Donna Montgomery
is a senior research fellow at Merck Research Laboratories where she has worked on several vaccine programs, including DNA vaccines for influenza and TB.

Tuberculosis vaccine design: Influence of the completed genome sequence

Donna L. Montgomery

Date received (in revised form): 13th June 2000

Abstract

Tuberculosis continues to be a major health problem, with more adults dying from *Mycobacterium tuberculosis* than any other pathogen world-wide. With the onset of the HIV epidemic and an increase in drug-resistant *M. tuberculosis* strains, the need for an improved vaccine has become an international priority. The recent completion of the genome sequences for two *M. tuberculosis* strains provides a wealth of information that can be used to design new strategies for vaccine development. The challenge comes in making rational choices from among the 4,000 genes of the most probable candidate immunogens or virulence genes. Thus, a well-designed screen is needed to reduce the number of candidates that must be tested. Presently, the most valuable role that bioinformatics can play is to provide such a screen.

Keywords: Mycobacterium tuberculosis *immunology,* Mycobacterium tuberculosis *vaccines, bacterial genome,* Mycobacterium tuberculosis *genetics*

Dr Donna Montgomery,
Merck & Co, Virus and Cell Biology, P O Box 4 – WP26-250, West Point, PA 19486, USA

Tel: +1 (215) 652 4953
Fax: +1 (215) 652 7320
E-mail: donna_montgomery@merck.com

INTRODUCTION

Tuberculosis continues to be a major health problem, with more adults dying from *Mycobacterium tuberculosis* than any other pathogen world-wide.[1] Although there has been widespread use of the bacille Calmette–Guérin (BCG) vaccine since it was first used in humans in 1921, its efficacy has varied from 0 to 80 per cent in different populations. In all populations studied, BCG was found to be more protective against meningeal and miliary TB, which are mostly seen in the young, than against pulmonary disease, which is seen in older adults.[2] To date, there is no evidence that repeated vaccination with BCG throughout life gives a boosting effect.[3] With the onset of the HIV epidemic and an increase in drug-resistant *M. tuberculosis* strains, the need for an improved vaccine has become an international priority.

The recent completion of the genome sequences for *M. tuberculosis* strains H37Rv[4] and CDC1551 (Fleischmann-R, TIGR) provides a wealth of information that can be used to design new strategies for vaccine development. *Mycobacterium tuberculosis* has about 4,000 genes, with functions assigned or predicted for about 60–70 per cent of the open reading frames (ORFs). The sequences and annotations for H37Rv can be accessed from the Sanger Center web site[5] and the sequence and annotations for CD1551 at The Institute for Genomic Research web site.[6]
The challenge comes in making rational choices from among the 4,000 genes of the most likely candidates for vaccine development. Presently, bioinformatics can be used to limit the number of candidates that need to be tested. Programs exist that can identify protein motifs predicted to be necessary for immunogenicity or for virulence.

IDENTIFICATION OF VACCINE CANDIDATE GENES

As yet, there are few published reports of vaccine candidates identified by using genomic sequence data. One such report is the recent publication by Pizza *et al.*[7] describing a genomic approach to identifying vaccine candidates for serogroup B meningococcus. The type of immune response needed for protection from this bacterium is known: bactericidal antibodies correlate to

protection in humans.[8] Most surface antigens that had previously been tested as vaccine candidates were found to be species-specific,[9] and therefore failed in clinical trials.

A bioinformatics approach was used to find ORFs that potentially encode conserved surface-exposed or exported proteins. An initial screen using BLASTX identified ORFs with defined cytoplasmic functions and these were excluded from further analysis. The remaining ORFs were further screened by Φ-BLAST, FASTA, MOTIFS, FINDPATTERNS and PSORT in combination with the ProDom, Pfam and BLOCKS databases to predict features typical of gram (–) bacterial surface-associated proteins. These features included transmembrane domains, leader peptides, homologies to known surface proteins, lipoprotein signature, outer membrane anchoring motifs and host cell binding domains. The final screen identified 570 ORFs out of a total of 2,158 predicted coding regions.[10] Since a screen that was designed to include all possible candidates was used, about 25 per cent of the total ORFs predicted for the organism had still to be tested.

host immune response

Th1-type T-cell response

The ORFs were generated by polymerase chain reaction (PCR) and 61 per cent successfully cloned and expressed in *Escherichia coli*. Recombinant proteins were injected into mice and the resulting antibodies tested for bactericidal activity and for reactivity to the surface of a set of diverse capsulated *Neisseria meningitidis* strains. Seven genes were identified that elicited the desired antibody response. These seven genes were then cloned from 31 *N. meningitidis* strains to determine the amino acid conservation, and five were found to be highly conserved.

The report demonstrates that candidate antigens can be identified from genomic sequence with the help of bioinformatics. However, a highly selective bioinformatics screen was not used and 15 per cent of the predicted ORFs in serogroup B meningococcus had to be tested to find five candidates. These candidates look very promising in mice, but must be tested in a human population to prove their value.

IMMUNOLOGY OF TB

A similar approach can be taken for any pathogen that has been completely sequenced and for which the protective immune response(s) are well-defined humoral responses. Unfortunately, the actual immune responses needed for protection from primary *M. tuberculosis* infection and from reactivation of latent infections are still not well understood. The host immune response to *M. tuberculosis* infection is better understood than the vaccine responses needed for protection in human populations. *Mycobacterium tuberculosis* is an intracellular pathogen: primary infection involves the uptake of *M. tuberculosis* by macrophages where the bacteria are able to survive and replicate. Therefore, a cellular immune response should occur early in infection and it is generally believed that a Th1-type of T-cell response is necessary for protection.[11,12] The prevailing hypothesis that protective antigens will be found in the secreted or exported proteins of bacillus has been supported by observations in animal models.[13]

Further evidence that a Th1-type T-cell response is important for protection comes from a recent study of BCG intradermal vaccination in humans. Hoft *et al.*[14] were able to demonstrate that significant increases in *Mycobacterium*-specific T-cell proliferation, type 1 cytokine responses, and an antibody profile characteristic of a type-1 cytokine bias were induced by BCG over placebo. This result suggests that protection by BCG correlates to Th1-type T-cell responses. However, it does not address the type of response needed to improve protection over BCG.

NEW VACCINE APPROACHES

A number of approaches to new vaccines are being investigated, including subunit

challenge models difficult

protein and DNA vaccines, new recombinant BCG, auxotrophic mycobacteria and avirulent mycobacteria as vectors to express *M. tuberculosis* genes.[11] Testing candidate vaccines in animals is a challenge with TB. Candidates can be initially screened in animals by testing for type-1 T-cell responses, but should be followed by challenge models with clinically relevant strains of *M. tuberculosis* in at least two animal models: mice for general protection[15] and guinea pigs or rabbits for lung pathology.[16,17] Challenge models require BL3 (physical containment Biosafety Level 3 (BL3) as defined in the 1998 NIH Guidelines for Research involving Recombinant DNA molecules) facilities and need to be followed for months. Thus, it becomes imperative to choose vaccine candidates wisely, since the ability to test them is limited by facilities and time. The use of genome sequences and bioinformatics tools is one approach to identifying a limited number of genes to test.

cell surface functions

bioinformatics search example

secreted proteins

SUBUNIT VACCINES: EITHER PROTEIN OR DNA

Developing an effective subunit vaccine requires finding specific gene products that elicit the type of immune response needed for protection. The prevailing hypothesis is that protective antigens will be found in the secreted or exported proteins. Thus, identification of extracellular ORFs could be the first step. SignalP, PSORT, MOTIFS and FINDPATTERNS can be used to search for protein motifs such as signal sequences, transmembrane domains and membrane anchoring domains. This will provide a list of proteins that are secreted by a hydrophobic signal sequence mechanism, but will miss any secreted proteins that do not contain a classic leader sequence. For example, proteins such as ESAT-6, which has no classical leader sequence yet is a potent T-cell antigen,[18] will be missed. However, until more is known about the mechanism by which proteins without a canonical leader sequence are exported, their identification by bioinformatic approaches will be impossible.

identification of extracellular ORFs

A search for protein functions that may be associated with the cell surface can also be done. The web sites for the sequence of both H37Rv and CD1551 contain listings of genes according to proposed functions. For H37Rv, there are 360 ORFs listed under cell envelope function: 65 lipoproteins; 39 surface polysaccharides, lipopolysaccharides, proteins and antigens; 28 murien saculus and peptidoglycan; 17 conserved membrane proteins and 211 other membrane proteins. A similar listing can be found for CD1551.

An example of a search for secreted proteins was done several years ago (J. Yuan and D. Montgomery, unpublished data). SignalP was used to search the H37Rv genome sequence for ORFs that may contain a signal sequence of <50 amino acids. The cut-off length of the signal sequence was chosen because Ag85A, previously shown to be a good vaccine candidate,[19,20] has a leader sequence that is >40 amino acids.[21] Genemark was used to identify all putative ORFs that encoded proteins longer than 20 amino acids. This screen yielded 122 ORFs. A similar search of the CD1551 sequence, which was not completed at the time, produced an additional 11 ORFs that were not found in the H37Rv sequence. The screen could not identify proteins such as ESAT-6 which do not contain signal sequences. In addition, it was unable to identify Ag85A because the actual reading frame for 85A is much larger than the predicted ORF, so the actual signal sequence was not part of the predicted ORF. Undoubtedly, other extracellular proteins were missed since the final list of 122 ORFs was about half the number of extracellular proteins found in culture filtrates.[22,23]

MHC I and II search

The resulting list of candidate ORFs by either search method (secreted proteins or cell surface function) will produce a group of 100–350 candidates. This can be reduced further by searching for major histocompatibility complex (MHC) class I- and II-restricted epitopes. A set of algorithms for T-cell epitope prediction, OptiMer and EpiMer, were developed by Meister *et al.*[24] These algorithms predict motifs for both MHC class I- and II-restricted epitopes and identify clusters of these motifs as the most likely candidates for epitope location. OptiMer combines searches for secondary structural characteristics (amphipathicity) and published MHC-binding motifs to predict amphipathic, promiscuous putative T-cell epitopes that contain multiple MHC alleles. EpiMer searches only for published MHC-binding motifs and predicts putative promiscuous epitopes based solely on the clustering of MHC-binding motifs. Meister *et al.*[24] found both algorithms to have nearly equivalent levels of predictive success. De Groot *et al.*[25] tested EpiMer with four *M. tuberculosis* genes and found that all of the predicted T-cell epitopes were recognised by *M. tuberculosis*-infected people.

The accuracy of predicting epitopes is limited by the actual number of known epitopes used to determine motifs. Initially, OptiMer and EpiMer were designed using 15 distinct class II and 19 distinct class I motifs, but both algorithms can be modified to include new motifs as they are published. As more epitopes are defined by traditional research, the accuracy may improve. Identifying predicted epitopes is just the first step in finding proteins that will elicit Th1-type T-cell responses in human populations. There is no guarantee that predicted epitopes will be available for MHC binding *in vivo*. Even with these limitations, the clustering of promiscuous epitopes and the density of these clusters on specific genes can be used to rank ORFs as candidates. The ORFs with the highest density of clusters could be tested first.

Recently, Sturniolo *et al.*[26] published a matrix-based algorithm (TEPITOPE) to predict human MHC class II epitopes using virtual matrices representing the majority of human leucocyte antigen– (HLA-DR: DR is the allele of Human leukocyte Antigen and DR is the predominant isotype of the human MHC class II) peptide-binding specificity. They combined TEPITOPE with DNA microarray technology to find candidate promiscuous T-cell epitopes in genes preferentially expressed in human colon cancer tissue. A similar approach could be done with a microarray analysis of *M. tuberculosis* expression. The technology now exists to isolate small amounts of RNA from tissue, expand it by reverse transcriptase (RT)-PCR and label it with *in vitro* transcription reactions.[27] This approach would allow one to find the proteins that are preferentially expressed in macrophages during early infection and in lung tissue during disease manifestation or reactivation. These ORF sequences could be searched for T-cell epitopes.

IMPROVED BCG OR NEW ATTENUATED MYCOBACTERIAL VACCINES

recombinant attentuated vaccines

Expression of foreign genes in *M. smegmatis*, BCG and *M. tuberculosis* has been possible for about a decade. Both phage and plasmid vectors have been developed that allow stable expression of genes and gene disruption in these mycobacteria.[28] There has been a great deal of investigation into using BCG as a bacterial vaccine vector, expressing genes from HIV and bacteria.[29] These molecular biology tools can be used to build new bacterial vaccines: either by the addition of genes to BCG to improve immunogenicity or the deletion of virulence genes from *M. tuberculosis* to attenuate it.

BCG is effective in eliciting protective MHC class II-restricted responses, but it is unclear how effectively it elicits MHC class I-restricted cytotoxic T lymphocyte (CTL) responses. *In vitro* infection of macrophages with BCG finds the bacilli

only in the phagolysosomes,[30] suggesting BCG does not reach the cytoplasm and its proteins are not available for MHC class I-restricted responses. The major role of CD8 T cells in controlling *M. tuberculosis* infections has been well demonstrated in mice,[31,32] so the inability of BCG to elicit long-lived protection in humans may be due to an inefficient CD8 T-cell response. However, a recent paper by Smith *et al*.[33] reports finding BCG-reactive CD8+ T cells in healthy BCG-vaccinated donors, so this may not be the reason BCG does not confer long-lived protection in humans.

Hess and Kaufmann[34,35] have constructed a recombinant BCG that secretes biologically active listeriolysin (Hly) fusion protein of *Listeria monocytogenes* to improve the ability of BCG proteins to escape the phagolysosome and have shown that the recombinant BCG has improved ability to stimulate CD8+ T cells. They suggest the improved BCG strain could be tested as a vaccine candidate, or could be used as a vector for expression of *M. tuberculosis* protective antigens that are not expressed in BCG.

DNA vaccine

virulence genes

expression-library immunisation

Whole genome comparisons of BCG strains and virulent *M. tuberculosis* have recently been done by two groups to discover the genes responsible for virulence and to discover the genes that have been deleted in various BCG strains. Behr *et al*.[36] used whole-genome DNA microarray hybridisation to define the differences between *M. tuberculosis, M. bovis* and various BCG daughter strains. They were able to compile a BCG historical genealogy and define the six additional deletions that occurred during passage of the original BCG isolate. They were also able to define the 38 ORFs missing from some or all BCG strains but present in virulent *M. bovis,* and the 91 ORFs found in H37rv but missing from one or more virulent *M. bovis* strains. Gordon *et al*.[37] did a similar analysis that extended the comparison of *M. tuberculosis, M. bovis* and BCG-Pasteur to *M. africanum* and *M. microti*, thus allowing an analysis of genes that might affect host range and virulence. The information obtained from these whole-genome comparisons adds greatly to the knowledge of virulence genes, and provides a list of genes that can be tested as subunit vaccines, expressed in recombinant BCG, or deleted from *M. tuberculosis* to give an attenuated strain. The genes identified by the Behr *et al*. study are listed in the web site.[38]

PRE-CLINICAL TESTING OF NEW VACCINE CANDIDATES

Once the choice of new vaccine candidate genes has been made, they will need to be tested in animal models for Th1-type immune responses and for protection against challenge in mice and guinea pigs/rabbits. Recent advances in technology should speed the process of testing candidate ORFs. DNA vaccines have been shown to provide Th1 immune responses and provide protection in mice and guinea pigs for several secreted antigens.[19,20,39,40] Cocktails of at least three genes have been shown to give additive effects.[41] Thus, DNA immunisation is a good way to screen candidates in animals without having to express and purify recombinant proteins. A novel screening procedure termed expression-library immunisation (ELI) has shown the feasibility of screening pools of plasmids directly *in vivo*. Barry *et al*.[42] have shown that pools of plasmids can be used to immunise mice and screen genes directly in an *in vivo* challenge model. This approach should be especially fruitful if the pools are kept small by using bioinformatics to form a select list of ORFs.

SUMMARY

The recently completed genome sequence of two strains of *Mycobacterium tuberculosis* provides a wealth of information that is just beginning to be used to advance our basic knowledge of this important disease organism. DNA microarray studies are providing insights into the genes responsible for virulence,

host specificity and attenuation of BCG daughter strains. *Mycobacterium tuberculosis* genes can be generated by PCR and their function or immunogenicity tested in animals. However, testing vaccine candidates in animals requires BL3 facilities and is time consuming. The average screen in mice or guinea pigs takes about six months. Thus, a well-designed screen is needed to reduce the number of candidates that must be tested *in vivo*. At the moment the most valuable role that bioinformatics can play is to provide such a screen.

bioinformatics screens

Bioinformatics screens can be designed to chose ORFs that are potentially extracellular and available for cell-mediated immune responses during macrophage infection by *M. tuberculosis*. Sequence annotations that propose protein function can also be used to identify ORFs that have cell surface or virulence functions. Several programs exist that predict MHC class I and II epitopes, and these can be used for further screening of potentially extracellular ORFs. Once a good vaccine candidate gene has been identified, a genomic analysis could be done to find similar genes (multigene families) or to locate genes in close proximity on the genome (possible related function). The final proof that these screens are valuable will come when one results in a successful human vaccine.

bioinformatics limitations

There are many important aspects about vaccine design that cannot be approached by bioinformatics. Genetic differences in human populations that confer different susceptibility to disease or different responses to antigens are beyond the scope of existing bioinformatics. Prediction of protein expression during different phases of a disease cannot yet be done. Bioinformatics is limited by the amount of proven data that can be used to develop the predictive algorithms, and there are many aspects of disease and immune responses that are not well defined. The hope is that our basic knowledge in these areas will increase with the aid of new technologies and the genome sequence, and allow more powerful predictive programs to be developed.

Acknowledgement
I thank Jeffrey Yuan, Bioinformatics Department, Merck Research Laboratories, for critically reading the manuscript.

References

1. Sudre, P., ten Dam, G. and Kochi, A. (1992), 'Tuberculosis: a global overview of the situation today', *Bull. World Health Org.*, Vol. 70, pp. 149–159.
2. Rodrigues, L. C., Diwan, V. K. and Wheeler, J. G. (1993), 'Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis', *Int. J. Epidemiol.*, Vol. 22, pp. 1154–1158.
3. Bloom, B. R. and Fine, P. E. M. (1994), 'The BCG experience: implications for future vaccines against tuberculosis', in Bloom, B. R., Ed., 'Tuberculosis: Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 531–557.
4. Cole, S. T., Brosch, R., Parkhill, J. *et al.* (1998), 'Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence', *Nature*, Vol. 393, pp. 537–544.
5. Sanger Centre URL: http://www.sanger.ac.uk/Projects/M_tuberculosis/
6. Institute for Genomic Research URL: http://www.tigr.org/tdb/CMR/gmt/htmls/SplshPage.html
7. Pizza, M., Scarlato, V., Masignani, V. *et al.* (2000), 'Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing', *Science*, Vol. 287, pp. 1816–1820.
8. Maiden, M. C., Bygraves, J. A., Feil, E. *et al.* (1998), 'Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms', *Proc. Natl. Acad. Sci. USA*, Vol. 95, pp. 3140–3145.
9. Poolman, J. T. (1995), 'Development of a meningococcal vaccine', *Infect. Agents Dis.*, Vol. 4, pp. 13–28.
10. Tettelin, H., Saunders, N. J., Heidelberg, J. *et al.* (2000), 'Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58', *Science*, Vol. 287, pp. 1809–1815.
11. Orme, I. M. (1997), 'Progress in the development of new vaccines against tuberculosis', *Int. J. Tuberc. Lung Dis*, Vol. 1, pp. 95–100.
12. Chan, J. and Kaufmann, S. H. E. (1994), 'Immune mechanisms of protection', in Bloom, B. R., Ed., 'Tuberculosis:

Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 389–415.

13. Andersen, P. (1994), 'Effective vaccination of mice against *Mycobacterium tuberculosis* infection with a soluble mixture of secreted mycobacterial proteins', *Infect. Immunol.*, Vol. 62, pp. 2536–2544.
14. Hoft, D. F., Kemp, E. B., Marinaro, M. *et al.* (1999), 'A double-blind, placebo-controlled study of *Mycobacterium*-specific human immune responses induced by intradermal bacille Calmette–Guerin vaccination', *J. Lab. Clin. Med.*, Vol. 134, pp. 244–252.
15. Orme, I. M. and Collins, F. M. (1994), 'Mouse model of tuberculosis', in Bloom, B. R., Ed., 'Tuberculosis: Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 113–134.
16. McMurray, D. N. (1994), 'Guinea pig model of tuberculosis', in Bloom, B. R. Ed., 'Tuberculosis: Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 135–147.
17. Dannenberg, A. M. (1994), 'Rabbit model of tuberculosis', In Bloom, B. R., Ed., 'Tuberculosis: Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 149–156.
18. Sorensen, A. L., Nagai, S., Houen, G. *et al.* (1995), 'Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*', *Infect. Immunol.*, Vol. 63, pp. 1710–1717.
19. Huygen, K., Content, J., Denis, O. *et al.* (1996), 'Immunogenicity and protective efficacy of a tuberculosis DNA vaccine', *Nature Medicine*, Vol. 2, pp. 893–898.
20. Baldwin, S. L., D'Souza, C., Roberts, A. D. *et al.* (1998), 'Evaluation of new vaccines in the mouse and guinea pig model of tuberculosis', *Infect. Immunol.*, Vol. 66, pp. 2951–2959.
21. Borremans, M., De Wit, L., Volckaert, G. *et al.* (1989), 'Cloning, sequence determination, and expression of a 32-kilodalton-protein gene of *Mycobacterium tuberculosis*', *Infect. Immunol.*, Vol. 57, pp. 3123–3130.
22. Andersen, P., Askgaard, D., Ljungqvist, L. *et al.* (1991), 'Proteins released from *Mycobacterium tuberculosis* during growth', *Infect. Immunol.*, Vol. 59, pp. 1905–1910.
23. Sonnenberg, M. G. and Belisle, J. T. (1997), 'Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, *N*-terminal amino acid sequencing, and electrospray mass spectrometry', *Infect. Immunol.*, Vol. 65, pp. 4515–4524.
24. Meister, G. E., Roberts, C. G. P., Berzofsky, J. A. and DeGroot, A. (1995), 'Two novel T cell epitope prediction algorithms based on MHC-binding motifs: comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences', *Vaccine*, Vol. 13, pp. 581–591.
25. De Groot, A. S., Roberts, C. G. P., Edelson, B. T. *et al.* (1996), 'A novel algorithm for the efficient identification of T-cell epitopes: prediction and testing of candidate tuberculosis vaccine peptides in genetically diverse populations', in 'Vaccines 96', Cold Spring Harbor Laboratory Press, NY, pp. 127–134.
26. Sturniolo, T., Bono, E., Ding, J. *et al.* (1999), 'Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices', *Nature Biotechnol.*, Vol. 17, pp. 555–561.
27. Lockhart, D. J., Dong, H., Byrne, M. C. *et al.* (1996), 'Expression monitoring by hybridization to high-density oligonucleotide arrays', *Nat. Biotechnol.*, Vol. 14, pp. 1675–1680.
28. Jacobs, W. R. Jr and Bloom, B. R. (1994), 'Molecular genetic strategies for identifying virulence determinants of *Mycobacterium tuberculosis*', in Bloom, B. R., Ed., 'Tuberculosis: Pathogenesis, Protection and Control', ASM Press, Washington, DC, pp. 253–268.
29. Cirillo, J. D., Stover, C. K., Bloom, B. R. *et al.* (1995), 'Bacterial vaccine vectors and bacillus Calmette–Guérin', *Clin. Infect. Dis.*, Vol. 20, pp. 1001–1009.
30. McDonough, K. A., Kress, Y. and Bloom, B. R. (1993), 'Pathogenesis of tuberculosis: interaction of *Mycobacterium tuberculosis* with macrophages', *Infect. Immunol.*, Vol. 61, pp. 2763–2773.
31. Flynn, J. L., Goldstein, M. M., Triebold, K. J. *et al.* (1992), 'Major histocompatibility complex class I-restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection', *Proc. Natl Acad. Sci. USA*, Vol. 89, pp. 12013–12017.
32. Ladel, C., Daugelat, S. and Kaufmann, S. H. E. (1995), 'Immune response to *Mycobacterium bovis* bacille Calmette Guérin infection in major histocompatibility complex class I- and II-deficient knock-out mice: contribution of CD4 and CD8 T cells to acquired resistance', *Eur. J. Immunol.*, Vol. 25, pp. 377–384.
33. Smith, S. M., Malin, A. S., Lukey, P. *et al.* (1999), 'Characterization of human *Mycobacterium bovis* bacille Calmette–Guérin-reactive CD8+ T cells', *Infect. Immunol.*, Vol. 67, pp. 5223–5230.
34. Hess, J. and Kaufmann, S. H. E. (1999), 'Live antigen carriers as tools for improved anti-tuberculosis vaccines', *FEMS Immunol. Med. Micro.*, Vol. 23, pp. 165–173.

35. Kaufmann, S. H. E. and Hess, J. (1999), 'Impact of intracellular location of and antigen display by intracellular bacteria: implications for vaccine development', *Immunol. Lett.*, Vol. 65, pp. 81–84.

36. Behr, M. A., Wilson, M. A., Gill, W. P. *et al.* (1999), 'Comparative genomics of BCG vaccines by whole-genome DNA microarray', *Science*, Vol. 284, pp. 1520–1523.

37. Gordon, S. V., Brosch, R., Billault, A. *et al.* (1999), 'Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays', *Mol. Microbiol.*, Vol. 32, pp. 643–655.

38. http://molepi.stanford.edu/bcg/

39. Tanghe, A., Lefevre, P., Denis, O. *et al.* (1999), 'Immunogenicity and protective efficacy of tuberculosis DNA vaccines encoding putative phosphate transport receptors', *J. Immunol.*, Vol. 162, pp. 1113–1119.

40. Lowrie, D. B., Tascon, R. E., Bonato, V. L. *et al.* (1999), 'Therapy of tuberculosis in mice by DNA vaccination', *Nature*, Vol. 15, pp. 269–271.

41. Kamath, A. T., Feng, C. G., MacDonald, M. *et al.* (1999), 'Differential protective efficacy of DNA vaccines expressing secreted proteins of *Mycobacterium tuberculosis*', *Infect. Immunol.*, Vol. 67, pp. 1702–1707.

42. Barry, M. A., Lai, W. C. and Johnston, S. A. (1995), 'Protection against mycoplasma infection using expression-library immunization', *Nature*, Vol. 377, pp. 632-635.