Eliminating latent tuberculosis

Douglas B. Young1,2, Hannah P. Gideon3 and Robert J. Wilkinson1,3,4

1 National Institute for Medical Research, Mill Hill, London, NW7 1AA, UK
2 Division of Investigative Science, Imperial College London, SW7 2AZ, UK
3 Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
4 Division of Medicine, Imperial College London, W2 1PG, UK

Tuberculosis is recognized as the world’s leading bacterial cause of death. Yet 95% of infection is believed to exist in an asymptomatic ‘latent’ form that is defined not by the identification of bacteria, but by the host immune response in the form of reactivity to tuberculosis proteins in the tuberculin skin test. It seems likely that clinically defined latent tuberculosis actually represents a spectrum that runs from elimination of live bacilli to subclinical disease: hence, it might be unhelpful to use a single term to describe all these conditions. To support this view, here we focus on recent increased understanding of the heterogeneity in both bacillary physiology and host immune response that potentially illuminates new therapeutic and diagnostic approaches to this condition.

Tuberculosis: a persistent challenge
Tuberculosis (TB) is a leading bacterial cause of death: in 2006, there were 9.2 million new cases and an estimated 1.5 million people died [1]. An increasing proportion of disease is caused by organisms insensitive to first line chemotherapeutic agents (multi-drug resistant TB (MDR-TB)) [2]. Furthermore, it is estimated that one-third of the world’s population harbours drug sensitive, and perhaps increasingly drug resistant, TB in the form of latent infection [3]. Thus, understanding latent TB infection is an important question: even if all cases of active disease could be detected and successfully treated, an immense future reservoir of disease exists.

Mycobacterium tuberculosis (the bacterium causing the disease) is transmitted as a highly infectious aerosol from the lungs of individuals with active disease (Figure 1). An encounter with M. tuberculosis is classically regarded to give rise to three outcomes. First, a minority of perhaps 5% (higher in infants or immunocompromised persons) develop rapidly progressive disease that is sometimes confirmed by isolation of M. tuberculosis from clinical samples (evidence of bacterial replication) and referred to as ‘active’ primary TB. Although an acquired immune response is almost invariably detectable, in these circumstances it is manifestly incompletely effective. By contrast, the majority of infected persons show no disease symptoms but develop an effective acquired immune response – they are referred to as having ‘latent’ TB, on the basis of a positive result in either a tuberculin skin test (TST), or an in vitro assay of T-cell function (interferon gamma release assay, IGRA [4]). As a third possible outcome, a proportion of those ‘latently’ infected (maybe 1–2% per year) might reactivate infection and develop the disease (known then as ‘postprimary’ TB). Its onset can also be triggered by immunosuppression of which the greatest identified cause is HIV-1 infection [5].

Although clearly a useful clinical concept, we consider that the current view of ‘latent’ TB as a single homogeneous entity is inadequate and potentially limits the basis for the rational development of drugs, vaccines and associated biomarkers. Here, we suggest that this concept needs to be re-assessed and perhaps replaced (or complemented) by alternative terminologies that more accurately demarcate important biological features and thereby research goals. We expect that the new conceptual framework will facilitate further research towards improved detection methods and treatments for latent TB.

Rethinking ‘latent’ TB
What is the basis for considering most TB infection in the world to be latent? By analogy, until the mid 1990s, Human Immunodeficiency Virus (HIV-1) infection was characterized as having three phases. After acute seroconversion there was resolution to a minimally symptomatic ‘latent’ phase. Then, after a variable interval of 4–15 years and for reasons that remained vague, an overt clinical AIDS phase of terminal profound immunodeficiency ensued, characterized by serious opportunistic conditions, of which the commonest worldwide today remains extrapulmonary and disseminated TB. The triple advent of technology to easily determine circulating viral load (quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)), highly effective multidrug treatment for HIV-1 and improvement in methods to quantitate HIV-1 antigen-specific T cells, however, led to a paradigm shift in understanding [6,7]. These studies and many subsequent analyses revealed that the asymptomatic latent phase of HIV-1 infection was actually characterized by extraordinarily high rates of both virion generation and turnover of T cells that destroyed virally infected cells: a highly dynamic equilibrium that the host, in time, was destined to lose [8]. The terminology ‘latent’ was recognized to misrepresent pathogenesis and disappeared rapidly from both literature and thinking on the subject.

By analogy, how true is the current paradigm that classifies TB into ‘active primary’, ‘latent’ and ‘postprimary’? It is acknowledged to be an oversimplification, leading to a plethora of other terminologies in the literature to qualify tuberculous infection variously as dormant, active, inactive, subclinical, acute, chronic, persistent, remote and quiescent. We propose the following spectrum of responses
to TB infection: innate immune, acquired immune, quiescent infection, active infection, disease (Figure 2). Innate immune is proposed for infection that is eliminated without the need for an acquired response, and acquired immune for infection that is eliminated by a combination of innate and acquired responses. Quiescent infection would represent infection that is controlled by the immune response but in which viable bacteria persist in a non-replicating form. Active infection would denote the presence of replicating bacilli nevertheless controlled without symptoms by the immune response, and clinical disease the development of symptoms and signs of TB caused by bacteria that replicate despite the ongoing immune response.

Is there evidence that these states exist? First, it has long been recognized that a proportion of repeatedly TB-exposed healthy people, for example health care workers, never convert to a positive TST [9–12]. Although this might simply reflect a sensitivity deficit of the TST it seems also plausible that exposure might not always result in an acquired immune response. Non-immune (e.g. via ciliary washing up and thereby expectoration) or innate mechanisms including natural killer (NK) cells, neutrophils or macrophages might eliminate bacilli without the need for an acquired response [13,14]. Even if an acquired response is triggered, TST conversion is not invariably. The advent of IGRA (which rely on a single immune parameter, unlike the TST [15]) has consistently identified persons who are IGRA-positive but TST-negative after exposure to TB [16–18]. In the study by Ewer [16], reversion of a positive IGRA to negative was noted: could this represent elimination of low-level infection by an acquired immune response insufficient to trigger a TST reaction? Conversely, another reason for failure to convert to positive in the TST is HIV-1 infection, which markedly decreases the sensitivity of TST testing [17]. Another intriguing group thrown up by comparisons between the TST and IGRA are persons who are IGRA-negative but strongly TST-positive. The recognized more dynamic nature of IGRA responses might reflect recent infection that stimulates effector responses to the antigens secreted, by actively dividing bacilli upon which these assays depend [19,20]. Could resolution of such responses and the development of a memory response to other antigens, expressed at low levels by non-replicating bacilli and only present in PPD, in highly protected people, result in this phenotype? Is it also feasible that strain variation in M. tuberculosis has differential effects on TST conversion and disease progression? The advent of discriminatory techniques to classify clinical strains of M. tuberculosis has facilitated correlation of virulence attributes with genotype [21]. An early investigation of an outbreak caused by strain CDC1551 noted a tendency to marked TST conversion, possibly related to its ability to induce a vigorous innate immune response [22,23]. Conversely, transmission of TB detected only by IGRA and not by TST has also been described [18,24].

Microbiology of latent TB

Is there a correlation between an active TB-specific immune response and the presence of viable M. tuberculosis? This question has plagued TB researchers for over a hundred years. A straightforward approach to answer it involves a comprehensive screen of tissue samples for the presence of bacteria by microbiological culture or by inoculation into highly susceptible guinea pigs. Extensive autopsy studies were carried out in the first part of the 20th century, generating an impressive although contradictory and bewildering literature [25–29].

Widely varying rates of positive cultures were obtained from individuals living in endemic areas but dying from causes unrelated to TB. Some investigators isolated and
cultured bacteria from histologically normal tissues; more commonly investigators focused on lesions, reporting different rates of positive cultures from different lesion types. Confounding factors in such studies include the existence of ‘dormant’ bacterial phenotypes that require especially careful and prolonged culture to generate bacterial colonies, and the possibility that, in highly endemic settings, positive cultures might represent recently inhaled bacteria rather than persistent infection or progressive disease. A general conclusion might be that individuals with latent TB frequently harbour viable bacteria; however, it is difficult to draw hard quantitative conclusions from the essentially anecdotal process of post-mortem analysis.

With respect to ‘dormancy’, there is clear evidence from experiments performed on in vitro cultures that M. tuberculosis can transcriptionally adapt to a wide variety of environmental conditions [30]. A classic system employed to better understand non-replicating persistence uses oxygen limitation to induce this state (the Wayne model [31]). Under these circumstances it is recognized that an early response is the coordinate upregulation of genes under the control of two sensor kinases (dosS and dosT) and a response regulator (dosR) [32,33]. The ‘dosR regulon’ includes genes involved in triglyceride synthesis: there is great interest in this field because part of the adaptation of M. tuberculosis to persistence clearly involves a shift in carbon source from glucose to fatty acids [34,35]. More recently it has been shown that a second wave of genes are induced by more prolonged hypoxia that only minimally overlap with the dosR regulon: this has been called the enduring hypoxic response and includes a large number of transcriptional regulators [36].

Modern technologies, especially nucleic acid amplification, provide the opportunity to advance our detailed understanding of the status of bacteria present in tissue samples. Two interesting studies have demonstrated the presence of M. tuberculosis DNA outside both cells and granulomas in latently infected tissue [37,38]. Determination of the expression profile of bacilli in latently infected tissue would be of particular interest. However, we have not so far been able to move from localized analysis towards a systemic measure of bacterial load. It seems feasible that bacterial components could be detectable in the circulation (for example, cell wall components) and the search for such biomarkers is an important research priority. An encouraging precedent is provided by the detection of phenolic glycolipid in blood and urine from leprosy patients [39–41], and the detection of mycolic acids persisting in archaeological specimens [42]. A truly quantitative measure of bacterial load, analogous to measures of viral load routinely used in the epidemiology and management of HIV, would transform the spectrum proposed in Figure 2 from a conceptual hypothesis to an experimental framework.

Immunology of latent TB
Can we use the immune response as a surrogate measure of bacterial load? Careful analysis of low-level antibody responses, sometimes overlooked as ‘background’ in the development of serodiagnostics, might be a useful indicator of the presence of antigen in latent TB [43]. Changes in the level of T-cell response in persons undergoing treatment of both active and latent infection indicate some relationship to antigen load [44,45], although measurements using peripheral blood are complicated by the sequestration of reactive cells to disease sites [46,47]. From a population perspective, the level of T-cell response associated with asymptomatic infection is not readily separable from that seen in patients. An indirect indication of variations in antigen load in asymptomatic infection is provided by clinical studies of the impact of antiretroviral treatment of HIV patients. Restoration of T-cell responses in these individuals can be associated with a sudden conversion from inapparent TB to an acute unmasking form of the immune reconstitution inflammatory syndrome (IRIS, a syndrome arising from rapid restoration of pathogen-specific immune responses causing either the deterioration of a treated infection or the new presentation of a previously subclinical infection) that seems likely to be related to the induction of an immune response in the context of a high pre-existing antigen load [48].

Aside from the issue of bacterial load, refinements in analysis of T-cell responses offer the potential to explore the proposed latent TB spectrum. Because gene expression profiles differ widely between cultures of M. tuberculosis, subject to differing conditions in vitro, we might anticipate these differences would result in distinct sets of antigens being available for T-cell recognition in the groups we refer to in Figure 2 as ‘acquired immune’, ‘quiescent infection’ and ‘active infection’. There is some evidence of infection-stage-specific recognition of M. tuberculosis antigens in humans, although the overlap between clinical categories is substantial [49–51]. Having a better understanding of the dynamics and markers of T-cell memory, effector and perhaps regulatory [52] responses might also help in the important distinction of individuals who have actually cleared the infection. Over time, after initial infection, one might expect a progressive shift from effector to central memory phenotypes in the ‘adaptive immune’ group, in contrast to a predominance of effector cells in groups with more frequent antigen exposure or ongoing low-level bacterial replication. Again, HIV infection points the way. Polyfunctional T cells that produce multiple cytokines and which retain the capacity to divide, correlate with better viral control [53], a finding that has already influenced what is measured in trials of TB vaccines in humans [54].

Interventions in latent TB
Current efforts to control TB focus primarily on attempts to reduce transmission of M. tuberculosis by prompt detection and effective treatment of infectious patients, and it is hoped that optimal implementation of this strategy will achieve the United Nations Millennium Development Goal of a 50% reduction in the global prevalence and mortality of TB by the year 2015. To achieve the more ambitious goal of TB elimination by 2050, it will be necessary to combine the transmission blocking approach with interventions that reduce the percent of infected individuals who progress to clinical disease [55] (Figure 1). The powerful impact of changes in the risk of disease progression is illustrated
by the consequences of the enhanced susceptibility of persons who are co-infected with HIV: the 5%–10% lifetime risk of disease is increased to a 5%–10% annual risk in association with HIV infection, resulting in dramatic increases in TB rates in sub-Saharan Africa [5]. In addition to efforts to enhance protective immunity by vaccination before infection, the size and elevated risk of the latent TB population indicate a crucial role for interventions that would reduce disease progression in individuals who have already been exposed to infection.

The focus on development of intervention strategies for latent TB highlights a need for a more careful definition of the biological condition, or conditions, that we are attempting to target. The presence of antigen-specific T cells is common to a spectrum of responses to *M. tuberculosis* that range from complete clearance of the pathogen, through containment in a benign but potentially virulent form, to progressive pre-clinical and clinical disease (Figure 2). The disease risk associated with latent TB is a combination of endogenous reactivation of an initial infection or failure to control progression of an exogenous reinfection: the relative importance of the two routes presumably depending on the risk of repeated exposure [56,57].

Treatment of latent TB with antitubercular drugs (‘preventive therapy’), most commonly daily isoniazid for 9 months, typically results in a risk reduction of around 50%, even in HIV-infected persons [58,59]. The duration of protection seems to be related to the risk of reinfection, being short in HIV-infected persons in high incidence environments but very long in the historic studies conducted in North America [59,60]. Amongst researchers aiming to develop improved treatment for latent TB, the results of these clinical trials prompt two interrelated questions. First, why does the treatment take so long? Treatment of latent TB involves elimination of a very much smaller bacterial population than treatment of clinical TB, and it might, therefore, be anticipated that this would require a much less extensive antibiotic regime. An attractive hypothesis is that the bacteria present in latent TB persist in a physiological state that renders them tolerant to the effect of drugs. Isoniazid acts by inhibiting mycobacterial cell wall synthesis and loses its efficacy when bacteria are not replicating. Second, why does it work at all? Two potential mechanisms could account for the efficacy of isoniazid in preventive therapy: non-replicating bacteria might have some residual requirement for low-level synthesis of cell wall precursors (for maintenance and repair, for example), or bacterial populations might cycle between replicating and non-replicating phases providing periodic windows of isoniazid sensitivity). The search for rational development of improved drugs for latent infection focuses on understanding the physiology of bacteria present in latent lesions, primarily using transcriptional profiling to identify patterns of gene expression by *M. tuberculosis* present in latent lesions obtained from lungs resected from patients with suspected lung cancer.

The search for improved drugs for preventive therapy, and the associated exploration of bacterial *in vivo* phenotypes, is part of a broader programme to identify rational strategies for shortening of treatment times for active TB. Although excretion of bacteria in sputum is rapidly reduced within days or weeks of initiation of treatment, a 6-month course is required to eliminate ‘persistent’ *M. tuberculosis* and prevent relapse of disease. It is believed that the persistence phenomenon observed during treatment of TB disease is also because of the presence of phenotypically drug-tolerant bacterial populations. However, the relationship (if any) of the bacteria that persist in the face of a primed immune response in latent TB and the bacteria that persist during chemotherapy of active TB is unknown. Maintenance in a non-replicating state presents an attractive hypothesis in both cases. A clearly formulated version of this hypothesis invokes an important role for non-replicating persistence in the hypoxic zone of granulomas that have become necrotic (known as caseous, i.e. cheese-like, necrosis), and substantial research efforts are currently focused on rigorous assessment of this hypoxia hypothesis [61]. Most animal models incompletely reproduce all the pathological features of human TB and latent TB is particularly hard to recreate. Exciting recent progress in this respect has been made in a primate model of TB. A wide spectrum of lesions characterize both active and inactive ‘latent’ infection in this model that closely recreates much of the pathology of human disease [62,63].

Consideration of whether a novel agent (or combinations of existing agents) might be effective in elimination of latent TB also poses a substantial problem in clinical evaluation. By definition, latent TB is asymptomatic and not characterised by gross radiographic abnormality. Clinical trials would need to be very large and prolonged unless conducted in a high incidence area, in which case reinfection might be an important confounder. Attention is, therefore, being paid to surrogates of potential efficacy, for example a change in the immune response [45]. Efforts to shorten treatment times for active disease face a similar challenge in measuring elimination of bacteria that persist after sterilisation of the population accessed in the sputum. Live imaging of lesions using a combination of high resolution computerised axial tomography (CT) and positron emission tomography (PET) scans is being explored in this context and might also have application for monitoring of novel interventions against latent infection.

The alternative to drug therapy would be to boost the natural immunity associated with TB infection by post-exposure vaccination. Although it is attractive to propose that boosting of the immune response will clear persisting infection, there is no clinical proof-of-concept, and very few examples of success in experimental models. In addition, inappropriate immune activation in the presence of infection clearly has the potential to exacerbate clinical symptoms [64]. The most optimistic hypothesis for vaccines against latent TB is to suggest that infection with *M. tuberculosis* subverts the natural immune response in a way that impairs its optimal efficacy. This could involve triggering signalling pathways involved in tolerance to commensal bacteria, for example, misleading the immune system as to the potential danger of the infection. Theoretically, vaccines and adjuvants that overcome this immune subversion, rather than the classical approach of reproducing natural infection, might provide an attractive approach to treatment of latent TB.
Opinion

Box 1. A wish list for future research on latent TB

- Identification of potential mechanisms of resistance to *M. tuberculosis* in the absence of T-cell priming
- Distinction of individuals who mount a stabilising immune response to initial infection
- Identification of latent TB groups at highest risk of progression to disease
- Understanding of the physiology of *M. tuberculosis* during human infection
- Improvement in imaging to determine the metabolic activity associated with ‘inactive’ or ‘old’ TB lesions
- Overall, an improved scientific platform for rational development of drugs and vaccines to reduce progression from asymptomatic infection to active disease

Refining ‘latent TB’ in the research lexicon?

Although the concept of latent TB, defined by the biomarker of antigen-specific T cells in the absence of clinical symptoms, is useful in a clinical context, it can be detrimental in the context of formulating a research agenda. Although the term is unlikely to fall into disuse, we would like to propose that it is subdivided as outlined in Figure 2. We do not envisage these as rigid boxes: individuals might well move back and forth between categories, and the concept of host-pathogen interactions mapped along a spectrum of outcomes seems attractive. A potential unifying theme across the spectrum could be cycling of bacteria between replicating and non-replicating states. The oscillation frequency might depend on the physiology of individual granulomas, but the net effect might be a relatively short frequency in active disease as compared to a long frequency in asymptomatic infection. Whether individual bacteria do oscillate between replicating states is conjectural, although recent advances in single cell analysis suggest this possibility (McKinney J: presentation to 7th International Congress on the pathogenesis of mycobacterial infections, Stockholm, June 2008). We anticipate that consideration of TB as a spectrum from immunity to disease will facilitate further research towards improved detection methods and treatments of latent TB (Box 1).

Acknowledgements

The ideas included in this review have been developed in the context of a consortium funded by the Bill and Melinda Gates Foundation and the Wellcome Trust to explore development of drugs for treatment of latent TB as part of the initiative to address Grand Challenges in Global Health. R.J.W also receives support from the European Union.

References

9 Committee, Tuberculosis Research (1932) Tuberculosis in South African natives with special reference to the disease amongst the mine labourers on the Witwatersrand. 430, South African Institute for Medical Research
23 Manca, C. et al. (1999) *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J. Immunol.* 162, 6740–6746
25 Loomis, H. (1890) Some facts in the etiology of tuberculosis, evidenced by thirty autopsies and experiments upon animals. *Medical Record* 38, 689–698
26 Opie, E. and Aronson, J. (1927) Tubercle bacilli in latent tuberculous lesions and lung tissue without tuberculous lesions. *Arch. Pathol.* 4, 1–21


