Tuberculosis (2008) **88**, 526–544

available at www.sciencedirect.com

ScienceDirect

journal homepage: http://intl.elsevierhealth.com/journals/tube

Identification of outer membrane proteins of *Mycobacterium tuberculosis*

Houhui Song [a,d], Reatha Sandie [b,d], Ying Wang [a,d], Miguel A. Andrade-Navarro [b,c], Michael Niederweis [a,*]

[a] *Department of Microbiology, University of Alabama at Birmingham, 609 Bevill Biomedical Research Building, 845 19th Street South, Birmingham, AL 35294, USA*
[b] *Bioinformatics Group, Ottawa Health Research Institute, 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada*
[c] *Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany*

Received 13 September 2007; received in revised form 8 February 2008; accepted 18 February 2008

KEYWORDS
Secondary structure;
Prediction;
Amphiphilicity;
Beta-strand;
Exported;
Inner membrane;
Periplasmic;
Secreted proteins

Summary
The cell wall of mycobacteria includes an unusual outer membrane of extremely low permeability. While *Escherichia coli* uses more than 60 proteins to functionalize its outer membrane, only two mycobacterial outer membrane proteins (OMPs) are known. The porin MspA of *Mycobacterium smegmatis* provided the proof of principle that integral mycobacterial OMPs share the β-barrel structure, the absence of hydrophobic α-helices and the presence of a signal peptide with OMPs of gram-negative bacteria. These properties were exploited in a multi-step bioinformatic approach to predict OMPs of *M. tuberculosis*. A secondary structure analysis was performed for 587 proteins of *M. tuberculosis* predicted to be exported. Scores were calculated for the β-strand content and the amphiphilicity of the β-strands. Reference OMPs of gram-negative bacteria defined threshold values for these parameters that were met by 144 proteins of unknown function of *M. tuberculosis*. Two of them were verified as OMPs by a novel two-step experimental approach. Rv1698 and Rv1973 were detected only in the total membrane fraction of *M. bovis* BCG in Western blot experiments, while proteinase K digestion of whole cells showed the surface accessibility of these proteins. These findings established that Rv1698 and Rv1973 are indeed localized in the outer membrane and tripled the number of known OMPs of *M. tuberculosis*. Significantly, these results provide evidence for the usefulness of the bioinformatic approach to predict mycobacterial OMPs and indicate that *M. tuberculosis* likely has many OMPs with β-barrel structure. Our findings pave the way to identify the set of proteins which functionalize the outer membrane of *M. tuberculosis*.

Abbreviations: OM, outer membrane; OMP, outer membrane protein; IM, inner membrane; IMP, inner membrane protein; wt, wild-type.

* Corresponding author. Tel.: +1 205 996 2711; fax: +1 205 934 9256.
E-mail address: mnieder@uab.edu (M. Niederweis).
[d] These authors contributed equally to this work.

1472-9792/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tube.2008.02.004

Introduction

The cell wall of mycobacteria is an intriguingly complex structure consisting of a great variety and a large amount of lipids.[1,2] Very long-chain fatty acids, the mycolic acids, are covalently bound to the arabinogalactan-peptidoglycan co-polymer and were proposed to form the inner layer of an asymmetric outer membrane while other lipids constitute the outer leaflet.[3] X-ray diffraction studies showed that the mycolic acids are indeed oriented in parallel and perpendicular to the plane of the cell envelope.[4] The presence of a second lipid bilayer outside of the cytoplasmic membrane in mycobacterial species has recently been visualized for the first time in a near native state by cryo-electron microscopy.[5] The discovery and the analysis of the porin MspA of *Mycobacterium smegmatis* provided the first conclusive evidence that functionally similar, but structurally completely different outer membrane proteins (OMPs) exist also in mycobacteria.[6–9] Despite the well-documented importance of OMPs for the import of nutrients, secretion processes and host–pathogen interactions in gram-negative bacteria,[10] surprisingly few OMPs of mycobacteria are known. The only two well-characterized examples of integral OMPs are the porin MspA of *M. smegmatis* and the channel-forming protein OmpA of *M. tuberculosis*.[11–14] By contrast, *E. coli* uses more than 60 proteins to functionalize its outer membrane,[15] none of which has significant sequence similarity to any *M. tuberculosis* protein.

Traditionally, OMPs have been discovered by isolating the cell envelope and then separating the inner from the outer membrane in sucrose gradients.[16–18] Due to the covalent linkages between the peptidoglycan, the arabinogalactan and the mycolic acid layer,[1] it is difficult to mechanically lyse mycobacterial cells.[19] This is usually achieved only by harsh conditions, which inadvertently leads to mixing of components of both membranes. This has hampered localization experiments and identification of *M. tuberculosis* OMPs so far.[20] Bioinformatic analysis of the genome of *M. tuberculosis* provides an alternative strategy, but OMPs are more difficult to identify by the amino acid sequence than inner membrane proteins, whose hydrophobic α-helices are predicted with accuracies exceeding 99%.[21] So far, all known OMPs are β-barrel proteins, which are characterized by a pattern of alternating hydrophobic and hydrophilic amino acids in the β-strands forming the β-barrel.[22] Such a pattern is recognizable in the protein sequences and has been exploited in recent years to develop programs for prediction of β-barrel proteins. For example, more than 10 previously unknown OMPs were predicted for *E. coli*.[23] Notably, a consensus method performed better than each individual prediction method for a set of 20 β-barrel OMPs whose structures are known at atomic resolution.[24] The success of these approaches motivated the application of one of these algorithms to predict OMPs of *M. tuberculosis*.[25] However, the usefulness of this analysis is limited for several reasons: (i) Pajon et al. did not exclude proteins with hydrophobic α-helices and therefore the list of predicted OMPs contains a large number of inner membrane proteins (IMPs). (ii) One of the two variables chosen as predictors of putative β-barrel structures was based on the prevalence of a C-terminal phenylalanine,[26] which is typical for OMPs of gram-negative bacteria but is not present in the two known mycobacterial OMPs MspA and OmpA.[6,7,11] (iii) Proteins with sequence homology to known cytoplasmic and periplasmic proteins or lipoproteins, which are anchored in membranes by their lipid moieties and not by a transmembrane β-barrel,[27] were not excluded from the list.

In this study, we have employed an algorithm entirely based on physical principles to predict OMPs of *M. tuberculosis*. In contrast to all other prediction methods so far, we provide a scoring system for the number and amphiphilicity of β-strands of a particular protein. By combining both parameters with biological knowledge we predicted 144 proteins as OMPs of *Mtb*. The subcellular localization of two of these proteins was determined by demonstrating their association with membranes and their surface accessibility to proteases. This alternative approach identified Rv1698 and Rv1973 as OMPs of *M. tuberculosis* and provided experimental evidence for the usefulness of the bioinformatic predictions.

Materials and methods

Genome-wide analysis to identify putative OMPs of *M. tuberculosis*

A FASTA file with 3991 protein sequences for the *Mycobacterium tuberculosis* H37Rv genome was obtained from ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Mycobacterium_tuberculosis_H37Rv/ [version: AL123456.2 date: 04/18/2006].[28] All sequences were scanned for sequence similarity matches against the Pfam database of protein motifs (version 20.0[29]) using hmmpfam.[30,31] To identify exported proteins of *M. tuberculosis* we predicted the presence of an N-terminal signal peptide using SignalP 3.0.[32,33] All proteins that had a Y_{max} score $\geq$0.51 were considered as exported proteins. These proteins were selected and the sequences corresponding to the signal peptide predicted by SignalP were removed. These shortened sequences should represent the mature proteins and were used for all further analyses. Next, the sequences were examined by TMHMM to predict transmembrane α-helices.[34,35] For all proteins that did not have a hydrophobic α-helix the secondary structure was predicted from the sequence using the Jnet algorithm[36] which gives the best performance among secondary structure prediction algorithms and achieves a 76.4% average accuracy on a large test set of proteins.[37] Predicted β-strands of a minimum of five consecutive residues were registered. Next, we computed the amphiphilicity of these β-strands. To this end, the mean hydrophobicity of one side of a β-strand $H_\beta(i)$ was calculated at position i in a sequence following Vogel and Jähnig[38] as $H_\beta(i) = 1/5 \times (h(i-4) + h(i-2) + h(i) + h(i+2) + h(i+4))$, where $h(i)$ is the hydrophobicity of the amino acid at position i. Note that in a sequence of amino acids from 1 to N, these values can only be computed for $i = 5, \ldots, N-4$. The values of hydrophobicity for the amino acids were taken from Sweet and Eisenberg.[39] Given the average value of hydrophobicity