SECTION 7

PRECLINICAL AND CLINICAL EVALUATION

‘Let’s arrive the hope of any color: green, red or black, but with love
Let’s Arrive the Hope
Silvio Rodriguez

‘I have new and old blood that is not hoping for more, which makes itself a single gigantic wave, a suspended wave that opens. Do you see? I have it; it’s here. Didn’t you kow? I didn’t know, and it was my blood!’
Revelation
Dulce Maria Loinaz

‘Windows’ from the series Windows
Pedro Pablo Oliva
Oil on canvas; 145 × 120 cm
CHAPTER 7.1

PRECLINICAL EVALUATION OF TB VACCINES

Ann Williams, Helen McShane

Human TB is considerably more complex than the standard animal models used for preclinical evaluation of TB vaccine candidates.

‘It is midday, everything is very dark and suddenly red from time to time.’

At the Corner of a Street
Jacques Prevert

Untitled
Adam Rabinowitz
Liquid acrylic, ink and oil on paper; 95 × 60 cm
INTRODUCTION

Preclinical studies to evaluate the safety, immunogenicity, and protective efficacy of candidate TB vaccines in animal models form an essential step in the development pathway for these vaccines. Formal safety and toxicology studies, performed under Good Laboratory Practice (GLP) conditions and using clinical grade material, are necessary before any new vaccine can be evaluated in early clinical trials. In addition, detailed immunological and protective efficacy studies are essential in determining which candidate vaccines should go forward into human testing. There are a variety of different species of animals used for such models, and each species has both advantages and disadvantages. In addition, there are increasingly complex animal models being developed as an attempt to more effectively model the human situation. These more complex models include pre-exposure to environmental (non-tuberculous) mycobacteria and helminths as well as a variety of models simulating both latent MTB infection and therapeutic vaccination (discussed in Chapter 5.11). There are many factors for consideration within all of these models which affect how well the model represents human disease and also how reproducible it is. The various different species, models, and the factors for consideration, will be reviewed in this chapter.

DIFFERENT ANIMAL MODELS

There are a wide range of animal species that can be experimentally infected with mycobacteria and many of these have been employed as animal models which mimic natural infection. Because of the wide host range of *M. bovis*, there are several animal species (cows, deer, possums, badgers) that are natural hosts and therefore make highly relevant experimental animal models of infection caused by *M. bovis* (1). Similarly, there are natural hosts of other mycobacterial species (armadillo — *M. leprae* (2), zebra fish — *M. marinum* (3)) and it has been argued that these animal models can be used to study mycobacterial infection in general and extrapolations can be made to human infection with MTB. This may be most relevant to hosts of *M. bovis* because of the close similarity of *M. bovis* to MTB (4).
However, researchers studying human TB have in general favoured animal species that can be successfully infected with MTB and which go on to develop disease that, at least partially, reproduces that seen in humans. The difficulty with this approach is that there is no single animal model that completely replicates human TB. The closest species are non-human primates, because they are natural hosts of MTB, but no experimental model will ever replicate the complexity of human behavioural and environmental factors that can influence the outcome of infection or therapeutic intervention. Despite this, studies in animal models of TB have been instrumental in advancing the field towards understanding pathogenic mechanisms of mycobacterial infection and in progressing novel vaccine development. There are several excellent reviews of the merits of various animal models of mycobacteria which should be studied to supplement the following summary of each of the commonly used species (5–9).

**Mouse**

A wide range of mouse strains have been used to study MTB infection and disease and, particularly, the immune response to mycobacteria, but there is no single mouse model for TB. There are two strains commonly used to determine the immune response and protective efficacy of new TB vaccines; these are C57BL6 and BALB/c mice. These mouse strains are inbred and have different haplotypes; the choice of strain is usually dictated by the vaccines being studied and the MHC-restriction of the antigens. Inbred strains have the advantage of being more reproducible in their response to either vaccination or challenge. Genetic manipulation of in-bred mice is widely used to study disease mechanisms and several strains of mice have proven to be useful for TB vaccine development. Gene knockout mice that lack key features of the cellular immune response have become a recognized means of demonstrating the safety of live attenuated vaccines in immunocompromised hosts. Severe Compromised Immuno Deficient (SCID) mice are the most commonly used and safety is measured in terms of survival of the mice following high dose, infection with the vaccine strain in comparison to BCG (10). Other genetically manipulated, or cross-breed strains of mice which ‘reproduce’ features of the host response in humans (e.g. Kramnik mice, (11)) have relevance to vaccine development but have not yet been incorporated into routine screening of vaccine efficacy. This may change as strains become more widely available.

Mice are generally considered to be more resistant to TB infection and in that respect may be viewed as more human-like (12). However, there are certain features of the infectious process in resistant strains of mice that are quite distinct from what
is seen in human disease. These include differences in granuloma structure and
development (mice do not develop organized granulomas with central necrosis and
caseation), and extremely high numbers of bacilli that are maintained in the lungs
and spleens of animals over extended periods of time (13). Mice also lack features
of the human immune system that are known to be important in the host response
to MTB and these differences are related to the different pathophysiological
aspects of TB disease in mice. Mice are often used as models in which to evaluate
the protective effects of vaccination, usually because of their advantages in terms
of cost and space. However, they are usually regarded as only being useful for
screening candidates in the early stages of development. Considering the large
numbers of putative vaccine candidates that have required screening in the past
and will require screening in the future, it is entirely appropriate to begin the
screening process in this species (14, 15). There are, however, examples of antigens
that appeared promising in mouse studies but which did not induce sufficiently
protective immune responses in more stringent animal models such as the guinea
pig and non-human primate (16, 17). Therefore, initial screening in mice should
ideally be followed by evaluation in at least one other animal model in order to
provide assurance that the vaccine candidate offers protective efficacy prior to
evaluation in humans.

Guinea pig
The guinea pig model of TB is the forerunner of all the experimental models, being
employed at the very beginning of research into human TB when Robert Koch first
described the tubercle bacillus and then went on to test possible therapies. The
most commonly used guinea pig strain is the outbred ‘Dunkin-Hartley’ or ‘Hartley’
strain. This animal is noted for being highly susceptible to extremely low doses of
MTB, particularly when delivered by the aerosol route and it is often suggested
that this model is too susceptible to be able to mimic human TB (18). It is wrong,
however, to assume that guinea pigs are overwhelmed by infection with MTB; the
acquired immune response that develops following low-dose aerosol challenge
will contain progression of disease for several months — a feature that can easily
confound vaccine evaluation studies (see relevant section below). There are indeed
several features of the disease in humans that are replicated in guinea pigs and it
has been proposed that aerosol challenge of guinea pigs models key features of TB
in children (19, 20).

The guinea pig model has an important role in the vaccine development process,
being a more stringent means of determining the potential of a novel vaccine to
prevent the development of severe disease than the mouse. Due to the similarities to humans described above, data on vaccine efficacy is generally considered to be more relevant from the guinea pig than that from other small animal models (21).

An often quoted disadvantage of the guinea pig is the lack of immunological reagents with which to study immune responses. This is undoubtedly a drawback when attempting to define immune correlates of vaccine protection or TB disease. Although the field is responding to the need to be able to measure at least some of the key immune parameters (22–24), it will be several years before the detailed immune response to TB in the guinea pig will be unravelled. Advances in immunology will continue to emerge from studies in humans, mice, and non-human primates where the vast array of existing reagents will be supplemented by the steady development of new reagents. The development of guinea pig-specific reagents will tend to follow this field and those using the guinea pig model could benefit from a more targeted approach to understanding the immunological basis of vaccine-induced protection. The key strength of the guinea pig model will continue to be its use as a ‘standard’ model for vaccine evaluation studies; the reproducibility of the protective effect of BCG vaccination allows for reliable evaluations of the efficacy of novel candidates relative to BCG.

Rabbit

The value of the rabbit model to inform the preclinical progression of novel vaccine candidates is unclear. Historically, the use of susceptible and resistant strains of rabbits enabled the characterization of key events in the pathogenesis of mycobacterial infection, such as the definition of the cellular mechanisms involved in granuloma formation (25), and the model was shown to reproduce important manifestations of human pulmonary disease such as the formation of cavities (6). The susceptible and resistant strains of rabbit are no longer available but a spectrum of disease has been demonstrated in rabbits by infection with mycobacterial strains of different virulence (26, 27). In particular, the characteristics of disease progression in New Zealand White rabbits following aerosol infection with strain HN878 have been extensively described (28). These characteristics include the development of cavities and the use of the model to follow the response to antibiotic treatment by PET-CT (29) imaging has been elegantly presented. Rabbits have been used to study detailed immune responses (28) and would be a suitable and relevant model to measure the impact of vaccination on pulmonary disease progression but rabbits are not widely used, probably due to their increased size and requirement for more advanced biocontainment systems compared to guinea pigs and mice.
Non-human primate

Non-human primates are the elite of the animal models of human TB. They are natural hosts and reproduce all of the main features of human infection with MTB, and have elements of the host response that are lacking in some of the small animal models, for example, the ability to respond to phosphoantigens via specific T cell subsets (30). Studies are conducted in either rhesus or cynomolgus macaques which are reported to have different susceptibilities to infection with MTB (31). Depending on the infectious dose, both species can develop a spectrum of clinical symptoms including the propensity to harbour latent bacilli that reactivate upon naturally or chemically induced stress (32, 33).

Many of the reagents that are used in humans are applicable to nonhuman primates and it is therefore possible to perform identical assays in the two species. This applies to the measurement of immunological and clinical parameters; when one adds to this the ability to sample sequentially from individual animals, the power of non-human primates to predict the outcome of vaccination in humans becomes obvious. It is, however, very clear that studies in these animals will be limited to only those vaccines that have shown promise in other preclinical models due to the costs of housing these animals under Biosafety Level 3 (BSL3) containment and the personnel costs of monitoring, sampling, and sample processing. There is no regulatory requirement to conduct non-human primate studies prior to entry into Phase I trials but the benefits of non-human primate data are clearly recognized (34) and many vaccine developers have opted to evaluate the immunogenicity, efficacy, and safety of their candidates in non-human primates either prior to or in parallel with their clinical development (35). Such studies additionally provide a valuable source of material for validation of putative immunological (or clinical) markers of vaccine-induced protection or which indicate likelihood of progression to disease post-infection. Assays can be applied retrospectively to stored samples from individuals who have a known protection status. Safety studies are particularly important for live attenuated vaccines such as modified BCG or MTB mutants because the introduction of such vaccines will involve stringent regulatory hurdles (36). Non-human primate safety data, both pre- and post-challenge, will be extremely valuable to assure developers of the likely safety profile in humans. The ability to monitor and quantify disease progression using advanced medical imaging techniques such as Magnetic Resonance imaging and computed tomography has been described for the evaluation of both drugs and vaccines (37, 38). When applied to animals ‘in-life’, these techniques have the potential to minimize animal use and enable more vaccines to be tested in this highly relevant species.
Cow

Bovine TB is a major economic problem throughout the developed and developing world and the requirement to undertake research into ways to control this disease has led to the establishment of several animal models of *M. bovis* infection (32–46). Cattle are one of the natural hosts of *M. bovis* and this animal species is therefore ideal for the study of the host–pathogen interactions. The relatively large size of these animals presents challenges in terms of housing at BSL3 containment but also offers advantages similar to those of the non-human primate in terms of frequent and sequential sample collection. Studies of the immune response to *M. bovis* infection are limited by the number of bovine-specific reagents available but it is possible to conduct a wide range of appropriate assays (more so than in the guinea pig for example) (47). The close similarity of *M. bovis* to MTB makes the cattle model relevant to the study of vaccines to combat human mycobacterial infection and there is much potential synergy in the efforts to find a vaccine to control human and bovine TB. A major advantage of the cattle model of *M. bovis* is the ability to conduct natural transmission studies, where it is possible to evaluate the ability of a vaccine to prevent disease caused by a naturally acquired infection (48 and Chapter 5.12).

Deer

Deer are used by several researchers studying *M. bovis* infection in cattle since these animals represent one of the wildlife hosts that are responsible for transmission and maintenance of bovine TB (49). They also represent a target host in communities where deer are farmed or managed for hunting and a TB-free status is important. The manifestations of *M. bovis* infection in cervids is largely focused on lymph node involvement and, apart from this model being used to evaluate vaccines targeted at reducing *M. bovis* infection in deer, there are few parallels with studies of MTB infection in humans.

Fish

*M. marinum* is a natural pathogen of fish, and experimental infection of zebra fish has been proposed as a means of rapidly screening mutants in order to understand the role of mycobacterial genes (50). This model is showing promise for high throughput screening of the efficacy of antibiotics but its role in the evaluation of vaccines has yet to be defined.
Pig

The pig is an animal species with many similarities to humans in terms of physiology, anatomy, metabolism, body size and body composition, and may be particularly useful for the study of vaccines which are applied via the skin. Infection of pigs with MTB results in pathological processes that resemble those observed in humans and is proposed as a relevant model of latent TB infection (51). The size of these animals makes them suitable for serial blood sampling and therefore monitoring of immune profiles following vaccination. Studies on neonatal pigs are also more feasible than in small animal species where samples are difficult to obtain and maturation of the immune system is very rapid. Only a few sites have established the pig model for TB studies but it has the potential to be useful for safety, immunogenicity and efficacy studies of vaccines aimed for use across the spectrum of age.

PRECLINICAL MODELS OF DIFFERENT CLINICAL SCENARIOS

The standard study design for the evaluation of a candidate TB vaccine in any of the above animal models is one of prophylactic vaccination. In this model, the animals are vaccinated with the candidate vaccine under evaluation, rested for a variable period of time, and then challenged with a virulent strain of mycobacteria (usually MTB; *M. bovis* is used for the cattle model). This experimental design simulates administration of a prophylactic vaccination scenario, whereby subjects are vaccinated prior to exposure. However, the clinical situation is considerably more complex than this somewhat simplified design, and the various preclinical animal species are increasingly being utilized to model more complex clinical scenarios. Some of these scenarios, and the models being developed to mimic them, are discussed below.

Prophylactic models

As discussed above, this is the commonest model used in preclinical evaluation and is the simplest and most standardized model. Variable factors include length of interval between vaccination and challenge, interval between challenge and study termination, read-outs used to define vaccine efficacy, all of which are discussed in the relevant sections below.
Models of latent MTB infection

It is estimated that one-third of the world’s population (approximately two billion people) is latently infected with MTB and is at risk of reactivation of this latent infection and thus the development of disease (52). Globally, this represents a huge reservoir of potential disease and any therapeutic intervention which reduced the risk of reactivation in this population would have enormous benefit. There are two potential ways in which a ‘post-exposure vaccine’ may act in this situation: first, a post-exposure vaccine might result in eradication of latent MTB infection; second, a vaccine may lead to increased immune containment of latent infection and thus reduce the risk of reactivation without eradication of infection. Additionally, prophylactic vaccines that specifically target the immune response to ‘latency-specific’ antigens are in the development pipeline. Increasingly, efforts are being made to develop animal models to model this clinical scenario and aid in the development of a vaccine for this important target population.

Early studies in mice have shown that the ability to contain MTB infection deteriorates with age, and that older mice are more likely to develop overwhelming disease (53). These data suggest that the mouse may be a representative model of human latent infection. The first and most well-characterized murine model of latent MTB infection was the Cornell model (54, 55). In this model, intravenous infection with MTB (using the common laboratory strain, H37Rv) is followed by treatment with pyrazinamide and isoniazid anti-mycobacterial chemotherapy for 12 weeks to reduce the bacterial load to very low or undetectable levels. In this model, no bacilli are detectable for several months after this period of treatment; but then in many mice, the latent infection spontaneously reactivates and causes active disease. The administration of cortisone reduced the time for 50 per cent of the mice to spontaneously reactivate from 7 months to 2.5 months (56). This model has since been adapted to use an aerosol route of delivery of MTB in order to mimic the natural route of infection in humans (53, 57). To develop an animal model of latency, a dose of MTB needs to be administered which is sufficient to infect all animals, but not sufficient to cause disease. It can be difficult to determine the correct challenge dose to use in this setting. Once all animals are infected, anti-tuberculous chemotherapy is usually administered for a variable period of time in order to reduce the bacterial load and prevent the development of disease. However, the length of antibiotic administration needs to be carefully determined as the aim is not to eradicate infection completely. Once a stable state of latent infection is thought to have been achieved, the candidate vaccine(s) under evaluation are administered. The animals are then left to rest for a variable period
of time before the latent infection is either allowed to spontaneously reactivate, or immunosuppressive agents such as corticosteroids or the nitric oxide synthase inhibitor, aminoguanidine, are administered to reactivate latent infection (58, 59). These models of latent disease are long and complex, and the variability much greater than in the standard (and shorter) prophylactic models. Variability between strains of mice used in these models suggests that not all are suitable for these latency models (60).

Although these models are potentially very useful in modelling human latent MTB infection, in order both to evaluate which immunological factors are important in control of such infection and to evaluate potential therapeutic options, these models are complex, variable, time-consuming, and expensive. The outcome of these models is dependent on certain key variables, which include antibiotic regimen and length of treatment period, dose of mycobacterial challenge used, and length of antibiotic-free rest period (61, 62). The limitations of these models include spontaneous reactivation, and difficulties in inducing reactivation and in generating phenotypically altered mycobacteria (61, 62).

A non-human primate model of latency is currently being developed, which would potentially be very relevant to the human scenario (32, 63). Forty per cent of cynomolgus macaques infected with a low dose aerosol challenge did not go on to get disease, but remained latently infected. The use of such a model is constrained by inter-group variability, which will determine group size, cost, and availability of appropriate BSL3 containment facilities.

**Therapeutic vaccination models**

Treatment of TB disease involves a minimum of six months anti-tuberculous chemotherapy. Usually the first two months is an intensive phase involving three or four drugs, and then there is a maintenance period of four months where two drugs are administered. This protracted period of treatment is one of the reasons why the development of resistance is common, as compliance with treatment often deteriorates once the patient starts to feel better, long before the end of the treatment period. An important potential use for a new TB vaccine would be as a therapeutic vaccine, administered as an adjunct to chemotherapy, perhaps to reduce the duration of treatment or to increase cure rates. Such a vaccine would be of enormous benefit. However, there are safety concerns with the use of new TB vaccines which stimulate potent antmycobacterial immunity in patients who have TB disease. In the 1890s, Robert Koch administered short-term culture filtrate
to people with TB disease as a potential cure or ‘remedy’ for TB. Unfortunately, the results were disastrous and some people died (64). This so-called ‘Koch phenomenon’ has been reproduced in animal models and appears to relate to bacterial burden. Mice with a heavy bacterial load, but not those ‘latently infected’, were shown to have increased immunopathology after therapeutic vaccination (65).

One of the new generation of TB vaccines in development has been evaluated in a therapeutic model and found to be effective (66). A DNA vaccine encoding hsp65 from *M. leprae* was found to reduce bacterial burden after high-dose intravenous challenge with MTB (66). However, using an aerosol challenge model, others have demonstrated that post-exposure vaccination of both mice and guinea pigs with similar vaccines can lead to a detrimental effect (65). This illustrates the importance of evaluating any one vaccine candidate in more than one animal model. Better animal models, which allow us to dissect the mechanism of the immunopathology seen in some of these studies, are required before we can begin to exploit the potential of a therapeutic vaccine in this setting. Recently, a therapeutic vaccine which consists of killed fragmented whole MTB, RUTI, is being evaluated in both latency and therapeutic models, with some success (67, 68).

**Neonatal TB vaccines**

BCG is currently administered at birth throughout the developing world. One important target population for the next generation of BCG replacement vaccines is neonates. There may be factors such as relative immaturity of the immune system which mean that evaluating vaccines in adults does not reflect the situation in neonates. There are also environmental factors which may differ between infants and adult populations. Adults and neonates will differ in the amount of exposure to environmental mycobacteria, a factor considered important in explaining some of the variability in protective efficacy of BCG. A neonatal animal model would be a useful tool in exploring some of these issues. Efforts have been made to use neonatal mice to model this situation (69, 70). Non human primates would be the ideal species in which to perform these studies but the opportunity to do this is restricted to sites that have a close association of facilities with an NHP breeding colony.
Environmental mycobacterial exposure

One of the most likely explanations for the variable efficacy of BCG vaccination is exposure to environmental mycobacteria. Environmental mycobacteria are ubiquitous in the soil, but exposure is increased with increasing proximity to the equator. They may interfere with BCG either by masking or by blocking an appropriate immune response. The term ‘masking’ refers to the situation where exposure to environmental mycobacteria induces some antimycobacterial immunity, and vaccination with BCG does not incrementally increase that response. Just such an effect was seen in an elegant series of studies of vaccination of BCG-naïve adolescents in the UK (where BCG is effective) and Malawi (where it is not) (71). In these studies, BCG-naïve adolescents in the UK were found to have minimal antimycobacterial immune responses, in contrast to the Malawian adolescents who had high baseline immune responses. In the UK, most adolescents responded well to BCG vaccination, with a significant incremental rise in immune responses to mycobacterial antigens. In contrast, most Malawi adolescents failed to mount a significant rise in immune response after BCG vaccination. Thus, pre-exposure to environmental mycobacteria may ‘mask’ the effect of BCG vaccination.

This masking effect was demonstrated in the guinea pig model of TB in 1966 (72). In this study, exposure to various strains of environmental mycobacteria conferred some protection against subsequent challenge with MTB. This has also been demonstrated in mice (73). In this work, pre-exposure to environmental mycobacteria did not inhibit the protective efficacy of BCG. These animal studies are consistent with the hypothesis that exposure to environmental mycobacteria confers some protective immunity, and vaccination with BCG may not incrementally increase that protection.

‘Blocking’ describes a more active mechanism whereby pre-existing immunity to environmental mycobacteria prevents BCG replication, resulting in removal of BCG and inhibiting vaccine ‘take’, a mechanism shown to occur in animals (74). In a similar model, pre-exposure to *M. avium* increased BCG clearance, but had less effect on clearance of recombinant BCG expressing genes from the RD1 region of MTB (75).

It is important to evaluate the effect of environmental mycobacterial exposure on the protective efficacy of the new vaccines currently in development. Interestingly, in the study described above (74), pre-exposure to environmental mycobacteria did not appear to inhibit the protective effect of a subunit protein vaccine, suggesting the new generation of subunit vaccines may overcome this limitation of BCG. Likewise,
environmental exposure had less effect on a new generation recombinant BCG strain than on the wild type strain (75). However, modelling exposure to environmental mycobacteria in animals is challenging given our lack of detailed understanding of what human exposure to environmental mycobacteria involves. It is likely that different strains of environmental mycobacteria differ in prevalence throughout the world. It is also not clear whether exposure to dead or live environmental mycobacteria is more relevant to this sensitization. Finally, human exposure to environmental mycobacteria is likely to involve low grade, repeated exposure, and potentially by many different routes of entry. Modelling this in animals is complex and results in long experiments which can be highly variable in outcome.

Helminths

Exposure to helminths is common throughout the developing world and is likely to be an additional factor which explains some of the variability in efficacy seen with BCG vaccination. Chronic helminth infection results in a Th2-biased immune response, characterized by the production of IL-4 and IL-5; in contrast, protective immunity against mycobacteria requires a predominantly Th1-type immune response, characterized by the production of IFN-γ and IL-12. It has been shown that chronic worm infection reduces the immunogenicity of BCG in humans (76). Interestingly, this reduction in immunogenicity was associated with increased TGF-β production but not with enhanced Th2 immune responses. As with exposure to environmental mycobacteria, the species of helminths that different populations are exposed to, together with varying magnitude of such exposure, will vary widely throughout the world. This means modelling such exposure in preclinical animal models is challenging. However, attempts to do so have yielded important results. In a murine model, pre-existing *Schistosoma mansoni* infection has been shown to reduce the protective efficacy of BCG vaccination against virulent MTB challenge (77). In this study, the *S. mansoni*-infected mice exhibited a Th2 bias and had lower IFN-γ immune responses to PPD (purified protein derivative from MTB), consistent with the hypothesis that the interference with BCG-induced protection may be in part attributable to a Th2 immune bias.

Immunosuppression

Given the geographical overlap between the TB and HIV epidemics, and the devastating synergy that these two pathogens have, perhaps one of the most important factors to model in the preclinical models is the interaction between these two pathogens and the effect that immunosuppression has on the efficacy of new vaccines. A recent WHO working party agreed that attenuated MTB vaccines
should be evaluated for safety in SCID mice as well as in guinea pigs (36). Two such attenuated strains of MTB have been demonstrated to be safe in both of these animal models (78, 79). Non-human primate models of simian immunodeficiency virus (SIV) and MTB coinfection would allow further analysis of the safety and efficacy of all the new candidate vaccines in this important target population. However, such models are complex and require expensive containment facilities (80, 81).

**Natural transmission**

The two ‘gold-standard’ animal model systems being used to replicate natural transmission of mycobacteria are: 1. Within-species cattle model in which *M.bovis* infected animals are allowed to naturally infect naïve or vaccinated calves (48) and 2. A between species model where guinea pigs are exposed to the extract air from a ward containing patients who are undergoing treatment for MDR-TB (82). In both of these systems, the recipient animals are exposed to realistic doses of mycobacteria which are in a natural physiological state. However, neither model can be easily tech-transferred nor could they be used to screen a wide range of novel vaccines. More tractable animal models which replicate key features of the cattle-cattle or human-guinea pig models are needed. The recent call for proposals by the Bill and Melinda Gates Foundation specifically addresses this topic and advances in this area will undoubtedly occur.

**FACTORS FOR CONSIDERATION**

**Evaluation of vaccine efficacy**

A critical factor for consideration in any of the animal models described above is the measurement of vaccine efficacy. There are several ways in which the potency of a vaccine may be described or measured and variations in these read-outs of vaccine efficacy can cause problems in discriminating candidates in order to select those that should progress further in the development process. The underlying issue is that there are no clear correlates of protective immunity for TB in humans and in the various animal models (83). The ability of a vaccine to prevent disease that progresses to a terminal endpoint is the most obvious measurement of protective efficacy and for this reason survival studies are frequently conducted in animal models. However, there are several issues with survival studies in terms of their relevance to the human situation and in terms of feasibility and reliability of the
data generated. Infection with MTB progresses slowly in the majority of the animal models, and the time taken to reach a severe end point is in the region of months rather than days in the unvaccinated controls and longer in the protected animals. Survival studies are therefore time and resource consuming and many developers choose not to conduct such studies, particularly in the early, screening, stage of candidate evaluation. The main issue with survival studies is the statistical power and reproducibility of the evaluations. Survival data are usually presented using a Kaplan Meier plot and differences between the treatment groups are analyzed statistically by various methods but usually by the Log Rank test. The variation in time taken to end point within a treatment group is the key factor and this can be a major issue in outbred animal species such as the guinea pig and the non-human primate. Statistical power calculations can be used for any read-out to estimate the number of animals per group required to differentiate the negative and the positive control groups. The same group size will then identify a vaccine candidate that has the same efficacy as either of the control groups but much larger group sizes are needed to discriminate a vaccine that is different from the control groups unless this difference is substantial. Using survival as a read-out, the group sizes identified by the power calculations are often prohibitively large and, too frequently, studies are conducted where the group size was not sufficient to accurately describe the efficacy of a vaccine candidate. Variations in the time taken to reach severe disease can be reduced, for example by increasing the challenge dose (84), using more susceptible species, or more virulent challenge strains, with the outcome that the unvaccinated controls and the BCG positive controls succumb to disease more rapidly. The candidate vaccine under evaluation can then be more readily distinguished from the control groups, provided that it can protect against the more stringent challenge.

A commonly used alternative to survival studies is to measure the bacterial load in target organs such as the lungs and spleens, and compare the mean or median number of viable bacilli (determined by plating organs on solid media and counting the CFUs) in the test and control groups. This analysis, generically described as ‘CFU data’, is conducted at a fixed endpoint, usually within two months of challenge, depending on the animal model. The issue with this read-out is that there are very few studies which have evaluated whether bacterial load in the early stages of infection is predictive of long-term disease outcome. There are currently no TB vaccines that generate an immune response which eliminates MTB from lungs or spleens in the early stages of infection in the various animal models and, until more clinical trial data emerge, we will not know if this type of sterilizing immunity occurs
in humans and thus the relevance of ‘early CFU data’ remains unclear. Purely within a preclinical model setting, there are studies in the guinea pig that have correlated early CFU with survival data, using BCG as the vaccine and a recent study showed that the same applied to sub-unit vaccines (85). However, there remains scepticism that this will be the case for all types of vaccine, particularly those vaccines which are reported to have their greatest impact in the later stages of disease (86). Early CFU is the read-out of choice for early stage screening, particularly in mouse models. Statistical power and group size needs to be considered for CFU but the objective and continuous nature of these measurements means that within-group variation is less than survival data and much smaller group sizes can discriminate vaccines from control groups.

Pathology scoring is another frequently used read-out of vaccine efficacy. The assumption with this is that a protective vaccine will reduce the number and size of visible lesions and/or limit the dissemination of pathology from the initial site of challenge. Pathology scoring is more relevant to some animal models than others, having a greater utility in the larger animal models such as cows and non-human primates where clearly defined pathology scoring systems can be used to discriminate test vaccines from the control groups. There are some key issues with such scoring systems that need to be taken into account in order to ensure that the scores generated will be informative. The most important consideration is the relevance of the parameters that are being scored. Pathological lesions arise from host–pathogen interactions and, where possible, lesions that indicate a favourable host response (e.g. small well-formed granulomas) should be differentiated from lesions that indicate an unfavourable pathogen effect (e.g. extensive caseous necrosis) (87). Thus, a qualitative measurement should be used to describe the pathology but statistical analysis of qualitative data is unreliable at best. Quantitative data is much more reliable but may be much less relevant — the total number of lesions for example may be reliably measured but may be of low relevance to the disease process. As previously mentioned, advanced imaging is now offering the potential to quantitatively measure disease burden. Being able to apply this to whole organs is more powerful than histological analyses which can only sample a minor fraction.

Analysis of changes in weight is a read-out that is not commonly used but can provide highly valuable data. Weight loss is a key feature of progressive mycobacterial disease in most of the animal species, including humans. Weight can be measured reliably and analyzed using robust statistical methods, and should be considered as an additional read-out in survival studies.
The ideal read-out of vaccine efficacy is one that is highly relevant to the disease process in the particular animal model and can be measured accurately and reproducibly to allow robust statistical analysis. The perfect read-out will additionally have strong relevance to the disease process in humans. Regrettably, none of the read-outs that are used in the preclinical screening of vaccines are perfect and few are ideal. Despite the best efforts made by researchers to perform robust evaluations, practical and financial issues often limit the value of the data generated.

**Comparative evaluations and variability between evaluators**

Much of the early-stage development of vaccines is conducted within an individual research group or company and there is general acceptance that the data generated does not require verification by others. It has become common practice for the most promising vaccines to be evaluated by independent testing facilities that have expertise in the animal models rather than in vaccine development. Such groups include the National Institutes of Health (NIH)supported TB Vaccine Testing and Research Materials contract (TBVTRM Contract — Colorado State University) which has testing centres at Colorado State University and Texas A&M University, and the EU funded TBVAC contract which supports a network of preclinical evaluation sites across Europe. These groups offer a provision of animal models that would not otherwise be available to individual researchers and in doing so they enable more rapid progression of candidates towards the clinic. The independent status of these evaluators is critical to ensure that unbiased data are generated and a substantial benefit of testing centres is the ability to conduct comparative evaluations where the efficacy of one candidate can be compared directly with another (84).

Unfortunately, these test sites have not yet found a mechanism to fully share the very large body of data that each has accumulated on all of the novel TB vaccine candidates evaluated over the past 15 years. Publishing all of these studies would be a considerable benefit to the field as a whole. There are several barriers in the way of achieving this, including non-scientific factors such as intellectual property (IP) and confidentiality; scientifically, the main barrier is differences in methodology employed at the various sites. These differences include challenge strains of MTB, strains of BCG used as controls (see below), challenge routes, methods of challenge delivery, sources and genetic background of the animals, and different read-outs to define efficacy. Comparisons between the sites would appear to be impossible and there is an often quoted study which involved several laboratories evaluating different vaccines and where no two laboratories agreed on the relative potencies
of the five vaccines, clearly illustrating the difficulties (88). However, the field of TB vaccine development has progressed and there are much stronger incentives to harmonize the global effort, which will hopefully improve the situation. Large stocks of MTB Erdman strain and BCG Danish 1331 are available through a WHO/Aeras initiative in an attempt to encourage standardization of the positive and negative control groups. This is a major step towards harmonization but there is no obligation to use these materials and individual laboratories often prefer to continue using their own strains in order to maintain continuity of their data. Some of the animal models are more reproducible than others and minor differences in challenge strains and BCG strains can be overcome provided that the analysis of the vaccine efficacy takes those differences into account. For example, CFU analysis in mice and guinea pigs can be expressed as the log10 difference (reduction or increase) in CFU relative to the control groups. If a test vaccine is evaluated this way in two test sites, with the odd exception (e.g. if challenge strains have markedly different virulence) the vaccine will perform comparably relative to the BCG and unvaccinated controls.

Harmonization can be very difficult to achieve, particularly where specialized equipment or facilities are being used at one site or if a highly specialized animal model is being used, but some of the perceived barriers to standardization (e.g. challenge route, or means of challenge delivery) have been demonstrated to have a relatively low if any impact on the overall performance of a vaccine at different sites. A multi-centre trial of two vaccines in the guinea pig model yielded the same result in terms of relative efficacy compared to the standard BCG and saline controls, despite a number of differences in the methodology (85). However, the efforts to harmonize preclinical evaluations must continue in order to maximize the benefits of data generated in animal models. A database of preclinical vaccine efficacy in several animal model systems would be a very powerful tool for determination of the most promising of the current vaccine candidates to go forward to expensive Phase III efficacy trials.

**Evaluating BCG boosting regimens**

New TB vaccines are being designed to either replace BCG or to be applied in addition to BCG. As a part of the evidence needed to predict whether these vaccines will have any additional benefit over the current BCG, there is a need to demonstrate significantly improved efficacy in animal models. This may be relatively straightforward in mouse models because the protection offered by BCG is not strong and the new vaccine or booster may significantly increase the ability to reduce bacterial burden or enhance survival. However, in other animal models such
as the guinea pig, the effect of BCG is much stronger; for example, bacterial burden may be reduced to such an extent that numbers of bacilli in spleens are close to the detection limit of the assay and survival is extended such that it is frequently not possible to significantly prolong this further. There are examples, however, of BCG-boost regimens showing significantly improved survival over the BCG control group in the guinea pig but given the stringency and limitations of survival as a read-out (see above), it is not rational to assume that this method of evaluation will predict the outcome in humans (89, 90). There are several BCG-boost regimens that have shown good efficacy in mice and other animal models but cannot be distinguished from BCG-alone in the guinea pig (21, 84). There is no simple solution to this problem in guinea pigs. One option is to increase the challenge dose or to use a more virulent challenge strain such that the BCG control group is less protective. This will not be a solution for all vaccines, however, because the increased stringency of the challenge may be overwhelming and, given the low relevance of high dose challenge to the human situation, failure to protect in this model does little to inform the clinical development of the vaccine. Another solution is to leave a prolonged interval between vaccination and challenge, which simulates the ‘late boost’ concept in humans where the waning immunity of BCG in adolescent or adults is boosted. This may be feasible in a mouse model but the strong potency of BCG in guinea pigs may result in an impractical time interval between BCG and boost. However, there are no recent published studies which have investigated this with the BCG and challenge strains currently in use. The BCG-boost strategy is an important concept in TB vaccine development, and it is important that a consensus is reached on the most appropriate means of evaluating these strategies in the preclinical models in order to inform the design of the clinical studies where the number of potential prime-boost intervals will be vast.

**Using BCG as a gold standard**

BCG is always used as a gold standard in preclinical vaccine evaluation and confers very consistent protection in these models. Any new vaccine must do at least as well as BCG in these models. However, as efforts to harmonize methodologies across preclinical groups become more focused, it is becoming increasingly clear that there are differences between the different strains of BCG in use throughout the world. It is therefore important for all preclinical groups to use the same strains of BCG, to allow comparison between different candidate vaccines. This becomes potentially more important where BCG booster regimens are concerned. There is considerable evidence of genetic differences between the different strains of BCG (91). In addition, differences in gene expression between the different strains have
also been demonstrated (92). There is some preclinical data to suggest that these differences can result in different levels of protective immunity (93). In this study, the protective efficacy of 10 BCG substrains against MTB challenge was compared. Differences in both CFU counts and pathology scores were seen between the different strains, and there was some evidence of differences in the cytokine profile induced by the different strains, although this did not predict the protective efficacy. There is more conflicting evidence from human studies with some but not all studies showing differences in immunogenicity between the different BCG strains (94–96). Different strains of BCG have been used in the many efficacy trials conducted with BCG. However, these efficacy trials have also been conducted in different populations and in different age groups and it is therefore difficult to determine how important a factor BCG strain is in the differences seen. A meta-analysis of 26 BCG human efficacy studies suggests that BCG strain is not a major factor (97). Furthermore, a head-to-head comparison of the protective efficacy of BCG Danish and BCG Pasteur against virulent M. bovis challenge in cattle showed no differences, despite inducing different levels of IFN-γ post vaccination (98).

More evidence is needed before this issue can be resolved. However, given the possibility that there are differences in protective efficacy between different BCG strains, it may be prudent to use a standard strain of BCG across all the preclinical models. However, it may also be important to use multiple strains of BCG in the preclinical studies, given that there are many different strains of BCG in clinical use throughout the world.

**Different strains of MTB**

Another important factor for consideration in preclinical animal challenge studies is which strain of MTB to use. Commonly, fully drug-sensitive laboratory strains such as H37Rv and Erdman are used. Efforts have been made to standardize the use of the Erdman strain for preclinical work (99). However, given increasing evidence of both genetic and phenotypic differences between MTB strains, it may be important to include some of the recent more virulent clinical isolates (100). MTB CDC1551 was first isolated in a TB outbreak in the USA, and was found to cause higher infectivity and virulence in humans than other, more commonly isolated strains (101). This strain was subsequently found not to be more virulent in mice, despite inducing a stronger host immune response (102). However, studies in guinea pigs have confirmed the increased virulence seen in the clinical setting (103). Members of the W-Beijing family of MTB isolates, such as HN878, have also been found to be more virulent in both the clinical setting and in laboratory animals.
Interestingly, the increased virulence observed in mice with this strain was associated with a failure of the host Th1 type immune response (104). The HN878 strain is also more virulent in the guinea pig model (103).

The issue for preclinical vaccine evaluation is whether the protective efficacy of the new candidate vaccines being developed will be affected by challenge strain. A recent paper comparing the protective efficacy of BCG against nine different MTB isolates, including four W-Beijing type isolates, has shown a comparable level of efficacy against all strains, suggesting that protective immunity, at least induced by BCG, may not be strain specific (99). A more recent study indicated that BCG-induced protection against W-Beijing strains was only transient and when evaluated at a later time point post-challenge, BCG did not protect against these strains (105). This needs to be investigated for the new generation of TB vaccines, but care is needed in the interpretation of such studies. Only a relatively small number of MTB isolates have been tested in animal models, compared to the number of different isolates causing human disease and therefore selection of isolates which represent human disease is not yet possible. Major differences in the pathogenicity of the isolates have been observed in the animal models yet all of them caused significant human disease. Thus, placing these strains in an experimental setting may cause them to behave differently. There is a bias towards the study of the more pathogenic (for animals) isolates but new vaccines should be tested against isolates across the spectrum of human disease, not just those which give dramatic results in animal models.

**Immunological correlates of protection**

The identification of immunological correlates of protective immunity would greatly facilitate the development of new TB vaccines (106). Ultimately such immunological correlates will need to be evaluated and validated in large scale clinical efficacy trials. However, potential correlates can be identified and evaluated in preclinical models. A detailed analysis of which aspects of immunity are contributing to vaccine-induced protection is essential in preclinical models. These immunological responses can then be evaluated in the early stage clinical trials, and this information may help in prioritizing which candidate vaccines move forward into larger scale efficacy trials.

It is straightforward to perform detailed immunological studies in parallel with challenge studies in mice, where a wide range of immunological reagents and knock-out strains are readily available. More limited reagents are available for the guinea pig. Given the differences between human and murine TB disease, it is
important that potential correlates identified in murine studies are evaluated in the larger non-human primate and cattle models, as well as in humans. It is possible that there are species-specific differences in these immunological correlates. It is also possible that the different vaccine types under development (e.g. whole organism or subunit) may confer protection by different immunological mechanisms. It is only by performing detailed immunological studies, with all the different types of candidate vaccines currently in development, and in all the animal species, that we will begin to understand more about which aspects of immunity are important for protection, and which animal models are most relevant to human disease.

CONCLUSION

Preclinical animal models form an essential step in the development of new TB vaccines. The different animal species are all useful at different stages in development and can provide complementary information. These animal models can be used to develop a more detailed understanding of the factors important in protective immunity. At present, we do not know which animal model(s) best represent the different stages of human disease. This requires an iterative process of preclinical and clinical studies. We can only validate the different animal models and evaluate which are most representative of human disease when we have a more effective vaccine in humans. Until then, we have to develop and refine the models, use them to identify potential immunological correlates, and use them as a screening tool to exclude some vaccines.

It is important that there is more harmonization of efforts to allow better preclinical comparison between candidates. There are very limited resources and facilities available for large scale efficacy testing in humans. The preclinical data provide essential evidence which, together with early stage clinical data, allows prioritization between different candidates for these limited resources. Furthermore, it is important to conduct independent comparative efficacy studies as such studies may not be possible in humans.

Once we have some human efficacy data, we must validate the different animal models and ensure that models/assays which are not informative are discontinued in order to focus efforts on the right model, and choose the readout that gives optimal balance between relevance to disease and accuracy of measurement. Given that preclinical models of TB have been around since Koch’s identification of the causative organism, perhaps we also should review and use the wealth of historical data better to prevent us repeating the experiments and mistakes of the past.
REFERENCES


