REPORTS ON CURRENT ISSUES AND WORKSHOPS

Report of the Workshop on Genome, Transcriptome and Proteome of *M. leprae*: Application to Basic and Clinical Questions

Richard A. Slayden, Chairperson, Rapporteur

In the Transactions from the Fifteenth International Leprosy Congress Dr. Stewart Cole described the *Mycobacterium leprae* genome as follows, "Everything that we need to know about the leprosy bacillus, from its biology to its behavior, is encoded in its genome, the corresponding proteins offer great potential as reagents for use in diagnostic skin tests and may even be involved in neuropathy." The knowledge and insight provided by knowledge of the genome has ushered *M. leprae* research into the post-genomic era and has provided the research community with a basic mission. Many of the aspects of current research into the leprosy bacillus are addressing the basic issue of metabolism and its relationship to disease. Although a paradigm shift is beginning to occur about how we look at leprosy and the tools we use, it is clear that this organism presents some unique challenges. Even though we have the complete genome sequence, we still struggle to answer questions asked decades ago: what are the mechanisms of disease and nerve damage?; what are the mechanisms of protective immunity?; and what is the route of transmission?

Genomic data integration is a key issue in the design of a modern research platform. Biology is a complex concert of interactions among many members; post-genomic strategies allow researchers to take a global approach to viewing these complex interactions. This workshop dealt with the fundamentals of micro-spot assays and current technology and sample preparation, how *in vitro*-based analysis can prioritize scientific direction and how genomic-based strategies as applied to *M. leprae* research. The workshop agenda was to address and clarify 5 key aspects integral to genomic-based studies.

Fundamentals of micro-spot assays, current state of technology and its application: Although microarrays seem to be a recently developed technology, it is actually based on a theory developed 2–3 decades ago. Micro-spot assays are based on the percent occupancy principle. This principle states, "under ambient assay conditions [fractional occupancy] of binding sites is independent on the number of probes and dependent only on the analyte concentration." This is the basis for miniaturization into micrarray platforms. These platforms, although not completely developed for leprosy research, are beginning to be developed, firstly with DNA and secondly with proteins. With the development of these arrays come their applications. How these arrays could be applied to research in terms of genome comparisons and transcript profiling with DNA-based arrays was discussed, and how proteome profiling and markers for disease could be addressed with protein-based arrays.

Genomic and proteomic sample preparation: A major hurdle in using post genomic approaches is the ability to obtain the biological molecules of interest. *M. leprae* presents unique challenges in terms of obtaining RNA and proteins because of the restriction of growth in artificial media. Many advances have been made recently on the purification of both DNA and RNA from purified bacteria obtained from complex growth conditions. Many of these methods are based on the classical nucleic acid purification protocols and have been applied to viable *M. leprae* from the mouse foot pad model or Schwann cell cultures. Semi-quantitative and differential display RT-PCR studies have been performed on large

numbers of genes to determine whether or not they are expressed in these models of infection. Although these studies are limited by the technological approach, they are state of the art for defining the transcriptome of the bacillus during a disease process. The advances made in these studies will be the foundation for future studies using more advanced technologies and methods. A major hurdle not completely overcome is the difficulty analyzing the entire proteome of *M. leprae*. The primary issue here is contamination from host proteins and the loss of true secreted proteins. It will only be possible to study the latter category of proteins if the bacillus can be grown in defined artificial medium or the bioinformatic power for subtractive comparisons can be achieved.

***In silico*-based analysis approaches:** The complete genome sequence offers the possiblity of using computational power to analyze the biology of the bacillus. Although *in silico* analysis has its limitations, it can be used to prioritize scientific approaches and refine scientific questions. The method of analysis has the ability to quickly reveal the potential limitations of the coding capacity of an organism; from this a putative metabolic paradigm can be established. Currently *in silico* analysis is being done to address a number of scientific questions. One example is the identification of secreted proteins. This approach has been useful to identify a number of putative proteins that are likely to be secreted. This analysis was performed in such a way as to bias the proteins to *M. leprae* and to identify potential T-cell epitopes. For reasons discussed above the identification of large numbers of secreted proteins *in vivo* is very difficult if not impossible given our inability to grow the bacillus. Therefore, this is one area of leprosy research that could prove very useful. The long-term goal of this approach is to identify proteins involved in direct interaction with the host during disease. These proteins can them be exploited in terms of vaccine development, surrogate markers of disease and chemotherapeutic intervention.

Expression profiling and other genomic-based applications: The hallmark feature of post-genomic research is the ability to perform a global analysis. The expert panel of scientist assembled demonstrated that expression profiling had multiple applications, each of which can be applied to address uniquely different questions. Currently much effort is applied to characterizing the *M. leprae* genome and the role of the encoded protein. Although the global analysis of the coding capacity of *M. leprae* is at the initial stages it is clear that there are clear experimental differences that confirm the *in silico* data. These approaches are being used to determine the origin of *M. leprae* and its unique evolution into contemporary strains, genetic variability and molecular epidemiology.

Bioinformatics: With global analysis come vast amounts of data. To be useful this data must be analyzed into useable bits of information. Oftentime results vary widely from experiment to experiment using global approaches. This necessitates validation experiments to more clearly define the results obtained originally. In addition to data validation there are many analytical considerations. These considerations range from appropriate controls and comparisons to sample preparation and manipulation. In the end the final issue is data mining. Simply, how is the relevant information extracted from the large background of data generated? Bioinformatics is quickly evolving and perhaps future genomic experiments will be performed more precisely and the data analyzed more efficiently.

Fourteen experts presented under these five broad headings to achieve the Workshop aim of improving our understanding of the fundamentals of micro-spot assays and the complexity inherent to sample preparation; how to approach scientific questions more effectively with *in silico* analysis; understanding experimental techniques used to address specific expression profiling questions and to improve the ability to evaluate and validate post-genomic data.

The following people were participants and made this workshop valuable to the *M. leprae* research community by presenting in their field of expertise and are acknowledged for their contribution: Dr. Goulart, Dr. Phetsuksiri, Dr. Williams, Dr. Gillis, Dr. Dharmalingham, Dr. Alves, Dr. Tempone, Dr. Schaible, Dr. Khanolkhar-Young, Dr. Suffys, Dr. Matsuoka, Dr. Maeda and Dr. Britton.