

# Analysis of post-translational modification of mycobacterial proteins using a cassette expression system

Jean Louis Herrmann\*, Robin Delahay<sup>1</sup>, Alex Gallagher<sup>2</sup>, Brian Robertson, Douglas Young

Department of Infectious Diseases and Microbiology, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place, London W2 1PG, UK

Received 11 March 2000

Edited by Pierre Jolles

**Abstract** A recombinant expression system was developed to analyse sequence determinants involved in *O*-glycosylation of proteins in mycobacteria. By expressing peptide sequences corresponding to known glycosylation sites within a chimeric lipoprotein construct, amino acids flanking modified threonine residues were found to have an important influence on glycosylation. The expression system was used to screen mycobacterial sequences selected using a neural network (NetOglyc) trained on eukaryotic *O*-glycoproteins. Evidence of glycosylation was obtained for eight of 11 proteins tested. The results suggest that sites involved in *O*-glycosylation of mycobacterial and eukaryotic proteins share similar structural features.

© 2000 Federation of European Biochemical Societies.

**Key words:** Mycobacterium; *O*-Glycoprotein; Post-translational modification; Tuberculosis

## 1. Introduction

Post-translational modification of proteins by covalent attachment of carbohydrate moieties is common amongst eukaryotes but, by comparison, relatively rare in bacteria. Bacterial glycoproteins were initially characterised from the S-layer of the cell wall of archaeobacteria [1], and their occurrence in eubacteria is now well-established. Several of the best-characterised prokaryotic glycoproteins are expressed by pathogens, prompting the suggestion that they might contribute to interactions with mammalian cells during infection [2]. Examples include pili present on the surface of pathogenic *Neisseria* [3,4] and antigens secreted into the growth medium of *Mycobacterium tuberculosis* [5–8]. Deglycosylation of a 45 kDa glycoprotein antigen of *M. tuberculosis* decreased its immunogenicity, possibly as a consequence of reducing interaction with carbohydrate recognition receptors on antigen-presenting cells [9], and mutagenesis of a proposed *O*-glycosylation site in an *M. tuberculosis* 19 kDa antigen promoted proteolytic conversion from a cell-associated lipoprotein to a non-acylated secreted antigen [10]. The conversion between acylated and non-acylated forms is of particular interest in

light of the recent observation that bacterial lipoproteins trigger cytokine signalling via toll-like receptors on mammalian cells [11,12]. In the context of the recent completion of the *M. tuberculosis* genome sequence [13], we were interested in developing techniques for the identification of further mycobacterial glycoproteins with a view to assessment of their possible role in infection. The aim of the present study was to investigate the possibility of defining common structural features that might assist in the prediction of *O*-glycosylation sites in mycobacterial proteins.

In contrast to the well-defined motif for N-linked glycosylation sites (Asn-X-Ser/Thr), it has proved difficult to define a consensus sequence for *O*-glycosylation of eukaryotic proteins. Strategies that have been used to address this question include the development of in vitro assays using synthetic peptides mixed with purified glycosyl transferases [14–16], recombinant expression systems measuring in vivo glycosylation of peptide cassettes inserted in chimeric reporter proteins [17] and the compilation of sequence data from known *O*-glycoproteins [18]. From these studies, it has been shown that glycosylation sites are characterised by threonine or serine residues flanked by sequences rich in proline, glycine and alanine residues. The flanking sequences were found to have an important, though complex effect on glycosylation activity. A neural network, NetOglyc, was developed as a predictive tool by initial training on a set of almost 200 known *O*-glycosylation sites [19]. Threonine residues implicated in *O*-glycosylation of the *M. tuberculosis* 45 kDa [5] and 19 kDa [10] antigens have a high predictive score using NetOglyc. To evaluate applicability of the eukaryotic algorithm as a starting point for the definition of a mycobacterial *O*-glycosylation motif, we have established a mycobacterial expression system, analogous to the eukaryotic system described by Nehrke et al. [17], in which we could assess the modification of peptide cassettes inserted in a recombinant lipoprotein backbone.

## 2. Materials and methods

### 2.1. Construction of *PhoA* fusions

The mycobacterial expression vector pCAS was prepared by ligation of the promoter region, signal sequence and two N-terminal amino acids of the *M. tuberculosis* 19 kDa antigen [20] to the *phoA* gene of *Escherichia coli* lacking its own signal sequence, using procedures described previously [10]. A *Bam*HI linker was included to facilitate insertion of peptide cassettes for glycosylation analysis.

Three approaches were used to insert cassettes into pCAS. For short inserts (up to 15 amino acids), complementary oligonucleotides matching the appropriate mycobacterial sequences were designed with flanking *Bam*HI linkers and ligated into *Bam*HI-digested pCAS, essentially as described by Janssen and Tommassen [21]. For longer inserts, 120 bp fragments were amplified from *M. tuberculosis* genomic DNA by polymerase chain reaction (PCR). Cycling conditions (95°C

\*Corresponding author. Present address: Laboratoire de Microbiologie, Centre Hospitalier, Universitaire Saint-Louis, 1 Avenue Claude-Vellefaux, 75475 Paris Cédex 10, France. Fax: (33)-1-42 49 9200. E-mail: jlh@chu-stlouis.fr

<sup>1</sup> Present address: Department of Biochemistry, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place, London W2 1PG, UK.

<sup>2</sup> Present address: MRC Molecular Pathogenesis Group, 32 Newark Street, London E1 2AA, UK.

for 54 s, 68°C for 30 s, 72°C for 30 s; 25 cycles) were optimised using forward and reverse oligonucleotide primers incorporating flanking *Bam*HI restriction sites. PCR products were subsequently digested, purified and ligated into pCAS following standard cloning methodology [22].

Ligation mixtures were transformed in *E. coli* DH5 $\alpha$ , grown in Luria–Bertani (LB) broth or agar medium with selection in the presence of hygromycin (Boehringer Mannheim) at 250  $\mu$ g/ml, and plasmid preparations were obtained using standard methodology [22]. The identity of each construct was confirmed by sequence analysis using an ABI 310 Genetic Analyzer and ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kits (Perkin Elmer Applied Biosystems Division). Plasmids were subsequently introduced into *Mycobacterium smegmatis* mc<sup>2</sup>/11-2c by electroporation, using mycobacteria grown in LB broth or on agar medium, with 0.05% tyloxapol (Sigma) and 50  $\mu$ g/ml hygromycin [23]. For visualisation of the alkaline phosphatase activity of recombinant colonies, mycobacteria were grown on similar plates supplemented with 5-bromo-4-chloro-3-indolyl-phosphate-*p*-toluidine (BCIP; Sigma) at 40  $\mu$ g/ml. Positive colonies were dark blue.

## 2.2. Analysis of concanavalin A (ConA)-binding

For electrophoretic analysis, recombinant *M. smegmatis* cultures were grown for 3 days in LB medium containing tyloxapol and hygromycin, harvested by centrifugation, resuspended in 1/50 of the original volume of phosphate-buffered saline (PBS) and sonicated on ice in bursts of 30 s over a 5 min period at 18  $\mu$ m amplitude (Soniprep 150, MSE). Alternatively, bacteria from a single colony were suspended in 200  $\mu$ l of PBS and sonicated. Sonicated preparations were boiled in the presence of sodium dodecyl sulphate (SDS) and  $\beta$ -mercaptoethanol and samples (2–4  $\mu$ g total protein) were analysed by SDS–polyacrylamide gel electrophoresis, followed by transfer to nitrocellulose membranes (Sartorius). The binding of peroxidase-conjugated ConA (Sigma) and recognition by rabbit polyclonal antibodies against PhoA (kindly provided by Juliano Timm, Institut Pasteur) were visualised as described previously [10]. Densitometric analysis of autoradiographs was performed with Scan Analysis software (Biosoft) installed on a Macintosh computer using a Studio Scan II flatbed scanner (Agfa).

## 2.3. DNA sequence analysis

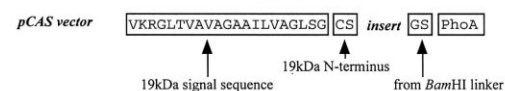
Sequence information from the *M. tuberculosis* H37Rv genome project was searched using the Prosite motif for prokaryotic lipoproteins (PS00013). Potential lipoproteins were further checked using the Psort programme (<http://www.psort.nibb.ac.jp>) and open reading frames predicted to encode lipoproteins were then analysed using the NetOglyc neural network for the identification of *O*-linked glycosylation sites (<http://www.cbs.dtu.dk/netOglyc/>) [19].

## 3. Results

### 3.1. A cassette system for identification of glycosylation sites

A DNA fragment encoding the mature form of *E. coli* alkaline phosphatase (PhoA) was fused in frame with a fragment encoding the signal sequence and two N-terminal amino acids from the *M. tuberculosis* 19 kDa antigen, generating the pCAS construct (Fig. 1A). Inclusion of the PhoA reporter facilitated identification of positive transformants as blue colonies on BCIP agar. Two additional amino acids (glycine and serine) were also present, as a result of insertion of a *Bam*HI linker between the two fragments. Introduction of this construct into *M. smegmatis* resulted in the expression of a PhoA fusion protein which showed no binding to the lectin ConA (Fig. 1B, lane 1). As previously reported [10], inclusion of a longer fragment of the 19 kDa N-terminal region-containing threonine clusters at positions 13–15 and 19–20-generated a ConA-positive fusion protein (Fig. 1B, lane 2). *Mycobacterium avium* and *Mycobacterium intracellulare* express proteins that share more than 70% identity with the *M. tuberculosis* 19 kDa antigen, but with serine-rich sequences in place of the

### A. Sequence of pCAS fusions



lane	construct	insert source (no insert)	insert sequence
1	pCAS		
2	pCAS-19-TB	<i>M. tuberculosis</i> 19kDa antigen	GSGETTTAAGTTASPG
3	pCAS-19-MA	<i>M. avium</i> 19kDa antigen	SAPSSASSSSSTSPSASS
4	pCAS-19-MI	<i>M. intracellulare</i> 19kDa antigen	SGTSSASSASSSGTSASAT
5	pCAS-45i	<i>M. tuberculosis</i> 45kDa antigen	DPEAPPPVPTTAAS
6	pCAS-45ii	<i>M. tuberculosis</i> 45kDa antigen	PPVPTTAAS
7	pCAS-45iii	<i>M. tuberculosis</i> 45kDa antigen	PTTAAS
8	pCAS-45iv	<i>M. tuberculosis</i> 45kDa antigen	TTAAS

### B. ConA binding to pCAS fusions

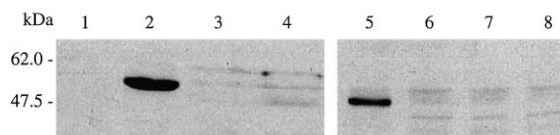


Fig. 1. Construction of pCAS fusions and assay for lectin-binding. (A) shows the sequence of the pCAS vector, together with peptides from 19 and 45 kDa antigens used to construct pCAS fusions. (B) illustrates ConA-binding of the fusion constructs listed in (A). In contrast to the threonine-rich N-terminal region of the *M. tuberculosis* 19 kDa antigen, serine-rich sequences from *M. avium* and *M. intracellulare* homologues were negative for ConA-binding. A peptide spanning a known glycosylation site of the 45 kDa antigen was positive for ConA-binding, dependent on inclusion of appropriate flanking residues.

threonine-rich N-terminal region [24,25]. PhoA fusion proteins containing the serine-rich sequences were both ConA-negative in the recombinant expression system (Fig. 1B, lanes 3 and 4).

We next examined the effect of inserting peptides spanning glycosylated threonine residues from the *M. tuberculosis* 45 kDa antigen in the pCAS vector. The relevant sequences were inserted using synthetic oligonucleotides designed to include *Bam*HI linkers. Fig. 1 shows the sequences of the fusion proteins and the results of ConA-binding analysis. Insertion of 13 N-terminal amino acids from the 45 kDa sequence generated a ConA-positive fusion. Fusion proteins with shorter inserts (5–10 amino acids) were ConA-negative, in spite of the fact that each contained threonine-10, a residue known to be glycosylated in the native protein [5]. Thus, the glycosylation site from the 45 kDa antigen can be substituted for the equivalent 19 kDa region in the chimeric lipoprotein but, as observed in earlier site-directed mutagenesis experiments [10], amino acids flanking the modified residues play an important role in glycosylation.

### 3.2. Prediction of *O*-glycosylation sites in mycobacterial lipoproteins

The ability to exchange glycosylation sequences between mycobacterial proteins in the cassette expression system suggests the possibility of a shared glycosylation motif. Threonine residues within the positive 45 and 19 kDa sequences are flanked by alanine and proline or glycine residues in a pattern resembling that found in *O*-glycosylation sites in eukaryotic proteins. The glycosylated threonines in the 45 kDa sequence and the threonine residues implicated in ConA-binding of the

19 kDa antigen each had a high predictive score when screened using the NetOglyc algorithm for eukaryotic *O*-glycosylation sites [19]. To test whether the analogy with the eukaryotic consensus could be extended to further mycobacterial proteins, we used NetOglyc to select additional predicted glycosylation sequences from the *M. tuberculosis* genome [13].

Threonine residues with the potential to act as glycosylation sites occur at a frequency of approximately 10 per 1000 amino acids within open reading frames in the *M. tuberculosis* genome. To restrict the NetOglyc screen to a manageable output, and with the aim of further exploring the glycosylation–acylation link observed with the 19 kDa antigen [10], an initial search was focused on lipoproteins. From 98 lipoprotein sequences in the *M. tuberculosis* genome (identified by the presence of a prokaryotic lipoprotein lipid attachment motif), 35 had two or more positively-predicted threonines within the N-terminal 40 amino acids. These are listed in Table 1.

### 3.3. Expression of predicted glycosylation sites

To evaluate the predictive value of the NetOglyc analysis, sequences from a subset of the predicted glycosylated lipoproteins were expressed in the PhoA cassette system and assayed for ConA-binding. Two PhoA fusion constructs were prepared using fragments from PhoS1, the 38 kDa antigen [26]. 38S-PhoA contained amino acids 15–31, spanning the NetOglyc-predicted threonines at positions 20, 21 and 28, as an

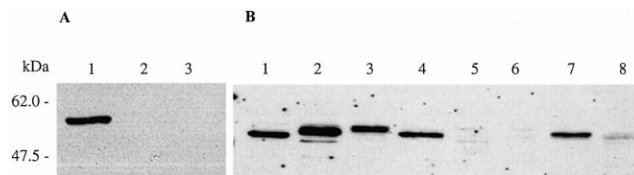


Fig. 2. Testing of NetOglyc predictions by ConA-binding. Sequences predicted using NetOglyc were expressed in the pCAS vector and screened for ConA-binding. (A) shows results with sequences from PhoS1, the 38 kDa antigen. pCAS-38i (lane 1) contained 42 amino acids from the N-terminus of the protein. Lanes 2 and 3 show negative constructs with shorter inserts: pCAS-38ii (PETGAGAGT-VATTP) and pCAS-38iii (GAGTVATTPASSPVTLA). (B) illustrates pCAS fusions containing 40 amino acid inserts from predicted lipoproteins. Lane 1, pCAS-19-TB (positive control); 2, LppN; 3, SodC; 4, LppQ; 5, LppB; 6, LppT; 7, GlnH and 8, LprI. Samples were considered negative when ConA-binding (assessed by scanning densitometry) was less than 10% of that in the positive control.

insert in pCAS. 38L-PhoA contained a longer fragment (43 amino acids) from the N-terminal region, as well as the native signal sequence of the 38 kDa protein. As had been observed in the case of the 45 kDa antigen, the longer fusion protein bound ConA but the short construct was negative (Fig. 2A). A further eight constructs were prepared from open reading frames containing from two to six NetOglyc-positive threonines. In each case the entire 40 amino acid N-terminal fragment was expressed in pCAS, with the level

Table 1  
*M. tuberculosis* genes encoding lipoproteins with predicted glycosylation sites

Genome location	Name	Description	NetOglyc hits
Rv0583c	LpqN	similar to <i>M. kansasii</i> antigen MK35	13
Rv0344c	LpqJ	unknown lipoprotein	10
<b>Rv3763</b>	<b>LpqH</b>	<b>19 kDa antigen</b>	<b>7</b>
Rv1228	LpqX	unknown lipoprotein	6
<b>Rv2270</b>	<b>LppN</b>	<b>unknown lipoprotein</b>	<b>6</b>
Rv1911c	LppC	lipoprotein, similar to Rv1910c	5
<b>Rv2341</b>	<b>LppQ</b>	<b>unknown lipoprotein</b>	<b>5</b>
Rv1016c	LpqT	similar to <i>M. kansasii</i> antigen MK35	5
Rv1899c	LppD	4-hydroxybutyrate dehydrogenase	4
<b>Rv2873</b>	<b>MPT83</b>	<b>antigen, related to MPT70</b>	<b>4</b>
Rv1899c	LpqI	beta-hexosaminidase precursor	4
Rv0835	LpqQ	unknown lipoprotein	4
Rv0838	LpqR	D-Ala/D-Ala dipeptidase	4
Rv1166	LpqW	antigen ORFA2-898	4
Rv1252c	LprE	unknown, similar to Rv3483c	4
Rv1922		probable penicillin-binding protein	4
<b>Rv0432</b>	<b>SodC</b>	<b>Cu–Zn superoxide dismutase precursor</b>	<b>4</b>
Rv2672		protease, similar to Rv2223c,2224c	3
Rv2330c	LppP	unknown lipoprotein	3
Rv3006	LppZ	unknown lipoprotein	3
<b>Rv0934</b>	<b>PhoS1</b>	<b>38 kDa antigen, phosphate uptake</b>	<b>3</b>
Rv0237	LppI	unknown lipoprotein	3
Rv1677	DsbF	thiol:disulphide interchange similar to antigen MPT53	3
Rv2394	GgtB	γ-glutamyltranspeptidase precursor	3
<b>Rv0411c</b>	<b>GlnH</b>	<b>glutamine-binding protein</b>	<b>2</b>
Rv2068c	BlaC	class A β-lactamase	2
Rv1881c	LppE	unknown lipoprotein	2
Rv2116	LppK	unknown lipoprotein	2
<b>Rv1541c</b>	<b>LprI</b>	<b>unknown lipoprotein</b>	<b>2</b>
Rv2138	LppL	unknown lipoprotein	2
Rv0399c	LpqK	possible penicillin-binding protein	2
Rv0132c		putative oxidoreductase	2
Rv1274	LprB	unknown lipoprotein	2
Rv1368	LprF	lipoprotein, similar to LprA, LprG, LppX	2
Rv1418	LprH	unknown lipoprotein	2

The number of NetOglyc-positive threonine residues is shown for the first 40 amino acids after the predicted acylation site. For entries shown in bold, there is evidence of post-translational modification of expressed products.

amino acid sequence	protein	ConA
DPEPAPPVPT TAAS .....T T...	45kDa antigen	+
GSGETTTAAG TTASPG .....TTT... TT....	19kDa antigen	+
CSSNGARGGI ASTNMNPTNP PATAETATVS PTPAQOSART .....T... ..T..T.T... ..T.....T	LppN	+
CGRSAGKTSE PTCPTPEIDA ADSSTTPDPS CVVRATEING .....T... ..T..T... ..TT.....	LppQ	+
CGPKSPDFQS ILSTSPPTSA VSTTTEVPVP LMKYLESVGV .....T... ..T..T... ..TT.....	LpqT	-
CSSTKPVSDQ TSPKPATSPA APVTTAAMAD PAADLIGRGC .....T... ..T..T... ..TT.....	MPT83	+
CTGQHTTIRA ASTTYTPHIK ASSQDVLDA INADEPGCSA .....T.T... ..T..T... ..TT.....	Rv1922	-
CSSPQHASTV PGTTPSIWGT SPAPSGLSGH DEESPGAQSL .....T... ..T..T... ..TT.....	SodC	+
CGSKPPSGSP ETGAGAGTVA TTPASSPVTL AETGSLLYP .....T... ..T..T... ..TT.....	PhoS1 (38kDa antigen)	+
CGHSETLGVE ATPILPLPTP VGMEIMPQP PLPPDSSSDQ .....T.T... ..T..T... ..TT.....	GlnH	+
CAANPPANIT SPTAGQSLDC TKPATIVQQL VCHDRQLTSL .....T... ..T..T... ..TT.....	LprI	+/-
CSDSGDNKPG ATIPSTPANA EGRHGPFFPQ CGGVSDQTVT .....T... ..T..T... ..TT.....	LprB	-

Fig. 3. Summary of NetOglyc predictions and ConA-binding. Peptide sequences containing potential glycosylation sites predicted by NetOglyc were expressed as pCAS fusions and tested for lectin-binding. Sequences are shown on the top line, with threonine residues matching the NetOglyc prediction indicated on the lower line.

of expression of each fusion assessed by Western blot analysis with an antibody to PhoA (data not shown). Strong ConA-binding was observed with four of the fusions (Fig. 2B). A fifth fusion, LprI, was expressed at low level and showed weak but detectable binding (Fig. 2B, lane 8). The remaining three fusions were clearly detectable by PhoA staining but failed to bind ConA. In a separate study (unpublished results in collaboration with S. Michell and G. Hewinson), the amino-terminal region of MPT83 was found to confer ConA-binding when expressed in pCAS. Overall, from 11 of the lipoprotein sequences predicted by NetOglyc, eight were found to be ConA-positive in the pCAS expression system (Fig. 3).

#### 4. Discussion

Although abundant in eukaryotic cells, glycoproteins are rare in bacteria and relatively little is known about the mechanisms and specificity of glycosylation in prokaryotes. In the present study we have tested whether a neural network trained on eukaryotic *O*-glycoproteins can be applied in the study of protein glycosylation in mycobacteria. The results indicate

that the eukaryotic neural network does provide a useful starting point for characterisation of mycobacterial glycosylation, with the majority of predicted regions proving positive in a lipoprotein expression system assessed by lectin-binding. This finding suggests that, despite the absence of an organised Golgi structure in prokaryotes, the fundamental enzymology of *O*-glycosylation in mycobacteria probably resembles that in eukaryotic cells. The absence of glycosylation in constructs in which serine-rich regions replaced the threonine-rich domain of the 19 kDa antigen suggest that, at least in the *M. smegmatis* expression system, threonine is the preferred residue for mycobacterial *O*-glycosylation.

There is considerable scope for further optimisation of a predictive approach for bacterial glycoprotein sequences. Several of the tested sequences were negative and the high frequency of *M. tuberculosis* open reading frames containing potential glycosylation sites vastly exceeds the likely number of glycoproteins. It can be anticipated that training of the neural network on bacterial *O*-glycoprotein sites would significantly enhance predictive potential, although this will be dependent on significant expansion of the currently available structural database for bacterial glycoproteins. Recombinant products of expression systems such as the pCAS vector described in the present study may provide a useful source of substrates for the necessary structural analysis. Regions of the protein surrounding potentially modified residues clearly have a major impact on glycosylation. The length of the peptide insert had a critical influence in the pCAS expression system, with peptides below 10 amino acids showing no evidence of modification, in spite of the inclusion of a NetOglyc-positive motif. The protein framework used for expression was also important. When 19 kDa peptides positive in pCAS fusions were cloned in the permissive loop of *M. tuberculosis* superoxide dismutase [27], there was no evidence of glycosylation, as assessed by lectin-binding or by mass spectrometry analysis (data not shown). It is probable that protein glycosylation in mycobacteria depends not just on the presence of an appropriate motif, but also on localisation of the protein to a particular part of the bacterial cell.

Sequence alignments of the predicted glycosylated lipoproteins with related open reading frames from *M. tuberculosis* itself or with functional homologues from other bacteria frequently highlight the threonine-rich domain as a divergent region. This is particularly clear when comparing MPT83 with the related MPT70 antigen, with the threonine-rich region standing out as an apparent 32 amino acid insertion (Fig. 4). Glycosylation domains in *M. tuberculosis* may have evolved together with gene duplication as a means for expressing functionally related proteins differing in their intracellular location. Conversion between cell-associated and secreted forms has previously been documented for lipoproteins produced by Gram-positive bacteria [28,29] and glycosylation may provide a means for regulating this in mycobacteria. The possible role of glycosylation in host cell interactions is illustrated by links between the panel of predicted glycoproteins and the immune repertoire. In addition to the 19 kDa antigen, the NetOglyc search identified PhoS1 (the 38 kDa antigen of *M. tuberculosis* [26]) and MPT83, an antigen extensively analysed in the context of bovine tuberculosis [30]. The thiol-disulphide exchange protein, DsbF, is closely related to the previously described antigen MPT53 from *M. tuberculosis* [31] and LpqW was recently identified during a screen

(N-terminus of mature proteins)	
MKVKNITIAAT SPAAGLAAL AVAVSPAAA G.....DLVGPQC	*****
MV QAKPAAAAAL AAIAIAFLAG CSSTKPVSDQ TSPKPATSPA APVTTAAMAD PAADLIGRGC	*****
AEYAAANPTG PASVQMSQD PVAVASNNP ELTTLTAALS GQLNPQVNLV DTLSNGQYTV FAPNTAAFSK	*****
AQYAAQNPTG PGSVAGMAQD PVATAASNNP MLSTLTALS GKLNPDVNLV DTLSNGEYTV FAPNTAAFDK	*****
LPASTIDELK TNSLLTSLI TYHVAGQTS PANVVGTRQT LQASVTVTG QGNSLKVNGA DVVCGGVSTA	*****
LPAAITDQLK TDA.LLSSIL TYHVIAQAS PSRIDGTHQT LQADLTIVG ARDDLNVNNA GLVCGGVHTA	*****
NATVYMIDSV LMPPA	Rv2875 (MPT70)
NATVYMIDTV LMPPAQ	Rv2873 (MPT83)

Fig. 4. Sequence alignment of MPT83 and MPT70. The threonine-rich region predicted as a potential glycosylation site in MPT83 has no counterpart in the closely related MPT70 antigen.

for mycobacterial secreted antigens [32]. LpqN and LpqT show sequence similarity both to each other and to a prominent antigen from *Mycobacterium kansasii* [33].

In summary, this study highlights similarities between *O*-glycosylation of proteins in bacterial and eukaryotic cells. The use of a predictive algorithm in combination with a recombinant expression system provides an approach to the identification of mycobacterial *O*-glycoproteins, generating results consistent with the existence of a family of glycosylated lipoproteins in *M. tuberculosis*.

**Acknowledgements:** This work was funded by the Glaxo Wellcome Action TB initiative. Support was also provided by a Wellcome Trust Research Career Development Fellowship (B.R.) and a Wellcome Trust Programme Grant (D.Y.). We are grateful to Dr. Ken Duncan and to Dr. Peadar O'Gaora for helpful discussions.

## References

- [1] Messner, P. and Sleytr, U.B. (1991) *Glycobiology* 1, 545–551.
- [2] Tuomanen, E.I. (1996) *J. Clin. Invest.* 98, 2659–2660.
- [3] Stimson, E. et al. (1996) *Biochem. J.* 316, 29–33.
- [4] Virji, M. (1997) *Gene* 192, 141–147.
- [5] Dobos, K.M., Khoo, K.H., Swiderek, K.M., Brennan, P.J. and Belisle, J.T. (1996) *J. Bacteriol.* 178, 2498–2506.
- [6] Espitia, C. and Mancilla, R. (1989) *Clin. Exp. Immunol.* 77, 378–383.
- [7] Ffifis, T., Costopoulos, C., Radford, A.J., Bacic, A. and Wood, P.R. (1991) *Infect. Immun.* 59, 800–807.
- [8] Garbe, T., Harris, D., Vordermeier, M., Lathigra, R., Ivanyi, J. and Young, D. (1993) *Infect. Immun.* 61, 260–267.
- [9] Romain, F., Horn, C., Pescher, P., Namane, A., Riviere, M., Puzo, G., Barzu, O. and Marchal, G. (1999) *Infect. Immun.* 67, 5567–5572.
- [10] Herrmann, J.L., O'Gaora, P., Gallagher, A., Thole, J.E. and Young, D.B. (1996) *EMBO J.* 15, 3547–3554.
- [11] Aliprantis, A.O. et al. (1999) *Science* 285, 736–739.
- [12] Brightbill, H.D. et al. (1999) *Science* 285, 732–736.
- [13] Cole, S.T. et al. (1998) *Nature* 393, 537–544.
- [14] Gooley, A.A., Classon, B.J., Marschalek, R. and Williams, K.L. (1991) *Biochem. Biophys. Res. Commun.* 178, 1194–1201.
- [15] O'Connell, B.C., Hagen, F.K. and Tabak, L.A. (1992) *J. Biol. Chem.* 267, 25010–25018.
- [16] Wilson, I.B., Gavel, Y. and von Heijne, G. (1991) *Biochem. J.* 275, 529–534.
- [17] Nehrke, K., Hagen, F.K. and Tabak, L.A. (1996) *J. Biol. Chem.* 271, 7061–7065.
- [18] Gupta, R., Birch, H., Rapacki, K., Brunak, S. and Hansen, J.E. (1999) *Nucleic Acids Res.* 27, 370–372.
- [19] Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) *Glycoconj. J.* 15, 115–130.
- [20] Ashbridge, K.R., Booth, R.J., Watson, J.D. and Lathigra, R.B. (1989) *Nucleic Acids Res.* 17, 1249.
- [21] Janssen, R. and Tommassen, J. (1996) in: *Methods in Molecular Medicine: Vaccine Protocols* (Robinson, R., Farrar, G.H. and Wiblin, C.N., Eds.), pp. 65–74. Humana Press, Totowa, NJ.
- [22] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [23] Garbe, T.R., Barathi, J., Barnini, S., Zhang, Y., Abou-Zeid, C., Tang, D., Mukherjee, R. and Young, D.B. (1994) *Microbiology* 140, 133–138.
- [24] Booth, R.J., Williams, D.L., Moudgil, K.D., Noonan, L.C., Grandison, P.M., McKee, J.J., Prestidge, R.L. and Watson, J.D. (1993) *Infect. Immun.* 61, 1509–1515.
- [25] Nair, J., Rouse, D.A. and Morris, S.L. (1992) *Mol. Microbiol.* 6, 1431–1439.
- [26] Andersen, A.B., Ljungqvist, L. and Olsen, M. (1990) *J. Gen. Microbiol.* 136, 477–480.
- [27] Hetzel, C. et al. (1998) *Infect. Immun.* 66, 3643–3648.
- [28] Jenkinson, H.F. (1995) *Trends Microbiol.* 3, 333–335.
- [29] Navarre, W.W., Daefler, S. and Schneewind, O. (1996) *J. Bacteriol.* 178, 441–446.
- [30] Hewinson, R.G., Michell, S.L., Russell, W.P., McAdam, R.A. and Jacobs, W.J. (1996) *Scand. J. Immunol.* 43, 490–499.
- [31] Wiker, H.G., Michell, S.L., Hewinson, R.G., Spierings, E., Nagai, S. and Harboe, M. (1999) *Microb. Pathog.* 26, 207–219.
- [32] Moran, A.J. et al. (1999) *FEMS Microbiol. Lett.* 177, 123–130.
- [33] Armoa, G.R., Rouse, D.A., Nair, J., Mackall, J.C. and Morris, S.L. (1995) *Microbiology* 141, 2705–2712.