Mol Gen Genet (1991) 228:385–392
0026892591002577

MGG

The *Mycobacterium tuberculosis* shikimate pathway genes: Evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases

Thomas Garbe[1], Spiros Servos[2,4], Alastair Hawkins[3], George Dimitriadis[4], Douglas Young[1], Gordon Dougan[2], and Ian Charles[2]

[1] MRC Tuberculosis and Related Infections Unit, RPMS, Hammersmith Hospital, Ducane Road, London W12 0HS, UK
[2] Department of Cell Biology, Wellcome Research Laboratories, Langley Park, Beckenham, Kent BR3 3BS, UK
[3] Department of Biochemistry and Genetics, Catherine Cookson Building, Medical School, University of Newcastle upon Tyne, NE2 4HH, UK
[4] Department of Biology, University of Patras, 26500 Patras, Greece

Received March 8, 1991

Summary. The *Mycobacterium tuberculosis* shikimate pathway genes designated *aroB* and *aroQ* encoding 3-dehydroquinate synthase and 3-dehydroquinase, respectively were isolated by molecular cloning and their nucleotide sequences determined. The deduced dehydroquinate synthase amino acid sequence from *M. tuberculosis* showed high similarity to those of equivalent enzymes from prokaryotes and filamentous fungi. Surprisingly, the deduced *M. tuberculosis* 3-dehydroquinase amino acid sequence showed no similarity to other characterised prokaryotic biosynthetic 3-dehydroquinases (bDHQases). A high degree of similarity was observed, however, to the fungal catabolic 3-dehydroquinases (cDHQases) which are active in the quinic acid utilisation pathway and are isozymes of the fungal bDHQases. This finding indicates a common ancestral origin for genes encoding the catabolic dehydroquinases of fungi and the biosynthetic dehydroquinases present in some prokaryotes. Deletion of genes encoding shikimate pathway enzymes represents a possible approach to generation of rationally attenuated strains of *M. tuberculosis* for use as live vaccines.

Key words: *aro* mutants – Dehydroquinate synthase – 3-Dehydroquinase – *Mycobacterium tuberculosis* – Quinic acid utilisation – Shikimate pathway

Introduction

Although effective chemotherapeutic agents have been developed, *Mycobacterium tuberculosis* (the etiologic agent of tuberculosis) continues to contribute to human morbidity and mortality in many parts of the world (Styblo 1989). In addition to chemotherapy, prophylactic vaccination against mycobacterial infection using BCG (*Mycobacterium bovis, B*acillus *C*almette-*G*uerin) has provided protection in some communities, but has failed in others (Fine 1988). In recent years considerable effort has been directed towards the identification and cloning of individual mycobacterial antigens involved in interactions with the immune system, with a view to developing "subunit" vaccines against mycobacterial disease (Engers et al. 1985, 1986; Young et al. 1985a, b; Young 1988). Construction of mutant strains of mycobacteria that have been rationally attenuated for virulence can be proposed as an alternative approach to development of novel vaccines, and the recent development of techniques for the genetic manipulation of mycobacteria (Snapper et al. 1988; Husson et al. 1990; Martin et al. 1990) raises the possibility of using modern genetic tools to accomplish such a goal. Mycobacterial pathogens resemble certain *Salmonella* spp. in their ability to replicate within cells of the host reticuloendothelial system and in the requirement for live vaccines to elicit optimal protective responses in animal models of disease (Collins 1974). *Salmonella* mutants carrying defects in genes encoding enzymes in the shikimate pathway are highly attenuated in vivo and are of potential use as live vaccines (Dougan et al. 1987; O'Callaghan et al. 1988). By analogy, construction of mycobacterial mutants with similar genetic defects represents a possible approach to development of live mycobacterial vaccines.

The shikimate pathway leading to biosynthesis of aromatic compounds (Fig. 1) is present in bacteria, fungi and plant cells, but is absent from mammalian cells. In addition, some fungi have a related pathway, the quinic acid utilisation pathway (Fig. 1), which catabolises quinate to protocatechuate thus rendering quinic acid available as a nutrient source. Interestingly, although the interconversion of dehydroquinic acid and dehydroshikimic acid is an essential step common to both the biosynthetic and catabolic pathways, it is catalysed by two quite distinct 3-dehydroquinase (DHQase) enzymes. The form of the enzyme used in the biosynthetic pathway (bDHQase) is one of five consecutive shikimate pathway steps catalysed by the arom polypeptide encoded by the complex *AROM* locus in *Aspergillus nidulans*, and is related to the dehydroquinase encoded

Offprint requests to: I. Charles

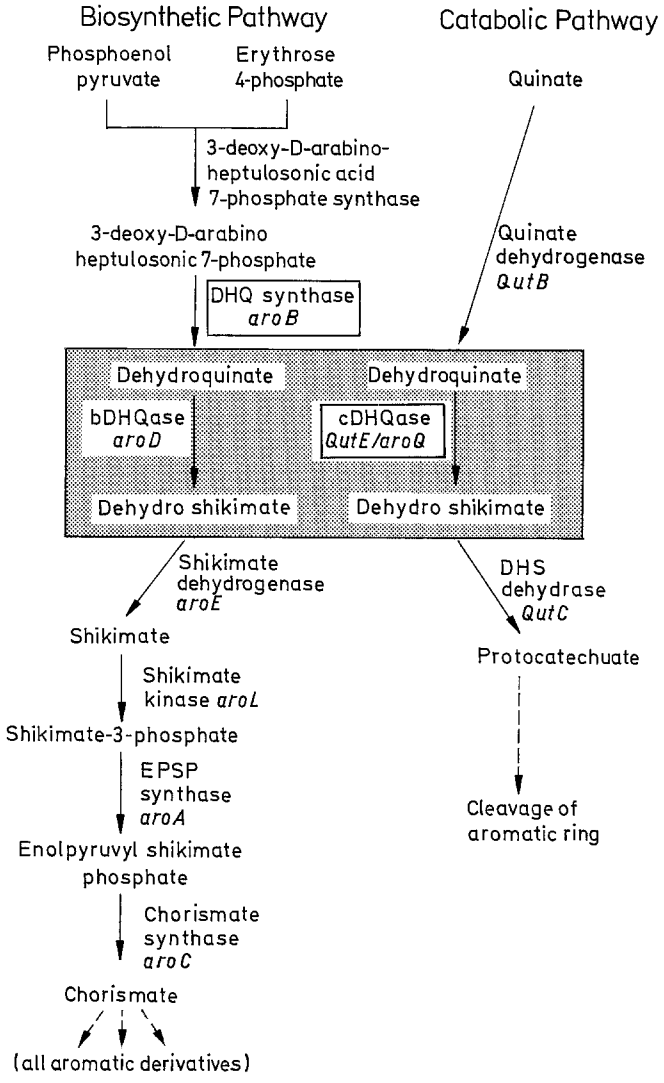

Fig. 1. Schematic comparison of the shikimate biosynthesis and the quinate utilisation pathways. The biosynthetic shikimate pathway is present in bacteria, fungi and plant cells, while the catabolic quinate utilisation pathway has been characterised in filamentous fungi. The 3-dehydroquinase (DHQase) reaction shown in the *shaded box* is common to both pathways but is catalysed by two different forms of the DHQase enzyme. *Open boxes* indicate the enzymes encoded by the *Mycobacterium tuberculosis* genes described here. Gene designations are according to the *Escherichia coli* genes for the biosynthetic pathway (Pittard and Wallace 1966; DeFeyter et al. 1986) and according to *Aspergillus nidulans* (Grant et al. 1988) for the catabolic pathway

by the *Escherichia coli aroD* gene (Charles et al. 1986; Hawkins 1987). The catabolic enzyme (cDHQase) on the other hand is encoded by the single function *Qut*E gene, one of seven genes in the quinic acid utilisation (*Qut*) gene cluster (Kinghorn and Hawkins 1982; Grant et al. 1988). Sequence analysis of cDHQase shows no relationship with the *E. coli* enzyme, and it has been proposed that the two dehydroquinase isozymes evolved by convergent evolution (Charles et al. 1985; Da Silva et al. 1986; Hawkins 1987; Hawkins and Roberts 1989; Beri et al. 1990).

As an initial step in the development of *M. tuberculosis* shikimate pathway mutants by rational gene deletion, we report here the cloning and DNA sequence of the *M. tuberculosis* genes encoding dehydroquinate synthase and 3-dehydroquinase enzymes. Analysis of the deduced protein sequences demonstrates that, whereas the mycobacterial dehydroquinate synthase is closely related to the equivalent fungal and prokaryotic shikimate pathway enzyme, the 3-dehydroquinase gene of *M. tuberculosis* is related in sequence to the *Qut*E gene of *A. nidulans* encoding the enzyme used in the catabolic pathway, and is unrelated to the previously described biosynthetic enzymes.

Materials and methods

Bacterial strains and plasmids. The *E. coli* strain GLW38 (*aroB*$^-$) was the kind gift of Dr. I. Hunter, Institute of Genetics, Glasgow University. The properties of the *aroD* mutant, *E. coli* strain AB2827, were described by Pittard and Wallace (1966). The genotypes and origins of other bacterial strains and cloning vectors, and of the *M. tuberculosis* λgt11 genomic DNA library have been described in detail previously (Garbe et al. 1990; Hawkins and Smith 1990).

Media and reagents. Media for the growth of phage lambda, *M. tuberculosis* and *E. coli*, and sources of materials were those described previously (Garbe et al. 1990). Di-deoxy chain terminator sequencing using [α^{35}S]dATP and buffer gradient gels was as previously described, using bacteriophage T7 polymerase, universal primer and specific oligonucleotides to overlap and complete the sequence on both strands (Sanger et al. 1977; Biggin et al. 1983; Charles et al. 1985; Tabor and Richardson 1987). Southern blot analysis was carried out as described by Garbe et al. (1990). Mini-cell preparation and SDS-polyacrylamide gel electrophoresis (SDS-PAGE), lysogenisation and complementation of *E. coli aro* mutant strains were as previously described (Dougan and Sherratt 1977; Dougan et al. 1987). Preparation of cell-free extracts and 3-dehydroquinase enzyme assays were carried out as described by Hawkins et al. (1984) and Hawkins and Smith (1991).

Results and discussion

Isolation of recombinant λgt11 clones able to complement an E. coli aro*D mutant*

As part of a programme to produce rationally attenuated strains of *M. tuberculosis* for use as candidate vaccines we wished to isolate the gene encoding the shikimate pathway bDHQase enzyme. A gene library of sheared *M. tuberculosis* DNA fragments in the cloning vector λgt11 (Young et al. 1985a) was used to produce lysogens of *E. coli* strain AB2827 (*aroD*$^-$; lacking bDHQase) that, in contrast to the original mutant, could grow on minimal medium without aromatic amino acid supplementation. Seven independently isolated lysogens

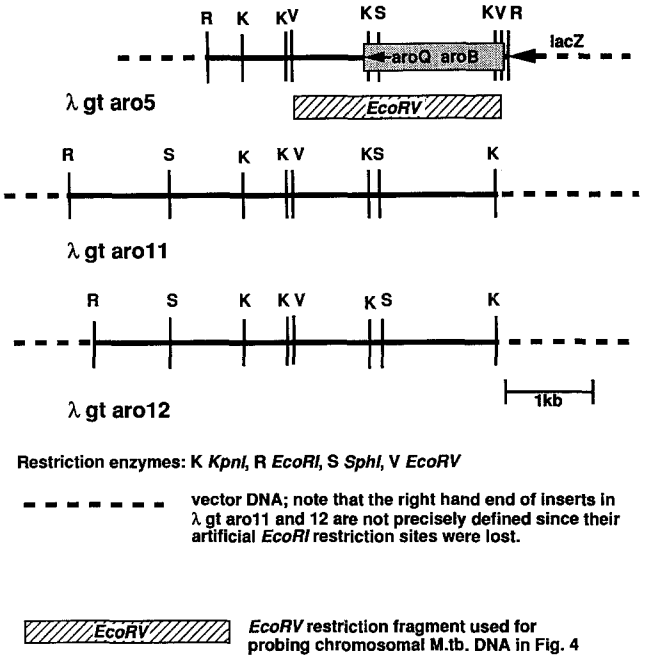

Fig. 2. Restriction map of insert DNA from recombinant λ phage able to complement an *Escherichia coli aroD* mutant. Insert DNA from seven complementing phage clones was analysed by restriction endonucleolytic digest. Three distinct patterns were observed with overlapping framents as shown: λgt aro5 was unique; two clones showed the λgt aro11 pattern; four clones showed the λgt aro12 pattern. Restriction enzyme abbreviations: K, *Kpn*I; R, *Eco*RI; S, *Sph*I; V, *Eco*RV. *Broken line*, vector DNA; note that the right-hand end of the inserts in λgt aro11 and 12 is not precisely defined since their artificial *Eco*RI restriction sites were lost. *Hatched bar*, *Eco*RV restriction fragment used for probing *Mycobacterium tuberculosis* chromosomal DNA in Fig. 4

were identified by this procedure and, upon induction, were found to contain recombinant λ phage that had three distinct restriction patterns but contained a single overlapping region of insert DNA (Fig. 2). One recombinant λ clone, designated λgt aro5, that contained a 3.4 kb *Eco*RI fragment of *M. tuberculosis* DNA was chosen for further detailed study. When the insert from λgt aro5 was subcloned into the *Eco*RI site of pBR322 efficient complementation of the *aroD* mutant was observed only with the gene in one orientation. As was found for the *M. tuberculosis aroA* gene (Garbe et al. 1990) therefore, optimal complementation of the *aroD* mutant is dependent on expression of a mycobacterial gene from exogenous promoter sequences provided by the vector rather than on recognition of endogenous mycobacterial sequences.

DNA sequence analysis

The nucleotide sequence of the entire *M. tuberculosis* DNA fragment in λgt aro5 was determined on both strands using suitable restriction fragments subcloned into M13 vectors as templates. A 1.8 kb subfragment of this sequence, starting at the artificial *Eco*RI recognition site, is shown in Fig. 3. Computer-aided analysis of this sequence leads to the following conclusions. (i) a DNA sequence extending from nucleotide 1185 to 1628 encodes the protein sequence for a dehydroquinase enzyme that is very similar to the cDHQase of filamentous fungi (Da Silva et al. 1986). The deduced protein sequence shows no significant similarity with the fungal or prokaryotic shikimate pathway bDHQase enzymes (Hawkins 1987). (ii) A DNA sequence extending from nucleotide 100 to 1188 encodes the protein sequence for a dehydroquinate synthase enzyme that shows close similarity to the equivalent enzyme in the shikimate pathway of filamentous fungi and prokaryotes (Charles et al. 1986; Millar and Coggins 1986). Although the two functions are present on a single polypeptide in fungi, close linkage between genes encoding dehydroquinate synthase and dehydroquinase enzymes (catalysing sequential steps in the shikimate pathway) has not been found in other prokaryotes. Southern blot analysis of chromosomal DNA from *M. tuberculosis* using the labelled *Eco*RV restriction fragment from λgt aro5 as probe (Fig. 4) did not identify any restriction fragments overlapping with those identified using the *aroA* gene (Garbe et al. 1990). It is possible that some of the genes encoding the shikimate pathway of *M. tuberculosis* are present as an "*aro* cluster" as has been described in *Bacillus subtilis* (Hoch and Nester 1973), but these results suggest that the *aroA* gene is excluded from the *aroB-aroQ* cluster. Recently a 3-dehydroquinase enzyme from *Streptomyces coelicolor* has been purified and on the basis of limited N-terminal sequencing has been shown to be related to the cDHQase enzymes of filamentous fungi as well (White et al. 1990).

Biological characterisation of the putative dehydroquinate synthase and 3-dehydroquinase genes of M. tuberculosis

In order to determine whether the DNA sequence encoding the putative dehydroquinate synthase enzyme was biologically active, the 3.4 kb *Eco*RI fragment was subcloned into the vector pUC9 to yield plasmid pMT1. This plasmid was used to transform *E. coli* strain GLW38 (*aroB*$^-$; lacking dehydroquinate synthase), transformants being selected by their resistance to ampicillin on LB agar. Transformed strains were able to grow on minimal medium without aromatic amino acid supplements, demonstrating that the putative dehydroquinate synthase gene did encode a biologically active enzyme that functioned in *E. coli*. Additional experiments, in which subclones with part of the 5′ coding region of the dehydroquinate synthase gene deleted were used to transform *E. coli* GLW38, failed to produce transformants capable of growth on minimal medium without aromatic supplements (data not shown). On the basis of these experiments we designate the *M. tuberculosis* gene encoding dehydroquinate synthase as *aroB*.

Plasmid pMT1 encoding dehydroquinate synthase and 3-dehydroquinase was transferred into the minicell producing strain DS410. Plasmid harbouring minicells were purified, labelled with [^{35}S]methionine, and analysed using SDS-PAGE and autoradiography. A photo-

```
GAATTCCTGCGAGTGGACACCAATGCCGCAACCCCGGGGGGCGGTGGTCCGCCATATCCTGTCGCGGCTGCAGGTTCCCAGCCCCAGCGA
        10        20        30        40        50        60        70        80        90

           M  T  D  I  G  A  P  V  T  V  Q  V  A  V  D  P  P  Y  P  V  V  I  G  T  G  L  L
GGCGGCCACATGACCGATATCGGCGCACCCGTGACCGTGCAGGTGGCCGTCGATCCGCCATACCCGGTGGTCATCGGTACCGGCCTGCTC
        100       110       120       130       140       150       160       170       180

  D  E  L  E  D  L  L  A  D  R  H  K  V  A  V  V  H  Q  P  G  L  A  E  T  A  E  E  I  R  K
GACGAGCTGGAAGACCTGCTGGCCGACCGGCACAAGGTCGCCGTCGTGCATCAGCCCGGACTAGCCGAGACCGCGGAAGAGATCCGAAAG
        190       200       210       220       230       240       250       260       270

  R  L  A  G  K  G  V  D  A  H  R  I  E  I  P  D  A  E  A  G  K  D  L  P  V  V  G  F  I  W
CGCTTGGCCGGCAAGGGCGTCGACGCGCACCGCATCGAGATCCCCGACGCCGAGGCCGGCAAGGACCTGCCCGTCGTGGGATTCATCTGG
        280       290       300       310       320       330       340       350       360

  E  V  L  G  R  I  G  I  G  R  K  N  A  L  V  S  L  G  G  G  A  A  T  D  V  A  G  F  A  A
GAGGTGTTGGGCCGCATCGGAATCGGCCGCAAAAACGCCCTGGTCAGCCTCGGCGGCGGGGCCGCCACCGACGTCGCCGGGTTCGCGGCG
        370       380       390       400       410       420       430       440       450

  A  T  W  L  R  G  V  S  I  V  H  L  P  T  T  L  L  G  M  V  D  A  A  V  G  G  K  T  G  I
GCCACCTGGCTGCGCGGCGTCTCGATTGTGCACCTGCCCACCACACTGCTGGGCATGGTCGATGCGGCCGTCGGCGGCAAGACCGGCATC
        460       470       480       490       500       510       520       530       540

  N  T  D  A  G  K  N  L  V  G  A  F  H  Q  P  L  A  V  L  V  D  L  A  T  L  Q  T  L  P  R
AACACCGACGCCGGCAAGAACCTGGTCGGGGCGTTTCATCAGCCGTTGGCGGTCCTGGTGGACCTGGCGACGCTGCAAACCTTGCCACGC
        550       560       570       580       590       600       610       620       630

  D  E  M  I  C  G  M  A  E  V  V  K  A  G  F  I  A  D  P  V  I  L  D  L  I  E  A  D  P  Q
GACGAAATGATCTGCGGCATGGCCGAAGTGGTCAAGGCCGGCTTCATCGCCGACCCGGTGATCCTGGATCTCATCGAAGCTGACCCGCAG
        640       650       660       670       680       690       700       710       720

  A  A  L  D  P  A  G  D  V  L  P  E  L  I  R  R  A  I  T  V  K  A  E  V  V  A  A  D  E  K
GCCGCACTCGACCCGGCCGGCGACGTGCTGCCCGAGCTGATCCGGCGCGCGATCACCGTCAAGGCCGAGGTGGTCGCCGCCGACGAAAAG
        730       740       750       760       770       780       790       800       810

  E  S  E  L  R  E  I  L  N  Y  G  H  T  L  G  H  A  I  E  R  R  E  R  Y  R  W  R  H  G  A
GAATCCGAGCTGCGCGAAATCCTCAACTACGGCCACACATTAGGCCACGCGATCGAGCGCCGGGAACGCTACCGGTGGCGCCACGGCGCC
        820       830       840       850       860       870       880       890       900

  A  V  S  V  G  L  V  F  A  A  E  L  A  R  L  A  G  R  L  D  D  A  T  A  Q  R  H  R  T  I
GCCGTGTCGGTGGGGCTGGTGTTCGCGGCCGAGCTGGCCAGGCTTGCCGGGCGGCTCGACGACGCGACCGCGCAGCGCCACCGCACCATC
        910       920       930       940       950       960       970       980       990

  L  S  S  L  G  L  P  V  S  Y  D  P  D  A  L  P  Q  L  L  E  I  M  A  G  D  K  K  T  R  A
CTGTCCTCGTTGGGATTGCCGGTCAGCTACGACCCGGACGCGCTGCCCCAGCTGCTGGAAATCATGGCCGGCGACAAGAAGACTCGGGCG
       1000      1010      1020      1030      1040      1050      1060      1070      1080

  G  V  L  R  F  V  V  L  D  G  L  A  K  P  G  R  M  V  G  P  D  P  G  L  L  V  T  A  Y  A
GGTGTGTTGCGGTTCGTGGTGCTCGACGGATTGGCCAAGCCGGGCCGAATGGTGGGACCGGACCCCGGTCTGCTGGTAACCGCCTACGCC
       1090      1100      1110      1120      1130      1140      1150      1160      1170

  G  V  C  A  P  *
                 M  S  E  L  I  V  N  V  I  N  G  P  N  L  G  R  L  G  R  R  E  P  A  V  Y  G
GGAGTTTGCGCCCCATGAGCGAACTGATCGTGAACGTGATCAACGGCCCCAACCTGGGCCGGTTGGGCCGGCGCGAGCCCGCCGTCTATG
       1180      1190      1200      1210      1220      1230      1240      1250      1260

     G  T  T  H  D  E  L  V  A  L  I  E  R  E  A  A  E  L  G  L  K  A  V  V  R  Q  S  D  S  E
GCGGCACCACCCACGACGAGCTGGTCGCTCTGATCGAGCGTGAGGCCGCCGAGCTCGGACTCAAAGCCGTTGTGCGGCAAAGTGATAGCG
       1270      1280      1290      1300      1310      1320      1330      1340      1350

     A  Q  L  L  D  W  I  H  Q  A  A  D  A  A  E  P  V  I  L  N  A  G  G  L  T  H  T  S  V  A
AAGCCCAGCTGCTGGACTGGATTCATCAGGCCGCCGATGCGGCCGAACCCGTGATCCTCAACGCCGGCGGTCTGACGCACACGTCGGTGG
       1360      1370      1380      1390      1400      1410      1420      1430      1440

     L  R  D  A  C  A  E  L  S  A  P  L  I  E  V  H  I  S  N  V  H  A  R  E  E  F  R  R  H  S
CACTGCGCGATGCCTGCGCGGAGCTGAGCGCTCCGCTGATCGAGGTGCATATCTCCAACGTGCATGCGCGTGAGGAGTTCCGCCGCCACT
       1450      1460      1470      1480      1490      1500      1510      1520      1530

     I  L  S  P  I  A  T  G  V  I  V  G  L  G  I  Q  G  Y  L  L  A  L  R  Y  L  A  E  H  V  G
CCATCCTCAGCCCCATCGCGACCGGGGTGATCGTCGGGCTCGGCATCCAGGGCTACCTGCTGGCCCTGCGGTACCTAGCTGAGCATGTCG
       1540      1550      1560      1570      1580      1590      1600      1610      1620

     T  *
GGACGTGAGGGTTAATCCTTCTTGGGCTCGTCGCCGGGCTTGGTCTGGTCCGCCTCGGTGTCGGTGCGGATAACCTCGGTGGGTGTATCG
       1630      1640      1650      1660      1670      1680      1690      1700      1710

CTTTCGGTGGTGCGGATCACCTCGGTGGGTGCCTCACGTTCGACCGTGGCGACGGCCGCGGTTTGTGCCTCAGGCTGCGCGACCTGCATC
       1720      1730      1740      1750      1760      1770      1780      1790      1800
```

Fig. 3. Nucleotide sequence of 1.8 kb fragment from λgt11 aro5. The nucleotide sequences of the *Mycobacterium tuberculosis* genes encoding dehydroquinate synthase and 3-dehydroquinase are shown with the appropriate open reading frames translated into the single letter amino acid code. Nucleotides 100 to 1188 specify dehydroquinate synthase, and nucleotides 1185 to 1628 specify 3-dehydroquinase

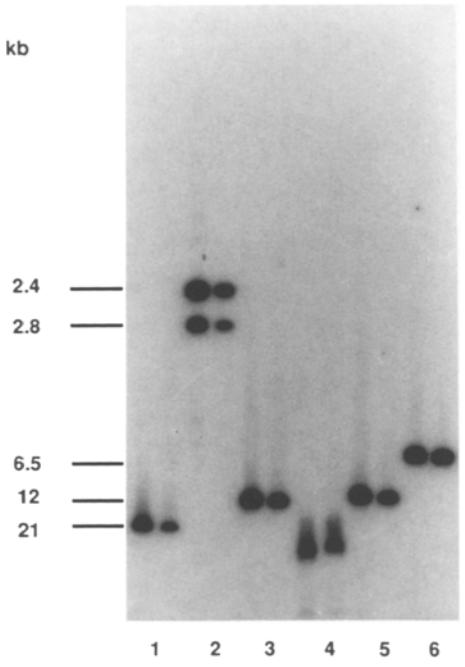

Fig. 4. Identification of the *aro*B-*aro*Q gene cluster on *Mycobacterium tuberculosis* genomic DNA. Two independent preparations of genomic DNA from *M. tuberculosis* (2 and 1.5 μg) were digested with restriction enzymes and fragments separated by agarose gel electrophoresis. The gel was blotted and then hybridised with the ^{32}P-labelled 2.34 kb *Eco*RV fragment from λgt11 aro5. The two DNA preparations are shown in adjacent lanes after digestion with (1) *Eco*RI; (2) *Sph*I; (3) *Xho*I; (4) *Hin*dIII; (5) *Nco*I; (6) *Not*I

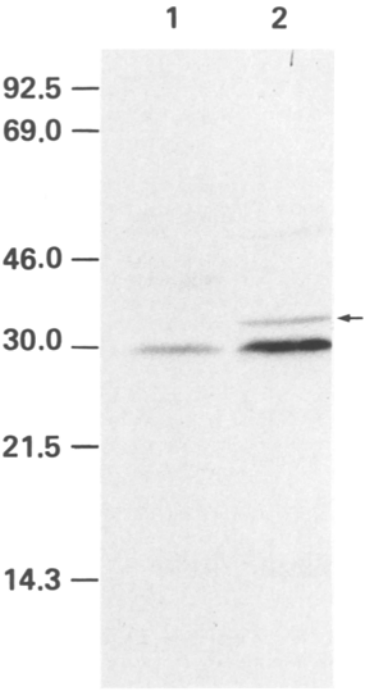

Fig. 5. Expression of polypeptides from pMT1 in minicells. An autoradiograph showing [^{35}S]methionine labelled minicell preparations separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Lane 1, pUC9 control; lane 2, pMT1. The positions of molecular weight markers are indicated by *bars* with the molecular weights shown in kDa. The extra band highlighted with an *arrow* in the pMT1 lane is in the position expected for dehydroquinate synthase

graph of the autoradiograph is shown in Fig. 5. Inspection of Fig. 5 reveals the presence of a single major unique protein of Mr 37 kDa in the experimental lane, which is in close agreement with the calculated Mr of 38 kDa for dehydroquinate synthase deduced from the DNA sequence. Four other unique proteins are also present in very low yield in the experimental lane, the fastest migrating of which is in the position expected for 3-dehydroquinase. The poor incorporation of label into the 3-dehydroquinase is consistent with the presence of only a single methionine residue (the initiation codon) in the derived amino acid sequence.

In order to assess the physical characteristics of the *M. tuberculosis* 3-dehydroquinase enzyme, plasmids pUC9 and pMT1 were transferred into *E. coli* mutant strain SK3430 (*aroD*$^-$; lacking bDHQase). The two plasmid harbouring strains were grown in appropriate minimal medium to an OD_{550} of 0.8, harvested by centrifugation, washed and the cells disrupted by sonication. Portions of the two cell free extracts were heated to 71° C for 10 min, the heat denatured proteins removed by centrifugation, and 3-dehydroquinase levels were measured in the extracts from each strain before and after the heat treatment. No linear dose-dependent conversion of the 3-dehydroquinic acid substrate was detected in either sample prior to the heat treatment. After heat treatment, however, linear dose dependent 3-dehydroquinase activity was found in extracts from the pMT1 transformant (specific activity 0.013 units/mg), while no substrate conversion was detected in extracts from the control pUC9 transformant. The heat treatment is presumed to inactivate a heat labile inhibitor of 3-dehydroquinase activity.

The data from the enzyme assay demonstrate that the *M. tuberculosis* 3-dehydroquinase enzyme is heat stable, a feature shared by the cDHQase of filamentous fungi (Hawkins et al. 1982b). Taking into account the strong similarity both in deduced amino acid sequence and physical characteristics between the mycobacterial dehydroquinase and fungal cDHQase, we designate the gene encoding the 3-dehydroquinase enzyme as *aroQ*. The designation Q (for quinate) is to denote that, in contrast to the *aroA* (enolpyruvyl shikimate-phosphate synthase) and *aroB* (dehydroquinate synthase) genes of *M. tuberculosis, aro* Q shows no detectable sequence similarity to the equivalent gene in the shikimate pathway of *E. coli* (*aroD*), but is instead clearly related to the gene encoding cDHQase in the quinic acid utilisation pathway of filamentous fungi (*QutE*).

Evolutionary relationships among the 3-dehydroquinate synthase and 3-dehydroquinase enzymes

Figure 6A shows an optimal manual alignment of the deduced amino acid sequence of dehydroquinate syn-

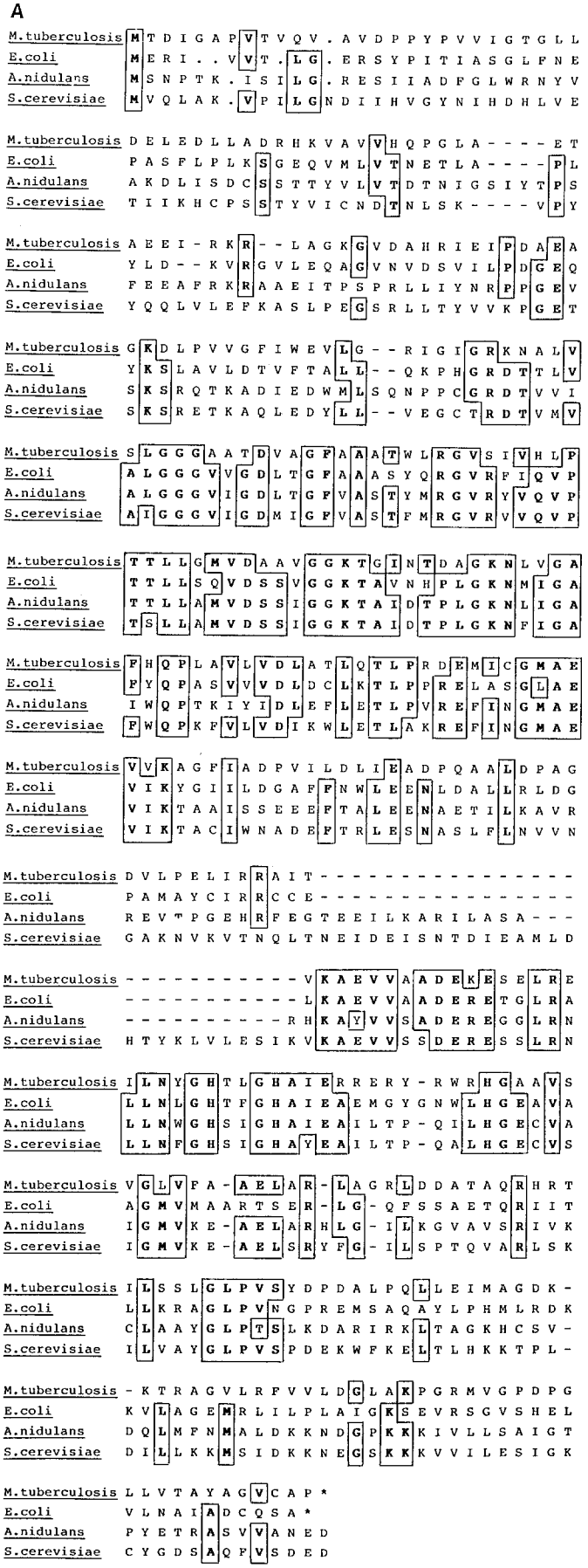

Fig. 6A and B. Amino acid sequence comparisons for dehydroquinate synthase and 3-dehydroquinase enzymes. **A** The deduced amino acid sequence of the *Mycobacterium tuberculosis* dehydroquinate synthase shown in optimal manual alignment with dehydroquinate synthase from *Escherichia coli, Aspergillus nidulans* and *Saccharomyces cerevisiae*. **B** The deduced amino acid sequence of the *M. tuberculosis* 3-dehydroquinase shown in optimal manual alignment with the catabolic 3-dehydroquinases (cDHQases) of *A. nidulans* and *N. crassa*. In each case residues that are identical in at least three sequences are *boxed*. There are numerous additional examples of conservative substitutions

thase from *M. tuberculosis* with the shikimate pathway dehydroquinate synthase enzymes of *E. coli* (*aroB*) (Millar and Coggins 1986), *A. nidulans* (Charles et al. 1985), and *Saccharomyces cerevisiae* (Duncan et al. 1987). Figure 6B shows an optimal manual alignment of the deduced amino acid sequence of 3-dehydroquinase from *M. tuberculosis* with the quinic acid utilisation pathway cDHQase enzymes of *A. nidulans* (*QutE*) (Da Silva et al. 1986) and *Neurospora crassa* (*QA-2*) (Hawkins et al. 1982b). Inspection of Figs. 6A and 6B reveals that *M. tuberculosis* dehydroquinate synthase has approximately 33% identity with the equivalent *E. coli, A. nidulans* and *S. cerevisiae* sequences, and the 3-dehydroquinase approximately 28% identity with the fungal cDHQases. In addition there are many examples of conservative substitutions and the overall level of similarity is equivalent to that seen between the *M. tuberculosis* EPSP synthase (the *aroA* gene product) and the equivalent enzymes in *E. coli* and fungi (Garbe et al. 1990).

The fact that the mycobacterial 3-dehydroquinase is related to the enzymes found in the catabolic pathway in fungi, rather than to the conventional biosynthetic enzymes represents an unexpected finding of this study. All of the recombinant λ phage capable of complement-

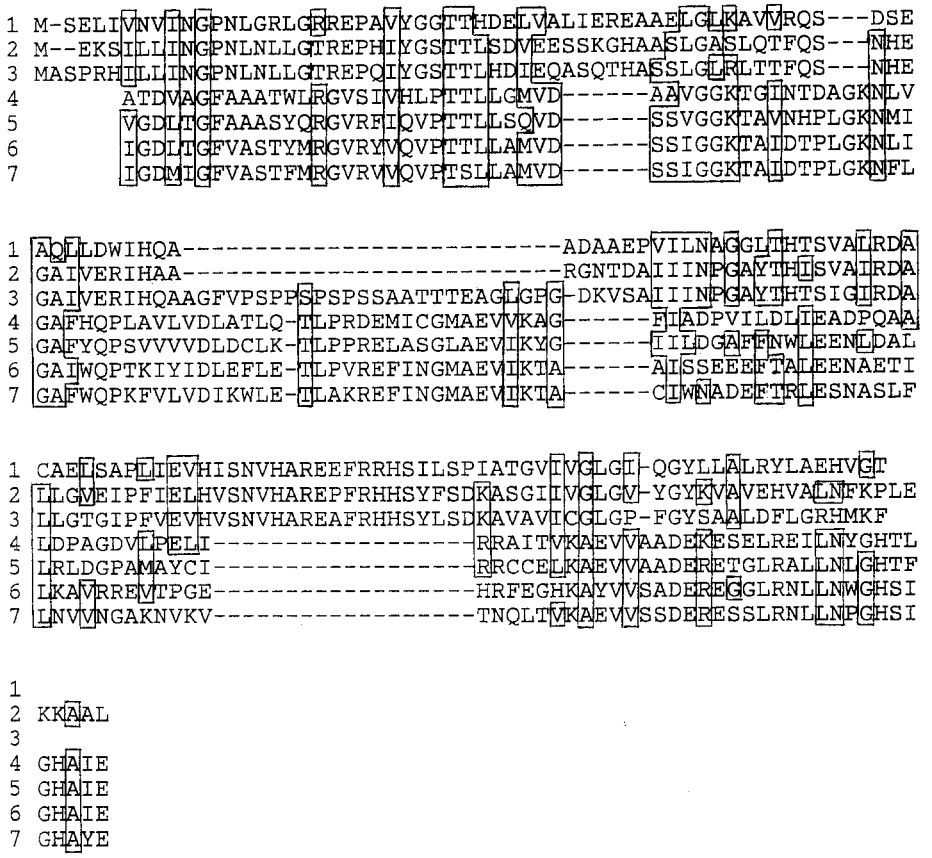

Fig. 7. Sequence similarities between dehydroquinate synthases and cDHQases. An optimal alignment of the entire *Mycobacterium tuberculosis* 3-dehydroquinase and fungal cDHQases with parts of the dehydroquinate synthase of prokaryotes and filamentous fungi. Identical or closely related amino acids are *boxed*. 1, *M. tuberculosis* 3-dehydroquinase; 2, *Aspergillus nidulans* cDHQase; 3, *Neurospora crassa* cDHQase; 4, *M. tuberculosis* dehydroquinate synthase, residues 109–256; 5, *Escherichia coli* dehydroquinate synthase, residues 106–256; 6, *A. nidulans* dehydroquinate synthase, residues 116–233 and 248–279; 7, *Saccharomyces cerevisiae* dehydroquinate synthase, residues 112–229 and 259–287

ing growth of an *E. coli aroD* mutant strain carried the same gene, and it therefore seems probable that *M. tuberculosis* has only a single dehydroquinase enzyme. In preliminary experiments we have found that *M. tuberculosis* is unable to utilise quinic acid as a carbon source and we propose that the dehydroquinase encoded by the *aroQ* gene does in fact function in the biosynthetic pathway of *M. tuberculosis*. The close linkage with the *aroB* gene – which encodes the enzyme catalysing the previous step in the shikimate pathway – lends further support to the conclusion that the dehydroquinase performs a biosynthetic rather than a catabolic role in *M. tuberculosis*.

On the basis of comparative sequence analysis it has been proposed that the cDHQase and bDHQase of fungi arose by convergent evolution (Hawkins 1987) and the finding of a cDHQase active in the shikimate pathway of *M. tuberculosis* provides a further clue to the evolutionary origin of the two isozymes. It is intriguing to note that there is a limited degree of sequence similarity between the dehydroquinate synthase and 3-dehydroquinase enzymes of *M. tuberculosis* (see Fig. 7). It has been suggested previously (Horowitz 1965) that biosynthetic pathways in contemporary organisms have been built progressively by "retro-evolution" from the final metabolite in the pathway, probably by gene duplication and subsequent divergence. It is possible that the sequence relationship between the enzymes catalysing sequential steps in the mycobacterial shikimate pathway represents an example of such retro-evolution. The optimal manual amino acid alignment shown in Fig. 7 indicates that this limited sequence relationship can in fact be seen to apply to the entire families of cDHQase and dehydroquinate synthase enzymes. These findings suggest that the fungal bDHQases and cDHQases both arose originally from enzymes that had evolved independently as part of the shikimate biosynthesis pathways in different prokaryotes. Sequence analysis further suggests a possible difference in catalytic mechanism between the different classes of dehydroquinase. The dehydroquinase in *E. coli* is related to the bDHQase of fungi (Hawkins 1987) with the catalytic reaction proceeding via a Schiff's base intermediate (Duncan et al. 1986). Inspection of Fig. 6B shows that, in spite of considerable identity at the amino acid sequence level between the 3-dehydroquinases, there is no lysine residue conserved amongst all three enzymes. It is unlikely therefore that the catalytic mechanism employed by the cDHQases involves a Schiff's base intermediate.

This study has resulted in isolation and characterisation of two additional genes from the shikimate biosynthesis pathway of *M. tuberculosis* which can now be considered as attractive targets for construction of rationally attenuated candidate vaccine strains. In addition, detailed sequence analysis of the genes provides novel fundamental information with regard to evolution of biosynthetic pathways both in *M. tuberculosis* and in other micro-organisms.

Acknowledgements. The bulk of this research was supported by a grant from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases awarded to D.Y., and in part by SERC grant GR/E 1055.5 awarded to A.H.

References

Beri R-K, Grant S, Roberts CF, Smith M, Hawkins AR (1990) Selective overexpression of the *QutE* gene encoding catabolic 3-dehydroquinase in multicopy transformants of *A. nidulans*. Biochem J 265:337–342

Biggin MD, Gibson TS, Hong CF (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 86:3963–3965

Charles IG, Keyte JW, Brammar WJ, Hawkins AR (1985) Nucleotide sequence encoding the biosynthetic dehydroquinase function of the pentafunctional *AROM* locus of *Aspergillus nidulans*. Nucleic Acids Res 13:8119–8128

Charles IG, Keyte JW, Brammar WJ, Smith M, Hawkins AR (1986) The isolation and nucleotide sequence of the complex *AROM* locus of *Aspergillus nidulans*. Nucleic Acids Res 14:2201–2213

Collins FM (1974) Vaccines and cell-mediated immunity. Bacteriol Rev 38:371–402

Da Silva AJS, Whittington H, Clements J, Roberts CF, Hawkins AR (1986) Sequence analysis and transformation by the catabolic 3-dehydroquinase (*QUT*E) gene from *Aspergillus nidulans*. Biochem J 240:481–488

DeFeyter RC, Davidson BE, Pittard J (1986) Nucleotide sequence of the transcription unit containing the *aroL* and *aroM* genes from *Escherichia coli* K-12. J Bacteriol 165:233–239

Dougan G, Sherratt DJ (1977) *Tna* as a probe for the structure and function of ColE1. Mol Gen Genet 151:151–160

Dougan G, Maskell D, Pickard D, Hormaeche C (1987) Isolation of stable *aroA* mutants of *Salmonella typhi* Ty2: properties and preliminary characterisation in mice. Mol Gen Genet 207:402–405

Duncan K, Chaudhuri S, Campbell MS, Coggins JR (1986) The overexpression and complete amino acid sequence of *Escherichia coli* 3-dehydroquinase. Biochem J 238:475–483

Duncan K, Edwards RM, Coggins JR (1987) The penta functional *arom* enzyme of Saccharomyces cerevisiae is a mosaic of monofunctional domains. Biochem J 246:375–386

Engers HD, Workshop Participants (1985) Results of a World Health Organisation sponsored workshop on monoclonal antibodies to *Mycobacterium leprae*. Infect Immun 48:718–720

Engers HD, Workshop Participants (1986) Results of a World Health Organisation sponsored workshop to characterise antigens recognised by mycobacteria-specific monoclonal antibodies. Infect Immun 51:718–720

Fine PEM (1988) BCG vaccination against tuberculosis and leprosy. Br Med Bull 44:691–703

Garbe T, Jones C, Charles I, Dougan G, Young D (1990) Cloning and characterisation of the *aroA* gene from *Mycobacterium tuberculosis*. J Bacteriol 172:6774–6782

Grant S, Roberts CF, Lamb H, Stout M, Hawkins AR (1988) Genetic regulation of the quinic acid utilisation (*QUT*) gene cluster in *Aspergillus nidulans*. J Gen Microbiol 134:347–358

Hawkins AR (1987) The complex *AROM* locus of *Aspergillus nidulans*: evidence for multiple gene fusions and convergent evolution. Curr Genet 11:491–498

Hawkins AR, Roberts CF (1989) In: Nevalainen H, Pentilla M (eds) Molecular biology of filamentous fungi, vol 6. Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, Finland, pp 85–100

Hawkins AR, Smith M (1991) Domain structure and interaction within the pentafunctional *arom* polypeptide. Eur J Biochem 196:717–724

Hawkins AR, Giles NH, Kinghorn JR (1982a) Genetic and biochemical aspects of quinate breakdown in the filamentous fungus *Aspergillus nidulans*. Biochem Genet 20:271–286

Hawkins AR, Reinert WR, Giles NH (1982b) Characterization of *Neurospora crassa* catabolic dehydroquinase purified from *N. crassa* and *Escherichia coli*. Biochem J 203:769–773

Hawkins AR, Da Silva AJS, Roberts CF (1984) Evidence for two control genes regulating expression of the quinic acid utilisation (*QUT*) gene cluster in *Aspergillus nidulans*. J Gen Microbiol 130:567–574

Hoch JA, Nester EW (1973) Gene-enzyme relationships of aromatic acid biosynthesis in *Bacillus subtilis*. J Bacteriol 116:59–66

Horowitz NH (1965) In: Bryson B and Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 15–23

Husson RN, James BE, Young RA (1990) Gene replacement and expression of foreign DNA in mycobacteria. J Bacteriol 172:519–524

Kinghorn JR, Hawkins AR (1982) Cloning and expression in *Escherichia coli* K12 of the biosynthetic dehydroquinase function of the *AROM* gene cluster from the eukaryote *Aspergillus nidulans*. Mol Gen Genet 186:145–152

Martin C, Timm J, Rauzier J, Gomez-Lus R, Davies J, Gicquel B (1990) Transposition of an antibiotic resistance element in mycobacteria. Nature 345:739–743

Millar C, Coggins JR (1986) The complete amino acid sequence of 3-dehydroquinate synthase of *Escherichia coli* K12. FEBS Lett 200:11–17

O'Callaghan D, Maskell D, Liew FY, Easmon CSF, Dougan G (1988) Characterisation of aromatic and purine-dependent *Salmonella typhimurium*: attenuation, persistence, and ability to induce protective immunity in Balb/c mice. Infect Immun 56:419–423

Pittard J, Wallace BJ (1966) Distribution and function of genes concerned with aromatic biosynthesis in *Escherichia coli*. J Bacteriol 91:1494–1508

Sanger F, Michelin S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Snapper SB, Lugosi L, Jekkel A, Melton RE, Kieser T, Bloom BR, Jacobs WR (1988) Lysogeny and transformation in mycobacteria: stable expression of foreign genes. Proc Natl Acad Sci USA 85:6987–6991

Styblo K (1989) Overview and epidemilogic assessment of the current global tuberculosis situation with an emphasis on control in developing countries. Rev Infect Dis 11, suppl 2:S339–S346

Tabor S, Richardson CC (1987) DNA sequencing analysis with a modified bacteriophage T7 polymerase. Proc Natl Acad Sci USA 84:4767–4771

White PJ, Young J, Hunter TS, Nimmo HG, Coggins JR (1990) The purification and characterization of 3-dehydroquinase from *Streptomyces coelicolor*. Biochem J 265:735–738

Young DB (1988) Structure of mycobacterial antigens. Br Med Bull 44:4767–4771

Young RA, Bloom BR, Grosskinsky CM, Ivanyi J, Thomas D, Davis RW (1985a) Dissecting *Mycobacterium tuberculosis* antigens using recombinant DNA. Proc Natl Acad Sci USA 82:2583–2587

Young RA, Mehra V, Sweetser D, Buchanan T, Clark-Curtiss J, Davis RW, Bloom BR (1985b) Genes for the major protein antigens of the leprosy parasite *Mycobacterium leprae*. Nature 316:450–452

Communicated by J. Gajewski