Microbes and Infection xxx (xxxx) xxx

Contents lists available at ScienceDirect

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf

Original article

A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex

Mathis Steindor [a, *], Vanesa Nkwouano [a], Anja Stefanski [b], Kai Stuehler [b], Thomas Richard Ioerger [c], David Bogumil [d], Marc Jacobsen [a], Colin Rae Mackenzie [e], Rainer Kalscheuer [f]

[a] Department of General Pediatrics, Neonatology, and Pediatric Cardiology, University Children's Hospital, Medical Faculty, Heinrich Heine University, Moorenstr. 5, 40225, Duesseldorf, Germany
[b] Molecular Proteomics Laboratory, Heinrich Heine University, Universitaetsstr. 1, 40225, Duesseldorf, Germany
[c] Department of Computer Science and Engineering, Texas A&M University, 77843-3112, TX, USA
[d] The Department of Life Sciences & The National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel
[e] Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Universitaetsstr. 1, 40225, Duesseldorf, Germany
[f] Institute of Pharmaceutical Biology and Biotechnology, Heinrich Heine University, Universitaetsstr. 1, 40225, Duesseldorf, Germany

ARTICLE INFO

Article history:
Received 17 August 2018
Accepted 30 October 2018
Available online xxx

Keywords:
Non tuberculous mycobacteria
Mycobacterium abscessus
Epitopes
Proteomics
Cystic fibrosis

ABSTRACT

The *Mycobacterium abscessus* complex can cause fatal pulmonary disease, especially in cystic fibrosis patients. Diagnosing *M. abscessus* complex pulmonary disease is challenging. Immunologic assays specific for *M. abscessus* are not available. In this study seven clinical *M. abscessus* complex strains and the *M. abscessus* reference strain ATCC19977 were used to find species-specific proteins for their use in immune assays. Six strains showed rough and smooth colony morphotypes simultaneously, two strains only showed rough mophotypes, resulting in 14 separate isolates. Clinical isolates were submitted to whole genome sequencing. Proteomic analysis was performed on bacterial lysates and culture supernatant of all 14 isolates. Species-specificity for *M. abscessus* complex was determined by a BLAST search for proteins present in all supernatants. Species-specific proteins underwent *in silico* B- and T-cell epitope prediction. All clinical strains were found to be *M. abscessus* ssp. *abscessus*. Mutations in MAB_4099c as a likely genetic basis of the rough morphotype were found in six out of seven clinical isolates. 79 proteins were present in every supernatant, of which 12 are exclusively encoded by all members of *M. abscessus* complex plus *Mycobacterium immunogenum*. *In silico* analyses predicted B- and T-cell epitopes in all of these 12 species-specific proteins.

The *Mycobacterium (M.) abscessus* complex (MABSC) comprises three species of rapidly growing non-tuberculous mycobacteria (NTM): *Mycobacterium abscessus*, *Mycobacterium massiliense* and *Mycobacterium bolletii*. A former proposal for the union of *M. bolletii* and *M. massiliense* into one subspecies seems obsolete against the background of their distinct genome sequences and clinically relevant features (e.g. regarding *erm*(41)-encoded inducible macrolide resistance) [1]. MABSC is one of the most frequently isolated NTM from human specimens, the most common clinical manifestations being pulmonary and skin infections [2]. Whereas soft tissue infection might occur in otherwise healthy subjects, MABSC pulmonary disease (MAPD) predominantly affects patients with a predisposing (airway) disease, especially cystic fibrosis (CF) [3]. The reported prevalence of NTM in CF varies widely, with MABSC being the most frequently isolated species in Caucasian populations [4]. Recent evidence for transmission of MABSC between CF patients has particularly drawn attention to this pathogen [5]. Pulmonary infection with MABSC is associated with a fast decline in lung function and fatal clinical courses in CF [6–8]. Moreover, it has a significant impact on the outcome of lung transplantation [9]. Current guidelines for the therapy of MAPD counsel for long-term antibiotic combination therapies against the background of constitutive and inducible multidrug resistance in MABSC, despite the absence of evidence for any therapy regime [3,10]. MABSC grown on agar plates show rough and/or smooth colony

* Corresponding author. Pediatric Pulmonology and Sleep Medicine, Children's Hospital, University of Duisburg-Essen, Hufelandstr. 55, 45147, Essen, Germany. Fax: +201 723 5517.
E-mail address: mathis.steindor@uk-essen.de (M. Steindor).

https://doi.org/10.1016/j.micinf.2018.10.006
1286-4579/

morphotypes according to the capacity for glycopeptidolipid (GPL) biosynthesis [11]. There is evidence that the phenotype and phenotype-changes of MABSC are associated with altered pathogenicity: The GPL-rich smooth isoforms tend to build biofilms whereas the GPL-lacking rough isoform has putatively increased immunogenicity and virulence [11–13].

The diagnosis of MAPD is based on clinical, radiological and microbiological criteria provided by the American Thoracic Society (ATS) in 2007 [3]. Diagnosis remains challenging especially in CF due to significant overlaps in clinical and radiological presentation of MAPD with other features of CF disease. The isolation of MABSC is hampered by overgrowth of mycobacterial cultures by other more rapidly growing bacteria often colonizing the lungs of CF patients, e. g. *Pseudomonas aeruginosa* [3]. There are reports of transient, intermittent or even persistent colonization of CF airways with MABSC without significant impact on the host's clinical course and thus presumably no necessity for treatment [14]. Additional diagnostic approaches pertinent to MAPD in particular to distinguish colonization from infection are not currently available. Serological tests for mycobacteria have been evaluated for MAPD in CF but lack specificity due to the use of non-specific antigenic targets [15]. In our own previous studies, we investigated a T-cell-based immunologic approach for the detection and immunologic characterization of MABSC infection in CF patients [16]. In a comparative assay we were able to show that the characteristics and level of T-cell immune response to different mycobacterial antigen compounds adds to the detection and understanding of MABSC infection. T-cells of MABSC-infected patients responded to antigens of MABSC, *Mycobacterium avium* and *Mycobacterium tuberculosis* but the different levels of immune response indicate the presence of MABSC specific immunogenic proteins. In the current study we aimed to identify MABSC-specific proteins for possible future use in immunologic analyses.

1. Materials and methods

1.1. Ethics statement

All samples (bacterial isolates) were retrieved from the local clinical microbiology department (Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Düsseldorf, Germany). All samples used in this study were anonymized. This study and the use of the samples were approved by the local ethics committee of the University Hospital Duesseldorf (Internal Study No. 4505).

1.2. Strains and culture conditions

Clinical MABSC isolates were retrieved from the local clinical microbiology department (Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Düsseldorf, Germany). These isolates were obtained from respiratory specimens of CF patients (n = 5) and patients with non-CF bronchiectasis (n = 2), of whom one patient was diagnosed with primary ciliary dyskinesia (the CF or non-CF origin of each strain is indicated in the study name of the strain, Table 1, left panel). *M. abscessus* ATCC19977 reference strain was purchased from LGC standards (LGC Standards GmbH, Wesel, Germany). For selection of rough and smooth colony morphotypes strains were inoculated onto Middlebrook 7H10 agar supplemented with 10% (v/v) OADC (oleic acid, albumin, dextrose-catalase) enrichment (Becton Dickinson Heidelberg, Germany) and 0.5% (v/v) glycerol and grown aerobically at 37 °C for 3–4 days. Rough and smooth colony morphotypes were selected visually (Fig. 1) and stored separately for further analyses. Liquid cultures for proteomic analyses were performed in Middlebrook 7H9 medium (Becton Dickinson) 0.5% (v/v) glycerol and 0.05% (v/v) Tyloxapol (without OADC enrichment to avoid interference of OADC with mass spectrometry).

1.3. Whole genome sequencing and phylogenic analyses

MABSC genomic DNA of clinical samples was extracted following the cetyltrimethylammonium bromide (CTAB) extraction protocol [17]. DNA samples were prepared for sequencing using the TruSeq whole-genome sample preparation kit (Illumina, Inc.; San Diego, CA) and sequenced on an Illumina HiSeq 2500, operated in paired-end mode, with a read length of 125 bp. The mean depth of coverage was 60× (for detailed sequencing statistics see Table S1). Genome sequences were assembled by a comparative-assembly method. First, reads were mapped to the genome of *M. abscessus* ATCC 19977 (NC_010397.1) as a reference sequence using Burrows-Wheeler Aligner (BWA) [18]. Thereafter, regions with insertions/deletions (indels) or clusters of single nucleotide polymorphisms (SNP) were identified and repaired by building local contigs from overlapping reads spanning these regions [19]. Polymorphisms (including SNPs and indels) were identified by aligning the genome sequence of each isolate to *M. abscessus* ATCC 19977 using the MUMmer sequence alignment package version 3.20 [20]. A phylogenetic tree was reconstructed by the maximum parsimony method using *dnapars* in Phylip 3.66 [21] based on a genome-wide set of 187 092 SNPs (excluding sites with low-coverage,

Table 1
Characteristics of MABSC isolates used for the study.

Isolate	Morphotype	Source	Polymorphisms
CF001	S + R	CF	MAB_4099c: -g in aa 2629*
CF022	R	CF	MAB_4099c: R2744Q, V2681I T2425A, A1794V, G1532E, D710E, E591D, M212V
CF023	R	CF	MAB_4099c: -a in aa 1721*, V2852I, T2733S, T2425A, T2129S, G2083D, G1263S, A1195V, D710E, E591D
CF029	S + R	CF	MAB_2210c: 233bp del
CF034	S + R	CF	MAB_4099c: Q1156*
NonCF01	S + R	PCD	MAB_1463: P204Q MAB_4099c: A2320P
NonCF02	S + R	Non-CF-BE	MAB_3507: G276E MAB_4099c: -ag in aa 863*
ATCC19977	S + R	Purchased	Not done

The left panel shows the colony morphotype (S smooth; R rough) and source of each strain used in the study (CF cystic fibrosis; PCD primary ciliary dyskinesia; Non-CF-BE non-cystic fibrosis-bronchiectasis). The right panel shows genome locus and respective mutations in rough morphotype isolates compared to the matched smooth isolate. For CF022 and CF029, for which a matched smooth isolate was not available, mutations in MAB_4099c as compared to ATCC19977 are shown. ATCC19977 was not sequenced in this study, so no information is given in the respective panel. Asterisks indicate putative loss-of-function mutations in the MAB_4099c gene known to control GPL production.

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006

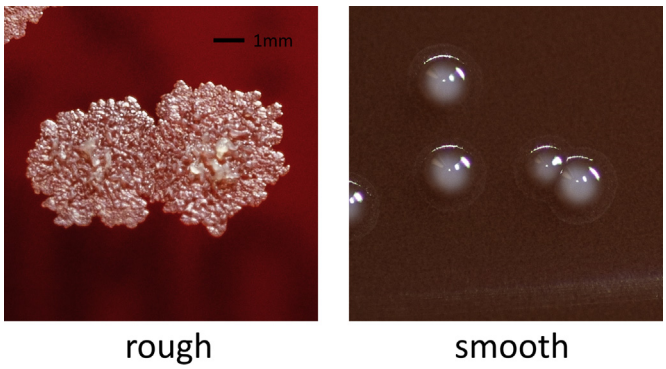


Fig. 1. MABSC colony morphotypes. Exemplary presentation of NonCF01 rough and smooth colony isolates on Columbia agar with 5% Sheep Blood.

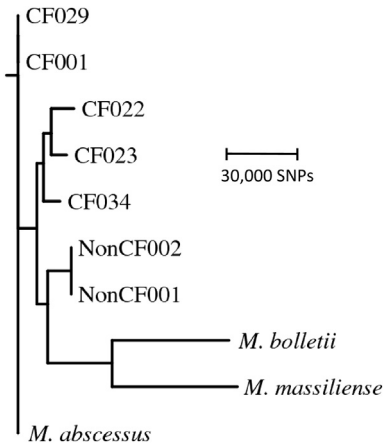


Fig. 2. Phylogenetic tree of MABSC strains. Maximum-parsimony phylogenetic tree of the clinical isolates based a genome-wide collection of SNPs in relation to the reference strains of *M. abscessus* ATCC19977 [39], *M. massiliense* GO 06 [40] and *M. bolletii* FLAC003 [41].

heterogeneous nucleotides, or indels). Genome sequences for *M. bolletii* FLAC003 (CP014950.1) and *M. massiliense* GO 06 (NC_018150.2) were included in the phylogeny for reference (Fig. 2).

1.4. Isolation of M. abscessus proteomes

MABSC isolates were grown aerobically in 10 ml of optimized protein-free 7H9 medium (see above) to an $OD_{600\ nm} = 0.6–1.0$ in a square bottle at 37 °C on shaking incubator at 80 rpm. MABSC liquid cultures were transferred to centrifugation tubes and centrifuged (4000 × *g*, 15 min) at 4 °C to sediment the bacteria, whilst aliquots of each bottle were plated out to confirm the desired colony morphotype. Culture supernatant was filtered through a 0.22 μm cellulose-acetate-filter membrane and stored at −80 °C until further analysis. Bacterial pellets were washed twice with chilled PBS/0.05% (v/v) Tween 80 (PBS-TW), the resulting bacterial sediments pooled in 1 ml chilled PBS-TW and then again collected by centrifugation (4000 × *g*, 15 min at 4 °C) and finally the bacterial pellet was re-suspended in 9 ml PBS and 1 ml proteinase inhibitor cocktail (10×, Sigma MS safe Protease-/phosphatase inhibitor cocktail, Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). 10 screw cap tubes each pre-filled with 0.3 g 0.1 mm glass beads (Sigma Aldrich) and 10–12 3 mm beads (Merck Millipore) were each filled with 1 ml of the bacterial suspension, followed by disruption in a bead-beater (50 Hz 4 × 1 min). Disrupted and homogenized samples were centrifuged (10 000 × *g*, 3 min) and the resulting bacterial lysates were collected by pipette. The samples (bacterial lysates) were separated from remaining beads by centrifugation (10 000 × *g*, 5 min) in Spin Basket-Tubes (Investigator Lyse&Spin Basket Kit, Qiagen, Hilden, Germany) and stored at −80 °C until further analysis.

1.5. Mass spectrometry analysis

The proteins in culture supernatants were precipitated for 1 h at 4 °C by addition of 50% (w/v) trichloroacetic acid and 0.1% sodium [dodecanoyl(methyl)amino]acetate in water, sedimented by centrifugation and washed with ice-cold acetone. After repeated centrifugation, the protein pellet was shortly air-dried at room temperature and resolved in 50 μL of lysis buffer (30 mM tris(hydroxymethyl)aminomethane, 2 M thiourea, 7 M urea, and 4% (w/v) 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, pH 8.5). The protein concentrations of supernatant and bacterial lysate samples were determined by the Pierce 660 nm Protein Assay (Fisher Scientific, Schwerte, Germany) and prepared for LC–MS analysis by short separation via a 4–12% polyacrylamide gel as described elsewhere [22]. Therefore, up to 10 μg per sample were loaded onto the gel (Novex NuPAGE, Thermo Scientific, Darmstadt, Germany) and run for about 10 min. After silver staining, protein bands were cut out, destained and washed. Proteins were reduced by 10 mM dithiothreitol and alkylated with 55 mM iodoacetamide. Subsequently, proteins were digested for 16 h at 37 °C with 0.1 μg of trypsin (Serva, Heidelberg, Germany) in 100 mM ammonium hydrogen carbonate in water. Tryptic peptides were extracted twice with a 1:1 (v/v) solution of acetonitrile and 0.1% trifluoroacetic acid and, after acetonitrile removal, resuspended in 0.1% (v/v) trifluoroacetic acid. After peptide extraction 1.25 μg were subjected to liquid chromatography for each LC–MS run. For peptide separation over a 120 min LC-gradient with 300 nL/min an Ultimate 3000 Rapid Separation liquid chromatography system (Thermo Scientific, Bremen, Germany) equipped with an Acclaim PepMap 100 C18 column (75 μm inner diameter, 25 cm length, 2 mm particle size from Thermo Scientific, Bremen, Germany) was used. MS analysis was carried out on a Q-Exactive plus mass spectrometer (Thermo Scientific) operating in positive mode and equipped with a nano electrospray ionization source. Capillary temperature was set to 250 °C and source voltage to 1.4 kV. Survey scans were carried out over a mass range from 200 to 2000 m/z at a resolution of 70 000 (at 200 m/z). The target value for the automatic gain control was 3 000 000 and the maximum fill time 50 ms. The 20 most intense peptide ions (excluding singly charged ions) were selected for fragmentation. Peptide fragments were analysed using a maximal fill time of 50 ms and automatic gain control target value of 100 000 and a resolution of 17 500 (at 200 m/z). Peptides already fragmented were excluded for fragmentation for 10 s.

Acquired spectra were searched using Mascot 2.4 within Proteome Discoverer version 1.4.1.14 against a *M. abscessus* UniProt database (release 2015_03, 4918 sequences). Within the software a decoy database search was performed using the Percolator node and peptides were filtered based on the peptide confidence score (FDR Settings: Threshold 0.01 with validation based on the q-Value). Carbamidomethyl at cysteine residues was set as a fixed modification and methionine oxidation was considered as variable modification. In addition, tryptic cleavage specificity (cleavage behind K and R) was set to a maximum of two missed cleavage sites.

Predefined values were used for other parameters including a false discovery rate of 1% on peptide level, a main search precursor

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006

mass tolerance of 10 ppm and mass tolerance of 0.4 Da for fragment spectra. Label-free quantification was performed with Progenesis QI for Proteomics (Version 2.0, Nonlinear Dynamics, Waters Corporation, Newcastle upon Tyne, UK).

1.6. Bioinformatic analyses

Proteins found in supernatants of all isolates were selected for further analyses. Amino acid sequences were obtained from the Uniprot database [23] and used in a BLAST search [24] against known or hypothetical proteins of clinically relevant NTM and the *M. tuberculosis* complex listed in the NCBI database of non-redundant sequences (for a list of mycobacterial species and their taxonomic ID see Table S2), applying an E-value threshold of 10^{-10} for a given BLAST hit to be considered significant. The data was subsequently transformed into a presence - absence matrix of proteins in mycobacterial genomes (Fig. 4). Mycobacterial phylogeny was reconstructed from 16S ribosomal RNA genes obtained from the NCBI GenBank database [25]. Sequences were aligned using the MAFFT tool version 7.310 [26] and maximum likelihood trees reconstructed using FastTree tool version 2.1.7 [27]. Protein candidates without relevant sequence homologies to proteins of other mycobacterial species were selected for further analyses. Signal peptides in selected proteins indicating active secretion to the extracellular department were identified via UniProt database and the SignalP 4.1 Server [28]. Similarity searches for selected proteins based on protein structure and domain composition were performed against KEGG [29], NCBI [30], UniProt [23] and HHpred [31] database. An E-value threshold of 10^{-3} for a given BLAST hit was considered significant.

1.7. Identification of immunogenic protein candidates

Stabilization matrix alignment method (SMM align) [32] was used for direct quantitative prediction of peptide MHC class II binding affinities for amino acid sequences of MABSC candidate proteins via Immune Epitope Database (IEDB, http://www.iedb.org/) [33]. Binding peptides for all 15 selectable alleles of the human HLA-DR locus were predicted and evaluated according to their binding affinities using IC50 values (high affinity binding: IC50 < 50 nM; intermediate affinity binding: IC50 < 500 nM; low affinity binding: IC50 < 5000 nM) as described by Gurung et al. [34].

For B-cell epitope prediction, the automated protein structure homology-modelling server SWISS-MODEL (https://swissmodel.expasy.org/) was used to generate 3D templates for each candidate protein [35]. The template with best GMQE (Global Model Quality Estimation) score was used to predict linear and discontinuous epitopes in ElliPro (http://tools.iedb.org/ellipro/) (minimum score 0.5, maximum distance 6 Å) [36].

2. Results

2.1. Bacterial culture

Bacterial culture yielded seven rough and five smooth colony isolates from seven clinical MABSC strains; five strains showed both phenotypes and two strains showed only rough phenotypes (Table 1, left panel, and Fig. 1). *M. abscessus* type strain ATCC 19977 showed both phenotypes as well. The colony morphotype is indicated as a suffix in the study name of each isolate (S smooth, R rough).

2.2. Whole genome sequencing

Whole genome sequencing showed a genetic clustering of clinical isolates with *M. abscessus* ATCC 19977 (Fig. 2). The identification of all clinical isolates as *M. abscessus* ssp. *abscessus* was confirmed based on 16S rRNA and *hsp65* sequences [37]. In a multiple alignment, the isolates exhibited greatest similarity to *M. abscessus* ssp. *abscessus* (54–26227 SNPs) and had at least 99 043 SNPs from *M. bolletii* and *M. massiliense* (Table S3). While CF001 and CF029 were nearly identical to ATCC 19977 (54 and 67 SNPs, respectively), the other isolates were distant more than 20 000 SNPs. These latter isolates also had several large-scale insertions and deletions relative to ATCC 19977. A maximum-parsimony tree based on genome-wide SNP analysis shows the phylogenetic relationships among the isolates (Fig. 2), indicating high genetic similarity between CF001 and CF0029, as well as nonCF001 and nonCF002, respectively. In fact, the two smooth non-CF isolates were identical, and the corresponding isolates differed by only 2 SNPs (outside of MAB_4099c). Though both CF001 and CF0029 are highly similar to ATCC 19977, each had several unique SNPs not found in the other (48 and 35, respectively). Despite the diversity among most strains from different patients, pairs of isolates (rough and smooth) from the same patient had only 1–2 differences between them (Table 1, right panel). 6 out of 7 rough morphotype isolates showed mutations in the MAB_4099c gene, which is known to control GPL production [38]. Three of these mutations are indels that produce a frameshift, and one is a nonsense mutation (stop codon), all presumably resulting in a loss-of-function (indicated by asterisks in the right panel of Table 1).

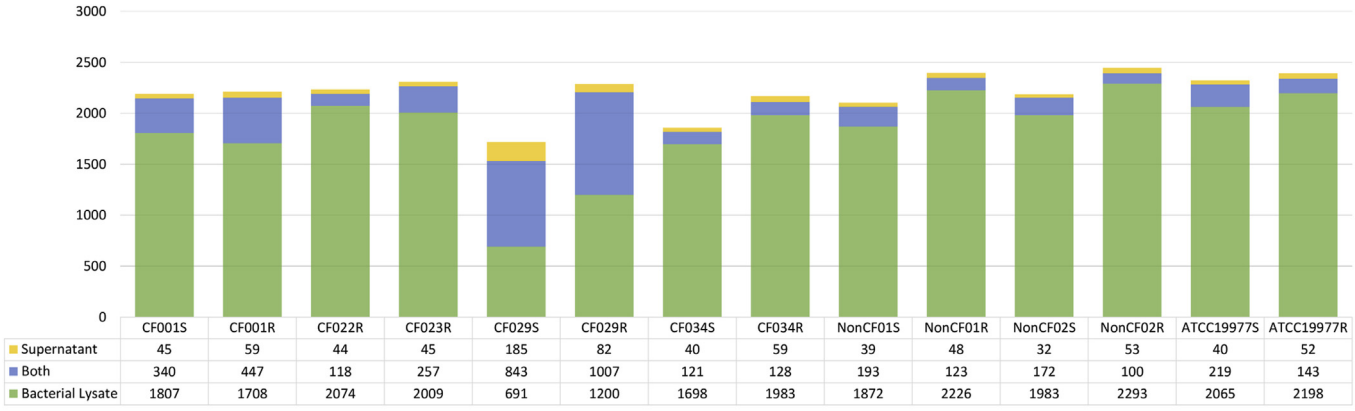
2.3. Proteomics

Mass spectrometry of the seven rough and five smooth clinical isolates and the two isolates of the ATCC19977 strain (rough and smooth, respectively) detected a total of 3137 different proteins (range 1719–2446 in the different strains, Fig. 3a). A total of 3084 different proteins were found in the bacterial lysates (range 1534–2 393, Fig. 3c) and 1302 in the supernatant (range 153–1 089, Fig. 3b), 1249 of those proteins were found in both compartments. The resulting total number of 3137 different detected proteins amounts to 63.5% of the 4940 proteins encoded both chromosomally (4 918) and on a plasmid (22) in the ATCC19977 genome [39]. 79 proteins were found in all analysed supernatants of which 12 only showed a significant homology *Mycobacterium immunogenum* in BLAST analyses (Fig. 4). All of these 12 proteins contain a putative signal peptide and are therefore predicted to be secreted according to the UniProt database and the SignalP algorithm. Four of these 12 secreted proteins were also found in every bacterial lysate indicating presence in the intracellular and extracellular compartment (Table 2, left panel). Three of these proteins are within the group of proteins that showed the highest abundance within the supernatant whereas nine proteins could be considered as medium or low abundant in the supernatant (Fig. S1). Notably, all proteins except one (MAB_0974) are less than 250 amino acids in length.

2.4. Epitope prediction

In-silico prediction of epitopes predicted MHC-II high-binders (IC50 < 50 nM) in 11 of these 12 proteins, including the four proteins found in all bacterial lysates and supernatants. Intermediate MHC-II binders (IC50 < 500 nM) as well as putative linear and discontinuous B-cell epitopes were found in all 12 proteins (Table 2, left panel). After the search for homologous proteins using BLAST [24] against the KEGG [29], Uniprot [23] and NCBI [42] databases did only yield hits annotated as “hypothetical protein”, we

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006



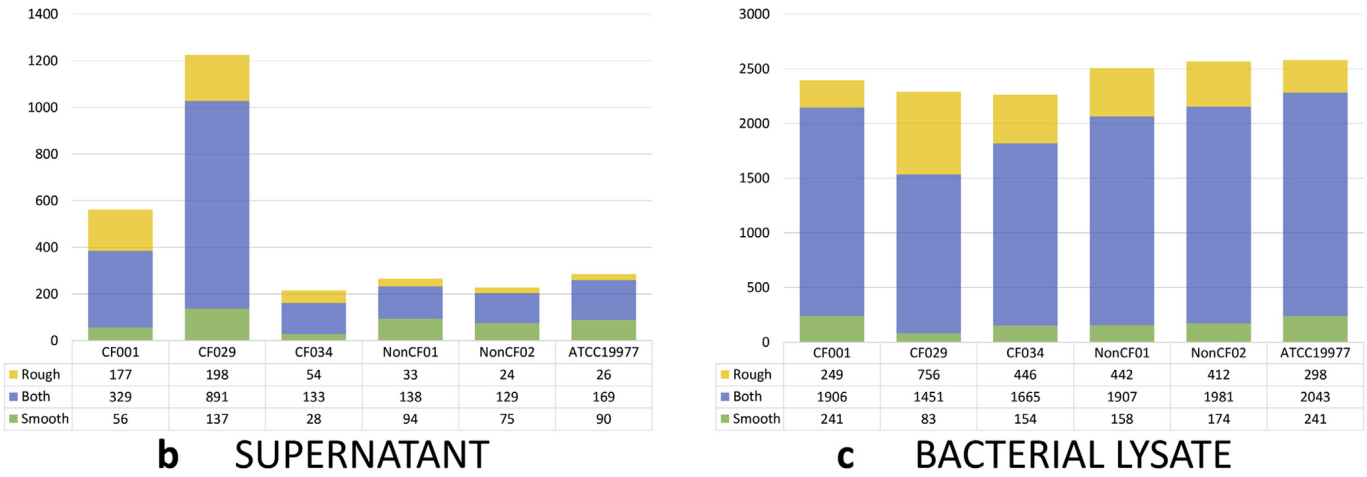


Fig. 3. Protein detection results of mass spectrometry, a Total number of detected proteins in each isolate, b Number of detected proteins in the culture supernatant of each isolate. c Number of detected proteins in the bacterial lysate of each isolate.

performed a similarity search based on protein structure and domain composition using HHpred [31]. Similarity search results are shown in Table 2 (right panel) alongside characteristics of the homologous proteins.

3. Discussion

Immunogenic proteins of MABSC would be very useful for studies on the immune response to MABSC-associated diseases and for their potential use in diagnostic immunoassays. In this current study, we aimed to identify specific T-helper cell-, and B-cell- epitopes of MABSC in clinical strains using an *in vitro* proteomic approach following *in silico* epitope prediction. We hypothesized that promising protein candidates should fulfil the following requirements: (1) they are expressed by all MABSC isolates, (2) they are secreted into the extracellular compartment and are thus easily available for interaction with the human immune system, and (3) they are species-specific for MABSC.

We used smooth and rough morphotype isolates of clinical strains of MABSC for our analyses as well as the well characterized *M. abscessus* reference strain ATCC 19977. Whole-genome sequencing of the clinical strains did not indicate patient-to-patient transmission between CF002, CF023, CF034, which is consistent with the results of VNTR (variable number of tandem repeats) analysis of these strains in previous studies [16]. CF001 and CF029 were very similar with 83 SNPs difference. Presuming a threshold of 38 SNPs to indicate possible patient-to-patient transmission as proposed by Bryant et al. [43], direct transmission can neither be proven nor ruled out in this case. Notably, both patients were treated at the same pediatric CF center over several years so at least a former transmission is conceivable. NonCF001 and nonCF002 are likely to be the same strain. Interestingly, all strains are isolated from patients of a single tertiary care pediatric hospital except from nonCF002, which was isolated from an elderly female with bronchiectasis treated in the internal medicine ward of the same hospital at the same time nonCF002 had the first culture positive for MABSC. The source of acquisition of the same MABSC strain in these two unrelated patients was not investigated in detail in this study. However, the ward of internal medicine and pediatrics are two separate (yet adjacent) buildings and both patients did not share therapists or medical equipment. Still, direct or indirect contact of these patients and consecutive transmission of the strain in other facilities of the hospital is possible. Six out of seven rough isolates showed SNPs or indels in the MAB_4099c gene known to regulate GPL production and thus determining rough or smooth colony morphotype of MABSC [13]. These results support previous reports indicating a genetic cause for MABSC phenotype changes [12,44]. Although spontaneous mutations in MABSC and other mycobacteria – e.g. in *M. tuberculosis* leading to loss of glycolipid virulence factors such as phenol phthiocerol dimycolate (PDIM) – under *in vitro* selection pressure have been described [11,19,45], we consider an *a priori* presence of both phenotypes *in vivo* likely. Howard et al. proposed an infection sequence for MABSC, in which smooth phenotypes are part of and contribute to the biofilm

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006

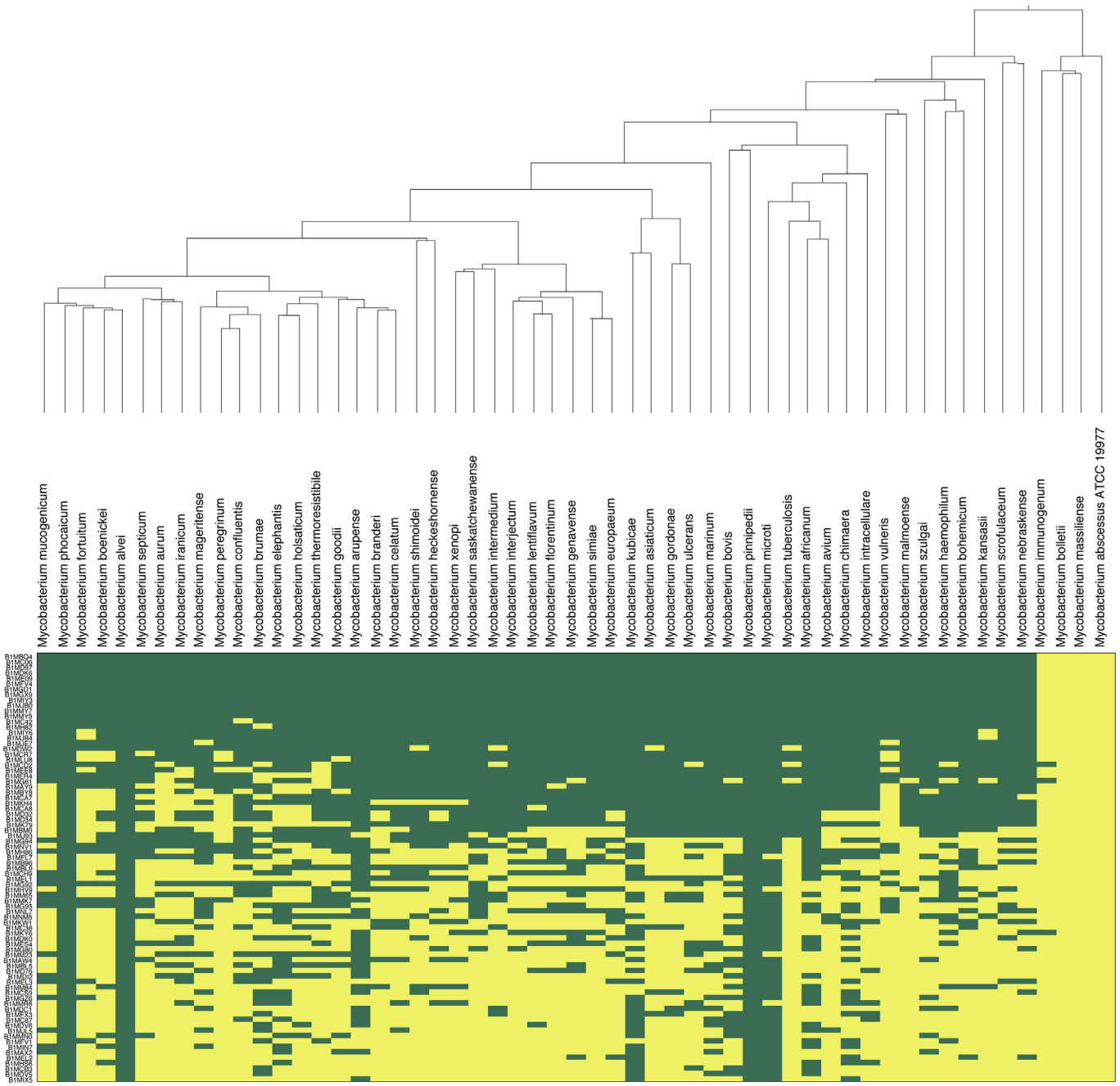


Fig. 4. Presence-absence matrix of candidate proteins in the mycobacterial genus. Presence (yellow) or absence (green) of 79 proteins found in all culture supernatants of the study isolates in mycobacterial genomes with mycobacterial phylogeny.

properties of infected (CF-) airways and a more virulent and invasive rough phenotype arises from this environment, necessitating a simultaneous presence of both phenotypes in the host [11]. Besides, morphotypes of our clinical isolates were determined during the first passage on agar plate after taking them from the stock, which again was stored directly after isolation from the patient without repetitive passages on culture, which further argues against phenotype switches *in vitro*.

Proteomic analyses found 79 proteins in all 14 culture supernatants of the analysed isolates, indicating constitutive expression and secretion of these 79 proteins. 12 of these 79 proteins were identified to be encoded in all MABSC genomes in the available open databases, whilst showing no significant sequence homologies to proteins of other clinically relevant mycobacterial species except *M. immunogenum*. These 12 proteins all carried signal peptides indicating active secretion rather than bacterial disintegration as cause for presence in the culture supernatant. The overlap in protein expression between MABSC and *M. immunogenum* is an expected finding, as *M. immunogenum* (causative agent of hypersensitivity pneumonitis) is a species of the former *Mycobacterium chelonae-M. abscessus* complex, which comprises all members of MABSC (see phylogenetic tree in Fig. 4) [46]. Overlaps in protein expression in clinically relevant mycobacterial species are common: *M. tuberculosis* proteins used as antigenic targets in commercial interferon gamma release assays (ESAT-6, CFP-10 and TB 7.7) are also present in other NTM species,

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006

Table 2
Candidate protein characteristics.

Protein	Presence in bacterial lysate	T-cell epitope prediction		B-cell epitope prediction				Nearest protein homology based on protein structure and domain composition			
		MHC-II intermediate binders	MHC-II high binders	Swissmodel template	GMQE-score	Linear B-cell epitopes	Discontinuous B-cell epitopes	Best Hit (PDB ID)	Probability [%]	E-value	Description (protein; characteristics; origin)
MAB_2594 (B1MBQ4)	5/14	172	0	4ILI.1.A	0.29	4	4	5CYB_A	64.96	8.6	Lipoprotein, lipocalin, PccL Virulence/transport *Streptococcus pneumoniae*
MAB_2697c (B1MC06)	7/14	247	5	4ZJM.6.A	0.41	6	5	4ZJM_B	99.93	5.1e-28	Lipoprotein LpqH Virulence *Mycobacterium tuberculosis* strain ATCC 25618/ H37Rv
MAB_3141 (B1MD97)	0/14	266	9	4UEJ.1.A	0.24	4	2	2LW3_A	95.76	0.019	Putative membrane protein mmpS4 Soluble domain *Mycobacterium tuberculosis*
MAB_3249 (B1MDK6)	10/14	381	92	5HHJ.1.A	0.18	2	4	4V4N_BT	42.36	23	Preprotein translocase subunit SecE Ribosomal *Methanococcus jannaschii*
MAB_3402 (B1ME09)	6/14	255	16	2IV9.1.A	0.10	4	3	5XMZ_A	50.39	49	Effector protein PevD1 Effector *Verticillium dahliae*
MAB_0405c (B1MFV4)	14/14	380	58	3ROB.1.A	0.29	7	4	3K7C_D	98.82	6.4e-10	Putative NTF2-like transpeptidase Putative NTF2-like transpeptidase *Campylobacter jejuni*
MAB_3801c (B1MGD1)	7/14	299	36	3OD9.1.A	0.12	5	1	4QTQ_A	98.76	2.3e-9	XAC2610 protein Beta-sandwich, calcium binding motif, Beta-propeller *Xanthomonas axonopodis pv. citri*
MAB_0565c (B1MGX9)	12/14	157	8	2LW3.1.A	0.47	5	4	2LW3_A	99.72	3e-19	Putative membrane protein mmpS4 MmpS4 soluble domain *Mycobacterium tuberculosis*
MAB_4281c (B1MIY3)	14/14	441	40	5K69.1.A	0.06	3	3	4ESQ_A	57.15	120	Serine/threonine protein kinase Membrane *Mycobacterium tuberculosis*
MAB_0974 (B1MJB0)	7/14	845	75	2F9Y.1.B	0.04	4	4	5Z0R_A	18.56	410	Extracellular solute-binding protein family 1 Capsid Zika virus
MAB_1614 (B1MMY7)	14/14	484	70	3CU3.1.A	0.27	6	5	3FKA_B	98.58	4.3e-9	Uncharacterized NTF-2 like protein Unknown *Silicibacter pomeroyi* DSS-3
MAB_1616 (B1MMY9)	14/14	250	27	3D9R.1.B	0.29	6	3	4OVM_C	99.1	2e-11	Uncharacterized protein SgcJ Neocarzinostatin biosynthesis *Streptomyces carzinostaticus* subsp. *neocarzinostaticus*

Characteristics of the 12 MABSC species-specific candidate proteins found in every culture supernatant with their respective NCBI GenBank accession number. The number of strains in which the protein was found in the bacterial lysate in the 14 analysed isolates is shown in column 2. The cumulative predicted number of MHC-II high-affinity binders (IC50 < 50 nM), intermediate-affinity binders (IC50 < 500 nM) is shown as predicted by the immune epitope database for all 15 selectable alleles of the human HLA-DR locus. Linear/discontinuous B-cell epitopes predicted according to the best matching Swissmodel template (by GMQE-score) are shown. The first four characters of the Swissmodel template ID represent the protein data bank (PDB) ID. Results of a similarity search of the candidate proteins based on protein structure and domain composition using HHpred are shown in the right panel. Protein structural profiles are compared to known structures from PDB and hits are labelled with their PDB ID.

namely *Mycobacterium kansasii*, *Mycobacterium szulgai, Mycobacterium marinum* and *Mycobacterium flavescens* (21). *M. immunogenum* – unlike MABSC – does not seem to target CF or non-CF-bronchiectasis patients rendering the protein expression overlap only a minor drawback in terms of the specificity of possible future immune-based diagnostic assays (22). All our 12 protein candidates were predicted to embody several T-helper-cell epitopes for human HLA-DR alleles *in silico*, rendering them potentially applicable for T-helper-cell-based immunoassays. Linear and discontinuous B-cell epitopes were predicted *in silico* by Ellipro for all proteins via Swissmodel templates. However, these templates only showed relatively poor homology to their original candidate proteins as indicated by low GMQE scores, potentially limiting the power of B-cell epitope prediction. The difficulties in generating B-cell epitope templates for our candidate proteins based on homologies to known proteins are likely caused by their (preconditioned) specificity for MABSC. Accordingly, the search for structural protein homologues for candidate proteins to provide information on protein characteristics failed in three standard protein databases (i.e. KEGG, NCBI, Uniprot). Still, six candidate proteins showed significant structural homologies to proteins annotated in the HHPred database. Of those, two proteins (i.e. MAB_2697c and MAB_0565c) were structurally homologue to proteins of *M. tuberculosis* (Table 2, right panel). Interestingly, these two mycobacterial proteins likely belong to protein families (lipoproteins and mmpS-membrane proteins respectively) that were recently linked to virulence mechanisms promoting intracellular survival of MABSC in amoebae and human macrophages, supporting their assumed relevance in host–pathogen interaction [47].

In conclusion, we provide 12 promising protein candidates for immunologic studies on MABSC. Their applicability needs to be confirmed further *in vitro* or *in/ex vivo* assays.

Acknowledgements

We thank Raquel Guadarrama-Gonzalez from the Institute of Medical Microbiology and Hospital Hygiene Düsseldorf for her assistance.

This study was funded by the Research Committee of the medical faculty of the Heinrich-Heine-University of Duesseldorf under the grant number of 2015-21. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micinf.2018.10.006.

Conflict of interest statement

All authors have declared that no conflicts of interests exist.

References

[1] Adekambi T, Sassi M, van Ingen J, Drancourt M. Reinstating *Mycobacterium massiliense* and *Mycobacterium bolletii* as species of the *Mycobacterium abscessus* complex. Int J Syst Evol Microbiol 2017;67:2726–30.

[2] Brown-Elliott BA, Griffith DE, Wallace Jr RJ. Diagnosis of nontuberculous mycobacterial infections. Clin Lab Med 2002;22:911–25. vi.

[3] Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 2007;175:367–416.

[4] Hill UG, Floto RA, Haworth CS. Non-tuberculous mycobacteria in cystic fibrosis. J R Soc Med 2012;105(Suppl. 2):S14–8.

[5] Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al. Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study. Lancet 2013;381:1551–60.

[6] Esther Jr CR, Esserman DA, Gilligan P, Kerr A, Noone PG. Chronic *Mycobacterium abscessus* infection and lung function decline in cystic fibrosis. J Cyst Fibros 2010;9:117–23.

[7] Sanguinetti M, Ardito F, Fiscarelli E, La Sorda M, D'Argenio P, Ricciotti G, et al. Fatal pulmonary infection due to multidrug-resistant *Mycobacterium abscessus* in a patient with cystic fibrosis. J Clin Microbiol 2001;39:816–9.

[8] Qvist T, Taylor-Robinson D, Waldmann E, Olesen HV, Hansen CR, Mathiesen IH, et al. Comparing the harmful effects of nontuberculous mycobacteria and Gram negative bacteria on lung function in patients with cystic fibrosis. J Cyst Fibros 2016;15:380–5.

[9] Gilljam M, Schersten H, Silverborn M, Jonsson B, Ericsson Hollsing A. Lung transplantation in patients with cystic fibrosis and *Mycobacterium abscessus* infection. J Cyst Fibros 2010;9:272–6.

[10] Haworth CS, Banks J, Capstick T, Fisher AJ, Gorsuch T, Laurenson IF, et al. British Thoracic Society guidelines for the management of non-tuberculous mycobacterial pulmonary disease (NTM-PD). Thorax 2017;72:ii1–64.

[11] Howard ST, Rhoades E, Recht J, Pang X, Alsup A, Kolter R, et al. Spontaneous reversion of *Mycobacterium abscessus* from a smooth to a rough morphotype is associated with reduced expression of glycopeptidolipid and reacquisition of an invasive phenotype. Microbiology 2006;152:1581–90.

[12] Kim BJ, Kim BR, Lee SY, Kook YH, Kim BJ. Rough colony morphology of *Mycobacterium massiliense* Type II genotype is due to the deletion of glycopeptidolipid locus within its genome. BMC Genom 2013;14:890.

[13] Nessar R, Reyrat JM, Davidson LB, Byrd TF. Deletion of the mmpL4b gene in the *Mycobacterium abscessus* glycopeptidolipid biosynthetic pathway results in loss of surface colonization capability, but enhanced ability to replicate in human macrophages and stimulate their innate immune response. Microbiology 2011;157:1187–95.

[14] Martiniano SL, Sontag MK, Daley CL, Nick JA, Sagel SD. Clinical significance of a first positive nontuberculous mycobacteria culture in cystic fibrosis. Ann Am Thorac Soc 2014;11:36–44.

[15] Qvist T, Pressler T, Taylor-Robinson D, Katzenstein TL, Hoiby N. Serodiagnosis of *Mycobacterium abscessus* complex infection in cystic fibrosis. Eur Respir J 2015;46:707–16.

[16] Steindor M, Nkwouano V, Mayatepek E, Mackenzie CR, Schramm D, Jacobsen M. Rapid detection and immune characterization of *Mycobacterium abscessus* infection in cystic fibrosis patients. PLoS One 2015;10:e0119737.

[17] Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs Jr WR. Genetic manipulation of *Mycobacterium tuberculosis*. Curr Protoc Microbiol 2007. Chapter 10: Unit 10A 2.

[18] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.

[19] Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, et al. Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. J Bacteriol 2010;192:3645–53.

[20] Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol 2004;5:R12.

[21] Felsenstein J. PHYLIP – phylogeny Inference package (version 3.2). Cladistics 1989;vol. 5:164–6.

[22] Seyfarth K, Poschmann G, Rozman J, Fromme T, Rink N, Hofmann A, et al. The development of diet-induced obesity and associated metabolic impairments in Dj-1 deficient mice. J Nutr Biochem 2015;26:75–81.

[23] The UniProt C. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2017;45:D158–69.

[24] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinf 2009;10:421.

[25] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:D48–53.

[26] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–80.

[27] Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26:1641–50.

[28] Nielsen H. Predicting secretory proteins with SignalP. Methods Mol Biol 2017;1611:59–73.

[29] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.

[30] Coordinators NR. Database resources of the national center for biotechnology information. Nucleic Acids Res 2017;45:D12–7.

[31] Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 2018;430:2237–43.

[32] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinf 2007;8:238.

[33] Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 2008;4:e1000048.

[34] Gurung RB, Purdie AC, Begg DJ, Whittington RJ. In silico identification of epitopes in *Mycobacterium avium* subsp. paratuberculosis proteins that were upregulated under stress conditions. Clin Vaccine Immunol 2012;19:855–64.

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006

[35] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014;42:W252–8.
[36] Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinf 2008;9:514.
[37] Adekambi T, Drancourt M. Dissection of phylogenetic relationships among 19 rapidly growing Mycobacterium species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. Int J Syst Evol Microbiol 2004;54:2095–105.
[38] Pawlik A, Garnier G, Orgeur M, Tong P, Lohan A, Le Chevalier F, et al. Identification and characterization of the genetic changes responsible for the characteristic smooth-to-rough morphotype alterations of clinically persistent *Mycobacterium abscessus*. Mol Microbiol 2013;90:612–29.
[39] Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, et al. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. PLoS One 2009;4:e5660.
[40] Raiol T, Ribeiro GM, Maranhao AQ, Bocca AL, Silva-Pereira I, Junqueira-Kipnis AP, et al. Complete genome sequence of *Mycobacterium massiliense*. J Bacteriol 2012;194:5455.
[41] Caverly LJ, Spilker T, LiPuma JJ. Complete genome sequence of *Mycobacterium abscessus* subsp. bolletii. Genome Announc 2016;4.
[42] Coordinators NR. Database resources of the national center for biotechnology information. Nucleic Acids Res 2018;46:D8–13.
[43] Bryant JM, Grogono DM, Rodriguez-Rincon D, Everall I, Brown KP, Moreno P, et al. Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. Science 2016;354:751–7.
[44] Bernut A, Viljoen A, Dupont C, Sapriel G, Blaise M, Bouchier C, et al. Insights into the smooth-to-rough transitioning in *Mycobacterium bolletii* unravels a functional Tyr residue conserved in all mycobacterial MmpL family members. Mol Microbiol 2016;99:866–83.
[45] Domenech P, Reed MB. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies. Microbiology 2009;155:3532–43.
[46] Wilson RW, Steingrube VA, Bottger EC, Springer B, Brown-Elliott BA, Vincent V, et al. Mycobacterium immunogenum sp. nov., a novel species related to *Mycobacterium abscessus* and associated with clinical disease, pseudo-outbreaks and contaminated metalworking fluids: an international cooperative study on mycobacterial taxonomy. Int J Syst Evol Microbiol 2001;51:1751–64.
[47] Laencina L, Dubois V, Le Moigne V, Viljoen A, Majlessi L, Pritchard J, et al. Identification of genes required for *Mycobacterium abscessus* growth in vivo with a prominent role of the ESX-4 locus. Proc Natl Acad Sci U S A 2018;115:E1002–11.

Please cite this article as: M. Steindor et al., A proteomics approach for the identification of species-specific immunogenic proteins in the *Mycobacterium abscessus* complex, Microbes and Infection, https://doi.org/10.1016/j.micinf.2018.10.006