Lepr Rev (1989) **60,** 89–93

Editorial

MOLECULAR BIOLOGY OF THE MYCOBACTERIA

The molecular biology of the mycobacteria is poised at the threshold of making major contributions to the understanding of the biochemistry and pathogenic mechanisms involved in mycobacterial infections. The application of molecular biology to the study of mycobacteria has recently begun, with preliminary studies on the nucleic acids of mycobacteria, cloning and expression of a number of mycobacterial genes and the development of mycobacteria themselves as gene cloning systems. In this review, we will discuss the progress that has been made so far and the likely direction of future work.

The nucleic acids of mycobacteria

Both DNA and RNA have been isolated from mycobacteria, including armadillo-grown *Mycobacterium leprae*. They belong to the high guanine plus cytosine (G+C) Gram-positive group of bacteria; the cultivable mycobacteria have G+C in the range 60–67%, while *M. leprae's* G+C content is somewhat lower, at 56%.[1] The genome size for *M. tuberculosis* is similar to that for *Escherichia coli* ($2{\cdot}5 \times 10^9$ M_r), while that for *M. leprae* is smaller ($1{\cdot}3$–$2{\cdot}2 \times 10^9$ M_r).[2]

Plasmids and phages have been isolated from cultivable mycobacteria, but not, probably for technical reasons, from *M. leprae*. There is a sugestion that plasmids isolated from members of the *M. avium, M. intracellulare* and *M. scrofulaceum* (MAIS) complex may be associated with antibiotic resistance[3] and virulence.[4] A plasmid carrying a gene associated with mercury resistance has been reported.[5] One plasmid (pAL5000) from *M. fortuitum* has been sequenced and is being developed as a vector for cloning DNA into mycobacteria (see below). Both lytic and temperate bacteriophages have been isolated from mycobacteria[6] and these are also being developed as cloning vectors.

In growing cells, more than 90% of extractable RNA is ribosomal RNA (rRNA) which is present in many thousands of copies.[7] The genes coding for rRNA are present in several copies in most eubacteria; thus *E. coli* has seven rRNA genes,[8] *Bacillus subtilis* has 10^9 and *Streptomyces lividans* has 6.[10] By contrast, slow-growing mycobacteria (*M. tuberculosis* and *M. lepraemurium*) have one such rRNA gene while rapid growers (*M. smegmatis* and *M. phlei*) have two.[11,12] Bacteria contain 3 species of ribosomal RNA (5S, 16S, and 23S). Sequence analysis of the 16S molecule has allowed preliminary taxonomic analysis which shows that *M. leprae* is clearly a mycobacterium and that mycobacteria conform to the Gram-positive pattern.[13,14]

Cloning of mycobacterial genes and expression of the gene products

One of the most powerful tools available to molecular biologists is the ability to take DNA from one type of cell and transfer it to another cell in which it can be replicated in limitless amounts, as part of

0305-7518/89/060089+05 $01.00

the host cell's own replication machinery. This construction of DNA 'libraries' is particularly important with difficult-to-grow organisms such as mycobacteria. Libraries of mycobacterial DNA have now been constructed in *E. coli* by a number of groups.[1,15–18] Ideally, libraries will be constructed such that the host cells, e.g. *E. coli*, contain segments of mycobacterial DNA, and that the entire mycobacterial genome is represented. Single *E. coli* cells which contain a particular mycobacterial gene of interest can then be grown up ('cloned') in large amounts, and the nucleic acid of that gene studied in detail or the corresponding gene product expressed.

Preliminary experiments indicated that mycobacterial promoters (sequences of DNA which are responsible for controlling the expression of genes) do not function very efficiently in *E. coli*.[1,19,20] Thus while the genes can be cloned they are not expressed (they do not produce the corresponding protein product). One way of overcoming this problem is to place the mycobacterial gene under the control of a foreign promoter which is known to work efficiently in *E. coli*. Such an approach has been successfully used to express the *M. leprae* gene which codes for citrate synthetase;[21] here the mycobacterial gene is expressed in *E. coli* under the control of a streptococcal promoter!

In what now has become a classic experiment, Young *et al.*[17] used a somewhat different approach. They took small fragments of mycobacterial DNA (*M. leprae* and *M. tuberculosis*) and inserted them into the middle of an *E. coli* gene (that encoding for the enzyme β-galactosidase). When this *E. coli* gene is switched on a hybrid (or fusion) protein is formed, consisting partly of a mycobacterial protein and partly of the *E. coli* protein. The 'vector' which is used to transfer the DNA into *E. coli* is a bacteriophage called λgt11. The λgt11 expression system does not usually allow for the high level production of the foreign protein, but once the gene has been identified and cloned in this way, other techniques can then be used to engineer *E. coli* (or other organisms) to produce large amounts of mycobacterial protein.

Expression of recombinant mycobacterial antigens and 'the heat-shock protein story'

Studies with monoclonal antibodies which are specific for mycobacterial proteins had identified a limited number of proteins (five in *M. leprae*) which were apparently 'immunodominant'. With the construction of libraries of mycobacterial DNA it then became possible to use these monoclonal antibodies to try to detect expression of the corresponding proteins in *E. coli*. Using the λgt11 approach described above, Young *et al*[17] were able to detect expression of the five *M. leprae* proteins which had been identified using monoclonal antibodies (in most, but not all cases, the *M. leprae* proteins were expressed, as fusions with β-galactosidase). To our knowledge the genes encoding three of these proteins, those with molecular weights of 65kDa,[22] 18kDa[23] and 28kDa[24] have been sequenced and the corresponding amino acid sequences of the proteins deduced. Recombinant mycobacterial proteins are becoming available in relatively large quantities,[25,26] and their immunological characteristics studied in detail. It now seems likely that the small number of proteins which have been concentrated upon does not adequately represent the immunogenicity of *M. leprae*; all of the original monoclonal antibodies were produced by immunizing one particular inbred strain of mouse. From work with *M. tuberculosis*[27] it appears that different mouse strains 'see' proteins other than the five originally described. For this reason a number of groups are attempting to identify important antigens using T cells rather than antibodies.[28,29]

One interesting aspect of the work in which immunodominant antigens have been studied using a molecular approach is that several of the mycobacterial antigens appear to be related to 'heat-shock' proteins found in other cells.[30] Heat-shock proteins are proteins whose production is increased when the cells are subjected to a sudden increase in temperature (or other forms of stress). It is thought that this response is a means by which cells protect themselves against stressful changes in environment. It has also become clear that stress proteins have been identified as being important in the immune response to other infectious agents,[31,32] and this has led to the suggestion that the process of infection stresses the infectious agents and leads to the overproduction of these proteins; hence their immunodominance.

One important outcome of the heat-shock protein story is that antibodies and T cells which recognize mycobacterial antigens (particularly the 65kDa protein, which is a major heat-shock protein) have been found in patients with apparently autoimmune phenomena. For example, T cells reactive to mycobacterial 65kDa protein are present in the synovial fluid of patients with active rheumatoid arthritis,[33,34] and antibodies to this protein are readily detected in such patients. One possible explanation for this is that there is a high degree of sequence conservation between stress proteins from different species, i.e. humans possess, for example, a protein equivalent to the mycobacterial 65kDa protein and which is very similar in terms of its amino acid sequence. Thus, if immune responses are directed towards parts of the protein which are similar in man and bacteria it could lead to an autoimmune response.

Molecular genetics of mycobacteria

Although the most obvious application of molecular biology is the production of recombinant proteins, perhaps the most exciting future applications will lie in the ability to manipulate genes in the mycobacteria themselves. Being able to delete, transfer or add genes to mycobacteria will enable a whole new approach to studies of virulence, pathogenicity, protective immunity and chemotherapy to be developed.

In order for such studies to become feasible it will be necessary to develop efficient methods for getting DNA into mycobacteria and for introducing genes into the mycobacterial chromosome. Jacobs *et al*[35] developed a vector for introducing DNA into mycobacteria by constructing a hybrid of an *E. coli* plasmid and a mycobacterial phage; this 'phasmid' was capable of growing as a phage in mycobacteria and as a plasmid in *E. coli*. The problem with this approach was that the phage lysed its mycobacterial host, whereas ideally one would wish to have the introduced DNA expressed in growing bacteria. This is being overcome by the use of lysogenic mycobacteriophages (phages which do not normally lyse their host, but whose DNA becomes stably integrated into the host chromosome).[36] An alternative approach has been to construct a hybrid plasmid from an *E. coli* plasmid and pAL 5000;[36] this hybrid can be introduced into mycobacteria by electroporation, and is reproduced. It is likely that other methods for stably introducing genes into mycobacteria, and for deleting specific genes will be developed in the near future.

Future applications of molecular biology in the study of the mycobacteria

The analysis of mycobacterial antigens at the molecular level has begun, but there is still much to be done. In addition, only two of the *M. leprae* recombinant proteins have been produced in sufficient quantities to carry out detailed studies on their immunogenicity. As more of these materials become available it will become possible to study their role in protective immunity. In addition, synthetic peptides based on known protein sequences could be developed as immunodiagnostic or immunoepidemiological tools. Enzymes which are potential targets for drugs could be produced by cloning and expressing the corresponding genes, thus enabling *in vitro* screens to be developed. The ability to transfer such genes to other mycobacteria will enable the effect of the permeability barrier posed by the mycobacterial cell wall to be investigated.

New methods for detecting small sequences of nucleic acids might enable us to detect the presence of *M. leprae* in tissue with a much greater degree of sensitivity than is currently possible. For example, a technique called the 'polymerase chain reaction' (PCR) involves the amplification of a single stretch of nucleic acid sequence to several thousand copies, which can then be readily detected; thus, in theory, it should be possible to detect the presence of a single bacillus. Information about nucleic acid sequences will enable much more rapid methods for species identification; for example the identification of species specific rRNA sequences has been used to rapidly differentiate

between *M. tuberculosis* and *M. avium intracellulare*.[37] Perhaps by using the PCR technique to amplify *M. leprae* specific rRNA sequences, it might be possible to detect strain differences between different isolates of *M. leprae*.

The use of molecular genetic techniques has been of great importance in understanding the basic mechanisms of pathogenicity with a number of bacterial pathogens. For example, the mechanisms of entry of *Yersinia pestis* into host cells has been defined at the protein level,[38] and genes which are responsible for virulence in a number of bacteria have been found to be closely associated. The ability to transfer genes between mycobacteria and to add or delete specific mycobacterial genes is likely to be of great importance in understanding the basic mechanisms by which mycobacteria invade and survive within their host cells.

Acknowledgment

We acknowledge the contribution of many of our colleagues, particularly the members of the Molecular Biology Subcommittee of IMMLEP and THELEP, who have discussed these issues with us on many occasions.

M J Colston and F I Lamb

National Institute for Medical Research
London NW7 1AA

References

[1] Clark-Curtiss JE, Jacobs WR, Docherty MA, Ritchie LR, Curtiss R. Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. *J Bacteriol*, 1985; **161:** 1093–102.
[2] Imaeda T, Kirchheimer WF, Barksdale L. DNA isolated from *Mycobacterium leprae* with other related bacteria as determined by optical DNA-DNA reassociation. *J Bacteriol*, 1982; **150:** 414.
[3] Franzblau SG, Takeda T, Nakamura M. Mycobacterial plasmids: screening and possible relationship to antibiotic resistance in *Mycobacterium avium/Mycobacterium intracellulare*. *Microbiol Immunol*, 1986; **30:** 903–7.
[4] Crawford JT, Bates JH. Analysis of plasmids in *M. avium intracellulare* isolates from persons with AIDS. *Am Rev Respir Dis,* 1986; **134:** 659–61.
[5] Meissner PS, Falkinham JO. Plasmid-encoded mercuric reductase in *Mycobacterium scrofulaceum*. *J Bacteriol*, 1984; **157:** 669–72.
[6] Grange JM. The genetics of mycobacteria and mycobacteriophages. In *The biology of the mycobacteria*, Vol. 1. Ratledge C, Stanford J, (eds), New York: Academic Press, 1982.
[7] Noller HF, Van Knippenberg PH. Structure and function of ribosomal RNA. In *Genes: Structure and expression*. Kroon AM. (ed), John Wiley and Sons, 1983.
[8] Kiss A, Sain B, Venetianer P. The number of rRNA genes in *Escherichia coli FEBS Lett*, 1977; **79:** 77–9.
[9] Kobayashi H, Osawa S. The number of 5S rRNA genes in *Bacillus subtilis*. *FEBS Lett*, 1982; **141:** 161–3.
[10] Baylis HA, Bibb JM. Organisation of the ribosomal RNA genes in Streptomyces coelicolor A3(2). *Mol. Gen Genet*, 1988; **211:** 191–6.
[11] Bercovier H, Kafri O, Sela S. Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochemical and Biophysical Research Communications,* 1986; **136:** 1136–41.
[12] Yasuhiko Suzuki, Tatsuo Mori, Yasuko Miyata, Takeshi Yamada. The number of ribosomal RNA genes in *Mycobacterium lepraemurium*. *FEMS Microbiology Letters*, 1987; **44:** 73–6.
[13] Smida J, Kazdar J, Stackebrandt E. Molecular genetic evidence for the relationship of *Mycobacterium leprae* to slow-growing pathogenic mycobacteria. *Int J Lepr*, 1988; **56:** 449–54.
[14] Estrada-G ICE, Colston MJ, Cox RA. Determination and evolutionary significance of nucleotide sequences near the 3′ end of 16S ribosomal RNA of mycobacteria. Submitted.
[15] Thole JER, Dauwerse HG, Das PK, Groothuis DG, Schouls LM, Van Embden JDA. Cloning of *Mycobacterium bovis* BCG DNA and expression of antigens in *Escherichia coli*. *Infec Immun*, 1985; **50:** 800–6.
[16] Khandekar PS, Munshi A, Sinha S, Sharma G, Kapoor A, Gaur A, Talwar GP. Construction of genomic libraries of mycobacterial origin: identification of recombinants encoding mycobacterial-specific proteins. *Int. J Lepr,* 1986; **54:** 416–22.

[17] Young RA, Mehra V, Sweetser D, Buchanan T, Clark-Curtiss J, Davis RW, Bloom BR. Genes for the major protein antigens of the leprosy parasite *Mycobacterium leprae. Nature*, 1985; **316:** 1 August 450–2.

[18] Husson RN, Young RA. Genes for the major protein antigens of *Mycobacterium tuberculosis*: the etiological agents of tuberculosis and leprosy share an immunodominant antigen. *Proc Natl Acad Sci USA,* 1987; **84:** 1679–83.

[19] Kieser T, Moss MT, Dale JW, Hopwood DA. Cloning and expression of *Mycobacterium bovis* BCG DNA in *Streptomyces lividans. J Bacteriol*, 1987; **168:** 72–80.

[20] Lamb FI, Colston MJ. Cloning of *Mycobacterium leprae* genes in *Streptomyces. Lepr Rev,* 1986; **57 (Suppl 2):** 225–9.

[21] Jacobs WR, Docherty MA, Curtiss R, Clark-Curtiss JE. Expression of *Mycobacterium leprae* genes from a Streptococcus mutans promoter in *Escherichia coli* K-12. *Proc Natl Acad Sci*, 1986; **83:** 1926–30.

[22] Shinnick TM, Sweetser D, Thole J, Van Embden J, Young RA. The etiologic agents of leprosy and tuberculosis share an immunoreactive protein antigen with the vaccine strain *Mycobacterium bovis* BCG. *Infec Immun*, 1987; **55:** 1932–5.

[23] Booth RJ, Harris DP, Love JM, Watson JD. Antigenic proteins of *Mycobacterium leprae*. Complete sequence of the gene for the 18-kDa protein. *J Immunol*, 1988; **140(2):** 597–601.

[24] Thangaraj H, Colston MJ, Lamb FI. (unpublished data).

[25] Thole JER, Keulen WJ, Kolk AH, Groothuis DG, Berwald LG, Tiesjema RH, Van Embden JDA. Characterisation, sequence determination and immunogeneicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli* K-12. *Infect Immun*, 1987; **55:** 1466–75.

[26] Lamb FI, Kingston AE, Estrada-GI, Colston MJ. Heterologous expression of the 65kDal antigen of *Mycobacterium leprae* and murine T-cell response to the gene product. *Infect Immun,* 1988; **56:** 1237–41.

[27] Ljungqvist L, Worsaae A, Heron I. Antibody responses against *Mycobacterium tuberculosis* in 11 strains on inbred mice; novel monoclonal antibody specificities generated by fusions using spleens from BALB.B10 and CBA/J mice. *Infect Immun,* 1988; **56:** 1994–8.

[28] Lamb JR, Rees ADM. Antigen specificity and function of human T lymphocyte clones reactive with mycobacteria. *Br Med Bulletin,* 1988; **44:** 600–10.

[29] Mustafa AS, Oftung F, Deggerdal A, Gill HK, Young RA, Godal T. Gene isolation with human T lymphocyte probes. Isolation of a gene that expresses an epitope recognised by T cells specific for *Mycobacterium bovis* BCG and pathogenic mycobacteria. *J Immunol*, 1988; **141:** 2729–33.

[30] Young D, Lathigra R, Hendrix R, Sweetser D, Young RA. Stress proteins are immune targets in leprosy and tuberculosis. *Proc Natl Acad Sci USA*, 1988; **85:** 4267–70.

[31] Vodkin MH, Williams JC. A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *E. coli*. *J Bacteriol,* 1988; **170:** 1227–34.

[32] Hedstrom R, Culpepper J, Schinski V, Agabian N, Newport G. Schistosome heat-shock proteins are immunogologically distinct host-like antigens. *Mol. Biochem Parasitol,* 1988; **29:** 275–82.

[33] Res PCM, Schaar CG, Breedveld FC, Van Eden W, Van Embden JDA, Cohen IR, de Vries RRP. Synovial fluid T cell reactivity against the 65kDa heat shock protein of mycobacteria in early onset of chronic arthritis. *Lancet*, 1988; **ii:** 478–80.

[34] Gaston JSH, Life P, Bailey L, Bacon PA. Synovial T cell responses to a 65kDa mycobacterial protein in inflammatory arthritis. In press.

[35] Jacobs Jr WR, Tuckman M, Bloom BR. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature*, 1987; **327:** 532–5.

[36] Snapper SB, Lugosi L, Jekkel A, Melton RE, Kieser T, Bloom BR, Jacobs WR. Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc Natl Acad Sci USA,* 1988; **85:** 6987–91.

[37] Musial CE, Tice LS, Stockman L, Roberts GD. Identification of mycobacteria from culture by using the Gen-probe rapid diagnostic system for *Mycobacterium avium* complex and *Mycobacterium tuberculosis* complex. *J Clin Microbiol*, 1988; **26:** 2120–3.

[38] Isberg RR, Voorhis DL, Falkow S. Identification of invasin: A protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell*, **50:** 769–78.