Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes

J. A. Soliveri,† J. Gomez,‡ W. R. Bishai and K. F. Chater

Author for correspondence: K. F. Chater. Tel: +44 1603 452571. Fax: +44 1603 456844. e-mail: chater@bbsrc.ac.uk

The *whiB* sporulation gene of *Streptomyces coelicolor* was shown [Davis, N. K. & Chater, K. F. (1992). *Mol Gen Genet* 232, 351–358] to encode a small, cysteine-rich putative transcription factor unlike any that had been described previously. The large database of DNA sequences of mycobacteria (like *Streptomyces*, members of the Actinomycetales) has revealed a family of genes encoding proteins related to WhiB. *Mycobacterium tuberculosis* contains at least six such genes (*whiB* homologues in *mycobacteria*: *whmA*–*F*) and a likely seventh, *whmG*. Using conserved features of Whm proteins, a PCR-based approach led to the discovery that *S. coelicolor* A3(2) contains several similar genes. Cloning and sequencing of these *whiB*-like (*wbl*) genes revealed likely orthologues of four of the *wbl* genes of *M. tuberculosis*. In all, *S. coelicolor* contains at least five *wbl* genes in addition to *whiB* itself. All five were shown by RT-PCR to be transcribed. A Southern blotting survey using each *wbl* gene as a probe showed that nearly all of a series of representatives of ten actinomycete genera (including morphologically simple organisms) contain close homologues of several *wbl* genes, suggesting that the ancient progenitor of all these organisms already contained a family of such genes, which have not been found in any other organisms.

**Keywords:** tuberculosis, leprosy, actinomycetes, prokaryotic phylogeny, sporulation in *Streptomyces*

**INTRODUCTION**

With the advent of 16S RNA sequencing, it has become clear that Gram-positive bacteria possessing high G+C DNA form a distinct phylogenetic class, and the term Actinomycetes has been appropriated from its earlier more morphologically based usage to describe these organisms (Embley & Stackebrandt, 1994). Two genera of actinomycetes, *Mycobacterium* and *Streptomyces*, have received particular attention: the genomes of *Mycobacterium tuberculosis* (Cole et al., 1998) and *Mycobacterium leprae* (Eiglmeier et al., 1993; Honoré et al., 1993; Fsihi et al., 1996; Smith et al., 1997), two of humankind’s greatest scourges, have been completely or extensively sequenced, and the genome of *Streptomyces coelicolor* A3(2), a representative of the industrially important antibiotic-producing streptomycetes, is currently the subject of a genome sequencing project, to be completed during 2000 (www.sanger.ac.uk/Projects/S_coelicolor/). This paper is concerned with the discovery of a family of genes in mycobacteria and streptomycetes that is also found in other actinomycetes, but is apparently absent from all other organisms studied.

Streptomycetes are developmentally complex (Chater & Losick, 1997; Chater, 1998). They form mycelial colonies in which different physiological functions take
place in different locations. Thus, most antibiotic synthesis takes place in the older parts of the substrate mycelium, whilst sporulation takes place in specialized aerial hyphae. Gene regulation and the possibility of development in mycobacteria are less well characterized, but there is some evidence that alterations in cell type may occur during infection by pathogenic mycobacteria (Chatterjee, 1976; Parrish et al., 1998).

The prototype of the gene family discussed in this paper, *whiB*, is a developmental regulatory gene identified and characterized in *S. coelicolor* as being essential for sporulation of aerial hyphae (Davis & Chater, 1992). It encodes an 87 amino acid polypeptide with attributes suggesting that it may be a DNA-binding protein. The sequence of WhiB contains four cysteines – a common motif in metal-coordinating DNA-binding proteins such as Zn-binding GAL4 and other zinc finger proteins (Schjerling & Holmberg, 1996) (including a class of small transcriptional activator proteins encoded by some genes (the consensus conserved amino-acid sequence alignments of the nucleotide and amino-acid sequence alignments of the *whiB* gene family) and the use of this gene family). Remarkably, this *whiB*-like gene downstream of the *groEL* operon of *M. tuberculosis* was detected during database searches (Chater, 1993). The partial sequencing of the *M. leprae* genome (Honore et al., 1993; Smith et al., 1997), and the recently completed sequencing of the *M. tuberculosis* genome (Cole et al., 1998) have provided an opportunity to scan large amounts of mycobacterial genomic DNA sequence for the presence of ORFs encoding homologues of WhiB. In this paper we describe multiple *whiB* homologues in mycobacteria (the *whm* gene family) and the use of this information to reveal their equivalents in *Streptomyces* (the *wbl*, for *whiB*-like, gene family). Remarkably, this substantial gene family appears to be present in most (perhaps all) actinomycetes, yet absent from all other sufficiently characterized organisms.

**METHODS**

**Strains, media and culture conditions.** The *Escherichia coli* host strain used for routine subcloning was DH15α (Hanahan, 1983). *S. coelicolor* M145, a plasmid-free prototrophic derivative of the wild-type strain A3(2) (Hopwood et al., 1985), was used as the source of DNA and of RNA for RT-PCR analysis. Media, conditions (37 °C, LB liquid medium, LBA solid medium) and media supplements (carbenicillin, 200 µg ml⁻¹; X-Gal, 40 µg ml⁻¹) for *E. coli* were as described by Sambrook et al. (1989). Standard media and conditions for *S. coelicolor* were as described by Hopwood et al. (1985).

**Genetic manipulation.** For *E. coli*, plasmid preparation and transformation of competent cells were as described by Sambrook et al. (1989). For *S. coelicolor*, genomic DNA preparations were as described by Hopwood et al. (1985). Standard methods were used for restriction enzyme digestion, ligation, Southern blotting, PCR and radiolabelling of DNA (Sambrook et al., 1989) unless stated otherwise.

**Screening of an *S. coelicolor* genomic library.** An *S. coelicolor* genomic library of ordered cosmids (Redenbach et al., 1996) was screened by DNA hybridization using different genes as probes (Table 1). Probes of PCR products were amplified from the following plasmids: *whmA*₃₀, pTA777F [containing *whmA*₃₀, amplified by PCR from *M. leprae* cosmid B57, in pCR1 (Invitrogen)]; *whmE*₃₀, pTA63.3 [containing *whmE*₃₀, amplified by PCR from *M. tuberculosis* genomic DNA, in pCR1]; *whmC*₃₀, pTA1937 [J. Gomez, unpublished]; *whmD*₃₀, pBS178 [J. Gomez, unpublished]; *whmE*₃₀, pJS5001 [encoded by PCR from *M. tuberculosis* genomic DNA and inserted into the Smal site of pIJ2926 (Jansen & Bibb, 1993)]; saf (≡ *wbl*), pUL300 (Daza et al., 1990); *wbl*, pJ538 (wbl* gene in pUC19; Davis & Chater, 1992). Probes were [x-3²P]dCTP-labelled using an Oligolabelling Kit (Amershams Pharmacia Biotech), following the supplier’s instructions. The cosmid library was probed under low-stringency conditions, prehybridization and hybridization being performed at 65 °C in a solution containing 5 × SSC, 5 × Denhardt’s solution (Hopwood et al., 1985), 0.5 % (w/v) SDS and 100 µg nonhomologous DNA ml⁻¹. Hybridization overnight was followed by two washes in 2 × SSC, 0.1 % SDS at 60 °C. After isolation, cosmid DNAs were further digested and a second Southern blot was performed.

**Southern analysis of DNA of diverse actinomycetes for *whm*- and *wbl*-like genes.** Southern blots of restriction digests of mycobacterial DNA and a collection of diverse actinomycete chromosomal DNAs were hybridized with *whm* and *wbl* probes. Labelling, hybridization and detection were carried out at moderate stringency, using conditions described by Soliveri et al. (1993) except in the case of the studies shown in Fig. 1, when washing was in 3 × SSC, 0.1 % SDS at 55 °C (3 × 20 min). The probe to screen for the presence of *whiB* homologues in *S. coelicolor* was both generated and [x-3²P]dCTP-labelled using a low-stringency PCR using genomic DNA from *S. coelicolor* as template and as primers, oligonucleotides WHB1 (5′-CTGGCAGAGGCGCGCCACTGTG) and WHB2 (5′-TCGCTCGCATTCGGACAGGCC), which were designed on the basis of strongly conserved regions from the nucleotide and amino-acid sequence alignments of *whm* and *wbl* genes (the consensus conserved amino-acid sequence alignments were, respectively, WQERALC and GLSERER). The primary PCR reaction mixture (100 µl) contained 59.5 µl double-distilled H₂O, 2 µl 5 mM dNTPs (A, G and T), 1 µl 100 µM dCTP, 5 µl [x-3²P]dCTP [3000 Ci mmol⁻¹ (111 TBq mmol⁻¹)]; Amersham], 1 µl (20 µM) of each primer, 20 µl 5 × Buffer Q (Qiagen), 10 µl 10 × PCR Buffer (Qiagen) and 0.5 µl (2.5 U) Taq Polymerase (Qiagen). In a Perkin-Elmer-Cetus DNA Thermal Cycler 9600, the amplification consisted of an initial 3 min denaturation step at 96 °C, and 32 cycles of 1 min at 96 °C, 1 min at 56 °C and 2 min at 65 °C. The final extension step was 10 min at 65 °C. The *wblG* probe was generated by PCR amplification using the same DNA as
Table 1. WhiB-like proteins of actinomycetes other than *Streptomyces* spp., identified through database searches

<table>
<thead>
<tr>
<th>Protein</th>
<th>Source</th>
<th>Length</th>
<th>EMBL accession number</th>
<th>Percentage identity to WhiB</th>
</tr>
</thead>
<tbody>
<tr>
<td>WhmA (<strong>whiB</strong>4, Rv3681c)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>118‡</td>
<td>AL022121</td>
<td>37</td>
</tr>
<tr>
<td>WhmA</td>
<td><em>Mycobacterium leprae</em></td>
<td>119</td>
<td>L01263</td>
<td>33</td>
</tr>
<tr>
<td>WhmB (<strong>whiB</strong>3, Rv3416)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>102</td>
<td>Z77165</td>
<td>32</td>
</tr>
<tr>
<td>WhmB</td>
<td><em>Mycobacterium leprae</em></td>
<td>102</td>
<td>U00020</td>
<td>33</td>
</tr>
<tr>
<td>WhmC†</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>96</td>
<td>AL021646</td>
<td>22</td>
</tr>
<tr>
<td>WhmC</td>
<td><em>Mycobacterium leprae</em></td>
<td>89</td>
<td>U00016</td>
<td>22</td>
</tr>
<tr>
<td>WhmD (<strong>whiB</strong>2, Rv3260c)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>123‡</td>
<td>AL021840</td>
<td>67</td>
</tr>
<tr>
<td>WhmE (<strong>whiB</strong>1, Rv3219)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>84</td>
<td>Z95120</td>
<td>41</td>
</tr>
<tr>
<td>WhmF (Rv3862c)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>96‡</td>
<td>Z83864</td>
<td>25</td>
</tr>
<tr>
<td>WhmG (Rv0022c)*</td>
<td><em>Mycobacterium tuberculosis</em></td>
<td>139</td>
<td>Z80233</td>
<td>22</td>
</tr>
<tr>
<td>WblX (ORF6)</td>
<td><em>Rhodococcus opacus</em></td>
<td>92</td>
<td>AF030176</td>
<td>34</td>
</tr>
</tbody>
</table>

* Designations of the corresponding genes given by Cole et al. (1998).
† This gene was not recognized as an ORF by Cole et al. (1998).
‡ These lengths differ from those allocated in the EMBL database. See Fig. 5 legend for further details.

Table 2. Oligonucleotides used for RT-PCR amplifications of *wbl* RNAs in *S. coelicolor*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sense oligonucleotide</th>
<th>Antisense oligonucleotide</th>
<th>Expected PCR product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>wbla</em></td>
<td>5'-GAA TAT GCC GCT GCC</td>
<td>5'-CCC CTC GTT CGT ACT CCG TCC</td>
<td>334</td>
</tr>
<tr>
<td><em>wblb</em></td>
<td>5'-CGT CCT TCG TCC ACT CCC C</td>
<td>5'-GTC CTC GCT GAG ACC ACC C</td>
<td>249</td>
</tr>
<tr>
<td><em>wble</em></td>
<td>5'-CGT TCA CAT TCA CAA GCA TCC</td>
<td>5'-GTT CCT TCT CCG ACA GGC</td>
<td>268</td>
</tr>
<tr>
<td><em>wblh</em></td>
<td>5'-GTT GGT CGT CAG TGC ACC</td>
<td>5'-TAG ACC CGT GCT GAC ACC</td>
<td>348</td>
</tr>
</tbody>
</table>

RT-PCR. Total RNA was isolated (Kelemen et al., 1996) from *S. coelicolor* exponential-phase cultures grown in YEME medium (Hopwood et al., 1985). To remove any residual DNA, RNA was treated with RNase-free DNase I (Boehringer Mannheim). The concentration of RNA was determined by absorbance measurements, and the quality of RNA was analysed on the 1% agarose gel. To control for the presence of contaminating DNA, samples containing 4 µg of the RNA preparation, 20 pmol of the same oligonucleotide and 1 µg of RNase A (DNase-free; Boehringer Mannheim) were incubated, in a 20 µl reaction volume, at 37 °C for 1 h. RT-PCR was performed using suitable oligonucleotides (Table 2) and a Titan One Tube RT-PCR Kit (Boehringer Mannheim) according to the supplier’s instructions (which include 10 rounds of reverse transcription and 25 rounds of PCR). Control samples containing genomic DNA instead of total RNA as the template were run as RT-PCR samples in parallel.

DNA sequencing and deduced protein analysis. The DNA sequence was determined with a *Thermus aquaticus* polymerase chain reaction sequencing kit and a model 373A DNA sequencing system (Applied Biosystems). Nucleotide sequences and derived amino acid sequences were analysed with the GCG software package (version 8.1; Genetics Computer Group, Madison, WI, USA). Likely coding regions were revealed by frame plot analysis (PSA) server (bmerc-www.bu.edu/cgi-bin/frameplot.pl). Protein secondary structure predictions were carried out using the Protein Sequence Analysis (PSA) server (bmerc-www.bu.edu/psa/) and the PredictProtein server, PHDsec (www.embl-heidelberg.de/predictprotein/predictprotein.html#P2SEC).
Phylogenetic analysis of whiB-like genes. Amino acid sequences of the various whm and wbl genes were aligned and analysed using the PROTDIST, FITCH and CONSENSE programs from the PHYLIP software package (Felsenstein, 1988). The dendrogram shown in Fig. 4 is the consensus tree of 500 runs of the FITCH program. Branch lengths are proportional to the genetic distance between proteins as assessed by PROTDIST.

Terminology (homologues, orthologues and paralogues). For the purpose of this paper, homologues (genes or their corresponding proteins that have an obvious common ancestry) can be broadly subdivided into orthologues (functionally equivalent homologues from different organisms) and paralogues (functionally related, but not equivalent, homologues arising from gene duplication within an evolutionary lineage) (Holland, 1999).

RESULTS

Database searches reveal a number of mycobacterial genes encoding WhiB-like proteins

Cole et al. (1998) identified four whiB-like genes in their annotation of the complete genome sequence of M. tuberculosis. We have identified a further three (Table 1). These seven whiB homologues of mycobacteria (whm genes) would encode proteins ranging in identity to WhiB from 22 to 67% (Table 1). The four M. tuberculosis whm genes most closely related to whiB were described by Cole et al. (1998) as whiB1, whiB2, whiB3 and whiB4 (see Table 1), but we suggest that the use of the whm terminology will be less confusing and conform better to the conventions of bacterial genetics. The alphabetical order of the gene names (whmA–G) mainly reflects the order in which we became aware of their existence. Orthologues of whmA, B and C are present in M. leprae (Honoré et al., 1993; Smith et al., 1997), and where necessary we distinguish between orthologues from different organisms by the use of subscripts. whmD is the most similar of the whm genes to whiB. No orthologues of whmD, whmE, whmF or whmG have yet appeared in the publicly available M. leprae genome sequence, but the Mycobacterium smegmatis whmD gene has been cloned and characterized (J. Gomez & W. R. Bishai, unpublished).

In most members of the Whm/Wbl family, a conserved set of cysteine residues is arranged as Cys-X(14–22)-Cys-X2-Cys-X5-Cys (see below), but in WhmG (≡Rv0022c in the M. tuberculosis genome) the spacing of the corresponding residues is Cys-X11-Cys-X3-Cys-X7-Cys. Nonetheless, the other similarities to conserved regions of the Whm/Wbl protein family are striking enough to warrant the designation of Rv0022c as whmG. Further sequence comparisons are made later in this paper.

Multiple whm genes are also present in a rapidly growing, saprophytic mycobacterium, M. smegmatis

M. smegmatis is saprophytic and has a significantly faster generation time than the pathogenic M. tuberculosis (3 vs 18–24 h in culture). Southern blotting experiments on M. smegmatis DNA gave unique and strong signals with probes from whmA, whmB and whmC (Fig. 1), whmE (not shown) and whiB (revealing the previously cloned whmD; see above). Similar results (but different fragment sizes) were also obtained with DNA of M. bovis BCG (see Fig. 1).

Multiple whiB-like genes are present in S. coelicolor

The presence of paralogues of whiB in mycobacteria suggested that a similar situation might prevail in Streptomyces spp., and, in particular, in the extensively studied strain S. coelicolor A3(2). This would not only extend (and be informed by) the large amounts of genetic and biological information available for this strain, but might also in future help to evaluate the roles of whiB-like genes in mycobacteria. A firm indication that S. coelicolor has multiple whiB-like genes came from the use of a probe obtained by PCR using S. coelicolor DNA as template. The oligonucleotides designed as primers for PCR were based on two segments conserved among the known whiB-like genes (see Methods). Thus, it was hoped that several whiB-like genes would be represented in the PCR-amplified DNA. Indeed, the PCR-generated probe hybridized
fairly strongly at six different positions to Southern blots of chromosomal DNA digests, one of which had the mobility expected for \textit{wbiB}.

Thus encouraged, we used PCR products corresponding to each of five \textit{whm} genes (\textit{whmA–whmE}; see Methods) to detect \textit{whm} orthologues in \textit{S. coelicolor} DNA. Each of these probes except the \textit{whmC} probe hybridized strongly to one band in Southern blots of \textit{S. coelicolor} DNA digests, and each band corresponded to one of the bands obtained in the preliminary experiment. (Of the two other bands, one was later identified as \textit{wblH}, which is described below, and the other remains unidentified.) The probes were then used to identify hybridizing cosmids in the \textit{S. coelicolor} library, and after suitable subcloning we sequenced the relevant regions to reveal a \textit{wbiB}-like (\textit{wbl}) gene in each case. With one exception (the \textit{whmD}/\textit{wblB} pair) the genes have been given letters to match their likely equivalent \textit{whm} genes. Thus, considering a conserved core segment corresponding to \textit{WhiB} residues 24–80, \textit{WblA} protein was 67\% identical to \textit{WhmA}; \textit{WblB} was 66\% identical to \textit{WhmB}; no \textit{WhmC} homologue was found (but see below); \textit{WhmD} was 87\% identical to \textit{WbiB}; and \textit{WblE} was 75\% identical to \textit{WhmE}. Furthermore, the ongoing \textit{S. coelicolor} genome sequencing project (www.sanger.ac.uk/projects/S_coelicolor/) revealed another \textit{wbl} gene in cosmid St4C6, which was not closely similar to any \textit{whm} genes and so was termed \textit{wblH} (at the time of completing this paper, approximately 80\% of the genome was available). \textit{wblH} is identified as St4C6.25 (PID e1490484). The \textit{WblH} protein is most similar to \textit{WhiB} (58\% identity) and \textit{WhmD} (59.5\% identity). In some cases, the corresponding \textit{whm} and \textit{wbl} genes had similar genetic contexts. Thus, \textit{whmA} and \textit{wblA} both diverge from genes encoding related putative penicillin-binding proteins; \textit{whmB} and \textit{wblB} both converge on \textit{groEL}, though \textit{wblB} is separated from the convergent \textit{groEL} operon by an intervening gene; and \textit{whmE} and \textit{wblE} both converge on genes encoding related putative membrane sensor kinases. \textit{WblH}, which has no close \textit{whm} homologue, converges on a gene encoding a putative exported desaturase.

In summary, \textit{S. coelicolor} contains close homologues (presumptive orthologues) of at least four of the seven known \textit{whm} genes of mycobacteria. The seven \textit{whm} genes of \textit{M. tuberculosis} are noticeably clustered in less than 1 Mb of the circular 4411529 base pair \textit{H37Rv} chromosome, between positions 3468401 (\textit{whmC}) and 27442 (\textit{whmG}) (position 1, located near the origin of replication, is defined as the first base in the \textit{dnaA} ORF); the homologues from \textit{S. coelicolor} are more widely scattered and in a quite different order (Fig. 2), illustrating an absence of large-scale synteny between these diverged actinomycetes.

\textbf{Examination of the \textit{saf} gene of \textit{Streptomyces griseus} and its use as probe reveals another \textit{wbl} gene in \textit{S. coelicolor} A3(2)}

Daza et al. (1990) identified a segment of \textit{S. griseus} DNA (termed \textit{saf}) that, when introduced into \textit{Streptomyces lividans} on a multicopy plasmid, elicited pleiotropic
changes in the production of exoenzymes and antibiotics. Surprisingly, routine database screening had revealed a wblB-like coding sequence extensively overlapping, but in the antisense strand of, the published saf coding sequence (Chater, 1993). Using a PCR-generated probe to this ‘anti-saf’ coding sequence, we cloned and then sequenced homologous DNA from the S. coelicolor cosmid library to reveal a gene that was designated wbl\textsubscript{Sc}. Both wbl\textsubscript{Sc} and the homologous ‘anti-saf’ sequence from S. griseus (which we term wbl\textsubscript{Sg}) had potential ribosome-binding sites. Their deduced products were 78% identical end-to-end, and the DNA flanking wbl\textsubscript{Sc} was closely similar to that in the S. griseus saf region, confirming that they were true orthologues. Frame plot analysis of the S. coelicolor sequence did not indicate that a saf-like gene was anti-
sense to \textit{wblI}$_{\text{Ac}}$: there was no suitable ribosome-binding site compared to that suggested for \textit{S. griseus}, and a stop codon was present in the middle of what might have been the \textit{safl}-like sequence of \textit{S. coelicolor}. We therefore suggest that the \textit{wblI} homologue of \textit{S. griseus} may be responsible for the effects in \textit{S. lividans} attributed to \textit{safl}.

The discovery of \textit{wblI} brought the total number of \textit{whiB} paralogues so far identified in \textit{S. coelicolor} to six (including \textit{whiB} itself).

**Evidence of transcription of \textit{wbl} genes**

To find out whether transcription occurred in \textit{wbl} genes during normal liquid culture (representing vegetative growth), a preparation of RNA isolated late in exponential growth of \textit{S. coelicolor} A3(2) strain M145 was evaluated by RT-PCR, using primers specific to each \textit{wbl} gene (Table 2). In each case, a product of the predicted size was readily obtained, whilst control experiments showed no signal (Fig. 3). Thus, all of the \textit{wbl} genes are transcribed.

**Evidence that the \textit{wbl} gene family is present in many different actinomycetes**

The presence of apparent orthologues of most of the \textit{wbl} genes in mycobacteria and \textit{Streptomyces} suggested that this gene family could be a general feature of actinomycetes. To test this, each \textit{wbl} gene (as well as \textit{whiB}) was used as a probe against Southern blots of DNA from representatives of nine other actinomycete genera as well as four other \textit{Streptomyces} spp. (Table 3). For nearly all DNA samples, at least one clearly hybridizing sequence was detected with most of the probes. Each sequence was located on a different restriction fragment, and each hybridized most strongly to one particular \textit{wbl} probe. Homologues of \textit{whiB} and all five \textit{wbl} genes were present in all four streptomycetes. There were two main exceptions to the general presence of all the \textit{wbl} genes among actinomycetes: \textit{Amycolatopsis mediterranei} hybridized strongly only with \textit{wblI}, and significant hybridization to \textit{wblII} was confined to the five \textit{Streptomyces} spp. tested. Undoubtedly, other \textit{wbl} genes remained undetected in this Southern blot analysis: for example, a gene (called here \textit{wblX}$_{\text{Ro}}$) encoding a protein with 34% identity to \textit{wblB} (and similarly distant from all known WhiB-like proteins) was discovered fortuitously in another actinomycete, \textit{Rhodococcus opacus}, during the study of a quite different nearby gene (Seibert et al., 1998). The \textit{wblX}$_{\text{Ro}}$ gene would not have been detected by any of the probes used in Table 3.

**DISCUSSION**

**Implications of \textit{whm}/\textit{wbl} genes for theories of the evolution and phylogeny of actinomycetes**

Although conventional searches have revealed no \textit{whiB}-like genes in any organisms other than actinomycetes, it is now clear that most actinomycetes contain several such genes. Molecular phylogenetic analysis (Fig. 4) of the genes discovered so far indicates that probable orthologues of at least four of the seven \textit{whm} genes found in \textit{M. tuberculosis} are present in \textit{Streptomyces}. A limited (but wide-ranging) Southern blot analysis indicated that orthologues of at least four of the \textit{wbl}/\textit{whm} paralogues (probes \textit{wblA}, \textit{wblB}, \textit{wblE}, \textit{whiB}) are also present in most actinomycetes (Table 3). This makes it highly likely that the last ancestor common to all these genera, which is estimated to have existed nearly one billion years ago (Embley & Stackebrandt, 1994), already possessed a set of these genes. Shared derived characters (synapomorphies) are the basis of phylogenetic theories, and \textit{wbl} genes can now join (at least) two other striking synapomorphies of actinomycetes: possession of a homologous approximately 100 nt insertion in their 23S rRNA genes (Roller et al., 1992), and utilization of the carbohydrate-modified cysteine derivative mycothiol as their major antioxidant thiol, instead of the tripeptide glutathione (Newton et al., 1996). It is therefore important for the further understanding of early bacterial evolution to establish whether the \textit{wbl} or mycothiol synapomorphies are shared with any of the non-actinomycete descendants of even earlier evolutionary radiations such as \textit{Thermotoga}, \textit{Thermus}, \textit{Deinococcus}, \textit{Thermomicrobium}, \textit{Atopobium}, \textit{Sphaerobacter} and \textit{Bifidobacterium} (the

---

**Fig. 4.** Phylogeny of \textit{whiB}-like proteins. The unrooted tree was constructed by using the SEQBOOT, PROTDIST, FITCH and CONSENSE programs from the PHYLIP software package. Nodes with bootstrap values >40 are indicated. Because only regions conserved among all WhiB-like proteins were included in the analysis, the \textit{Rhodococcus} \textit{WblX} protein (here termed ORF6 \textit{Ro}) does not cluster with any of the \textit{Whm}/\textit{Wbl} proteins, despite the apparent similarities in length and sequence between residues 33–48 of \textit{WblX}, the \textit{WhmB}/\textit{WblB} proteins and the \textit{Wbl} proteins. The analysis includes the \textit{WhiB}-like product of gene \textit{gp49} of the mycobacterial phage TM4 (Ford et al., 1998).
Fig. 5. Alignment of known Whi8-like proteins and their predicted structural features. Shown are the deduced amino acid sequences of 3 Whi8 proteins and 19 other Whi8-like proteins. In three cases, the choice of the N-terminal methionine residue of WmMt proteins differs from that suggested in the EMBL annotation of the *M. tuberculosis*.
order of this list reflects increasing closeness to actinomyces; Embley & Stackebrandt, 1994). Notably, the *wbl* family seems to be confined to *Streptomyces* spp. Possibly, this may indicate that it fulfils a role peculiar to *Streptomyces*.

It is interesting to note the presence of a *whiB*-like gene, gp49, in a recently sequenced mycobacterial phage genome (Tm4; Ford *et al*., 1998: Figs 4 and 5), which raises the possibility that comparatively recent phage-mediated horizontal transfer of *whiB*-like genes might have occurred; however, this could not readily account for the presence of similar sets of several such genes in nearly all actinomyces. Small transcriptional activator proteins containing four cysteine residues, albeit with quite different spacing (Cys-X2-Cys-X23-Cys-X4-Cys) from the array of conserved cysteines in the WhiB family (and showing no other similarities to WhiB) are specified by several phages of Gram-negative bacteria (Julien *et al*., 1998).

### Structure–function predictions for WhiB-like proteins

The WhiB-like proteins are aligned in Fig. 5. They are all small (from 87 to 130 amino acid residues). They all possess a high overall hydrophilicity suggestive of a cytoplasmic location. Although they all possess an overall negative charge, they have positively charged regions near their carboxy-termini. All the proteins are predicted to contain extensive α-helical structure with a central β-sheet region between the first and second α-helices. The predicted α-helical regions all show periodic conservation of residues consistent with a face-of-the-helix arrangement, such that the conserved features (mostly hydrophobic and including the four conserved cysteines) are on one face, and the non-conserved (more often hydrophilic) are on the other (the changed spacing of cysteine residues in WhmG$_{Mt}$ would still permit this arrangement). Such arrangements suggest that the conserved hydrophobic faces may be involved in interactions within the protein, and the more variable hydrophilic faces may make contact with other molecules in the surrounding cytoplasm. The most C-terminal α-helix, which always contains a segment rich in basic residues, is a candidate for DNA-binding (Suzuki, 1993), but probably with different targets among the different members of the family, since this α-helical region is not very highly conserved in primary sequence between paralogues (in contrast to its much greater conservation between Whm and Wbl orthologues). A short loop region between the two most C-terminal α-helical regions contains a few residues closely similar within orthologues but differing among paralogues, followed by a universally conserved turn region (GV/IWGG). This loop between α-helical regions is likely to be exposed, and the conserved turn region is a prime candidate for an interaction with another conserved cellular component (perhaps RNA polymerase, bound adjacent to a Wbl protein at a promoter).

The presence of four conserved cysteine residues suggests that the Whm/Wbl proteins may be sensitive to redox changes, perhaps through a bound metal atom or through direct sensitivity to oxidation via disulphide bond formation, as in the OxyR transcription factor of *E. coli* (Zheng *et al*., 1998). To date, these possibilities have been difficult to address experimentally because no biochemical activity or cellular target for the activity of any of the proteins has been defined. Nevertheless, the possession of these genes by actinomyces, and the reliance of actinomyces on mycothiol in place of glutathione, may be no coincidence; both may perhaps reflect a specialization coinciding with biologically very significant increases in oxygen concentration believed to have occurred in the earth’s atmosphere around 700–800 million years ago, and to have permitted the emergence of fully aerobic bacteria (Ochman & Wilson, 1987).

In summary, we speculate that the proteins are transcription factors that activate different target genes, possibly sensing redox changes that could originate from environmental change or be generated internally during metabolic shift-down (for example, at the end of vegetative growth or when aerial hyphae switch from growth to sporulation) (Nyström, 1999).

Studies of the effects of *wbl* mutations in the genetically manipulable *S. coelicolor* A3(2) may help to illuminate the functions of *whm* genes in pathogenic mycobacteria. A particular hope is that the epidemiologically important, but little understood phenomenon of mycobacterial dormancy within hosts may share regulatory features with the extensively studied process of sporulation in *Streptomyces*, in which *whiB* plays a very important part. Intriguingly, it has been demonstrated that *M. tuberculosis* is capable of achieving a dormant state in vitro following a gradual shift to anaerobic conditions (Lim *et al*., 1999). The presence of *whm* genes in *M. smegmatis*, a non-pathogen, suggests that the *whm* genes do not play a specific role in mycobacterial virulence, although indirect or contributory roles in virulence cannot be ruled out. At least one *whm* gene (*whmD$_{Ms}$*) is essential for colony formation by *M. smegmatis* (J. Gomez & W. R. Bishai, unpublished), whereas a very recent paper (Hutter & Dick, 1999)
reported that a whmB homologue of M. smegmatis did not affect growth or dormancy.

NOTE ADDED IN PROOF

Mulder et al. (1999) provide evidence that whmB_Mi (≡ whiB3) is expressed throughout growth, but expression from each of three promoters is growth-phase-dependent.

ACKNOWLEDGEMENTS

We thank J. A. Gil (Universidad de León, Spain) for plasmid pUL300 and helpful information about the saf gene; Tobias Kieser for help with preparing some of the figures; and Jose Aimsa, Mark Buttner, David Hopwood, Gabriella Kelemen, Tobias Kieser and Sebastien Mouz for comments on the manuscript. This work was supported in part by CICYT (BI095-1558-C02-02), Spain. J. Soliveri was supported by Ministerio de Educación y Ciencia, ‘Estancias de Investigadores Españoles en Centros de Investigación Extranjeros (PR95-382)’, Spain. Work at Johns Hopkins University was supported by NIH grants AI 36973 and AI 37856. Work at the John Innes Centre was funded by a competitive strategic grant from the Biotechnological and Biological Sciences Research Council and by the John Innes Foundation.

REFERENCES


Fshi, H., De Rossi, E., Salazar, L. & 7 other authors (1996). Gene arrangement and organization in a ~ 76 kb fragment encompassing the orf C region of the chromosome of Mycobacterium leprae. Microbiology 142, 3147–3161.


Janssen, G. R. & Bibb, M. J. (1993). Derivatives of pUC18 that have BglII sites flanking a modified multiple cloning site and that retain the ability to identify recombinant clones by visual screening of Escherichia coli colonies. Gene 124, 133–134.


Neidhardt and others. Washington, DC: American Society for Microbiology.


Received 21 July 1999; revised 15 October 1999; accepted 27 October 1999.