


RESEARCH ARTICLE

Open Access

Comparison of carbohydrate ABC importers from *Mycobacterium tuberculosis*



Lilia I. De la Torre^{1,2,3}, José G. Vergara Meza^{3,4}, Sindy Cabarca^{1,2,3}, André G. Costa-Martins⁵ and Andrea Balan^{1,2,6*} 

Abstract

Background: *Mycobacterium tuberculosis*, the etiological agent of tuberculosis, has at least four ATP-Binding Cassette (ABC) transporters dedicated to carbohydrate uptake: LpqY/SugABC, UspABC, Rv2038c-41c, and UgpA/EBC. LpqY/SugABC transporter is essential for *M. tuberculosis* survival in vivo and potentially involved in the recycling of cell wall components. The three-dimensional structures of substrate-binding proteins (SBPs) LpqY, UspC, and UgpB were described, however, questions about how these proteins interact with the cognate transporter are still being explored. Components of these transporters, such as SBPs, show high immunogenicity and could be used for the development of diagnostic and therapeutic tools. In this work, we used a phylogenetic and structural bioinformatics approach to compare the four systems, in an attempt to predict functionally important regions.

Results: Through the analysis of the putative orthologs of the carbohydrate ABC importers in species of *Mycobacterium* genus it was shown that Rv2038c-41c and UgpA/EBC systems are restricted to pathogenic species. We showed that the components of the four ABC importers are phylogenetically separated into four groups defined by structural differences in regions that modulate the functional activity or the interaction with domain partners. The regulatory region in nucleotide-binding domains, the periplasmic interface in transmembrane domains and the ligand-binding pocket of the substrate-binding proteins define their substrates and segregation in different branches. The interface between transmembrane domains and nucleotide-binding domains show conservation of residues and charge.

Conclusions: The presence of four ABC transporters in *M. tuberculosis* dedicated to uptake and transport of different carbohydrate sources, and the exclusivity of at least two of them being present only in pathogenic species of *Mycobacterium* genus, highlights their relevance in virulence and pathogenesis. The significant differences in the SBPs, not present in eukaryotes, and in the regulatory region of NBDs can be explored for the development of inhibitory drugs targeting the bacillus. The possible promiscuity of NBDs also contributes to a less specific and more comprehensive control approach.

Keywords: Carbohydrate uptake, ABC transporters, Phylogeny, Multitask NBDs, Structure-function, Protein-protein interaction

* Correspondence: abalan@usp.br

¹Department of Microbiology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil

²Genetics and Molecular Biology Postgraduate Program, Institute of Biology, State University of Campinas, São Paulo, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Mycobacterium tuberculosis is the causative agent of tuberculosis, one of the top causes of human death worldwide from a single infectious agent. About a quarter of the world's population has been infected by *M. tuberculosis* and thus at risk of developing tuberculosis disease [1]. Although there are many studies about the ability of *M. tuberculosis* to persist inside the host cells under a variety of adverse conditions including oxidative stress, hypoxia, and nutrient starvation [2], some aspects of the mechanisms behind are poorly understood. The upregulation of different nutrient uptake responsive genes at different stages of infection indicates that *M. tuberculosis* utilizes a set of nutrient sources from early to persistent phase, that include ions, amino acids, lipids, carbohydrates, and others required for many biological processes [3–5]. Different mechanisms to obtain essential nutrients from microenvironments and the broad range of substrate specificities, allow bacteria to quickly adapt to and colonize challenging environments. Some of these nutrient acquisition mechanisms are described as key virulence determinants used by pathogens to mediate disease [6, 7].

Carbohydrates have traditionally been considered an important source of carbon and energy supply in bacteria. Particularly, *M. tuberculosis* prefers host lipids, as evidenced in an over-representation of genes in the genome that encode enzymes for fatty acid metabolism, and upregulation of such genes during macrophage infection [8, 9]. Even though these studies suggest that lipids are the main source of carbon and energy for *M. tuberculosis*, other yet to be identified carbon sources also have an important role to play [9, 10]. In this sense, *M. tuberculosis* is equipped with five putative importers of carbohydrates: one belonging to the major facilitator superfamily and four members of the ATP-Binding Cassette (ABC) transporter family, one of the largest families of paralogous proteins present in the bacillus [11]. Genes encoding ABC transporters account for about 2.5% of *M. tuberculosis* genome, being reported 20 importers and 14 exporters [12]. In comparison with *Escherichia coli*, *Bacillus subtilis*, and even *Mycobacterium smegmatis*, there is a significant reduction of genes encoding ABC transporter components in *M. tuberculosis* genome, particularly evident for the transporters involved in carbohydrate uptake [13]. ABC transporters dedicated to carbohydrate transport have been related to virulence and pathogenesis in bacteria, but the role of these transporters in *M. tuberculosis* still needs to be explored.

ABC transporters type importers are responsible for the translocation of the substrate into the cell and they were identified until today in prokaryotes. Structurally, they consist of oligo protein assemblies with two

hydrophobic transmembrane domains (TMDs) that form the transport channel, two cytoplasmic nucleotide-binding domains (NBDs), which are responsible for the breakdown of ATP and provision of energy for the transport process, and an additional periplasmic substrate-binding protein (SBP) or domain (SBD) exposed to the periplasm of the cell [14].

The four operons that encode ABC transporters dedicated to carbohydrate uptake in *M. tuberculosis* genome are: *lpqY-sugABC*, *rv2038c-41c*, *uspABC*, and *ugpAEBC* [12], where *lpqY*, *rv2041*, *uspC* and *ugpB* encode the SBPs, SugAB, Rv2039c-40c, UspAB and UgpAE encode the heterodimeric TMDs, and *sugC*, *rv2038c* and *ugpC* encode the NBDs. Genetic and cellular approaches applied to the study of LpqY/SugABC transporter demonstrated that it was essential for virulence of *M. tuberculosis* in vivo and potentially involved in recycling of trehalose monomycolate, a cell wall glycolipid [15]. This transporter also raises interest for the detection of *M. tuberculosis* in sputum samples, since it can probably be the pathway for uptake of a solvatochromic trehalose probe [16]. The three-dimensional structure of the *M. smegmatis* full transporter LpqY/SugABC was resolved in different states evidencing important secondary structures and residues for the trehalose transport mechanism [17]. Additionally, the interaction of the SBP LpqY of *M. tuberculosis* and *M. thermoresistibile* with different ligands was also explored [18]. The transporter UspABC consists of the SBP UspC and two TMDs (UspA and UspB) but lacks the NBD domain. The three-dimensional structure of the UspC was solved by X-ray crystallography (PDB codes 5K2X and 5K2Y), and binding studies with different putative substrates showed a higher affinity for carbohydrates containing an amino group at the C2 or C3 position, like D-glucosamine-6-phosphate and chitobiose than *sn*-glycero-3-phosphocholine, D-glucose or α,α -D-trehalose [19]. The UgpAEBC ABC transporter is predicted to be involved in scavenging of glycerophospholipids [20], which are carbon or phosphate sources that could be available for *M. tuberculosis* inside of macrophage or another cell host. The crystal structure of the substrate-binding protein UgpB was resolved in presence of glycerophosphocholine (PDB code 6R1B), but functional studies revealed that the protein also could bind other glycerophosphodiester [21]. The less known *M. tuberculosis* ABC importer is Rv2038c-41c. Studies with the SBP Rv2041c showed increased expression under conditions that are similar to those in a phagocytic environment (low pH and hypoxia) [22]. Immunological studies with a cocktail of five commonly used serological antigens for tuberculosis diagnostic (CFP-10, ESAT-6, HSP-X, Ag85 complex, and PstS1), showed increased sensitivity for TB diagnostic when Rv2041c was added in the mixture,

indicating the capability of this protein to induce the cellular immune response, and highlighting its potential for the development of a vaccine candidate against *M. tuberculosis* [23]. Table 1 shows a resume of the most available information for each component.

In this work, we made a comprehensive comparative analysis of the four ABC transporters type importer from

M. tuberculosis involved in carbohydrate uptake. The phylogenetic relationship among the components and their conservation in different *Mycobacterium* species revealed that Rv2038c-41c and UgpAEBC systems are exclusive of pathogenic *Mycobacterium* species and that the LpqY/SugABC system was possibly the first paralog to diverge in the evolution of *M. tuberculosis*. The

Table 1 Resume of available data regarding the components of carbohydrate ABC transporters from *Mycobacterium* species

| Mycobrow ser identifier | Gene name | Possible function (Mycobrow ser) | Phenotypes of mutants in different conditions and results of different studies of proteomics, biophysics, and others | References |
|-------------------------|-------------|---|--|----------------------|
| Rv1235 | <i>lpqY</i> | Active transport of sugar across the membrane (import) | | |
| Rv1236 | <i>sugA</i> | Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane. | Mutant is essential for in vitro growth and in macrophages; experiments of infection in mice and guinea pig also showed relevance of this gene; shown differential expression under hypoxia, dormancy, and anoxic conditions; influence on biofilm formation | [15-18] |
| Rv1237 | <i>sugB</i> | Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane. | ^a Rv1238 dimerizes in presence of ATP, maximum activity in pH 7.5 and Km of 0.15; presence of regulatory domain as shown for <i>E. coli</i> MalK. | |
| Rv1238 ^a | <i>sugC</i> | Active transport of sugar across the membrane (import). Responsible for energy coupling to the transport system. | | |
| Rv2038c | -- | Active transport of sugar across the membrane (import). Responsible for energy coupling to the transport system. | Mutant is essential for growth in vitro, upregulated under acidic and nutritive stress; experiments of infection in mice and guinea pig also showed relevance of this gene under hypoxia, dormancy and anoxic conditions and influence for biofilm formation. | [18,20,21,24,25] |
| Rv2039c | -- | Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane. | Mutant is essential for in vitro growth. | [20,21] |
| Rv2040c | -- | Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane. | Mutant is essential for in vitro growth, upregulated under acidic and nutritive stress, under hypoxia, dormancy, and anoxic conditions; showed influence for biofilm formation. | [12, 13, 16, 22, 23] |
| Rv2041c ^b | -- | Active transport of sugar across the membrane (import). | Mutant is essential for in vitro growth, upregulated under acidic and nutritive stress, under hypoxia, dormancy and anoxic conditions, showed influence for biofilm formation; ^b <i>ex vivo</i> and <i>in vivo</i> models of <i>M. tuberculosis</i> infection | [22-29] |

phylogeny associated with the structural analysis of the *M. tuberculosis* carbohydrate ABC components allowed us to identify that the segregation of the components is mainly based on sequence and structural features related to their functions. Exploring the characteristics and conservation of the interface between permeases and NBDs, we showed that the absence of a NBD in the UspABC system might be compensated by other NBD present in one of the three *M. tuberculosis* carbohydrate ABC transporters.

Results

The co-occurrence and similarity of the operons encoding carbohydrate ABC transporters type importer of *Mycobacterium tuberculosis* in different taxa

The co-occurrence of components of the four carbohydrate ABC importer of *M. tuberculosis* in different taxa was analysed using String server (Fig. 1). The genes are represented by arrows, whose colors were defined according to their functions. The data show that putative

orthologs of the *lpqY/sugABC*, *rv2038c-41c*, and *ugpAEBC* operons are predominantly present in most of the taxa evaluated, except in Eukaryota and *Rickettsia rickettsii* the causative agent of the tick-borne disease named Rocky Mountain spotted fever (RMSF). *uspABC* operon may be misrepresented due to the lack of an evident NDB component, and *ugpCBEA* is highly conserved in the Actinobacteria group. The group of *Corynebacterium diphtheriae* has no conservation of the systems found in *M. tuberculosis*, despite the great relevance of carbohydrates for the metabolism in the genus [24]. *Nocardia brasiliensis*, an actinobacteria that causes pulmonary disease as *M. tuberculosis*, shows conservation of all components related to the Rv2038c-41c and UgpAEBC systems. This bacterium encodes five times more ABC components than *M. tuberculosis* (516 and less than 100, respectively) and in this sense, more resembles a soil bacterium than a pathogenic bacterium [25]. On the other hand, genes related to *lpqY/sugABC* were significantly represented in *Rhodococcus fascians*, a

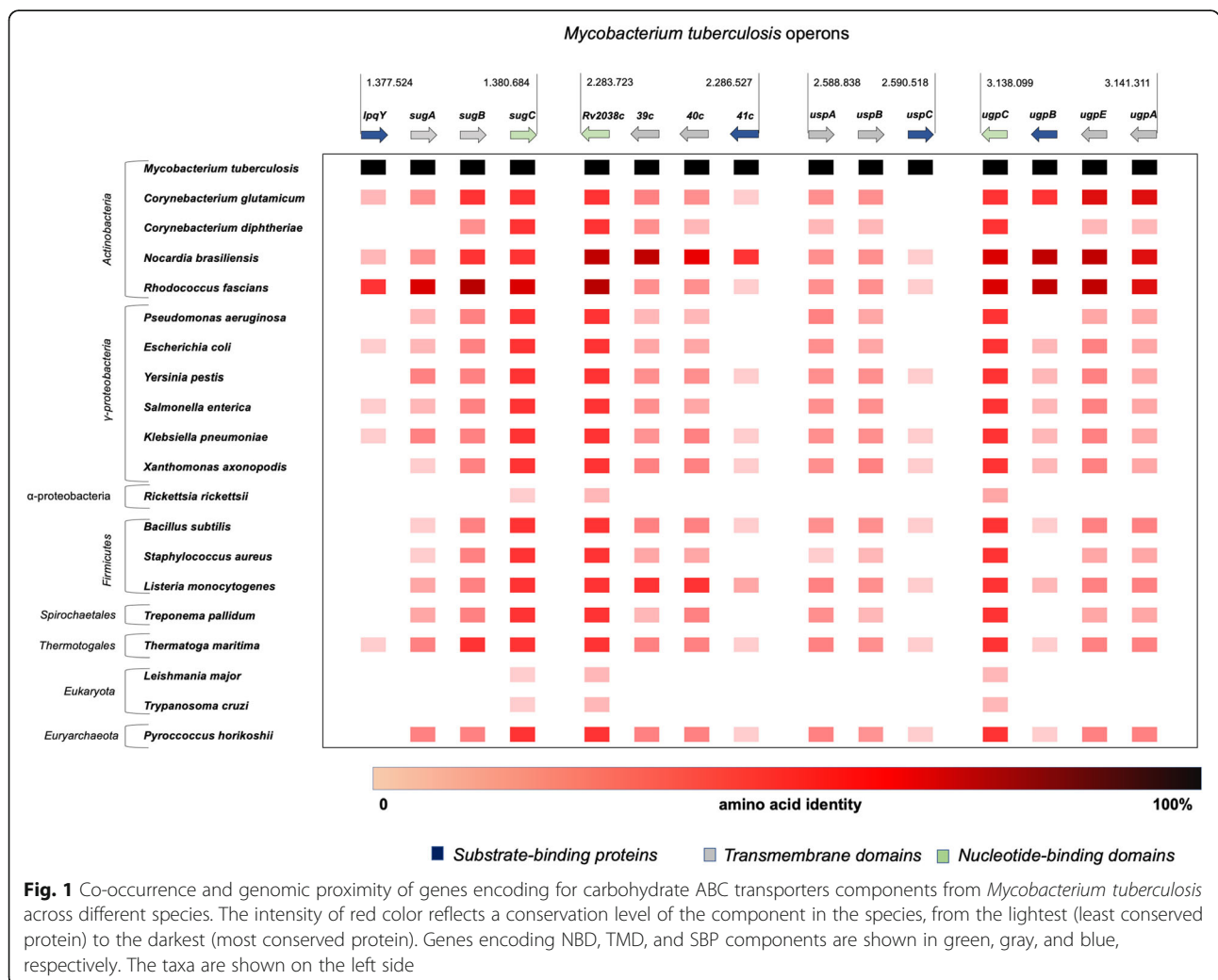


Fig. 1 Co-occurrence and genomic proximity of genes encoding for carbohydrate ABC transporters components from *Mycobacterium tuberculosis* across different species. The intensity of red color reflects a conservation level of the component in the species, from the lightest (least conserved protein) to the darkest (most conserved protein). Genes encoding NBD, TMD, and SBP components are shown in green, gray, and blue, respectively. The taxa are shown on the left side

phytopathogenic bacterium, that elicits an accumulation of the disaccharide trehalose during the early stage of plant infection [26]. Similarly, putative orthologs of *lpqY/sugABC* were identified in the oligotrophic bacterium *R. erythropolis* that is able to survive in a completely inorganic medium with no additional carbon source [27].

Rv2038c-41c and UgpAEBC are restricted to pathogenic species of the *Mycobacterium* genus

In addition to the previous studies, we conducted an analysis of components of putative *M. tuberculosis* carbohydrate ABC importers in 14 different species of *Mycobacterium* genus (Table 2). The choice of species was carefully based on their clinical and biological relevance, including species from well-characterized groups [28]. They consist of human pathogens that belong to the *M. tuberculosis* complex (MTBC) (three *M. tuberculosis* strains, *M. africanum*, and two *M. bovis* strains), pathogenic species non-belonging to the MTBC (*M. avium*, *M. intracellulare*, *M. ulcerans*, *M. marinum*, *M. abscessus*, and *M. leprae*) and the non-pathogenic and environmental species, *M. smegmatis*. The reference list for all of them is presented in Table S1 (Additional file 1).

The results revealed that all species of MTBC conserved orthologs of the four ABC transporters studied in this work (amino acid sequence identity of 96 to 100%) and that UgpAEBC is exclusive of MTBC members and *M. marinum*. Although *M. marinum* is not pathogenic for humans, it is responsible for tuberculosis like infections in fishes [29]. We highlight the results for *M. abscessus* (pathogenic but non-belonging to the MTBC) and *M. smegmatis* (non-pathogenic) that lost UgpAEBC and Rv2038-41c transporters. Furthermore, although *M. smegmatis* and other pathogenic species (except *M. leprae*) have a greater number of ABC transporters than MTBC, they do not include the UgpAEBC system (revised in Transport DB 2.0), suggesting this transporter can be related to the uptake of substrates only available in *Mycobacterium* species that cause tuberculosis or tuberculosis like diseases.

Phylogeny and protein variation evidenced in the components of the four *Mycobacterium tuberculosis* carbohydrate ABC transporters

In order to compare the orthologs identified in *Mycobacterium* species, amino acid sequences of all the

Table 2 Presence of putative orthologs of the carbohydrate ABC transporters from *Mycobacterium tuberculosis* identified in mycobacterial species. Mycobacteria species are grouped as *M. tuberculosis* Complex (MTBC), other pathogens, and *M. smegmatis*, a non-pathogenic species. The protein sequences were obtained using BLASTp analysis against each strain at NCBI using *M. tuberculosis* H37Rv homologs as the query sequence. The cut-off used was taken using coverage > 90% and amino acid sequence identity > 60%

| | LpqY/SugABC | | | | Rv2038c-41c | | | | UspABC | | | UgpAEBC | | | | | | |
|-------------------------------|-------------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|---------------------------|---------------------------|---------------------------|----------------------------|---------------------------|---------------------------|----------------------------|---------------------------|-----------------------------|---------------------------|---------------------------|---------------------------|--------------------------|
| | LpqY | SugA | SugB | SugC | Rv2038c | Rv2039c | Rv2040c | Rv2041c | UspA | UspB | UspC | UgpC | UgpB | UgpE | UgpA | | | |
| MTBC | <i>M. tuberculosis</i> H37Rv | Fv1235 100% | Fv1236 100% | Fv1237 100% | Fv1238 100% | Fv0338c 100% | Fv2039c 100% | Fv2040c 100% | Fv2041c 100% | Fv2316 100% | Fv2317 100% | Fv2318 100% | Fv2832c 100% | Fv2833c 100% | Fv2834c 100% | Fv2835c 100% | | |
| | <i>M. tuberculosis</i> H37Ra | MFA R306560 100% | MFA R306565 100% | MFA R306570 100% | MFA R306575 100% | MFA R310790 100% | MFA R310795 100% | MFA R310800 100% | MFA R310805 100% | MFA R312310 100% | MFA R312315 100% | MFA R315055 100% | MFA R315060 100% | MFA R315065 100% | MFA R315070 100% | MFA R315075 100% | MFA R315080 100% | |
| | <i>M. tuberculosis</i> CDC1551 | MT R306530 100% | MT R306535 100% | MT R306540 100% | MT R306545 100% | MT R310670 100% | MT R310675 99% | MT R310680 100% | MT R310685 100% | MT R312140 100% | MT R312145 100% | MT R312150 100% | MT R312155 100% | MT R314855 100% | MT R314860 100% | MT R314865 100% | MT R314870 100% | |
| | <i>M. africanum</i> GM041182 | MAF R306545 100% | MAF R306550 100% | MAF R306555 100% | MAF R306560 100% | MAF R310625 100% | MAF R310630 100% | MAF R310635 100% | MAF R310640 100% | MAF R312095 100% | MAF R312100 100% | MAF R312105 100% | MAF R314745 100% | MAF R314750 100% | MAF R314755 100% | MAF R314760 100% | MAF R314765 100% | |
| | <i>M. bovis</i> AF2122/97 | BQ2027 ME1267 100% | BQ2027 ME1268 100% | BQ2027 ME1269 100% | BQ2027 ME1270 100% | BQ2027 ME2064C 100% | BQ2027 ME2065C 99% | BQ2027 ME2066C 100% | BQ2027 ME2067C 99.6% | BQ2027 ME2343 100% | BQ2027 ME2344 100% | BQ2027 ME2345 99.77% | BQ2027 ME2856C 100% | BQ2027 ME2857C 99.31% | BQ2027 ME2858C 100% | BQ2027 ME2859C 100% | BQ2027 ME2860C 100% | |
| | <i>M. bovis</i> BCG Pasteur 1173P2 | BC0 R306705 100% | BC0 R306710 100% | BC0 R306715 100% | BC0 R306720 100% | BC0 R310595 100% | BC0 R310600 99.64% | BC0 R310605 100% | BC0 R310610 100% | BC0 R312065 99.66% | BC0 R312070 100% | BC0 R312075 99.77% | BC0 R314680 99.72% | BC0 R321385 100% | BC0 R314695 100% | BC0 R314700 100% | BC0 R314705 100% | |
| | <i>M. avium</i> 104 | MAV R306590 76.92% | MAV R306595 84.25% | MAV R306600 87.96% | MAV R306605 85.64% | MAV R311815 82.35% | MAV R311820 80.71% | MAV R311805 79.67% | MAV R311800 75.83% | MAV R309985 84.14% | MAV R309990 85.77% | MAV R309975 78.86% | - | - | - | - | - | |
| | <i>M. avium</i> subsp. paraTB-10 | MAP R312995 76.71% | MAP R312990 85.59% | MAP R312985 88.52% | MAP R312980 85.90% | MAP R308985 80.71% | MAP R308990 80.71% | MAP R308995 79.67% | MAP R309000 75.80% | MAP R310625 75.80% | MAP R310630 85.79% | MAP R310635 86.13% | MAP R310635 86.13% | MAP R310635 86.13% | MAP R310635 86.13% | MAP R310635 86.13% | MAP R310635 86.13% | MAP R310635 86.13% |
| | <i>M. intracellulare</i> ATCC 13950 | OCU R312165 78.21% | OCU R312170 84.01% | OCU R312175 87.96% | OCU R312180 85.64% | OCU R310620 82.87% | OCU R310625 81.43% | OCU R310630 81.88% | OCU R310635 77.80% | OCU R310640 82.41% | OCU R310645 82.41% | OCU R310650 86.50% | OCU R310655 86.50% | OCU R310660 79.99% | OCU R310665 79.99% | OCU R310670 79.99% | OCU R310675 79.99% | OCU R310680 79.99% |
| | <i>M. ulcerans</i> ACY99 | MUL R323385 No sequence | MUL R323380 No sequence | MUL R323375 No sequence | MUL R323370 No sequence | MUL R311890 82.02% | MUL R311895 84.29% | MUL R311900 85.53% | MUL R311905 85.53% | MUL R306750 85.17% | MUL R306745 83.21% | MUL R306740 81.82% | MUL R306735 81.82% | MUL R306730 81.82% | MUL R306725 81.82% | MUL R306720 81.82% | MUL R306715 81.82% | MUL R306710 81.82% |
| <i>M. marinum</i> | MMAF R321065 76.50% | MMAF R321060 82.99% | MMAF R321055 89.78% | MMAF R321050 84.99% | MMAF R321045 82.20% | MMAF R321040 83.56% | MMAF R321035 83.67% | MMAF R321030 83.67% | MMAF R315040 75.57% | MMAF R315035 83.17% | MMAF R315030 85.84% | MMAF R315025 81.82% | MMAF R309415 81.11% | MMAF R318025 81.82% | MMAF R309410 75.92% | MMAF R309405 84.00% | MMAF R309400 83.28% | |
| <i>M. leprae</i> TN | ML R305255 77.25% | ML R305260 79.28% | ML R305265 89.42% | ML R305270 79.74% | ML R307205 79.49% | ML R307210 80.00% | ML R307215 79.67% | ML R307220 78.77% | ML R309200 81.72% | ML R309205 85.84% | ML R309210 85.84% | ML R309215 85.84% | ML R309220 85.84% | ML R309225 85.84% | ML R309230 85.84% | ML R309235 85.84% | ML R309240 85.84% | |
| <i>M. abscessus</i> ATCC19977 | MAB 1372 64.53% | MAB 1373 73.49% | MAB 1374 80.29% | MAB 1375 72.56% | - | - | - | - | MAB 1717C 71.53% | MAB 1716C 73.43% | MAB 1715C 75.68% | MAB 1714C 64.20% | - | - | - | - | - | |
| <i>M. smegmatis</i> mc2155 | MSMEG 5061 69.76% | MSMEG 5060 78.83% | MSMEG 5059 85.04% | MSMEG 5058 73.83% | - | - | - | - | MSMEG 4466 81.28% | MSMEG 4467 79.17% | MSMEG 4468 70.25% | MSMEG 4469 70.25% | MSMEG 4470 70.25% | MSMEG 4471 70.25% | MSMEG 4472 70.25% | MSMEG 4473 70.25% | MSMEG 4474 70.25% | |

proteins belonging to the same functional group (NBDs, TMDs, and SBPs) were firstly aligned using Clustal Omega and then submitted to MEGA-X software for phylogenetic analyses. For inferences of phylogeny of conserved proteins, we used the maximum likelihood method [30]. In parallel, we use the available three-dimensional structures or built structural models for all components of the *M. tuberculosis* carbohydrate transporters and use them to map the differences found in the alignments. The models were built using SWISS-MODEL server [31] or Modeller program [32], and the information regarding templates, identities, and model quality details are listed in Table S2 (Additional file 1).

The nucleotide-binding domains (NBDs)

The phylogenetic analysis of NBDs showed that they segregate into three main groups (Groups 1 to 3) each one containing SugC, Rv2038c and UgpC orthologs, respectively (Fig. 2A). Orthologs that were grouped with UgpC (Group 2) belong to the MTBC group and *M. marinum*, and they were closely related to the orthologs of Rv2038c (Group 1). The most distant group of those mentioned above was formed by orthologs from SugC (Group 3). Similarly, the groups 1 and 3 (Rv2038c and SugC NBDs) were divided in two branches, one that

consisted of MTBC orthologs, and other that included orthologs of pathogenic mycobacteria and the environmental *M. smegmatis* (Fig. 2A). To assess the differences between the NBDs, we built structural models of the proteins and located the variable regions shown in the alignment after superimposition of them (Fig. 2B, spheres). The *M. tuberculosis* NBDs showed the conserved catalytic domain, similar to the core structure found in many RecA-like motor ATPases [33], and an additional small C-terminal domain that is unique in some ABC transporters, including that related to carbohydrate and ion uptake [34–36]. The amino acid sequence alignment revealed that the catalytic domains were highly conserved but significant differences were found in the regulatory domains, as evidenced by the low sequence identity and presence or absence of specific regions in the proteins (Additional file 2A). The matching and differences between each two proteins is presented in the Additional file 2B. Orthologs from group 3, including SugC, are larger and have at least four (I to IV) additional regions unidentified in the other proteins (Fig. 2B, forest green spheres). It is noted in the alignment that the differences between UgpC and Rv2038 orthologs, which lose the extra SugC regions, are also located in the same region, although the number

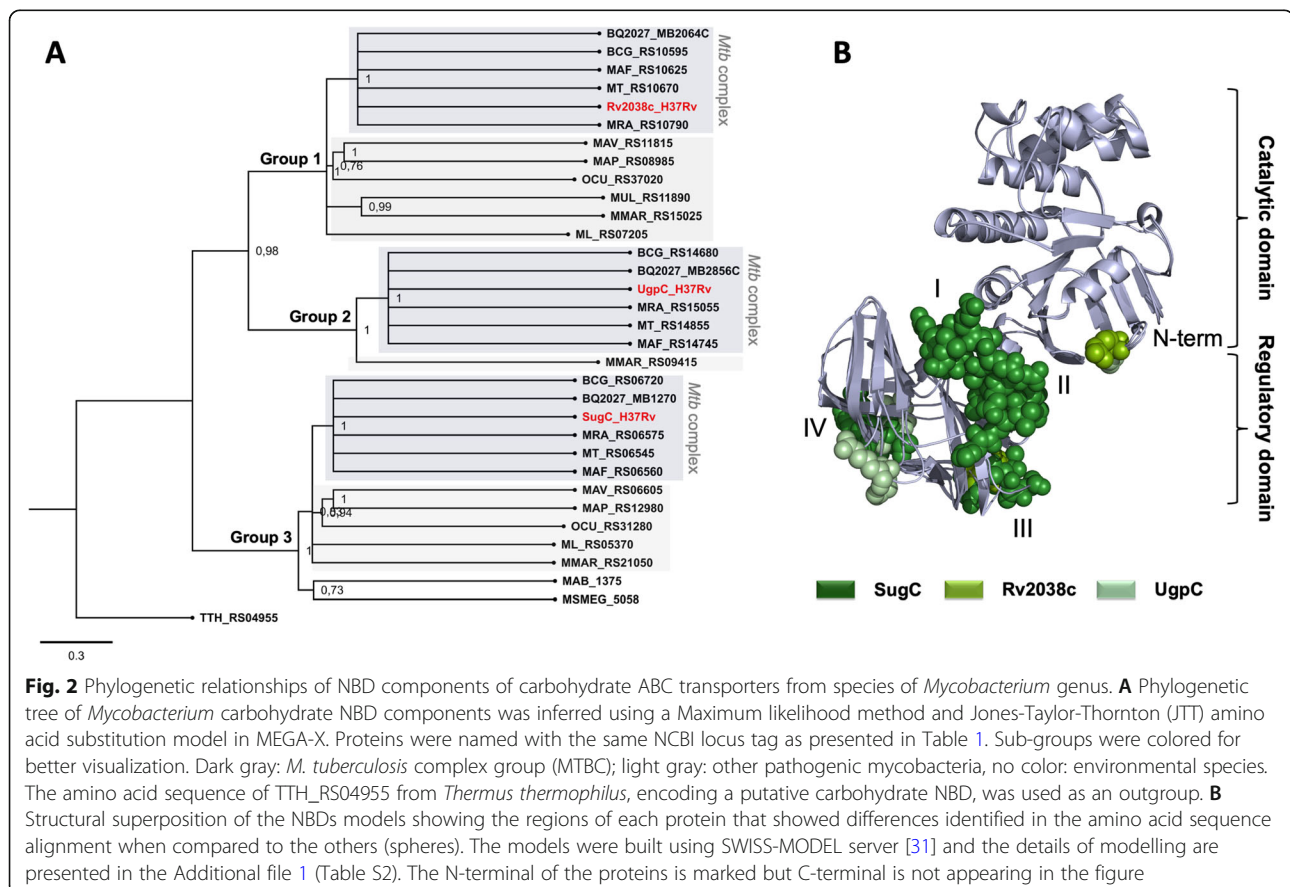


Fig. 2 Phylogenetic relationships of NBD components of carbohydrate ABC transporters from species of *Mycobacterium* genus. **A** Phylogenetic tree of *Mycobacterium* carbohydrate NBD components was inferred using a Maximum likelihood method and Jones-Taylor-Thornton (JTT) amino acid substitution model in MEGA-X. Proteins were named with the same NCBI locus tag as presented in Table 1. Sub-groups were colored for better visualization. Dark gray: *M. tuberculosis* complex group (MTBC); light gray: other pathogenic mycobacteria, no color: environmental species. The amino acid sequence of TTH_RS04955 from *Thermus thermophilus*, encoding a putative carbohydrate NBD, was used as an outgroup. **B** Structural superposition of the NBDs models showing the regions of each protein that showed differences identified in the amino acid sequence alignment when compared to the others (spheres). The models were built using SWISS-MODEL server [31] and the details of modelling are presented in the Additional file 1 (Table S2). The N-terminal of the proteins is marked but C-terminal is not appearing in the figure

of differential residues is reduced (Fig. 2B, split pea and pale green spheres, respectively).

The transmembrane domains (TMDs)

TMD components of ABC importers are responsible for important functions of the transport system, including interaction with the SBPs, formation of the translocation pore through the inner membrane, and interaction with NBDs. The four carbohydrate ABC transporters studied in this work are heterodimers constituted by two different TMDs each. To get phylogenetic insights of the proteins, monomers of each TMDs heterodimer were analyzed separately, forming two groups that we called group 1 and group 2, each one containing one member of each transporter (Additional file 3). Alignments with all amino acid sequences were generated in Clustal Omega for each protein group and used as inputs in MEGA-X software for the building of a rooted tree (Fig. 3A). Group 1 was formed by SugB, Rv2039c, UspB, and UgpE and group 2 by SugA, Rv2040c, UspA, and UgpA. Putative orthologs of SugAB permeases formed a separated group from the three other systems, as shown in SugC NBD group (Fig. 2A). According to the alignments, proteins from group 1 have 5 to 30 additional residues than those from group 2 and seem to be the most variable component in the architecture of *M. tuberculosis* carbohydrate ABC transporters, once alignments by pair revealed large insertions/deletions (indels) regions (Additional file 3A). To identify the location of possible variation regions in the proteins, a structural model of each *M. tuberculosis* TMD component was built (Additional file 1, Table S2), and proteins were compared by pairs (Additional file 3B). In general, the main differences among the proteins are located in the N-terminal that faces the NBDs and in the loop between helices TM1 and TM2, which in all models consists of a region that is more exposed to the periplasm and might be accessed by the SBP. Main differences were found in group 1 of TMDs, mostly including helices 3 and 4. (Fig. 3B, green spheres). No differences were identified in the coupling helices (Additional file 3C) as well as in the helices that form the pore. The SugA orthologs conserved the most residues that in *M. smegmatis* LpqY-SugABC transporter interact with trehalose, but differently, *M. smegmatis* SugB His118 is not conserved (Additional file 3A, in red).

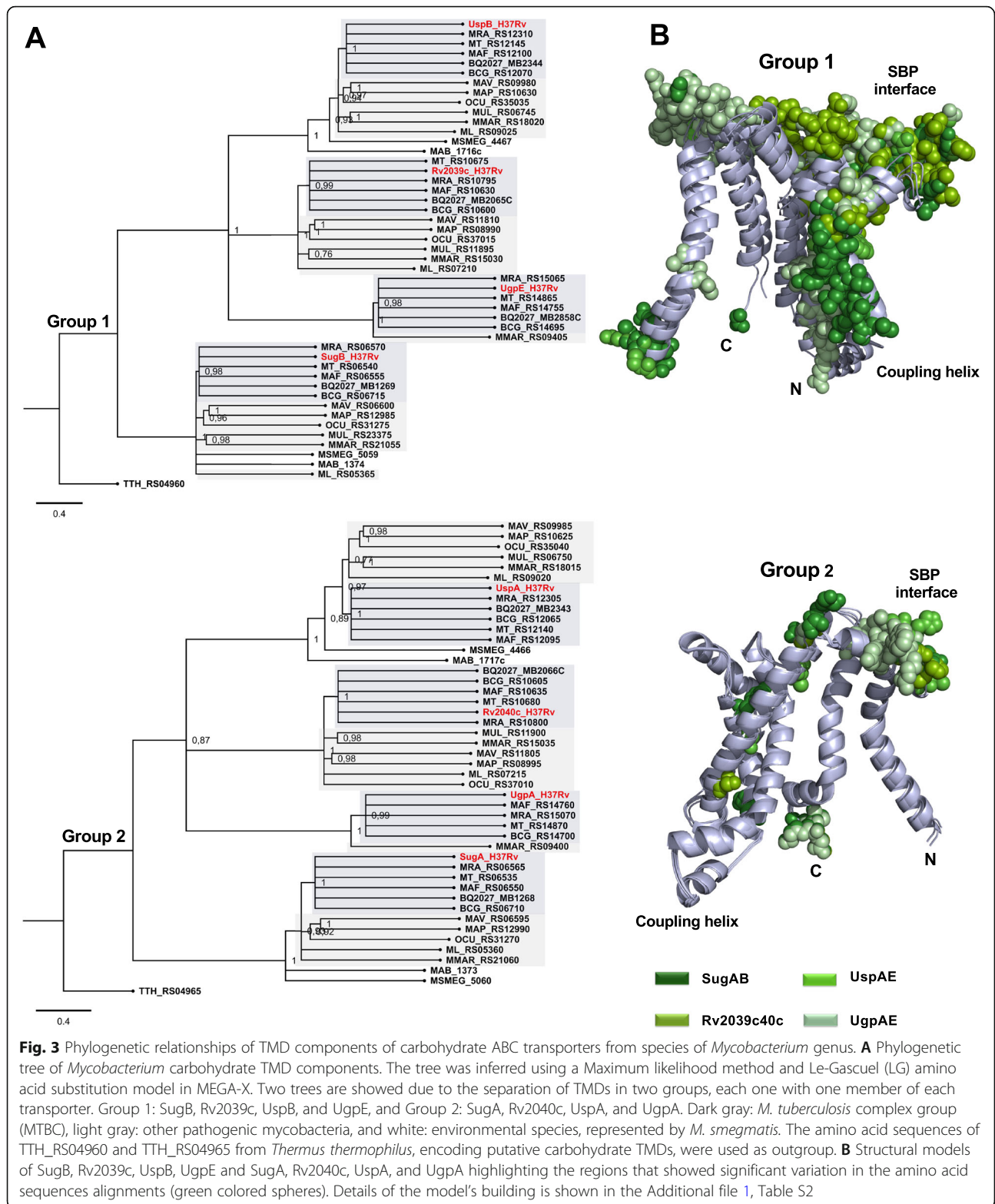
The substrate-binding proteins (SBPs)

The role of SBPs in ABC importers is of great relevance since they are the components responsible for the affinity and specificity of the transport systems. They perform the substrate uptake and transference to the TMDs for the translocation. The interaction between SBPs and TMDs triggers structural movements that will result in

the change of resting to an active state of the transporter [37]. The phylogenetic tree built with the amino acid sequences alignment of carbohydrate SBPs from *Mycobacterium* species showed that each component separated in a unique group with their orthologs (Fig. 4A). UgpB orthologs, exclusive of MTBC and *M. marinum*, are closer to the orthologs of LpqY than UspC and Rv2041c. The available structures of *M. tuberculosis* UgpB (PDB code: 4MFI) [20], UspC (PDB code: 5K2X) [19], and the structural models of LpqY and Rv2041c (Additional file 1, Table S2) were used for mapping the differences evidenced in the amino acid sequence alignments (Fig. 4B). Structurally, the *M. tuberculosis* carbohydrate SBPs consist of two globular domains, N-terminal (domain I) and C-terminal (domain II) that are connected by a hinge, in which interface it is located the substrate-binding site. A comparison among the protein groups in the alignment (Additional file 4) allowed us to identify specific regions with amino acids indels, as shown in Fig. 4B. UspC is the shortest protein and, differently from the three others, it has a deletion of two sets of amino acids, respectively in domain I (opposite to the entrance of binding pocket) and domain II (Fig. 4B, yellow spheres). LpqY, Rv2041c, and UgpB show amino acid insertions in domains I and II (Fig. 4B, spheres). These regions are not involved in the carbohydrate-binding site but can indirectly affect the structure of the binding-pockets. UgpB is the protein that shows more sites of variability, in both domains, including regions that directly affect the substrate-binding pocket (Fig. 4B, Additional file 4).

Substrate-binding pocket comparison of the *M. tuberculosis* H37Rv SBPs

The previous phylogenetics analysis showed that the four *M. tuberculosis* SBP segregate in different groups. We compared the available three-dimensional structures of the *M. tuberculosis* UspC and UgpB and the molecular models of LpqY and Rv2041c. The general structure consisted of two alpha-beta domains (or lobes), characteristic of periplasmic-binding proteins, where domain I is smaller than domain II and it is more conserved (Fig. 5A). The structures conserved the three-dimensional folding, but clearly revealed substrate-binding pockets with significant differences, which is in accordance with phylogenetic analyses that separate them in four groups (Fig. 5B, Additional file 5A). *Mtb* LpqY model was based on the *M. smegmatis* LpqY three-dimensional structure bound to trehalose (PDB 7CAF_E) that shared 69% of amino acid sequence identity (94% coverage). The comparison of model and structure revealed a ligand-binding pocket very similar. From six residues that interact with the sugar in *M. smegmatis* protein [18], *M. tuberculosis* LpqY conserves four (Asp97, Asn151,



Glu288 and Arg421). Residues Asn41 and Glu42 of *M. smegmatis* LpqY are replaced by Asp 41 and Thr42 in the *M. tuberculosis* protein, strongly suggesting it also can bind trehalose. LpqY folding is also

closely related to the trehalose-binding protein of *T. litoralis* (TMBP, PDB code: 1EU8) [34] and the maltose-binding protein of *T. maritima* (MalE3, PDB code: 6DTQ) [38]. The structural model of mature

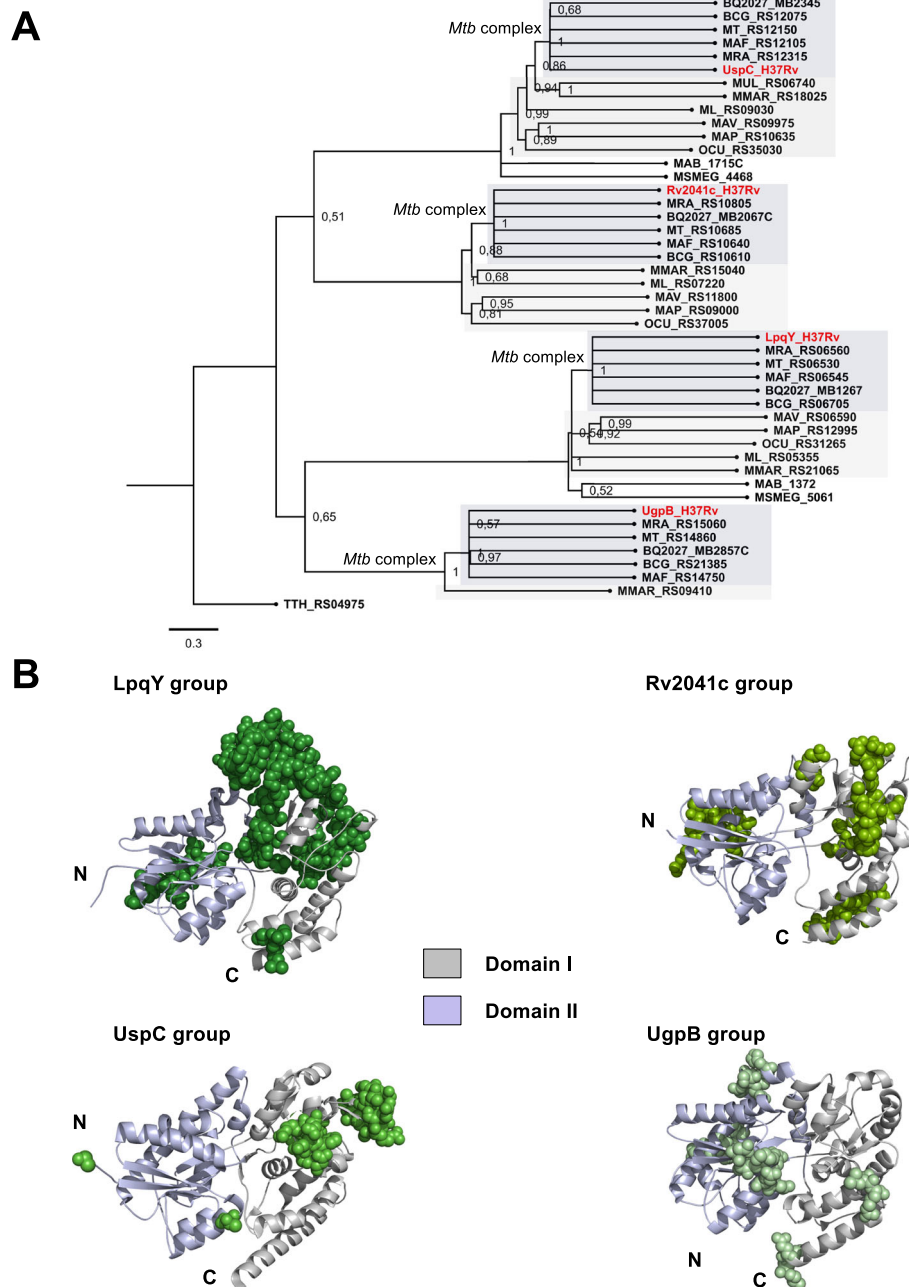
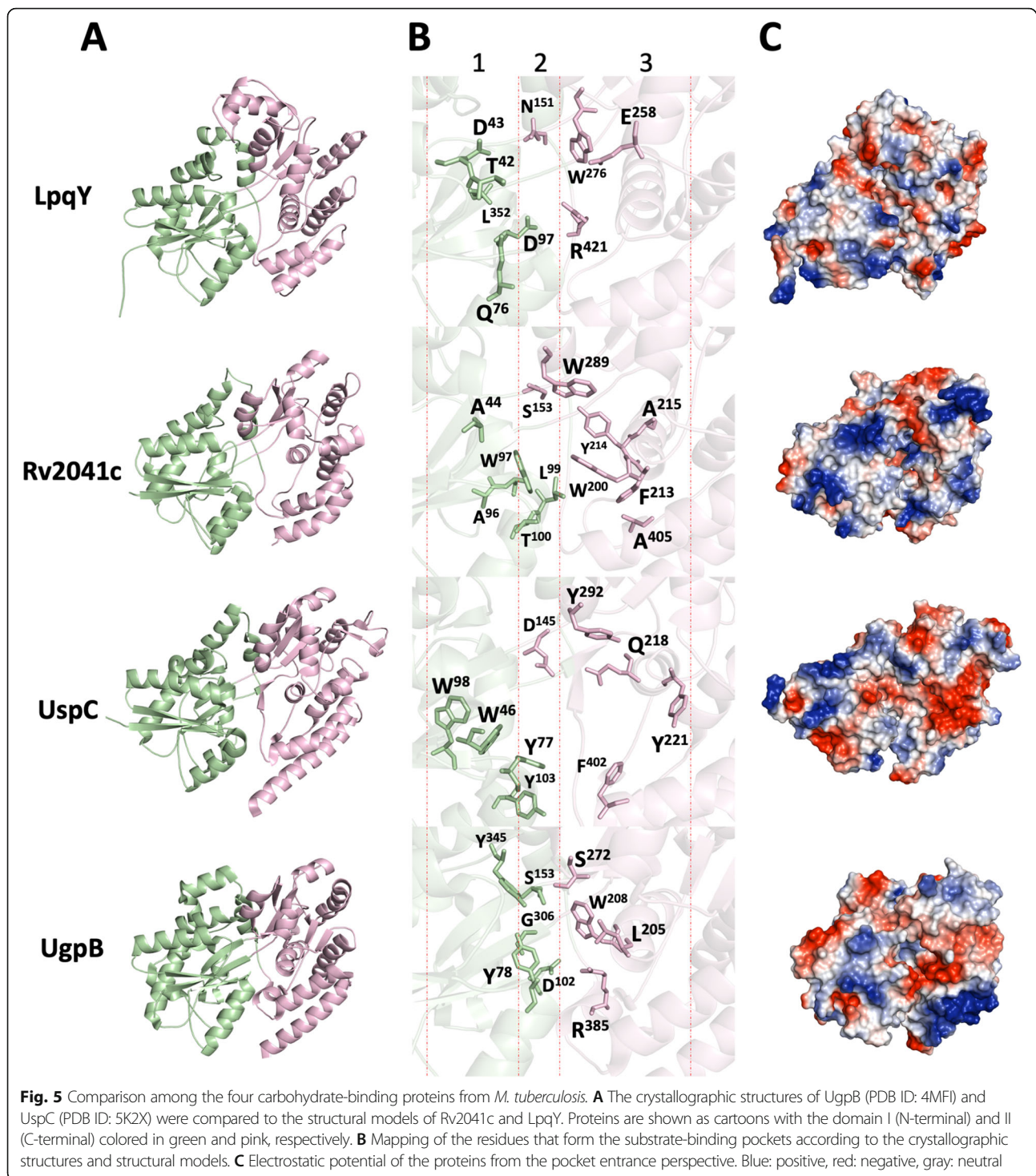


Fig. 4 Phylogenetic relationships of SBP components of carbohydrate ABC transporters from species of *Mycobacterium* genus. **A** Phylogenetic tree of *Mycobacterium* carbohydrate SBP components. The tree was inferred using a Maximum likelihood method and Whelan and Goldman (WAG) amino acid substitution model in MEGA-X. Dark gray: *M. tuberculosis* complex group (MTBC), light gray: other pathogenic mycobacteria, no color: environmental species. The amino acid sequence of TTH_RS04975 from *Thermus thermophilus* was used as an outgroup. **B** Crystallographic structures of UspC (PDB code: 5K2X) and UgpB (PDB code: 4MFI), and structural models of LpqY and Rv2041c were used for representation of regions that showed variation. Regions of amino acid insertion/deletion are represented in colored spheres

protein Rv2041c was generated by SWISS-MODEL server [31] based on the structural similarity with *Listeria monocytogenes* Lmo0181 (PDB code: 5F7V) [39] (Additional file 1, Table S2). However, Rv2041c did not conserve the pocket residues and apparently, there is no possibility for interaction with

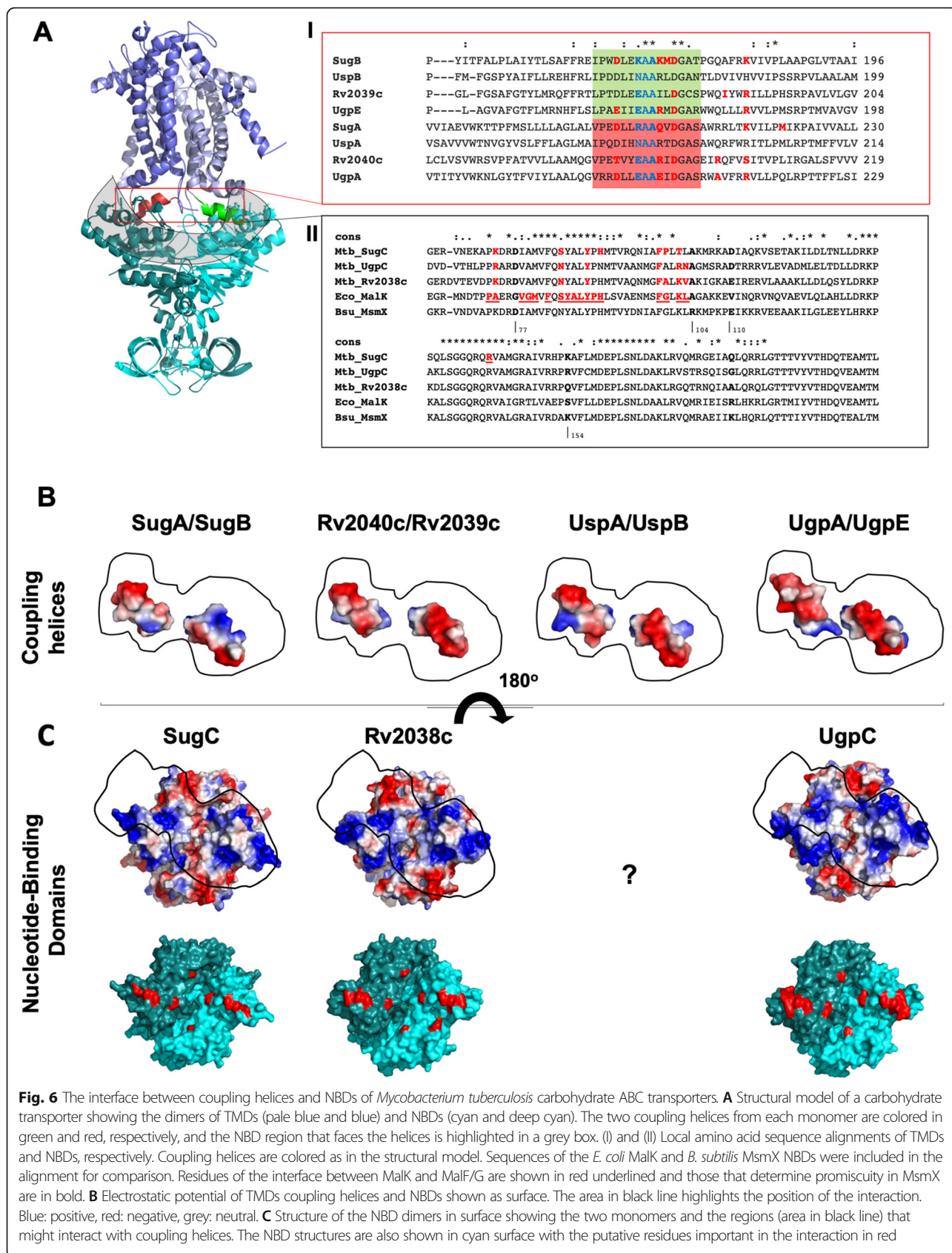
cycloaltnan, the substrate of Lmo0181. The SBPs superimposed well and revealed three subsites (Fig. 5B). Subsites 1 and 3 are rich in aromatic residues, and subsite 2 shows hydrophilic residues, also conserved in the alignment (Additional file 5A). Similar to what we observed in the ligand pockets, the



electrostatic potential of the proteins revealed differences in the entrance of the pocket and TMDs interface regions (Fig. 5C), in accordance with the phylogenetic analysis and supporting the variability of substrates. The structural alignment of the proteins revealed five conserved regions, four of them probably involved in the structural folding (Additional file 5B).

Characterization of the interaction between TMD coupling helices and NBDs

TMDs in ABC transporters are responsible for the pore formation and they play an essential role in the activation of the NBDs during the transport, mediating hydrophobic and hydrophilic interactions with the energy domains to stabilize the complex in the different states



of activity. The absence of a NBD in UspABC arised the questioning if could exist some promiscuity among NBDs of the aforementioned *M. tuberculosis* transporters. Incomplete ABC transporters missing ATPase domains were identified in many genomes of different species as well as orphan ATPase domains [40]. Multi-task NBDs, such as *Bacillus subtilis* MsmX, capable to energize more than one transporter seem to be common in the CUT1 subfamily (di-, tri- and oligosaccharides) [40, 41]. In *E. coli*, MalK and UgpC NBDs were functionally changeable maintaining the functions of the transporters but not the regulatory function [42]. To evaluate this possibility, we firstly compared in the TMDs, the characteristics of the predicted coupling helices, which were localized between TM4 and TM5 (Fig. 6A, green and red helices, and Additional file 3). It was observed that the coupling helices show variation in the first residue of the EAA motif [(E/N/R/K)AA]. UspA/B have an asparagine (non-polar), Rv2039/40c and UgpA/E glutamic acid, and SugA/B lysine or arginine (positively charged), indicating differences in the electrostatic potential of the helices (Fig. 6A-I, in blue). Using GREMLIN server analysis for complexes [43], we predicted the residues performing interactions between TMD and NBD (Additional file 1, Table S3). In relation to TMD residues, it was evident that the first and third residues of the EAA motif play an important role, as well as a set of charged residues, including a highly conserved aspartate (Fig. 6A-I). These residues essentially perform interactions with similar residues in the NBDs. To support our analysis, the putative coupling helices/NBD interfaces were mapped in the structural models (Fig. 6A, gray region) and had their amino acid sequences aligned. The amino acid sequences of *E. coli* MalK and *B. subtilis* MsmX, whose interactions with TMDs were largely explored, were included in the alignment. The MalK residues that interact with MalF and MalG were underlined in Fig. 6A-II, including Phe81 to Tyr87 that accommodate the MalF and MalG helices. The comparison of this region with *M. tuberculosis* sequences reveals high conservation of residues that form an hydrophobic environment in the models. Although the GREMLIN analyses did not identify as many residues as in MalK, the identified residues from *M. tuberculosis* proteins were structurally aligned with MalK. We also looked at the conservation of residues of *M. tuberculosis* NBDs in comparison with *B. subtilis* MsmX. Residues Asp77, Arg104, Glu110 and Lys154 are conserved in NBDs that can complement MsmX activity but they are replaced by hydrophobic residues in MalK [44] (Fig. 6A-II, in black bold). Similarly to MsmX, *M. tuberculosis* proteins conserve an aspartate in the equivalent position

of Asp77 and residues with negative and positive charges such as Glu110 and Lys154, respectively. The only change in *M. tuberculosis* proteins was a replacement of Arg104 by an alanine, as observed in *E. coli* MalK. Not directly interacting with the coupling helices, MsmX Arg104 is responsible for adding charge to the environment and it is supposed to maintain the needed structure of the region for transport [44]. Rv2038c and UgpC proteins seem to share more residues in common than with UspC (Fig. 6A-II). The electrostatic potential of the *M. tuberculosis* coupling helices and NBDs was evaluated and revealed a complementary interface consisting of prominent negative charges in the helices (Fig. 6B), and a set of 8 bunches of positive charges that are spread along the interface generated by the NBD dimers (Fig. 6C).

Discussion

The cellular and molecular mechanisms involved in *M. tuberculosis* nutrition have been largely studied and despite many uncertainties, the importance of carbohydrates and lipids has arised. ABC transporters have a clear contribution during colonization of the host environment, through nutrient scavenging, and evasion or resistance to host defenses. *M. tuberculosis* has four ABC transporters dedicated to carbohydrate transport, which studies have demonstrated they play important role in recycling of trehalose, transport of amino sugars and glycerolphosphocholine, but also in biofilm formation, virulence and immunogenicity [12, 19–22]. In this work, we made a comparative analysis of these transporters detaching the phylogenetic, structural and structural aspects of their components. In general, the analysis of co-occurrence of *M. tuberculosis* carbohydrate ABC transporter components reveals they are poorly represented with sequences that show low similarity. Actinobacteria is the unique group in which these components are more representative, mainly LpqY-SugABC, Rv2038-41c and UgpAEBC systems. There is a consensus for the presence of NBDs, reflecting the high level of conservation of these proteins in ABC transporters. In addition, there is a sporadic absence of periplasmic components and some of the TMDs, which is partly due to their low conservation of the amino acid sequence. In the case of permeases, it is possible that the absence of a component is indicative of the formation of homodimers of TMDs, as evidenced in other ABC systems. Interestingly, the complete absence of these carbohydrate ABC components was evidenced in the *Rickettsia rickettsia* and Eukariota group, which was represented by the two human parasites *Leishmania major* and *Trypanosoma cruzi*. *R. rickettsia* is a strict intracellular pathogen associated with arthropods [45] that has a reduced genome. The identified ABC transporters present in this

microorganism are mainly dedicated to heme acquisition, lipids, toluene tolerance and drug extrusion suggesting that carbohydrates don't represent a significant source of nutrients. The distribution of the four ABC transporters in representative species inside the *Mycobacterium* genus revealed they conserved at least orthologs of two ABC transporters, LpqY/SugABC and UspABC. Interestingly, was the genomic comparison between *M. smegmatis* and *M. tuberculosis*. *M. smegmatis* is a fast-growing non pathogenic species that can transport via ABC transporter systems, a variety of sugars: β -glucosides such as chitobiose, α -galactosides (melibiose), β -xylosides (xylobiose), xylose, arabinose, and sugar alcohols. Differently, *M. tuberculosis* has a reduced number of ABC transporters dedicated to carbohydrate uptake, and from the available data until the moment, both species do not seem to share substrates. These differences reflect the lifestyles of *M. smegmatis* and *M. tuberculosis* in their natural habitats, the soil, and human body, respectively [13]. Additionally, our results additionally showed that Rv2038c-41c and UgpAEBC are not only restricted to *M. tuberculosis* species, but also distributed in other pathogenic species of *Mycobacterium* genus. The presence of the UgpAEBC system exclusively in species of *M. tuberculosis* complex group and *M. marinum* reveal the relevance of this system for the bacilli. UgpB, the SBP component, is a substrate of twin-arginine translocation (Tat) pathway, conserved in different bacterial species [20, 21], upregulated during infection, and essential for virulence and survival in several pathogens [46, 47]. Although *M. marinum* has fish as its official host, it is also capable of infecting humans causing cutaneous tuberculosis. The conservation of UgpAEBC transporter in this species could be important for carbohydrate and phosphate sources uptake during the infection in humans [20, 21]. Indeed, NMR studies of the metabolomic profiling of intact lung tissue at various stages of *M. tuberculosis* infection have revealed a significant increasing of the UgpAEBC substrate (GPC) during the early stages of infection [48]. Rv2038-41c, the closest system to UgpEABC, is also conserved only in pathogenic species suggesting an importance in the infection and pathogenesis processes. Indeed, although the substrate of this transporter is not clear, the SBP Rv2041c was related to intracellular adaptation within the host and considered relevant for the pathogen biology and virulence, since it was upregulated in phagosome acidic and hypoxic conditions [22].

Based on the amino acid sequence alignments and evaluation of structural parameters of the protein models, we observed that the components from the four transporters keep clustering in four groups, separated by functional and structural differences. Evidence for that is highlighted in the differences found in the regulatory

domains from NBDs, in the substrate-binding pockets from periplasmic-binding proteins, and in the interface of TMDs. All NBDs analyzed conserve a catalytic subdomain, while the main differences are observed in the C-terminal regulatory region. Despite that NBDs mostly segregate according to sequence differences in a region that lies between Walker A and B motifs and includes the structurally more diverse helical subdomain, [49], our phylogenetic analyses showed that the differences on the regulatory domains of *M. tuberculosis* NBDs play an essential role in their segregation. The regulatory subdomain of several carbohydrate NBDs conserved short sequence motifs such as FVAXFIGSP, G ψ RPE, and ExxG, (ψ is an apolar residue and x is any amino acid) and C-terminal glycine and phenylalanine residues that can serve as signatures [50]. These motifs and residues are present in *E. coli* MalK and UgpC, *Thermococcus litoralis* MalK, and others, and also in *M. tuberculosis* SugC, Rv2038c, and UgpC. SugC presents a larger regulatory domain and previous phylogenetic analysis demonstrated that all the mycobacterial SugC proteins cluster together depicting the high sequence conservation. Interestingly, these proteins branched out together with homologous proteins from *Pseudomonas syringae* and *Klebsiella pneumoniae*, which are plant and animal pathogens, respectively [27].

The analysis of the TMDs showed predominant differences in the interfaces with SBPs and NBDs, especially that from group 2, indicating their relevance in the recognition and transport mechanism. Our analysis suggests that *M. tuberculosis* carbohydrate TMDs could have diverged from an ancient SugB in group B, and SugA in group A. Comparative studies of TMDs from ABC importers show they conserve six transmembrane segments (TMs) that could be originated from a duplication of a primordial protein containing 3 TMs [51]. The electrostatic potential of the pores formed by the TMDs in the four systems is mainly apolar with positive charges at the entrance and end of the pore, emphasizing the similar characteristics of the substrates.

The periplasmic components showed the highest diversity indicating they select different sets of substrates. On the contrary of NBD and TMD components, the phylogenetic analysis resulting from SBPs showed a different pattern of divergence. It could be explained by the quite discussed diversity of SBP components. The structural comparisons show that the diversity regions could not be necessarily associated with the binding site, but additional regions of the proteins that play distinct roles in the transport system and the bacterium physiology. The structural similarities, the evidence of solvent-accessible aromatic side chains in the binding cleft, and the characteristic acidic molecular surface corroborate their role as carbohydrate transporters. Previous works

have proposed the ability of *M. tuberculosis* to exchange carbohydrates and lipids as energy sources, an essential feature for *M. tuberculosis* survival in macrophages [9, 52]. Carbon and phosphate can be acquired by the action of bacillus phospholipases and glycerol phosphodiesterases on the host's phospholipids. The proximity of the periplasmic proteins to the amino-sugars of the cell wall could facilitate the acquisition of substrates like GPC, trehalose, and chitobiose by UgpB, UspC and LpqY respectively [15, 19, 20].

Structural details shared by all SBPs components of carbohydrate ABC transporters were analyzed and they exhibited a topology of subcluster DI in the structural classification of substrate-binding proteins [53]. Although the structure of Rv2041c is based on a model, the apparent topology and its molecular weights above 40 kDa, suggest it belongs to the same subcluster. Despite apparent conservation in the structure of these proteins, the electrostatic potential of the ligand-binding cleft does not reflect this conservation, indicating they have preferences and affinities for different substrates. Finally, in an attempt to evaluate the putative promiscuity of the *M. tuberculosis* carbohydrate NBDs, our study compared amino acids and charges in the coupling helices and NBDs of the carbohydrate transporters of *M. tuberculosis*. The sequences of the *B. subtilis* MsmX, a multitask NBD, and *E. coli* MalK, were included in the alignment for comparison. The results of this analysis showed the conservation of residues that form the hydrophobic cleft in NBDs for interaction with the coupling helices (Phe81-Tyr87, in *E. coli* MalK). Still, it was possible to notice that from the four residues of *B. subtilis* MsmX (Asp77, Arg104, Glu110 and Lys154) evaluated as signature of multitask NBDs [44], the *M. tuberculosis* proteins conserved the aspartate in position 77 in all the sequences, and Glu110 in Rv2038c, but replaced by an aspartate in SugC and UgpC. Although not identical, the residues in these key positions contribute to the more charged environment of this region when compared to MalK. The presence of charged residues in the vicinity of the coupling helices promotes conformational changes and greater plasticity in this region, favoring a greater number of interactions. Thus, the set of results from this comparison suggests that the carbohydrate NBDs could alternate between transporters in *M. tuberculosis*. This observation, however, depends on further studies to be confirmed.

Conclusions

The results showed that the segregation of the components of the *M. tuberculosis* carbohydrate transporters was determined by sequential alterations that culminated in structural changes, important for determination of their specificity and function. Still, significant differences

were found in the regulatory domains of the NBDs and in the ligand-binding pockets of the SBPs, both defining the function of the transporters. The conservation of charges and residues suggests that there is promiscuity among the NBDs of the analyzed ABC transporters. All similarities and differences addressed in this work can serve as a basis for further studies of development of inhibitors, using mainly the transporters Rv2038-41c and UgpAEBC as targets, since they were conserved in pathogenic species and are not identified in eukaryotes.

Methods

Searching for orthologs

The genes coding carbohydrate ABC transporters components of *M. tuberculosis* were obtained from Mycobrowser (<https://mycobrowser.epfl.ch/>) [54] or KEGG (<https://www.genome.jp/kegg/>) [55] servers. The co-occurrence analysis across different taxa was evaluated using String server (<https://string-db.org/>) [56]. The comparative genomic analysis across *Mycobacterium* species was described by Machowski and collaborators [57]. Briefly, we choose 15 reference genome sequences of *Mycobacterium* strains deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Additional file 1, Table S1), and searched for homologs of the carbohydrate ABC transporter components of *M. tuberculosis*. Identified amino acid sequences were used for BLASTp (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) at NCBI (<https://www.ncbi.nlm.nih.gov/>). Amino acid sequence alignments and identity of each sequence related to *M. tuberculosis* ortholog were performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [58].

Phylogenetic analysis

The alignments and phylogenetic analysis of ABC transporter components (NBDs, TMDs, or SBPs) were visualized and performed using MEGA-X [59]. The alignments were edited manually, using visual inspection of possible conserved and non-conserved regions. Models with the lowest BIC scores (Bayesian Information Criterion) were considered to describe the best substitution pattern. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value ($\ln L$), and the number of parameters (including branch lengths) were calculated. The phylogeny was established from analysis of Maximum Likelihood. As an external group, the components of a putative carbohydrate ABC transporter from *Thermus thermophilus* (NCBI locus tag: TTH_RS04955–70/TTH_RS04975) were used. The robustness of the inferred trees was tested by bootstrap analyses (500 replicates). All trees generated were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Construction of the structural models and interactions

Molecular models of the ABC transporter components were generated using SWISS-MODEL server [31] or Modeller program [32], which used structural coordinates of ortholog proteins deposited in the Protein Data Bank (PDB), as described in detail in Table S2 (Additional file 1). All models were subsequently validated for their stereochemical quality using the program MolProbity [60]. The final model to be used for further analysis was chosen based on the geometrical parameters. All figures were generated using the program PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) [61].

SBP characterization and binding site prediction

Signal peptide sequences of the substrate-binding proteins were predicted using SignalP-5.0 Server [62]. Rv2041c substrate-binding site was predicted after the structural superimposition of the three *M. tuberculosis* carbohydrate-binding proteins with the Rv2041c model (Additional file 1, Table S3). The pocket volumes of the proteins were calculated using the program CASTp [63] with a default probe radius of 1.4 Å and MetaPocket [64].

Characterization of the TMDs and prediction of interfaces with NBDs

The validation of the TMD models was performed in conjunction with the prediction of transmembrane helices using the programs TMHMM Server v. 2.0 [65] and TOPCONS [73] and analysis of the position, residue composition (EAA conservation) and predicted structure of the coupling helices. The Complexes program from Gremlin server (<http://gremlin.bakerlab.org/>) [43], was used for prediction of the interface between TMDs (coupling helices) and NBDs, based on evolutionary information. Since the region of interaction was larger than 60 residues, we used the e-value threshold of E-06 and number of iterations with Jackhammer to 4. We accepted interprotein residue pairs with a scaled score ≥ 1.30 and a probability > 0.88 as co-varying pairs (evolutionary couplings, ECs).

Abbreviations

ABC: ATP-Binding Cassette; MTBC: *Mycobacterium tuberculosis* Complex; NBD: Nucleotide-Binding domain; PDB: Protein Data Bank; SBP: Substrate-Binding Protein; TM: Transmembrane; TMD: Transmembrane domain

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07972-w>.

Additional file 1: Table S1. List of *Mycobacterium* species used in this work. **Table S2.** Proteins used as templates for structural modelling of the carbohydrate ABC transporter components of *M. tuberculosis* H37Rv. SBP: substrate-binding protein; TMD: transmembrane domain; NBD:

nucleotide-binding domain. I-TASSER server or Modeller program were used for structural modeling. **Table S3.** Paired alignment of carbohydrates NBDs and TMDs from *M. tuberculosis* H37Rv. The analysis was performed using Gremlin complexes. NBD: nucleotide-binding domain, TMD: transmembrane domain. Scaled score: "normalized coupling strength", a coupling strength larger than one indicates higher than average coupling between two residues. Probability: P (contact | scaled_score, seq/len). I_probability: P (contact | scaled_score, seq/len, top_inter_score).

Additional file 2. Structural and amino acid sequence differences found in the *M. tuberculosis* carbohydrate NBD components. A Amino acid sequence alignment of NBDs. The alignment made with Clustal Omega only shows the regulatory domains where the main differences are observed. B Structural comparison of NBDs and the variable positions identified in the amino acid alignment. Differences between every two proteins can be evidenced by the colored spheres. The structural models show regions of amino acid insertion/deletion identified when two proteins are compared. The percentage in each box represents the amino acid sequence identity between two orthologues. Structural models of Rv2038c, UgpC, and SugC were built as described in Additional file 1, Table S2.

Additional file 3. Sequential and structural differences found in the *M. tuberculosis* H37Rv carbohydrate ABC transporters TMDs. A Amino acid sequence alignment of TMDs produced with the Clustal Omega program. Proteins can be divided in group 1 (SugB, UspB, UgpB and Rv2039c) and group 2 (SugA, UspA, UgpA and Rv2040c, coloured in gray). The EAA residues of the coupling helix are shown in yellow. Residues forming the interface between SugA and SugB that interact with trehalose in the structure of *M. smegmatis* LpqY-SugABC transporter are shown in red. *M. tuberculosis* SugB residues in cyan are the corresponding in *M. smegmatis* SugB forming the scoop loop. B Structural comparison of TMDs. Comparison can be made for each two models and the coloured residues represent the position of variable regions between two proteins. The percentage in each box represents the amino acid sequence identity between the two proteins. Models of all proteins were built using the SWISS-MODEL program (see Additional file 1, Table S2). C Prediction of topology of the TMDs components from *M. tuberculosis* H37Rv carbohydrate ABC transporters. Amino acid sequences of the proteins were submitted to the TOPCONS program. The red bars highlight the position of coupling helices.

Additional file 4. Sequential and structural position of the main differences between the *M. tuberculosis* sugar transporters SBPs. A Amino acid sequence alignment of the SBPs showing the four groups (bold letters). The alignment was made using Clustal Omega. Residues in red are involved in the ligand binding. B Proteins in cartoon represent the modelled structures of all SBPs, except for UgpB and UspC that have crystallographic structures (PDB IDs: 4MFI and 5K2X, respectively). After the amino acid sequence alignment of each two structures different regions were shown by coloured spheres and can be compared. Models of LpqY and Rv2041c were built using the SWISS-MODEL program [31] (see Additional file 1, Table S2).

Additional file 5. Amino acid sequence alignment of the *M. tuberculosis* H37Rv sugar binding proteins and structural information. A Alignment of UgpB, Rv2041c, LpqY and UspC obtained with the Expresso program in the T-coffee server (<http://tcoffee.org.cat/apps/tcoffee/do:expresso>). Amino acids that form the ligand-binding pocket are highlighted in green. The region of the three subsites is shown. B Three-dimensional structure of UgpB in cartoon showing the five regions with highest conservation (left). The putative function of each region is represented in the structure in color surface (right).

Acknowledgments

We thank Professor Luís Carlos de Souza Ferreira from the Vaccine Development Laboratory at the Department of Microbiology (University of São Paulo) for the facility assistance.

Authors' contributions

LID performed all analysis and wrote all sections of the manuscript. JGVM and AGCM performed the phylogenetic analysis. SCB performed molecular

modeling. AB supervised, wrote, reviewed, and edited the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), process number 2018/20162–9; Conselho Nacional de Desenvolvimento Científico (CNPq), process number 401505/2016–2; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and COLCIENCIAS (PhD fellowship for LID and SCB). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its additional files. Genomic sequences of *Mycobacterium* species were obtained from the National Center for Biotechnology Information – NCBI (<https://www.ncbi.nlm.nih.gov/>). All NCBI accession numbers were listed in the Additional file 1, Table S1. All *Mycobacterium* spp. protein sequences were obtained from Mycobrowser (<https://mycobrowser.epfl.ch/>). Sequences of other microorganisms were obtained in the Kyoto Encyclopedia of Genes and Genomes - KEGG (<https://www.genome.jp/kegg/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Microbiology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil. ²Genetics and Molecular Biology Postgraduate Program, Institute of Biology, State University of Campinas, São Paulo, Brazil. ³Biomedical Research Group, University of Sucre, Sucre, Colombia. ⁴Department of Parasitology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil. ⁵Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil. ⁶Laboratory of Applied Structural Biology, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1374; Cidade Universitária, São Paulo, Brazil.

Received: 20 October 2020 Accepted: 3 September 2021

Published online: 19 November 2021

References

- World Health Organization. Global tuberculosis report 2019. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO
- Ehrt S, Schnappinger D, Rhee KY. Metabolic principles of persistence and pathogenicity in *Mycobacterium tuberculosis*. *Nat Rev Microbiol*. 2018;16(8):496–507. <https://doi.org/10.1038/s41579-018-0013-4>.
- Timm J, Post FA, Bekker LG, Walther GB, Wainwright HC, Manganeli R, et al. Differential expression of iron, carbon, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc Natl Acad Sci U S A*. 2003;100(24):24–14326. <https://doi.org/10.1073/pnas.2436197100>.
- Niderweis M. Nutrient acquisition by mycobacteria. *Microbiology*. 2008;154(3):3–692. <https://doi.org/10.1099/mic.0.2007/012872-0>.
- Lin W, De Sessions PF, Teoh GHK, Mohamed ANN, Zhu YO, Koh VHQ, et al. Transcriptional profiling of *Mycobacterium tuberculosis* exposed to in vitro lysosomal stress. *Infect Immun*. 2016;84(9):9–2523. <https://doi.org/10.1128/IAI.00072-16>.
- Jeckelmann JM, Erni B. Transporters of glucose and other carbohydrates in bacteria. *Pflugers Arch*. 2020;472(9):742–1153. <https://doi.org/10.1007/s00424-020-02379-0>.
- Tanaka KJ, Song S, Mason K, Pinkett HW. Selective substrate uptake: The role of ATP-Binding Cassette (ABC) importers in pathogenesis. *Biochim Biophys Acta Biomembr Elsevier*. 2018;1860:4.
- Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393(6685):537–44. <https://doi.org/10.1038/31159>.
- Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, et al. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J Exp Med*. 2003;198(5):5–704. <https://doi.org/10.1084/jem.20030846>.
- Pandey AK, Sasseti CM. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A*. 2008;105:11.
- Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev*. 2000;24(4):4–467. <https://doi.org/10.1111/j.1574-6976.2000.tb00550.x>.
- Oliveira MCB, Balan A. The ATP-binding cassette (ABC) transport systems in *Mycobacterium tuberculosis*: structure, function, and possible targets for therapeutics. *Biology*. 2020;9(12):443. <https://doi.org/10.3390/biology9120443>.
- Titgemeyer F, Amon J, Parche S, Mahfoud M, Bail J, Schlicht M, et al. A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *J Bacteriol*. 2007;189(16):16–5915. <https://doi.org/10.1128/JB.00257-07>.
- Ter Beek J, Guskov A, Slotboom DJ. Structural diversity of ABC transporters. *J Gen Physiol*. 2014;1:43–4.
- Kalscheuer R, Weinrick B, Veeraraghavan U, Besra GS, Jacobs WR. Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*. 2010;107(50):50–21766. <https://doi.org/10.1073/pnas.1014642108>.
- Kamariza M, Shieh P, Ealand CS, Peters JS, Chu B, Rodriguez-Rivera FP, et al. Rapid detection of *Mycobacterium tuberculosis* in sputum with a solvatochromic trehalose probe. *Sci Transl Med*. 2018;10(430):eaam6310. <https://doi.org/10.1126/scitranslmed.aam6310>.
- Liu F, Liang J, Zhang B, et al. Structural basis of trehalose recycling by the ABC transporter LpqY-SugABC. *Sci Adv*. 2020;6(44):eabb9833.
- Furze CM, Delso I, Casal E, Guy CS, Seddon C, Brown CM, et al. Structural basis of trehalose recognition by the mycobacterial LpqY-SugABC transporter. *J Biol Chem*. 2021;18:100307. <https://doi.org/10.1016/j.jbc.2021.100307>.
- Fullam E, Prokes I, Fütterer K, Besra GS. Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biol*. 2016;6(6). <https://doi.org/10.1098/rsob.160105>.
- Jiang D, Zhang Q, Zheng Q, Zhou H, Jin J, Zhou W, et al. Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J*. 2014;281(1):1–341. <https://doi.org/10.1111/febs.12600>.
- Fenn JS, Nepravishta R, Guy CS, Harrison J, Angulo J, Cameron AD, et al. Structural basis of glycerophosphodiester recognition by the *Mycobacterium tuberculosis* substrate-binding protein UgpB. *ACS Chem Biol*. 2019;14(9):9–1887. <https://doi.org/10.1021/acscchembio.9b00204>.
- Kim SY, Lee BS, Sung JS, Kim HJ, Park JK. Differentially expressed genes in *Mycobacterium tuberculosis* H37Rv under mild acidic and hypoxic conditions. *J Med Microbiol*. 2008;57(12):12–1480. <https://doi.org/10.1099/jmm.0.2008/001545-0>.
- Shin SJ, Kim SY, Shin AR, Kim HJ, Cho SN, Park JK. Identification of Rv2041c, a novel immunogenic antigen from *Mycobacterium tuberculosis* with serodiagnostic potential. *Scand J Immunol*. 2009;70(5):457–64. <https://doi.org/10.1111/j.1365-3083.2009.02324.x>.
- Blombach B, Seibold GM. Carbohydrate metabolism in *Corynebacterium glutamicum* and applications for the metabolic engineering of L-lysine production strains. *Appl Microbiol Biotechnol*. 2010;86(5):1313–22. <https://doi.org/10.1007/s00253-010-2537-z>.
- Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Ocampo-Candiani J. Complete genome sequence analysis of *Nocardia brasiliensis* HUJEG-1 reveals a saprobic lifestyle and the genes needed for human pathogenesis. *PLoS One*. 2013;8(6):e65425. <https://doi.org/10.1371/journal.pone.0065425>.
- Depuydt S, Trenkamp S, Fernie AR, Elftieh S, Renou JP, Vuylsteke M, et al. An integrated genomics approach to define niche establishment by *Rhodococcus fascians*. *Plant Physiol*. 2009;149(3):1366–86. <https://doi.org/10.1104/pp.108.131805>.

27. Matsuoka T, Yoshida N. Functional analysis of putative transporters involved in oligotrophic growth of *Rhodococcus erythropolis* N9T-4. *Appl Microbiol Biotechnol.* 2019;103(10):4167–75. <https://doi.org/10.1007/s00253-019-09714-1>.
28. Fedrizzi T, Meehan CJ, Grottola A, Giacobazzi E, Fregni Serpini G, Tagliazucchi S, et al. Genomic characterization of nontuberculous mycobacteria. *Sci Rep.* 2017;7(1):45258. <https://doi.org/10.1038/srep45258>.
29. Aubry A, Mougari F, Reibel F, Cambau E. *Mycobacterium marinum*. *Microbiol Spectr.* 2017;5(2).
30. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet.* 2012;13(5):303–14. <https://doi.org/10.1038/nrg3186>.
31. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296–303. <https://doi.org/10.1093/nar/gky427>.
32. Sali A, Blundell TL. Comparative modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234(3):3–815. <https://doi.org/10.1006/jmbi.1993.1626>.
33. Vetter IR, Wittinghofer A. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. *Q Rev Biophys.* 1999;32(1):1–56. <https://doi.org/10.1017/S0033583599003480>.
34. Diederichs K. Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. *EMBO J.* 2000;19(22):22–5961. <https://doi.org/10.1093/emboj/19.22.5951>.
35. Kadaba NS, Kaiser JT, Johnson E, Lee A, Rees DC. The high-affinity *E coli* methionine ABC transporter: Structure and allosteric regulation. *Science.* 2008;321(5886):5886–253. <https://doi.org/10.1126/science.1157987>.
36. Gerber S, Comellas-Bigler M, Goetz BA, Locher KP. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science.* 2008;321(5886):5886–250. <https://doi.org/10.1126/science.1156213>.
37. Thomas C, Tampé R. Structural and mechanistic principles of ABC transporters. *Annu Rev Biochem.* 2020;89(1):605–36. <https://doi.org/10.1146/annurev-biochem-011520-105201>.
38. Shukla S, Bafna K, Gullett C, Myles DAA, Agarwal PK, Cuneo MJ. Differential substrate recognition by maltose binding proteins influenced by structure and dynamics. *Biochemistry* 2018; 57: 5864–5876, 40, DOI: <https://doi.org/10.1021/acs.biochem.8b00783>.
39. Light SH, Cahoon LA, Halavaty AS, Freitag NE, Anderson WF. Structure to function of an α -glucan metabolic pathway that promotes *Listeria monocytogenes* pathogenesis. *Nat Microbiol.* 2016;2.
40. Ferreira MJ, Mendes AL. S-Noguesal. The MsmX ATPase plays a crucial role in pectin mobilization by *Bacillus subtilis*. *PLoS One.* 2017;12(12):1–22. <https://doi.org/10.1371/journal.pone.0189483>.
41. Marion C, Aten AE, Woodiga SA, King SJ. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. *Infect Immun.* 2011;79(10):4193–200. <https://doi.org/10.1128/IAI.05290-11>.
42. Hekstra D, Tommassen J. Functional exchangeability of the ABC proteins of the periplasmic binding protein-dependent transport systems Upp and mal of *Escherichia coli*. *J Bacteriol.* 1993;175(20):6546–52. <https://doi.org/10.1128/jb.175.20.6546-6552.1993>.
43. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife.* 2014;3. <https://doi.org/10.7554/eLife.02030>.
44. Leisico F, Godinho LM, Gonçalves IC, Silva SP, Carneiro B, Romão MJ, et al. Multitask ATPases (NBDs) of bacterial ABC importers type I and their interspecies exchangeability. *Sci Report.* 2020;10(1):19564. <https://doi.org/10.1038/s41598-020-76444-0>.
45. Renesto P, Ogata H, Audic S, Claverie J-M, Raoult D. Some lessons from *Rickettsia* genomics. *FEMS Microbiol Rev.* 2005;29(1):99–117. <https://doi.org/10.1016/j.femsre.2004.09.002>.
46. McDonough JA, McCann JR, Tekippe EME, Silverman JS, Rigel NW, Braunstein M. Identification of functional tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol.* 2008;190(19):6428–38. <https://doi.org/10.1128/JB.00749-08>.
47. Chandravanshi M, Gogoi P, Kanaujia SP. Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Upp ATP-binding cassette transporter. *Gene [Internet] Elsevier BV.* 2016;591:260–8.
48. Somashekar BS, Amin AG, Ritner CD, Trout J, Basaraba R, Izzo A, et al. Metabolic profiling of lung granuloma in *Mycobacterium tuberculosis* infected guinea pigs: *ex vivo* 1H magic angle spinning NMR studies. *J Proteome Res.* 2011;10(9):4186–95. <https://doi.org/10.1021/pr2003352>.
49. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev.* 2008;72(2):2–364. <https://doi.org/10.1128/MMBR.00031-07>.
50. Zheng WH, Västermark Å, Shlykov MA, Reddy V, Sun EI, Saier MH. Evolutionary relationships of ATP-binding cassette (ABC) uptake porters. *BMC Microbiol.* 2013;13(1). <https://doi.org/10.1186/1471-2180-13-98>.
51. Rengarajan J, Bloom BR, Rubin EJ. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. 2005; 102(23):23–8332. <https://doi.org/10.1073/pnas.0503272102>.
52. Scheepers GH, Lycklama A, Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* 2016;590(23):23–4401. <https://doi.org/10.1002/1873-3468.12445>.
53. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis Elsevier Ltd.* 2011;91:1.
54. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):1–30. <https://doi.org/10.1093/nar/28.1.27>.
55. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res Oxford University Press.* 2019;47:D1.
56. Machowski EE, Senzani S, Ealand C, Kana BD. Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. *BMC Microbiol.* 2014;14.
57. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol.* 2011;7(1):1. <https://doi.org/10.1038/msb.2011.75>.
58. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; 35:6, 6, 1549, DOI: <https://doi.org/10.1093/molbev/msy096>.
59. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Cryst.* 2010;D66(1):–21. <https://doi.org/10.1107/S0907444909042073>.
60. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 2.0; 2015.
61. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420–3. <https://doi.org/10.1038/s41587-019-0036-z>.
62. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018;46:W1.
63. Huang B. Metapocket: a meta-approach to improve protein ligand binding site prediction. *Omi A J Integr Biol.* 2009;13(4):4–330. <https://doi.org/10.1089/omi.2009.0045>.
64. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305(3):3–580. <https://doi.org/10.1006/jmbi.2000.4315>.
65. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 2015;43(W1):W1–W407. <https://doi.org/10.1093/nar/gkv485>.
66. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* 2011;7(9):9. <https://doi.org/10.1371/journal.ppat.1002251>.
67. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. *MBio.* 2017;8(1):1. <https://doi.org/10.1128/mBio.02133-16>.
68. Schubert OT, Ludwig C, Kogadeeva M, Kaufmann SHE, Sauer U, Schubert OT, et al. Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. 2015;18(1):1–108. <https://doi.org/10.1016/j.chom.2015.06.001>.
69. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. 2003;100(22):22–12994. <https://doi.org/10.1073/pnas.2134250100>.

70. Kruh NA, Trout J, Izzo A, Prenni J, Dobos KM. Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PLoS One*. 2010;5(11):11. <https://doi.org/10.1371/journal.pone.0013938>.
71. Ferreira MJ, S-Nogueira IA. Multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*. *J Bacteriol*. 2010;192(20):5312–8. <https://doi.org/10.1128/JB.00832-10>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

