**, 38–58.

Millar, A. J., Carré, I. A., Straer, C. A., Chua, H.-H. & Kay, S. A. 1995 Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163.

Miller, J. F., Mekalanos, J. J. & Falkow, S. 1989 Coordinate regulation and sensory transduction in the control of bacterial virulence. *Science* **243**, 916–922.

Moxon, E. R. 1992 Molecular basis on invasive *Haemophilus influenzae* type b disease. *J. Infect. Dis.* **165**, S77–S81.

Moxon, E. R. 1997 Applications of molecular microbiology to vaccinology. *Lancet* **350**, 1240–1244.

Moxon, E. R. & Anderson, P. 1979 Meningitis caused by *Haemophilus influenzae* in infant rats: protective immunity and antibody priming by gastro-intestinal colonization with *Escherichia coli*. *J. Infect. Dis.* **140**, 471–478.

Moxon, E. R. & Kroll, J. S. 1990 The role of bacterial polysaccharide capsules as virulence factors. In *Bacterial capsules* (ed. K. Jann & B. Jann), pp. 65–85. Berlin: Springer.

Moxon, E. R. & Murphy, P. A. 1978 *Haemophilus influenzae* bacteraemia and meningitis resulting from survival of a single organism. *Proc. Natl Acad. Sci. USA* **75**, 1534–1536.

Moxon, E. R. & Ostrow, P. T. 1977 *Haemophilus influenzae* meningitis in infant rats: role of bacteremia in pathogenesis of age-dependent inflammatory responses in cerebrospinal fluid. *J. Infect. Dis.* **135**, 303–307.

Moxon, E. R. & Schwartz, A. D. 1980 Heterotopic splenic autotransplantation in the prevention of *Haemophilus influenzae* meningitis and fatal sepsis in Sprague–Dawley rats. *Blood* **56**, 842–845.

Moxon, E. R. & Wills, C. 1999 DNA microsatellites: agents of evolution? *Sci. Am.* **280**, 72–77.

Moxon, E. R., Smith, A. L. & Averill, D. 1979 Brain carbohydrate metabolism during experimental *Haemophilus influenzae* meningitis. *Pediatr. Res.* **13**, 52–59.

Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. 1994 Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33.

Myerowitz, R. L. 1981 Mechanism of potentiation of experimental *Haemophilus influenzae* type B disease in infant rats by influenza A virus. *Lab. Invest.* **44**, 434–441.

Ostrow, P. T., Moxon, E. R., Vernon, N. & Kapko, R. 1979 Pathogenesis of bacterial meningitis: studies on the route of meningeal invasion following *Haemophilus influenzae* inoculation of infant rats. *Lab. Invest.* **40**, 678–685.

Pallen, M. J. 1999 Microbial genomes. *Mol. Microbiol.* **32**, 907–912.

Pascopella, L., Collins, F. M., Martin, J. M., Lee, M. H., Hatfull G. F., Stover, C. K., Bloom, B. R. & Jacobs, W. R. 1994 Use of *in vivo* complementation in *Mycobacterium tuberculosis* to identify a genomic fragment associated with virulence. *Infect. Immun.* **62**, 1313–1319.

Peitu, G., Alibert, O., Guichard, V., Lamy, B., Bios, F., Leroy, E., Mariage-Samxon, R., Houlgatte, R., Soularue, P. & Auffray, C. 1996 Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* **6**, 492–503.

Pettersson, J., Nordfelth, R., Dubinina, E., Bergman, T., Gustafsson, M., Magnusson, K. E. & Wolf-Watz, H. 1996 Modulation of virulence factor expression by pathogen target cell contact. *Science* **273**, 1231–1233.

Plaut, A. G., Gilbert, J. V. & Wistar, R. 1977 Loss of antibody activity in human immunoglobulin A exposed extracellular immunoglobulin A proteases of *Neisseria gonorrhoeae* and *Streptococcus sanguis*. *Infect. Immun.* **17**, 130–135.

Quirke, P. & Mapstone, N. 1999 The new biology: histopathology. *Lancet* **354**, SI26–SI31.

Razatos, A., Ong, Y.-L., Sharma, M. M. & Georgiou, G. 1998 Molecular determinants of bacterial adhesion monitored by atomic force microscopy. *Proc. Natl Acad. Sci. USA* **95**, 11 059–11 064.

Read, A. F. (and 10 others) 1998 What can evolutionary biology contribute to understanding virulence. In *Evolution in health and disease* (ed. S. Stearns), pp. 205–215. Oxford University Press.

Richter-Dahlfors, A., Buchan, A. M. J. & Finlay, B. B. 1997 Murine salmonellosis studied by confocal microscopy: *Salmonella typhimurium* resides intracellularly inside macrophages and exerts a cytotoxic effect on phagocytes *in vivo*. *J. Exp. Med.* **186**, 569–580.

Roche, R. J. & Moxon, E. R. 1995 Phenotypic variation in *Haemophilus influenzae*: the interrelationship of colony opacity, capsule and lipopolysaccharide. *Microb. Pathogen.* **18**, 129–140.

Rothbaun, R., McAdams A., Gianella, R. & Partin, J. 1982 A clinico-pathogenic study of enterocyte-adherent *Escherichia coli*: a cause of protracted diarrhea in infants. *Gastroenterology* **83**, 441–454.

Rothbaun, R., Partin, J., Sallfield, K. & McAdams, A. 1983 An ultrastructural study of enteropathogenic *Escherichia coli* infection in human infants. *Ultrastruct. Pathol.* **4**, 291–304.

Rubin, L. G. & Moxon, E. R. 1985 The effect of serum-factor induced resistance to somatic antibodies on the virulence of *Haemophilus influenzae* type b. *J. Gen. Microbiol.* **131**, 515–520.

Rubin, L. G., Zwahlen, A. & Moxon, E. R. 1985 Role of intravascular replication in the pathogenesis of experimental bacteraemia due to *Haemophilus influenzae* type b. *J. Infect. Dis.* **152**, 307–314.

Seiler, A., Reinhardt, R., Sarkari, J., Caugant, D. A. & Achtman, M. 1996 Allelic polymorphism and site-specific recombination in the *opc* locus of *Neisseria meningitidis*. *Mol. Microbiol.* **19**, 841–856.

Shaw, S., Smith, A., Anderson, P. & Smith, D. H. 1976 The paradox of *Haemophilus influenzae* type b bacteremia in the presence of serum bactericidal activity. *J. Clin. Invest.* **58**, 1019–1029.

Smith, H. 1972 The little-known determinants of microbial pathogenicity. In *Microbial pathogenicity in man and animals. Twenty-second symposium of the Society for General Microbiology, Imperial College, April 1972*, pp. 1–17. Cambridge University Press.

Smith, H. 1996 What happens *in vivo* to bacterial pathogens? *Ann. NY Acad. Sci.* **797**, 77–92.

Smith, H., Williams, A. E., Pearce, J. H., Keppie, J., Harris-Smith, P. W., Fitzgeorge, R. B. & Witt, K. 1962 Foetal erythritol: a cause of the localisation of *Brucella abortus* in bovine contagious abortion. *Nature* **193**, 47–49.

Stern, A., Brown, M., Nickel, P. & Meyer, T. F. 1986 Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**, 61–71.

Strauss, E. J. & Falkow, S. 1997 Microbial pathogenesis: genomics and beyond. *Science* **276**, 707–712.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. 1966 Frameshift mutations and the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 77–84.

Suk, K., Das, S., Sun, W., Jwang, B., Barthold, S. W., Flavell, R. A. & Fikrig, E. 1995 *Borrelia burgdorferi* genes selectively expressed in the infected host. *Proc. Natl Acad. Sci. USA* **92**, 4269–4273.

Sykes, K. F. & Johnston, S. A. 1999 Linear expression elements: a rapid, *in vivo*, method to screen for gene functions. *Nat. Biotech.* **17**, 355–359.

Topley, W. W. C. 1919 The spread of bacterial infection. *Lancet* (July 5) **ii**, 1–5.

Valdivia, R. H. & Falkow, S. 1996 Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. *Mol. Microbiol.* **22**, 367–378.

Valdivia, R. H. & Falkow, S. 1997 Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* **277**, 2007–2011.

Vasmatzis, G., Essand, M., Brinkmann, U., Lee, B. & Pastan, I. 1998 Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proc. Natl Acad. Sci. USA* **95**, 300–304.

Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. 1995 Serial analysis of gene expression. *Science* **270**, 484–487.

Virji, M., Weiser, J. N., Lindberg, A. A. & Moxon, E. R. 1990 Antigenic similarities in lipopolysaccharides of *Haemophilus* and *Neisseria* and expression of digalactoside structure also present on human cells. *Microb. Pathogen.* **89**, 441–450.

Wachtel, M. R. & Miller, V. L. 1995 *In vitro* and *in vivo* characterization of an *ail* mutant of *Yersinia enterocolitica*. *Infect. Immun.* **63**, 2541–2548.

Weiser, J. N. & Pan, N. 1998 Adaptation of *Haemophilus influenzae* to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. *Mol. Microbiol.* **30**, 767–775.

Weiser, J. N., Love, J. M. & Moxon, E. R. 1989 The molecular mechanism of phase-variation of *Haemophilus influenzae* lipopolysaccharide. *Cell* **59**, 657–665.

Weiser, J. N., Shchepetov, M. & Chong, S. T. 1997 Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. *Infect. Immun.* **65**, 943–950.

Weller, P. F., Smith, A. L., Smith, D. H. & Anderson, P. 1978 Role of immunity in the clearance of bacteremia due to *Haemophilus influenzae*. *J. Infect. Dis.* **138**, 427–436.

Wells, D., Sherlock, J. K., Handyside, A. H. & Delhanty, J. D. 1999 Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation. *Nucl. Acids Res.* **27**, 1214–1218.

Winson, M. K (and 12 others) 1995 Multiple *N*-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **92**, 9427–9431.

Winzeler, E. A. (and 24 others) 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906.

Zwahlen, A., Rubin, L. G., Connelly, C. J., Inzana, T. J. & Moxon, E. R. 1985 Alteration of the cell wall of *Haemophilus influenzae* type b by transformation with cloned DNA: association with attenuated virulence. *J. Infect. Dis.* **152**, 485–492.

Zwahlen, A., Rubin, L. G. & Moxon, E. R. 1986 Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. *Microb. Pathogen.* **1**, 465–473.

Zwahlen, A., Kroll, J. S., Rubin, L. G. & Moxon, E. R. 1989 The molecular basis of pathogenicity in *Haemophilus influenzae*: comparative virulence of genetically-related capsular transformants and correlation with changes at the capsulation locus *cap*. *Microb. Pathogen.* **7**, 225–235.