UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

LILIA JUDITH IRIARTE DE LA TORRE

CARACTERIZAÇÃO ESTRUTURAL E FUNCIONAL DOS TRANSPORTADORES DO TIPO ABC DE AÇÚCARES EM *Mycobacterium tuberculosis*

FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF SUGAR ABC TRANSPORTERS IN *Mycobacterium tuberculosis*

CAMPINAS

2020

LILIA JUDITH IRIARTE DE LA TORRE

CARACTERIZAÇÃO ESTRUTURAL E FUNCIONAL DOS TRANSPORTADORES DO TIPO ABC DE AÇÚCARES EM *Mycobacterium tuberculosis*

FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF SUGAR ABC TRANSPORTERS IN *Mycobacterium tuberculosis*

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Doutora em Genética e Biologia Molecular, na área de Genética de Microorganismos.

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Genetics of Microorganisms.

Orientador: PROFA. DRA. ANDREA BALAN FERNANDES

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA LILIA JUDITH IRIARTE DE LA TORRE E ORIENTADA PELA PROFA. DRA. ANDREA BALAN FERNANDES.

CAMPINAS

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

D375c

De la Torre, Lilia Iriarte, 1990-
Caracterização estrutural e funcional dos transportadores do tipo ABC de açúcares em *Mycobacterium tuberculosis* / Lilia Judith Iriarte De la Torre. – Campinas, SP : [s.n.], 2020.

Orientador: Andrea Balan Fernandes.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Açúcar - Transporte. 2. Transportadores de cassetes de ligação de ATP. 3. *Mycobacterium tuberculosis*. 4. Filogenia. 5. Interação proteína-proteína. I. Balan, Andrea. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Functional and structural characterization of sugar ABC transporters in *Mycobacterium tuberculosis*
Palavras-chave em inglês:
Sugar - Transportation
ATP-binding cassette transporters
Mycobacterium tuberculosis
Phylogeny
Protein-protein interaction
Área de concentração: Genética de Microorganismos
Titulação: Doutora em Genética e Biologia Molecular
Banca examinadora:
Andrea Balan Fernandes [Orientador]
Adriana Santos Soprano
Carla Cristina Polo
Rita de Cássia Café Ferreira
Jörg Kobarg
Data de defesa: 27-11-2020
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-8802-5469
- Currículo Lattes do autor: http://lattes.cnpq.br/7624996776012633

Campinas, 27 de novembro de 2020.

COMISSÃO EXAMINADORA

Dra. Andrea Balan Fernandes

Dra. Adriana Santos Soprano

Dra. Carla Cristina Polo

Dra. Rita de Cássia Café Ferreira

Prof. Dr. Jörg Kobarg

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Genética e Biologia Molecular da Unidade do Instituto de Biologia.

DEDICATÓRIA

A mi abuelo Julio, que me enseñó la poesía y el alma mansa.

AGRADECIMENTOS

A Deus, porque mesmo com minha indisciplina espiritual, Ele me conforta em cada momento.

Aos meus pais, Enrique Iriarte e Libia De la Torre, as pessoas mais sensatas que conheço. Espero desenvolver algum dia um pouquinho do bom juízo que eles têm.

Ao meu namorado José Gabriel, porque a sua paciência, amor e cuidado é um dos melhores presentes que Deus me deu até hoje.

Um agradecimento especial à minha orientadora Andrea Balan, a quem admiro em muitos aspectos. Além de ser minha guia no mundo acadêmico, ela é uma importante referência feminina para mim.

A Gobernación de Sucre (Sucre, Colômbia) por ter me concedido uma bolsa de estudos no exterior (Convocatoria 678 de 2014).

As agências de fomento, CAPES e FAPESP com números de processo 1/2020 e 2018/20162-9, respectivamente, pelo suporte financeiro do meu projeto de doutorado.

Ao ICB-USP, IF-USP e CNPEM por me permitir realizar experimentos nos seus laboratórios.

Ao Dr. Marko Hyvönen da Universidade de Cambridge, pelos cursos, pelas ideias, e por toda a ajuda metodológica.

Ao antigo e novo LBEA: Sindy Cabarca, Marcelo Barreto, Andreia Navarro, Cristiane Tambascia, Aline Sampaio, Cindy Lee, Gabriel Guerra, Brenda Oliveira, Pamela Pena, Gerardo Libreros, e o grupo LBEA-Fármacos.

Ao Dr. Gabriel Vignoli do IF-USP, pela paciência nos experimentos de fluorescência, e pela confiança.

Ao Prof. Cristiano Oliveira e o Dr. André Senhem do IF-USP, pelo apoio nos experimentos de SAXS.

Ao Dr. André Guilherme Costa-Martins, pelo apoio nas análises filogenéticas.

Ao grupo do Prof. Luís Carlos de Souza Ferreira, pelo apoio nos experimentos imunológicos.

Aos membros do Laboratório de Taxonomia e Filogenia de Tripanosomatídeos, pela amizade nestes anos.

Aos membros da Banca de Qualificação: Dra. Ana Carolina Migliorini, Dr. Jörg Kobarg e Dr. Renato Vicentini.

Aos membros da Banca de Defesa pela disponibilidade para a arguição do trabalho: Dra. Adriana Santos Soprano, Dra. Carla Polo, Dr. Jorg Köbarg, e Dra. Rita de Cássia Café Ferreira.

Ao programa de Pós-Graduação em Genética e Biologia Molecular e a UNICAMP pela oportunidade de realizar os meus estudos de doutorado.

A todos os que contribuíram para a realização dessa tese.

Este projeto foi desenvolvido no Departamento de Microbiologia do Instituto de Ciências Biomédicas II da Universidade de São Paulo. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001.

MUITO OBRIGADA!

... para descubrir con casi cuarenta años de retraso los privilegios de la simplicidad.

Gabriel García Márquez sobre el coronel Aureliano Buendía en Cien años de Soledad.

RESUMO

Os transportadores ABC (*ATP-Binding Cassette*) formam uma grande superfamília de complexos de proteínas que promovem o transporte de uma ampla gama de substratos através de membranas biológicas em eucariotos e procariotos. Estruturalmente consistem em dois domínios transmembrana hidrofóbicos (TMDs) e dois domínios citoplasmáticos de ligação ao nucleotídeo (NBDs). No caso dos importadores, uma proteína periplasmática ou domínio de ligação ao substrato (SBP) adicional é responsável pela afinidade e a especificidade do transporte. Os importadores ABC de açúcares são de grande importância na nutrição e sobrevivência de bactérias patogênicas e a identificação e caracterização destes sistemas é um passo para o conhecimento sobre a fisiologia do microrganismo. Este trabalho, teve como objetivo principal o estudo de quatro sistemas ABC relacionados ao transporte de açúcares em *M. tuberculosis*: LpqY/SugABC, Rv2038c-41c, UspABC e UgpAEBC, por meio de análises comparativas genômicas, filogenéticas e estruturais. Os resultados evidenciaram que os transportadores ABC Rv2038c-41c e UgpAEBC são exclusivos de espécies patogênicas do gênero *Mycobacterium* e podem estar relacionados à infecção e patogenicidade. A comparação das SBPs, TMDs e NBDs de *M. tuberculosis* com ortólogos putativos e análises filogenéticas mostraram que os componentes se dividem em quatro grupos de acordo com a função dos transportadores. Análises estruturais associadas às de filogenia, revelaram que os componentes de membrana são os mais diversificados e que as principais diferenças nesses TMDs se encontram na região de interface com a proteína periplasmática. Os NBDs são os mais conservados e as SBPs divergem principalmente na região do sítio de interação com os ligantes. Estudos de modelagem e docagem molecular sugerem que LpqY seja uma proteína ligadora de trealose e Rv2041c de açúcares cíclicos. Adicionalmente, avaliamos a possibilidade de estudar as interações entre TMDs e NBDs dos transportadores ABC utilizando-se peptídeos miméticos às hélices de acoplamento. Para tal, foram usados peptídeos das hélices de acoplamento dos componentes transmembrana MalF e MalG e o componente citoplasmático MalK, do sistema de transporte de maltose $MalEFGK_2$ de *Escherichia coli.* Ensaios biofísicos de fluorescência (Fluorescência Diferencial de Varredura, DSF; Termoforese em microescala, MST) e estudos de espalhamento de raios-X a baixo ângulo (SAXS) mostraram que MalK se liga aos peptídeos com um K_d entre 47.8 e 20.9 $\mu mol\ L^{-1}$, e que MalK sofre desestabilização térmica na presença de $MgCl_2$ e ATP, e do peptídeo MalF*ch*. As análises de SAXS mostraram que os dois peptídeos desencadeiam mudanças conformacionais na proteína MalK. A análise da fluorescência intrínseca do triptofano mostrou que a fluorescência dos dois únicos triptofanos em MalK não muda na presença de ambos os peptídeos. Este trabalho evidenciou o papel dos transportadores ABC de açúcares em *M. tuberculosis*, e a partir de uma abordagem multidisciplinar mostrou o papel destes componentes na família de transportadores ABC. O uso de peptídeos codificando as hélices de acoplamentos dos componentes transmembrana, se mostrou como uma alternativa interessante para estudos de interação proteína-proteína nestes transportadores, e essa abordagem pode servir como modelo para o estudo destas interações em outros sistemas de membrana e futuros estudos de inibição destes transportadores em bactérias patogênicas.

ABSTRACT

ABC (ATP-Binding Cassette) transporters form a large superfamily of protein complexes that mediate the transport of a wide range of substrates across biological membranes in eukaryotes and prokaryotes. Structurally, they consist of two hydrophobic transmembrane domains (TMDs) and two cytoplasmic nucleotide-binding domains (NBDs). In the case of importers, an additional periplasmic substrate-binding protein or domain (SBP) is responsible for the affinity and specificity of the transport. Sugar ABC importers are of great importance for the nutrition and survival of pathogenic bacteria and the identification and characterization of these systems is a step towards knowledge about the physiology of the microorganism. This work had as main objective the study of four ABC systems related to sugar transport in *M. tuberculosis*, LpqY/SugABC, Rv2038c-41c, UspABC and UgpAEBC, through comparative genomic, phylogenetic, and structural analyzes. The results showed that the ABC transporters Rv2038c-41c and UgpAEBC are exclusive to pathogenic species of *Mycobacterium* genus and may be related to infection and pathogenicity. The comparison of *M. tuberculosis* SBPs, TMDs and NBDs with putative orthologs and phylogenetic analyzes showed that the components are divided into four groups according to the function of the transporters. Structural analyzes associated with those of phylogeny revealed that the membrane components are the most diversified and the main differences in these TMDs are found in the interface region with the periplasmic protein. NBDs are the most conserved and SBPs differ mainly in the region of the site of interaction with the substrates. Molecular modeling and docking studies suggest that LpqY and Rv2041c are a trehalose-binding protein and a cyclic sugar-binding protein, respectively. Additionally, we evaluated the possibility of studying the interactions between TMDs and NBDs of ABC transporters, using peptides mimetizing to the coupling helices. For this purpose, peptides from the coupling helices of the transmembrane components MalF and MalG and the cytoplasmic component MalK from the $MalEFGK_2$ maltose ABC transport system of *Escherichia coli* were used. Biophysical fluorescence assays (Differential Scanning Fluorescence, DSF; Microscale thermophoresis, MST), and small angle X-ray scattering (SAXS) studies showed that MalK binds to peptides with a K_d between 47.8 and 20.9 μmol L^{-1}, and that MalK undergoes thermal destabilization in the presence of $MgCl_2$ and ATP, and of the MalFch peptide. SAXS analyzes showed that both peptides trigger conformational changes in the MalK protein. Analysis of the intrinsic fluorescence of tryptophan showed that the fluorescence of the only two tryptophans in MalK does not change in the presence of both peptides. This work highlighted the role of sugar ABC transporters in *M. tuberculosis*, and from a multidisciplinary approach showed the role of these components in the ABC transporter family. The use of peptides encoding the coupling helix of the transmembrane components, proved to be an interesting alternative for studies of protein-protein interaction in these transporters, and this approach can serve as a model for the study of these interactions in other membrane systems and future transporter inhibition studies in pathogenic bacteria.

LISTA DE ILUSTRAÇÕES

INTRODUÇÃO

Figura 1. O envelope celular do *Mycobacterium tuberculosis*.

Figura 2. O domínio de ligação ao nucleotídeo (NBD) de transportadores ABC.

Figura 3. Topologia dos domínios transmembrana (TMDs) em importadores ABC tipo I e II.

Figura 4. As hélices de acoplamento dos transportadores ABC.

Figura 5. Os tipos de proteínas de ligação ao substrato (SBPs).

Figura 6. Sítio de interação com maltose da proteína SBP MalE de *E. coli*.

Figura 7. Estados conformacionais do transportador ABC de maltose de *E. coli*.

CAPÍTULO I

Figure 1. Co-occurrence and genomic proximity of genes encoding for carbohydrate ABC transporters components from *Mycobacterium tuberculosis* across different species.

Figure 2. Phylogenetic relationship of NBD components of carbohydrate ABC transporters from species of *Mycobacterium* genus.

Figure 3. Phylogenetic relationship of TMD components of carbohydrate ABC transporters from species of *Mycobacterium* genus.

Figure 4. Phylogenetic relationship of SBP components of carbohydrate ABC transporters from species of *Mycobacterium* genus.

Figure 5. Comparison among the four carbohydrate-binding proteins from *M. tuberculosis.*

Figure 6. The substrate-binding pockets of *M. tuberculosis* LpqY and Rv2041c in comparison with possible orthologs.

Figure 7. The interface between coupling helices and NBDs of *M. tuberculosis* carbohydrate ABC transporters.

RESULTADOS COMPLEMENTARES – CAPÍTULO I

Figura 1. Arvore filogenética das proteínas UgpC de diferentes espécies de bactéria.

Figura 2. Características funcionais presentes na estrutura da proteína UgpC de *M. tuberculosis*.

Figura 3. Purificação das proteínas *full* UgpC e *short* UgpC.

Figura 4. Ensaios biofísicos da proteína *full* UgpC de *M. tuberculosis*.

Figura 5. Análises do potencial imunogênico da proteína *full* UgpC.

CAPÍTULO II

Figure 1. Identification of the coupling helices in *E. coli* MalF and MalG TMDs.

Figure 2. Interaction interface of MalK NBD with MalF and MalG coupling helices.

Figure 3. Purity and molecular weight assessment of synthetic coupling helices peptides MalFch and MalGch by RP-HPLC and mass spectrometry, respectively.

Figure 4. Overexpression and purification steps of *E. coli* MalK.

Figure 5. DSF assays of MalK NBD in presence of MalFch and MalGch peptides.

Figure 6. Determination of the affinity of MalFch and MalGch peptides for MalK NBD by microscale thermophoresis (MST).

Figure 7. Microscale Thermophoresis parameters of MalK in function of MalFch peptide concentration.

Figure 8. Microscale Thermophoresis parameters of MalK in function of MalGch peptide concentration.

Figure 9. SAXS analysis of *E. coli* MalK and upon addition of MalFch and MalGch peptides.

Figure 10. Absorption espectra of MalK as a function of MalFch and MalGch peptides concentration.

Figure 11. Intrinsic fluorescence measurements of *E. coli* MalK as a function of MalFch and MalGch peptides concentration.

Figure 12. Acrylamide quenching measurements of *E. coli* MalK.

LISTA DE TABELAS

CAPÍTULO I

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

Table 2. Presence of putative orthologs of the carbohydrate ABC transporters from *Mycobacterium tuberculosis* identified in mycobacterial species.

RESULTADOS COMPLEMENTARES – CAPÍTULO I

Tabela 1. Oligonucleotídeos utilizados para a clonagem do gene *ugpC* de *M. tuberculosis*.

Tabela 2. Testes de expressão e solubilidade da proteína UgpC de *M. tuberculosis*.

Tabela 3. Atividade enzimática obtida para *M. tuberculosis* UgpC em comparação com outros NBDs previamente caracterizados na literatura.

CAPÍTULO II

Table 1. MalF and MalG coupling helices peptides from *E. coli* maltose/maltodextrinas ABC transporter $MalEFGK_2$.

Table 2. Volume fraction of the three oligomeric states in samples of MalK protein in presence of MalFch and MalGch peptides, calculated with OLIGOMER.

LISTA DE ABREVIATURAS E SIGLAS

ABC	ATP-Binding Cassette
AICc	Akaike Information Criterion, corrected
ATP	Adenosina Trifosfato
BLASTp	Basic Local Alignment Search Tool protein
CUT1	Carbohydrate Uptake Transporter-1
CUT2	Carbohydrate Uptake Transporter-2
CD	Circular Dichroism
CV	Column Volume
DSF	Differential Scanning Fluorescence
DOPE	Discrete Optimized Protein Energy
E. coli	*Escherichia coli*
ELISA	Enzyme-Linked Immunosorbent Assay
HPLC	High Performance Liquid Chromatography
IF	Inward-Facing
K_d	Constante de dissociação
MalFch	MalF coupling helix
MalGch	MalG coupling helix
MDR-TB	Multi-Drug-Resistant Tuberculosis
MST	MicroScale Thermophoresis
MTBC	*Mycobacterium tuberculosis* Complex
M. tuberculosis	*Mycobacterium tuberculosis*
NBD	Nucleotide-Binding Domain
NBT/BCIP	nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate
NCBI	National Center for Biotechnology Information
OMS	Organização Mundial da Saúde
OF	Outward-Facing
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Pi	Fosfato inorgânico
PVDF	Polyvinylidene fluoride
RMSD	Root-Mean-Square Deviation
SAXS	Small-Angle X-ray Scattering
SBP	Substrate-Binding Protein
TM	Transmembrane
T_m	Temperatura de *melting*
TMD	Transmembrane domain
Tat	Twin-arginine translocation
XDR-TB	Extensively Drug-Resistant Tuberculosis

SUMÁRIO

INTRODUÇÃO 16

Mycobacterium tuberculosis e a Tuberculose 16

Papel dos açúcares em *M. tuberculosis* 18

Os transportadores ABC de açúcares de *M. tuberculosis* 19

Estrutura e função dos transportadores ABC 23

- Domínios de ligação ao nucleotídeo (*Nucleotide Binding Domains* – NBDs) 23
- Domínios transmembrana (*Transmembrane Domains* – TMDs) 26
- Hélices de acoplamento 27
- Proteínas de ligação ao substrato (*Substrate Binding Proteins* - SBPs) 28

O transportador ABC de maltose de *Escherichia coli* 30

OBJETIVOS 34

Objetivo geral 34

Objetivos específicos 34

RESULTADOS 35

CAPÍTULO I 37

RESULTADOS COMPLEMENTARES – CAPÍTULO I: A proteína UgpC de *M. tuberculosis* 71

CAPÍTULO II 85

DISCUSSÃO 110

CONCLUSÕES 113

REFERÊNCIAS 115

APÊNDICES 115

Apêndice A. Arquivos adicionais - CAPÍTULO I 118

ANEXOS 141

Anexo 1 - Manuscrito sobre a proteína PotF de *X. citri* 141

Anexo 2 - Artigo sobre peptídeos antimicrobianos sintéticos 181

Anexo 3 - Declaração de Bioética e/ou Biossegurança 194

Anexo 4 - Declaracao referente a direitos autorais 199

INTRODUÇÃO

Mycobacterium tuberculosis e a Tuberculose

A tuberculose é um flagelo antigo que atormentou a humanidade ao longo da história provocando grandes epidemias, para posteriormente, regredir e comportar-se como outras doenças infecciosas, porém, em uma escala de tempo que desafia as explicações aceitas para os ciclos epidêmicos (Daniel, 2006). A tuberculose é causada pelo bacilo *Mycobacterium tuberculosis*, uma das principais causas de morte no mundo e a principal causa de morte por infecção com um único agente infeccioso, acima do HIV/AIDS. Em 2018, estima-se que 10 milhões de pessoas adoeceram com tuberculose ao redor do mundo (variação: 9.0–11.1 milhões), um número relativamente estável nos últimos anos (OMS, 2019). A distribuição geográfica da tuberculose está intimamente relacionada aos índices socioeconômicos das diversas regiões. As taxas de incidência são baixas nas regiões desenvolvidas (Europa, América do Norte e Oceania) e muito altas naquelas regiões cuja população está sujeita a desnutrição e más condições de habitação (Ásia, África e América do Sul). O Brasil é considerado um dos trinta países "*high TB burden*" desde 1998, com uma incidência total entre 81.000 e 110.000 em 2018 (OMS, 2019).

Quase 90% dos casos anuais de tuberculose ocorrem nos 30 países com alta taxa de incidência. O surgimento de cepas de *M. tuberculosis* com múltipla (MDR-TB) e extensa (XDR-TB) resistência a fármacos agrava o problema. A OMS estima que só em 2016 ocorreram 600.000 novos casos com resistência à rifampicina (o fármaco de primeira linha no tratamento da doença), dos quais 490.000 tiveram MDR-TB (OMS, 2017). O fim da epidemia da tuberculose até 2030 é um dos objetivos de saúde dos Objetivos de Desenvolvimento Sustentável (OMS, 2018). Neste sentido, as respostas prioritárias, a detecção rápida da resistência aos medicamentos antituberculose, o uso de esquemas apropriados para o tratamento e o desenvolvimento de novos medicamentos são de suma importância.

M. tuberculosis é um bacilo identificado por Robert Koch em 1882, não formador de esporos, sem flagelos e que não produz toxinas. *M. tuberculosis* é uma espécie aeróbica estrita, necessitando de oxigênio para sua sobrevivência. Por ser capaz de sobreviver e de se multiplicar no interior de células fagocitárias, é considerado um parasita intracelular facultativo. O gênero *Mycobacterium* compreende mais de 170 espécies, a maioria das quais são organismos ambientais (Fedrizzi et al., 2017). Tradicionalmente, espécies de

Mycobacterium têm sido divididas pelo tipo de crescimento, rápido ou lento, com três principais patógenos de humanos, o complexo *Mycobacterium tuberculosis* (MTBC), *M. leprae* e *M. ulcerans*, pertencentes ao grupo de crescimento lento (Rogall et al., 1990). Diversas micobactérias não tuberculosas (NTMs, *non-tuberculous mycobacteria*), podem causar doença em indivíduos imunocomprometidos, entre elas *M. abscessus* (de crescimento rápido), e *M. avium*, *M. marinum*, *M. xenopi*, *M. gordonae* e *M. kansassi* (também de crescimento lento). A tuberculose humana é causada principalmente por membros do MTBC, estreitamente relacionados (>99% de identidade na sequência de nucleotídeos), com uma baixa variação na sequência comparada com outras bactérias (Achtman, 2008). Os membros do MTBC são patógenos adaptados a humanos sem nenhum reservatório ambiental ou animal conhecido, e causam a doença por meio da transmissão entre indivíduos (Brites & Gagneux, 2012), o que é diferente de muitos outros patógenos em que a virulência não está diretamente ligada à sua transmissão (Ebert & Bull, 2003; Gagneux, 2018).

O envoltório celular de *M. tuberculosis* é único (Figura 1), desde o ponto de vista da composição molecular como do arranjo de seus componentes: uma membrana citoplasmática composta principalmente de fosfolipídios aniônicos em um arranjo de bicamada com proteínas integrais; a membrana externa chamada micomembrana é uma bicamada lipídica assimétrica de ácidos graxos micólicos de cadeia longa na parte interna, e glicolipídeos intercalantes livres e componentes cerosos na parte externa; e por último, uma camada mais externa conhecida como cápsula, diferente entre espécies patogênicas e não patogênicas, e composta de polissacarídeos, proteínas, e pequenas quantidades de lipídeos (Bailo et al., 2015). Entre a membrana externa e a membrana interna é formado um espaço periplasmático, com a presença de uma fina camada de peptideoglicano covalentemente ligada a arabinogalactano e lipoarabinomanana, estes últimos ligados aos ácidos micólicos. O peptidoglicano, arabinogalactano e os ácidos micólicos formam o esqueleto da parede celular (Bailo et al., 2015).

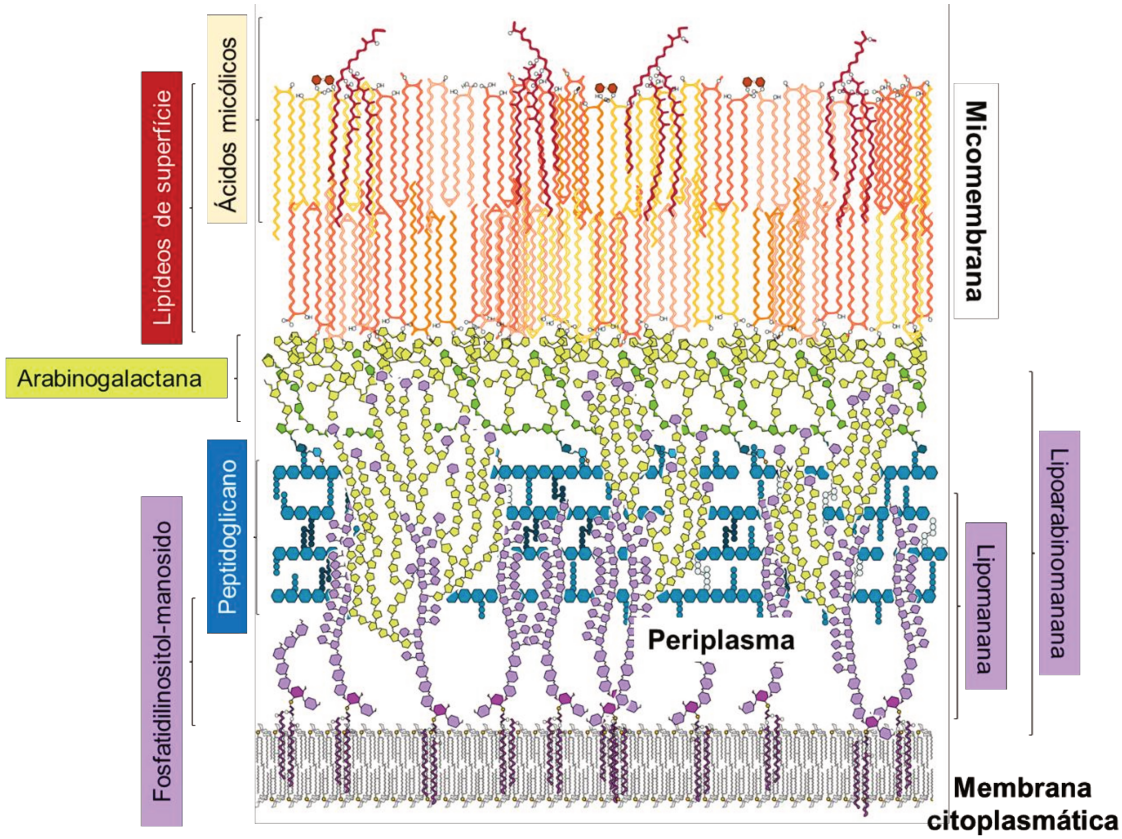


Figura 1. O envelope celular do *Mycobacterium tuberculosis*. O fosfatidilinositol-manosideo, lipídio de cadeia curta, é mostrado em roxo escuro (inositol) e roxo claro (manose). Lipomanana são derivados do fosfatidilinositol-manosideo e são mostrados em roxo claro. Lipoarabinomanana são semelhantes ao lipomanana, mas têm cadeias de arabinose (verde claro) ramificadas. No peptidoglicano, as ligações cruzadas são mostradas em azul escuro e claro. Arabinogalactana tem a galactana mostrada em verde escuro e o arabinano em verde claro. Trealose é representada por dois hexágonos laranja mais escuros e está ligada aos ácidos micólicos em monomicolato de trealose (laranja claro) e dimicolato de trealose (laranja escuro). Adaptado de Dulberger et al., 2020.

Papel dos açúcares em *M. tuberculosis*

Os carboidratos complexos são um dos principais blocos estruturais de todos os organismos vivos, e as suas funções variam desde proteção das paredes celulares bacterianas e vegetais a tecidos conjuntivos de animais. Os carboidratos também servem como fontes primárias de armazenamento de carbono e energia sintetizada por e para quase todas as células vivas. Em *M. tuberculosis*, os carboidratos possuem funções que vão desde componentes do envelope celular até virulência e resistência a fármacos. O envelope celular, por exemplo, é dominado por lipídios e carboidratos que fornecem uma barreira permeável contra drogas hidrofílicas e são cruciais para a sobrevivência e virulência da bactéria. A grande estrutura macromolecular, denominada complexo micolil-arabinogalactano-peptidoglicano, e os lipoglicanos à base de fosfatidil-mio-inositol, são as principais características da parede celular micobacteriana (Jankute et al., 2015). Um exemplo clássico

de como os açúcares podem ser essenciais na resistência de *M. tuberculosis* a fármacos, é o papel de arabinofuranosiltransferases codificadas pelo operon *embCAB*, envolvidas na polimerização de resíduos de arabinofuranosil a partir de DPA (decaprenilfosforil-β-D-arabinose), a componentes de arabinana do arabinogalactana e lipoarabinomanana das paredes celulares. Tanto evidências experimentais como epidemiológicas sugerem que mutações no operon *embCAB* são responsáveis pela resistência ao etambutol, um fármaco de primeira linha no tratamento da tuberculose (Safi et al., 2013).

No entanto, o papel dos açúcares não está restrito à proteção das paredes celulares e resistência a fármacos. Várias linhas de evidência sugerem fortemente que *M. tuberculosis* pode usar açúcares como fonte de carbono e fosfato (Jiang et al., 2014). *M. tuberculosis* muda seus requerimentos de carboidratos para lipídios após o início da resposta imune adaptativa. Os estudos bioquímicos sugerem que, em tecidos pulmonares cronicamente infectados, os ácidos graxos podem ser a principal fonte de carbono e energia para *M. tuberculosis* (Wheeler et al., 1990), no entanto, durante os primeiros 10 dias de infecção em camundongos, *M. tuberculosis* requer o transportador de açúcar LpqY/SugABC para a sobrevivência (Sassetti & Rubin, 2003). Posteriormente, enzimas como a isocitrato liase, malato sintases e fosfolipases são essenciais para a virulência (Le Chevalier et al., 2015; McKinney et al., 2000). O bacilo prolifera no fagossomo de células fagocíticas, nesse ambiente tem acesso restrito a nutrientes. Apesar da hipótese dominante dos lipídios do hospedeiro serem a principal fonte de carbono e energia de *M. tuberculosis,* a disponibilidade limitada de carboidratos levanta a importância de ao menos 5 transportadores putativos de carboidratos codificados pela bactéria, e como os açúcares internalizados por estes são essenciais para o modo de vida de *M. tuberculosis*.

Os transportadores ABC de açúcares de *M. tuberculosis*

Os genes que codificam transportadores ABC (*ATP-binding cassette*) ocupam em torno de 2.5% do genoma de *M. tuberculosis*. Em sua forma mais bem conhecida os sistemas de transporte do tipo ABC acoplam a hidrólise da adenosina trifosfato (ATP) à translocação de diversos substratos através da membrana celular (Dahl et al., 2004), e consistem de dois domínios transmembrana (TMDs) que formam o poro de passagem na membrana citoplasmática, e dois domínios de ligação ao nucleotídeo (NBDs) que geram a energia para o transporte. No caso de importadores ABC, uma proteína periplasmática de ligação (SBPs) captura o substrato no meio extracelular para ser entregue aos TMDs. Ao menos 37 transportadores, entre completos e incompletos, têm sido identificados em *M. tuberculosis*

com base em similaridades estruturais das proteínas componentes com proteínas ortólogas presentes em outros microrganismos. Os transportadores ABC de *M. tuberculosis* têm sido agrupados em diferentes subfamílias, das quais as subfamílias 2, 4, 5, 8 e 11 podem estar envolvidas na importação de peptídeos, amino ácidos, açúcares, ferro e ânions, respectivamente; por outro lado as subfamílias 3, 6, 7 e 12 podem estar envolvidas na exportação de macrolídeos, múltiplas drogas, antibióticos e substratos ainda desconhecidos, respectivamente. É interessante o fato de que quando comparados com o número de sistemas importadores de *E. coli* e *Bacillus subtilis*, o genoma de *M. tuberculosis* mostra uma redução significativa. Isto é particularmente evidente nos transportadores envolvidos na captação de carboidratos (Braibant et al., 2000). Neste sentido, o conhecimento da estrutura e função destes transportadores pode ser de grande importância no entendimento de aspectos chaves como a nutrição e sobrevivência da bactéria até virulência e resistência a fármacos. Finalmente, a caracterização destes transportadores pode ser de grande importância no desenvolvimento de novas estratégias para o controle da doença.

Em 2000, Milton Saier descreveu as proteínas transmembrana envolvidas no transporte de açúcares, entre elas a superfamília ABC (Saier, 2000). Das 18 famílias de transportadores envolvidas na captação de açúcares, só duas delas são específicas para carboidratos simples: a família de transportadores ABC para captação de carboidratos 1 (*carbohydrate uptake transporter*-1, CUT1), e a família de transportadores ABC para captação de carboidratos 2 (CUT2). A família CUT1 transporta uma variedade de di-, tri- e oligossacarídeos (entre eles, maltose, trealose, lactose etc.), assim como glicerol-fosfato e polióis. O transportador de maltose/maltodextrina presente em *E. coli* e *Salmonella typhimurium* é um dos representantes da família bastante estudado do ponto de vista genético, bioquímico e estrutural, e representa um paradigma não apenas na família CUT1, mas na superfamília de transportadores ABC. A família CUT2 transporta só monossacarídeos, representados pelo transportador Mgl de *S. typhimurium*, e o transportador de ribose de *E. coli* (Schneider, 2001).

Uma análise bioinformática do genoma de *M. tuberculosis* H37Rv revelou quatro transportadores do tipo ABC e um transportador da classe MFS (*Major Facilitator Superfamily*) envolvidos na captação de carboidratos (Titgemeyer et al., 2007), os quais já tinham sido descritos em uma análise global do genoma de *M. tuberculosis* (Braibant et al., 2000). Análises *in silico* dos componentes dos transportadores demostraram uma maior

homologia com o transportador ABC de maltose de *E. coli* (Braibant et al., 2000; Titgemeyer et al., 2007).

O sistema de transporte LpqY/SugABC foi previsto ser essencial para a virulência de *M. tuberculosis* em camundongos com base em experimentos de hibridização do sítio do transposon (TraSH) (Sassetti & Rubin, 2003). Estudos iniciais sugeriram que este sistema podia transportar maltose ou maltodextrinas (Borich et al., 2000; Braibant et al., 2000), no entanto, as semelhanças da proteína de ligação ao substrato LpqY correspondente aos transportadores de maltose e proteínas periplasmáticas de ligação a maltose MalE of *E. coli* e *Streptomyces coelicolor* são muito baixas (identidade de 25%). Assim, é questionável se a maltose é o substrato de LpqY/SugABC. Essa conclusão é sustentada pelo fato de que nem *M. smegmatis*, que possui um sistema LpqY/SugABC altamente semelhante, nem *M. tuberculosis* crescem com maltose como uma única fonte de carbono (Edson, 1951). Por outro lado, é sugerido que o transportador LpqY/SugABC pode ter um papel importante na captação e reciclagem do dissacarídeo trealose, um açúcar ausente em mamíferos e, portanto, indisponível como fonte de nutriente para *M. tuberculosis*. Neste sentido, é sugerido que a proteína de ligação ao substrato LpqY, pode ser um alvo para compostos que exibem atividade inibitória na superfície da parede celular (Kalscheuer et al., 2010).

O transportador Rv2038c-41c é o menos conhecido até a data, e consiste em uma proteína citoplasmática Rv2038c, duas proteínas transmembrana Rv2039c e Rv2040c, e uma proteína periplasmática de ligação ao substrato Rv2041c. O substrato da proteína Rv2041c ainda não é conhecido, no entanto, esta proteína é regulada positivamente em condições acídicas e de hipóxia, similares as que *M. tuberculosis* pode experimentar *in vivo* (especialmente dentro do fagossomo), sugerindo um papel na adaptação intracelular dentro do hospedeiro (Su Young Kim et al., 2008, 2009). Por outro lado, o tratamento de macrófagos com Rv2041c resultou em expressão elevada de citocinas pró-inflamatórias (TNF-α, IL-6 e IL-12p40) (Su Young Kim et al., 2008, 2009) e o tratamento de linfócitos de camundongos com tuberculose latente ou ativa com esta proteína resultou em um aumento da secreção de IFN-γ e TNF-α. Adicionalmente, foi descrito que Rv2041c resulta em uma resposta de anticorpos positiva tanto em camundongos infectados com *M. tuberculosis,* como em pacientes com tuberculose ativa (S. Y. Kim et al., 2009). Também foi estabelecido que esta proteína fornece uma alta sensibilidade no diagnóstico sorológico de infeção ativa por *M. tuberculosis*, semelhante à de antígenos sorológicos usados tradicionalmente (CFP-10, ESAT-6, HSP-X, complexo Ag85 e PstS1).

O transportador UspABC consiste em duas proteínas transmembrana UspA e UspB, e uma proteína periplasmática de ligação ao substrato UspC. O transportador carece de uma proteína citoplasmática de ligação ao ATP, o que levanta o questionamento de se outras proteínas de ligação ao ATP poderiam fornecer a função no transportador (Braibant et al., 2000). Recentemente, foi demonstrado que UspC prefere açúcares que contêm um grupo amino na posição C2 ou C3, o que pode sugerir um papel deste sistema na reciclagem de componentes chaves do peptideoglicano da parede celular micobacteriana (Fullam et al., 2016).

O sistema de transporte UgpABCE de *M. tuberculosis* é formado por uma proteína citoplasmática UgpC, duas proteínas transmembrana de UgpA e UgpE, e uma proteína periplasmatica UgpB. O transportador tem semelhança com o transportador de maltose de *E. coli*. Como consequência, é agrupado na família CUT1, e inicialmente foi predito como um transportador de sn-glicerol-3-fosfato (Schneider, 2001; Wuttge et al., 2012). Ao contrário dessas hipóteses, nenhuma ligação detectável foi observada para a proteína UgpB em presença de sn-glicerol-3-fosfato ou maltose (Jiang et al., 2014). A comparação da estrutura cristalográfica da UgpB de *M. tuberculosis* com o ortólogo de *E. coli* mostrou que o resíduo Trp169, essencial para a ligação a sn-glicerol-3-fosfato em *E. coli*, é substituído por Leu205 em *M. tuberculosis*. Além disso, os pesquisadores sugeriram que Leu205 foi um fator determinante para a ligação a glicerofosfocolina (Fenn et al., 2019; Jiang et al., 2014). Este estudo especifica o papel da proteína UgpB na importação de glicerofosfocolina, que pode atuar como fonte carbono e fosfato. Adicionalmente, UgpB também é o substrato da via de translocação de dupla arginina (Tat), essencial para virulência e sobrevivência de vários patógenos (McDonough et al., 2008; Sutcliffe & Harrington, 2004). No entanto, mais estudos são necessários para estabelecer-se seu papel na virulência e sobrevivência do *M. tuberculosis*.

Neste sentido, estudos focados nos sistemas envolvidos na captação de açúcares em *M. tuberculosis* podem revelar informações importantes sobre, por exemplo, as fontes de carbono disponíveis no fagossomo dos macrófagos durante a infecção, as bases estruturais e funcionais dos transportadores e ferramentas para o desenvolvimento de novos fármacos contra a tuberculose.

Estrutura e função dos transportadores ABC

Os transportadores ABC constituem uma das maiores e mais antigas superfamílias de complexos de proteínas encontradas em todos os organismos vivos. Estes sistemas são máquinas moleculares que acoplam a ligação a ATP-Mg^{2+}, hidrólise de ATP e liberação de ADP/fosfato ao transporte de substratos (íons, açúcares, amino ácidos, vitaminas, peptídeos, polissacarídeos, hormônios, lipídeos e xenobióticos) através de membranas biológicas (Jones & George, 2002; Ter Beek et al., 2014; Thomas & Tampé, 2020). Com base na direção da translocação do substrato, os transportadores ABC são comumente classificados como importadores e exportadores (Davidson et al., 2008). Os importadores, muito comuns em procariotos e ausentes em eucariotos, permitem o movimento de substratos do meio extracelular para o citoplasma, e estão envolvidos principalmente na captação de nutrientes hidrofílicos como peptídeos, íons e açúcares. Ao contrário, os exportadores permitem o movimento do substrato do citoplasma para o meio extracelular, e estão presentes tanto em procariotos como eucariotos (Moussatova et al., 2008; Ter Beek et al., 2014). Todos os transportadores ABC compartilham uma organização modular: 1) Dois domínios de ligação ao nucleotídeo (*nucleotide-binding domain*, NBD) ou domínios ABC, altamente conservados e característica unificadora dos sistemas ABC. Os NBDs são os motores que transformam a energia potencial química do ATP em mudanças conformacionais das proteínas do transportador. 2) Dois domínios transmembrana (*transmembrane domain*, TMD), responsáveis pela formação do canal de passagem e bastante variáveis. Em procariotos, os importadores geralmente contêm além uma proteína periplasmática (*substrate-binding protein*, SBP) ou domínio de ligação ao substrato (*substrate-binding domain*, SBD), que transfere o substrato ao transportador, e garante a especificidade e afinidade do transporte (Davidson & Chen, 2004; Jones & George, 2002; Ter Beek et al., 2014).

- **Domínios de ligação ao nucleotídeo (*Nucleotide Binding Domains* – NBDs)**

Todos os NBDs contém dois subdomínios: um subdomínio maior similar ao núcleo encontrado em muitas ATPases RecA-*like* suplementado com uma folha β antiparalela específica de transportadores ABC (ABCβ), e um subdomínio menor helicoidal (ABCα), também específico de transportadores ABC (Figura 2AB). O subdomínio RecA-*like* consiste em duas folhas β e seis hélices α, e inclui os motivos Walker A (GxxGxGKS/T, onde x é qualquer aminoácido), Walker B (ϕϕϕϕD, onde ϕ é um resíduo hidrofóbico), *loop* D e H-

switch. O subdomínio helicoidal consiste em três ou quatro hélices, e inclui o motivo assinatura, também conhecido como LSGGQ ou *loop* C. Os dois subdomínios estão unidos por dois *loops* flexíveis, um dos quais contém um resíduo de glutamina, altamente conservado, conhecido como *loop* Q (Figura2C). Nas estruturas de transportadores ABC completos, foi mostrado que o *loop* Q está envolvido na interação entre as subunidades NBDs e TMDs (Davidson & Chen, 2004; Thomas & Tampé, 2020).

Como todas as ATPases RecA-*like*, a hidrólise requer oligomerização. Todos os transportadores ABC contém dois NBDs, e a ligação ao ATP é requerida para obter um estado dimérico na maioria dos NBDs estudados. Duas moléculas de ATP são ligadas na interface do dímero, interagindo com resíduos do motivo Walker A de uma subunidade e do motivo LSGGQ da outra. O anel de adenosina do ATP é estabilizado por uma interação *ring-stacking* com um resíduo aromático do *loop* A. O resíduo de lisina conservado no motivo Walker A forma ligações de hidrogênio com os átomos de oxigênio dos fosfatos α e γ, mantendo os dois fosfatos em uma orientação definida. Um íon de magnésio Mg^{2+} é coordenado por átomos de oxigênio dos fosfatos β e γ, e resíduos do motivo Walker A. O motivo Walker B fornece um aspartato que auxilia na coordenação do íon Mg^{2+}, e o glutamato seguinte polariza o ataque a molécula de água durante a hidrólise do ATP. Um resíduo de histidina altamente conservado localizado no H-*switch* forma uma ligação de hidrogênio com o fosfato γ e é requerido para a hidrólise. As cadeias laterais da serina e o esqueleto de grupos amida dos resíduos de glicina no motivo LSGGQ coordenam o fosfato γ. Além da ligação a nucleotídeos, o resíduo de histidina conservado também interage com resíduos através da interface do dímero no motivo Walker A e o *loop* D, uma sequência conservada seguinte ao motivo Walker B, sugerindo um acoplamento estreito entre a ligação ao ATP e a formação do dímero (Davidson et al., 2008).

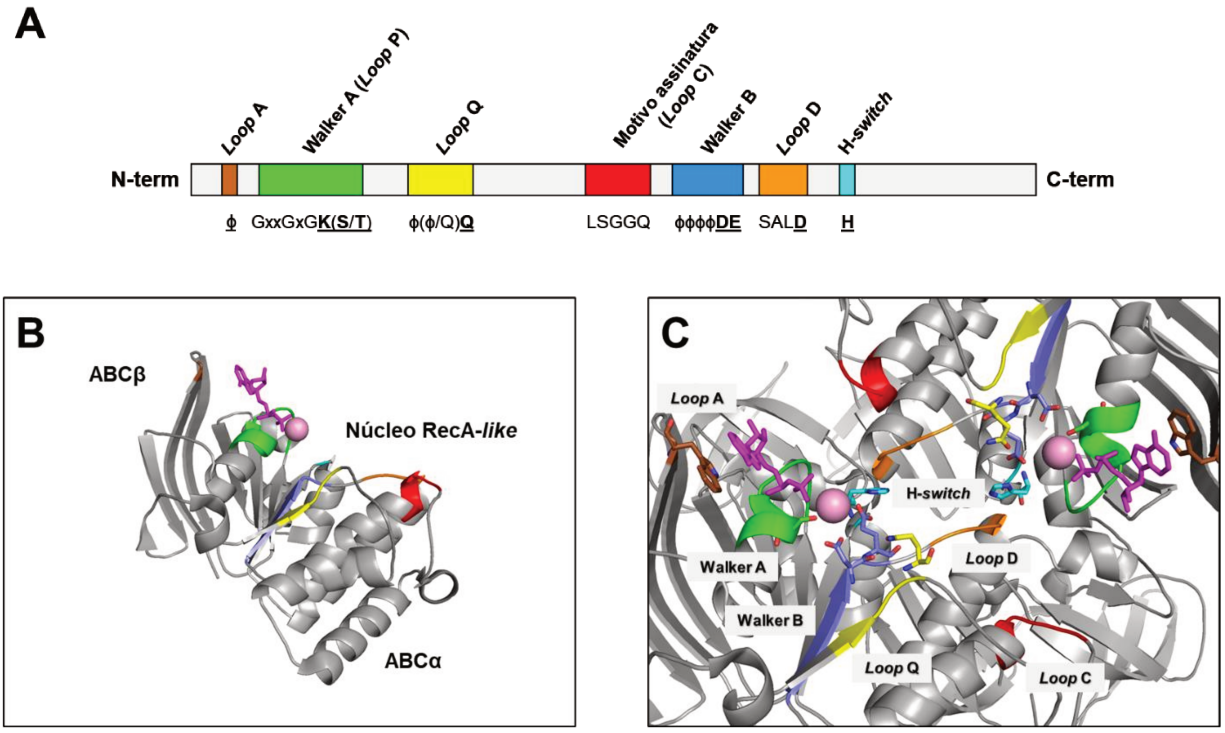


Figura 2. O domínio de ligação ao nucleotídeo (NBD) de transportadores ABC. A O arranjo linear dos motivos nos NBDs. **B** Vista de um NBD isolado, mostrando os três subdomínios que o formam. O Mg^{2+} é mostrado em esfera rosa e o ATP é mostrado em palito rosa. **C** Vista de dois NBDs dimerizados a partir da membrana. Os motivos são representados por cores no painel A. Adaptado de Thomas & Tampé, 2020.

Uma das funções essenciais dos transportadores ABC é aproveitar a ligação/hidrólise do ATP para o trabalho mecânico. Esta função parece ser alcançada através de mudanças conformacionais do transportador. A dimerização dos NBDs induzida pelo ATP ocorre na forma *yin-yang* (Thomas & Tampé, 2020). Como consequência, duas moléculas de ATP são sequestradas entre os NBDs dimerizados. A natureza dinâmica dos ABCs foi revelada pela comparação de estruturas cristalográficas de NBDs isolados em formas apo, ligadas a ATP e ADP, e é apoiada por análises bioquímicas. Na ausência de nucleotídeos e TMDs, os NBDs mostram alto grau de flexibilidade intrínseca. Em comparação com a forma apo, as estruturas ligadas a ATP na forma monomérica mostram que o subdomínio helicoidal gira em direção ao subdomínio RecA-*like* logo da ligação ao ATP, movendo o *loop* LSGGQ para a posição onde ele interage com o nucleotídeo através da interface do dímero (Davidson & Chen, 2004). Em muitos transportadores ABC, as mudanças no estado de dimerização dos NBDs fazem os TMDs alternar entre conformações *inward-* (IF) e *outward-facing* (OF), fenômeno conhecido como "acesso alternado". Essa alternância entre as conformações IF e OF está associada ao acesso ao poro de translocação do substrato através dos dois lados da membrana. A conformação OF permite que os importadores liguem o substrato para ser internalizado, e a conformação IF permite que os exportadores liberem o substrato para fora da célula (Thomas

& Tampé, 2020). Elementos importantes da comunicação entre NBDs e TMDs são as hélices de acoplamento, hélices curtas nos TMDs que interagem com os NBDs. As hélices de acoplamento são a única parte na interface TMD-NBD que é estruturalmente conservada entre os diversos TMDs (Locher, 2016), e são acopladas em uma cavidade na superfície dos NBDs, a qual localiza-se na região sensível ao fosfato γ na interface entre o núcleo de ligação ao ATP RecA-*like* e o subdomínio ABCα (Thomas & Tampé, 2020).

- **Domínios transmembrana (*Transmembrane Domains* – TMDs)**

Em todos os tipos de transportadores ABC, os TMDs constituem uma via de translocação, que é acessada alternadamente do lado *cis* e *trans* da membrana para o transporte do substrato. Os dois TMDs de importadores do tipo I são idênticos (homodímeros) ou estruturalmente similares (por exemplo, as duas TMDs do transportador de maltose $MalEFGK_2$ compartilham apenas 13% de identidade de sequência, mas estão estruturalmente relacionadas), com uma topologia de membrana central de cinco hélices TM por TMD (Figura 3A). Em muitos casos, uma hélice N-terminal adicional pode estar presente, totalizando 12 hélices TM. No entanto, algumas TMDs contêm até oito hélices TM. O caminho de translocação está localizado na interface entre as duas TMDs (Ter Beek et al., 2014).

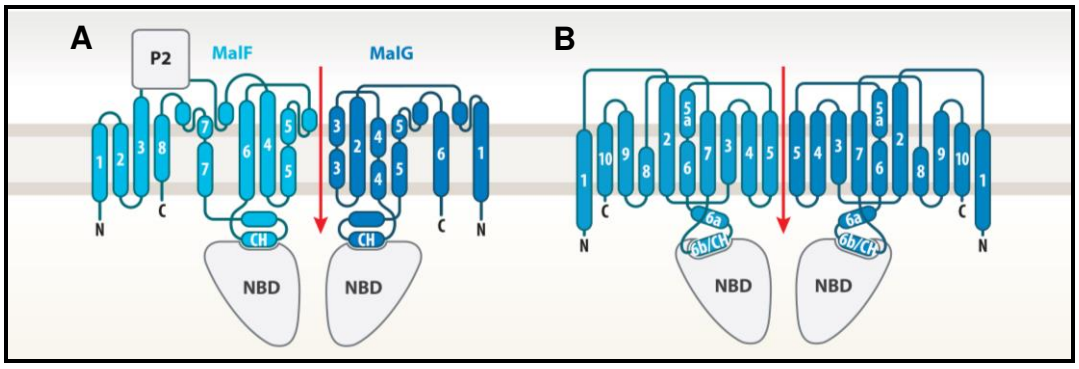


Figura 3. Topologia dos domínios transmembrana (TMDs) em importadores ABC tipo I e II. A Arranjo das hélices TM no importador de maltose de $MalFGK_2$ de *E. coli* (tipo I). **B** Arranjo das hélices TM no importador de vitamina B_{12} $BtuC_2D_2$ de *E. coli* (tipo II). Adaptado de Thomas & Tampé, 2020.

Os importadores de ABC do tipo II têm dois TMDs idênticos, cada um composto por 10 hélices TM (Figura 3B). No enovelamento do Tipo II, os TMDs estão alinhados um ao lado do outro. Em cada TMD, há uma simetria pseudo-dupla entre os segmentos que contêm as hélices TM 2–5 e as hélices TM 7-10. Esses dois subdomínios têm um empacotamento

helicoidal semelhante, mas com orientação oposta vista para a membrana. As hélices de um único TMD estão bem juntas, e os dois TMDs alinham o poro de translocação na interface (Ter Beek et al., 2014).

- **Hélices de acoplamento**

Uma questão crucial sobre o mecanismo do transporte em transportadores ABC é como o acesso alternado nos TMDs é acoplado às mudanças conformacionais nos NBDs quando ocorre a ligação e hidrólise do ATP e a liberação do Pi e ADP. Identificadas nos TMDs dos exportadores ABC e importadores de Tipo I e II (Dawson et al., 2007), as chamadas hélices de acoplamento são responsáveis pela interação dos TMDs aos NBDs (Figura 4A), e são uma extensão no citoplasma de hélice TM que interage diretamente em uma cavidade do NBD. Dessa maneira, as alterações conformacionais nos NBDs ou TMDs são sentidas de ambos os lados e se traduzem em diferentes conformações do transportador. Algumas hélices de acoplamento contêm uma sequência conservada, motivo EAA (Mourez et al., 1997), mas na maioria dos casos, não existe similaridade de sequência nas hélices de acoplamento. As hélices de acoplamento são encontradas entre as hélices TM 3 e 4 nos TMDs dos importadores do Tipo I. Nos importadores do Tipo II, elas estão localizadas entre as hélices TM 6 e 7, e nos exportadores ABC, a região da hélice de acoplamento é encontrada no circuito intracelular (ICL) entre as hélices TM 4 e 5. Nos exportadores em que os TMDs são fundidos aos NBDs, a hélice de acoplamento de um TMD interage com o NBD que está vinculado à outra subunidade. Embora os arranjos sejam diferentes nos diferentes tipos de transportadores, todas as hélices de acoplamento interagem de maneira semelhante com os NBDs. A região dos NBDs que interage com a hélice de acoplamento contém o *loop* Q. O sulco para as hélices de acoplamento nos NBDs está localizado exatamente na interface entre o subdomínio α helicoidal e o subdomínio do RecA-*like* (Figura 4B), que giram um em direção ao outro em resposta à ligação de ATP para hidrólise de ATP (Ter Beek et al., 2014).

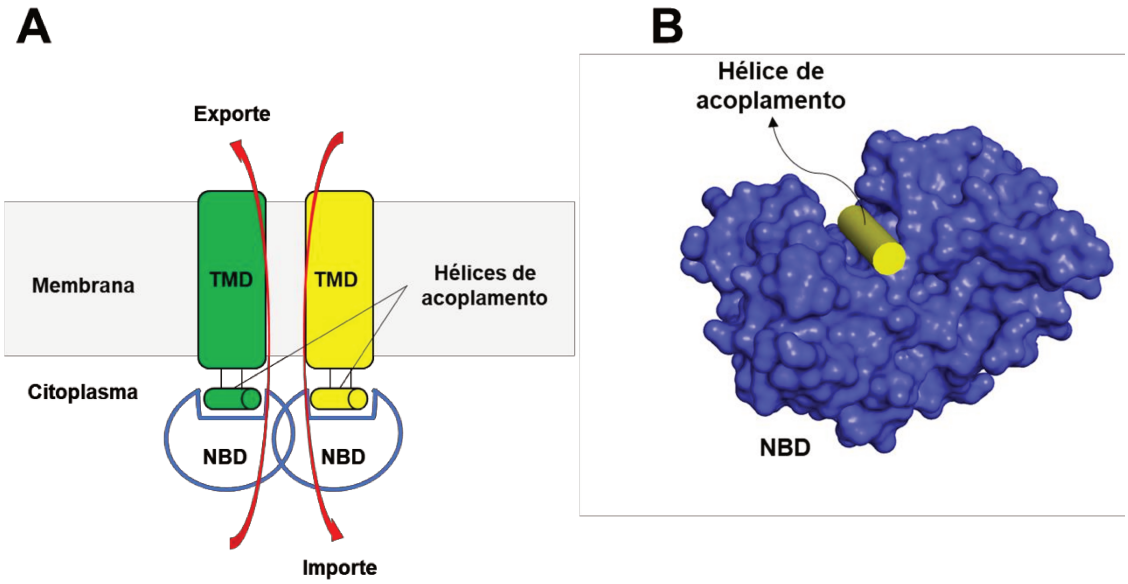


Figura 4. As hélices de acoplamento dos transportadores ABC. A Arranjo dos domínios nos transportadores ABC. As hélices de acoplamento transmitem mudanças conformacionais entre os NBDs e TMDs. **B** Esquema de um NBD isolado. Um sulco na superfície do NBD forma uma interface de contacto com a hélice de acoplamento do TMD. Embora a hélice de acoplamento não seja o único contato entre NBDs e TMDs, é o único contato conservado topologicamente entre os distintos enovelamentos dos TMDs caracterizados. Adaptado de Locher, 2016.

- **Proteínas de ligação ao substrato (*Substrate Binding Proteins* - SBPs)**

As proteínas ou domínios de ligação ao substrato (SBPs ou SBDs) são associadas a uma ampla variedade de complexos de proteínas (Berntsson et al., 2010). Em transportadores ABC, estas são responsáveis pela captura e entrega do substrato aos domínios transmembrana, e são ligadas através de um polipeptídio ao TMD do transportador, conectadas à membrana por meio de uma âncora lipídica ou hélice TM, ou difusíveis livremente no periplasma (o último apenas em bactérias Gram-negativas) (Biemans-Oldehinkel et al., 2006). As SBPs têm um enovelamento estrutural altamente conservado, apesar de não ter grande similaridade na sequência de aminoácidos nem nas funções. Estas proteínas consistem em dois domínios α/β globulares com uma folha β central (de 4 a 6 fitas β), flanqueado por hélices α. Estes são chamados lóbulos N e C, uma vez que contêm os extremos N-terminal e C-terminal. Os dois domínios estão conectados por uma região dobradiça, que compreende uma a três cadeias interconectadas. Na conformação livre de substratos, os domínios são bem separados, com um bolsão aberto, acessível por solvente. O substrato é ligado no bolsão entre os dois domínios (Holland et al., 2003), e a ligação causa que a SBP se feche ao redor do substrato, similar a planta apanha-moscas (*Venus Fly-trap*). Esta alteração conformacional induzida pelo substrato é central no mecanismo de translocação. As SBPs foram classificadas baseadas na

similaridade de sequência e os arranjos das folhas β (Fukami-Kobayashi et al., 1999). Mais tarde, foram reclassificadas com base nos alinhamentos das estruturas tridimensionais disponíveis das proteínas, o que resultou em 6 classes, que são descritas como clusters de A a G (Berntsson et al., 2010; Scheepers et al., 2016) (Figura 5). A diversidade de substratos ligados pelas SBPs é grande; proteínas do cluster C, por exemplo, ligam substratos entre peptídeos, açúcares e íons metais, e proteínas do cluster D ligam açúcares, íons inorgânicos, vitaminas, metais e poliaminas. É notável que todos os importadores de tipo I para os quais as estruturas cristalográficas foram determinadas fazem uso exclusivo de SBPs dos clusters B ou D, enquanto todos os importadores de Tipo II usam SBPs do grupo A. É possível que o uso de SBPs de diferentes clusters se correlacione com o uso de distintos *folds* de TMDs (Ter Beek et al., 2014).

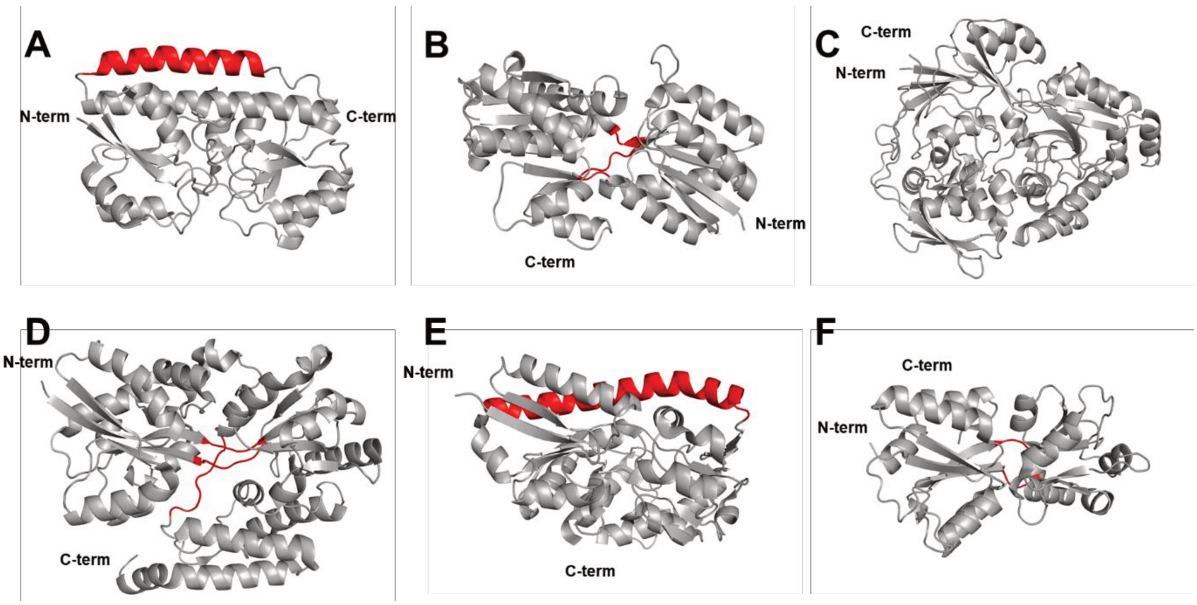


Figura 5. Os tipos de proteínas de ligação ao substrato. A Cluster A contém proteínas que tem uma única conexão entre os dois domínios na forma de uma hélice rígida. **B** Cluster B contém SBPS com três segmentos conectando os dois domínios. **C** Cluster C contém SBPs que tem um domínio extra e são significativamente maiores quando comparadas com os demais clusters. **D** Cluster D contém SBPs com dois *loops* relativamente curtos. **E** Cluster E contém SBPs associadas com transportadores TRAP, e todas contém uma longa hélice que funciona como a região de dobradiça. **F** Cluster F contém SBPs com dois *loops*, similares ao cluster D, porém com quase o dobro de comprimento gerando maior flexibilidade para a SBP. Adaptado de Berntsson et al., 2010.

As SBPs ligam seus substratos com alta afinidade, na faixa de 0.01 a 1 μM (Davidson et al., 2008). Essa ligação de alta afinidade é claramente responsável pela eficiência dos transportadores ABC em baixas concentrações de substrato; as células podem concentrar nutrientes em até 106 vezes quando os nutrientes estão presentes em concentrações sub-

micromolares no meio externo (Dippel & Boos, 2005). No entanto, as SBPs ainda são essenciais para o transporte, mesmo em altas concentrações de substrato, como demonstrado pela exclusão do gene que codifica a SBP de maltose em *E. coli*. As SBPs localizadas no espaço periplasmático das bactérias gram-negativas podem ser liberadas por um choque osmótico frio, procedimento que leva a inativação do transporte devido à perda das proteínas. Em qualquer circunstância, o transporte pode ser restaurado se as SBPs forem introduzidas de volta no espaço periplasmático (Brass et al., 1981).

O transportador ABC de maltose de *Escherichia coli*

Escherichia coli é uma bactéria Gram-negativa, anaeróbia facultativa e um dos mais importantes microrganismos para a engenharia metabólica, devido a seu rápido crescimento e as bem conhecidas ferramentas genéticas disponíveis (Pontrelli et al., 2018). Dado o amplo interesse em aplicações biotecnológicas, as vias de captação de açúcares como glicose, maltose, e lactose têm sido bastante estudadas (Luo et al., 2014). Adicionalmente, *E. coli* contém transportadores ABC envolvidos na captação de arabinose, galactose/glicose, alose, xilose, maltose e substratos ainda não identificados, estes últimos menos estudados. O transportador ABC de maltose é o mais caracterizado até a data (Jeckelmann & Erni, 2020). Apesar de informações bioquímicas, genéticas, celulares e estruturais serem disponíveis, existem poucas informações sobre como interagem os componentes do transportador, usando outras abordagens além da cristalografia de raios-X.

O transportador de maltose de *E. coli* tem sido um protótipo clássico para o estudo dos mecanismos moleculares de transportadores ABC. Ao longo dos anos, a genética deste transportador tem fornecido uma grande quantidade de informação sobre mutantes. O transportador funcional tem sido constituído em proteolipossomos e nanodiscos, e ensaios de atividade bem estabelecidos, utilizando-se principalmente técnicas de fluorescência e métodos de espectroscopia de ressonância paramagnética no intuito de avaliar mudanças conformacionais. A estrutura cristalográfica do transportador de maltose foi determinada em três estados funcionais, permitindo uma análise detalhada de eventos moleculares/estruturais, de grande auxílio no desenho de provas de conceito para estudos de interação entre os componentes (Chen, 2013). O transportador de maltose ($MalFGK_2$) é composto por dois TMDs, MalF e MalG, e duas ATPases citoplasmáticas, MalK, formando um homodímero. A proteína periplasmática de ligação à maltose (MalE) entrega maltose e outras maltodextrinas

ao transportador, e desencadeia a atividade da ATPase. O importe do substrato ocorre de maneira unidirecional devido à mudanças conformacionais promovidas pela hidrólise de ATP no homodímero $MalK_2$, que permitem o acesso alternado do sítio de ligação ao substrato em MalF para cada lado da membrana (Mächtel et al., 2019).

A maltose e as maltodextrinas entram no periplasma de *E. coli* através da proteína de membrana externa (*Outer Membrane Protein*, OMP) LamB, uma proteína trimérica, e cada monômero é uma folha β composta de 18 fitas em um motivo toroidal (barril β). Maltose e maltodextrinas de até sete monômeros de glicose são transportadas especificamente por LamB e não por porinas genéricas. O papel de LamB é crucial na difusão de maltodextrinas, especialmente as que tem um limite de exclusão maior, garantindo que a taxa de transporte não seja limitada por sua difusão passiva através da membrana externa. MalE, a SBP, é o receptor primário para maltose/maltodextrinas e se difunde livremente no periplasma, 20-40 vezes em excesso com relação ao transportador $MalFGK_2$. Maltose é ligada na MalE com uma constante de dissociação aparente de K_d de 3.5 μM. Uma comparação da MalE e outras SBPs que ligam monossacarídeos, dissacarídeos ou trissacarídeos revelou a presença de quatro sub-sítios que ligam unidades individuais do anel de glicose (Figura 6).

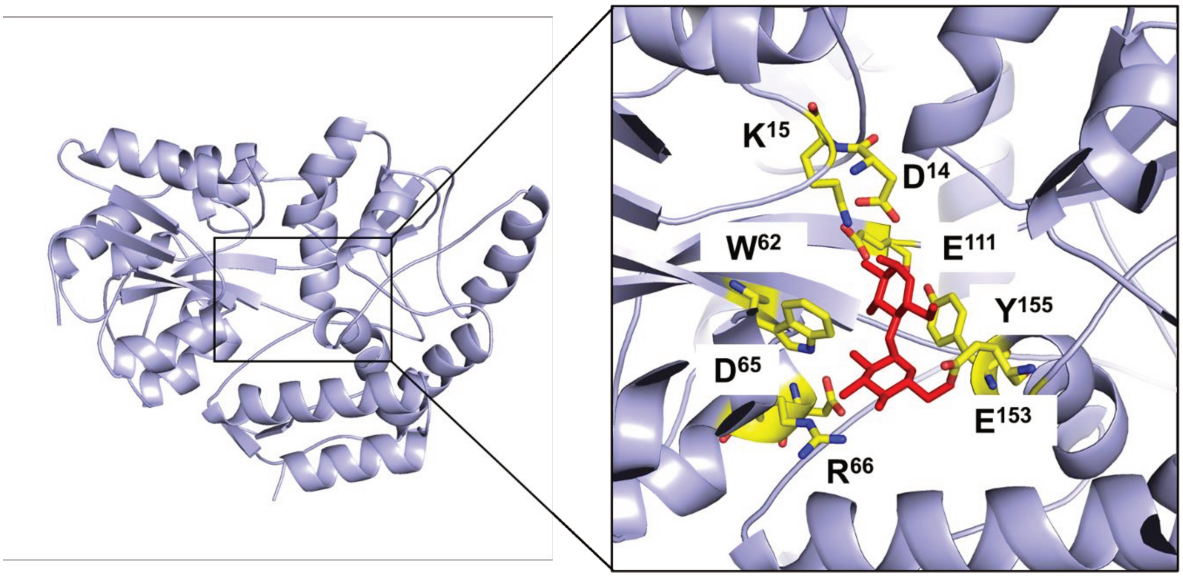


Figura 6. Estrutura tridimensional da proteína MalE de *E. coli* e detalhe do bolsão de interação com o respectivo açúcar. O sítio de interação com maltose mostra os diferentes subsítios. Subsítio B (D14, K15, E111). Subsítio B e C (E153, Y155). Os subsítios A e D não estão envolvidos na interação com maltose, e sim com outras maltodextrinas. (Código PDB: 1ANF).

O sub-sítio A forma ligações de hidrogênio com o primeiro anel de glicose em maltose e maltodextrinas. É ocluído em MalE pelos resíduos Asp14 e Lys15, que contribuem, com Glu111, para o subsítio B. Glu153 e Tyr155 formam contatos com o segundo anel de glicose e são encontrados em ambos os subsítios B e C. O subsítio D pode interagir com substratos maiores, por exemplo, maltotriose, mas não é usado para maltose.

As mudanças conformacionais que conduzem a translocação de uma molécula de maltose por $MalFGK_2$, e como eles são acoplados à ligação de ATP e hidrólise por MalK, são compreendidos no nível atômico (Oldham et al., 2007; Khare et al., 2009; Oldham e Chen, 2011). O transportador repousa em uma conformação voltada para dentro, onde um sítio de ligação de substrato fechado conformado por MalF e MalG é exposto ao citoplasma e os NBDs no dímero MalK estão abertos. A interação da proteína MalE ligada à maltose com as permeases MalF/MalG, no lado periplasmático, impulsiona a atividade dos NBDs MalK, levando o transportador a mudanças conformacionais que direcionam o transporte pela membrana celular (Figura 7).

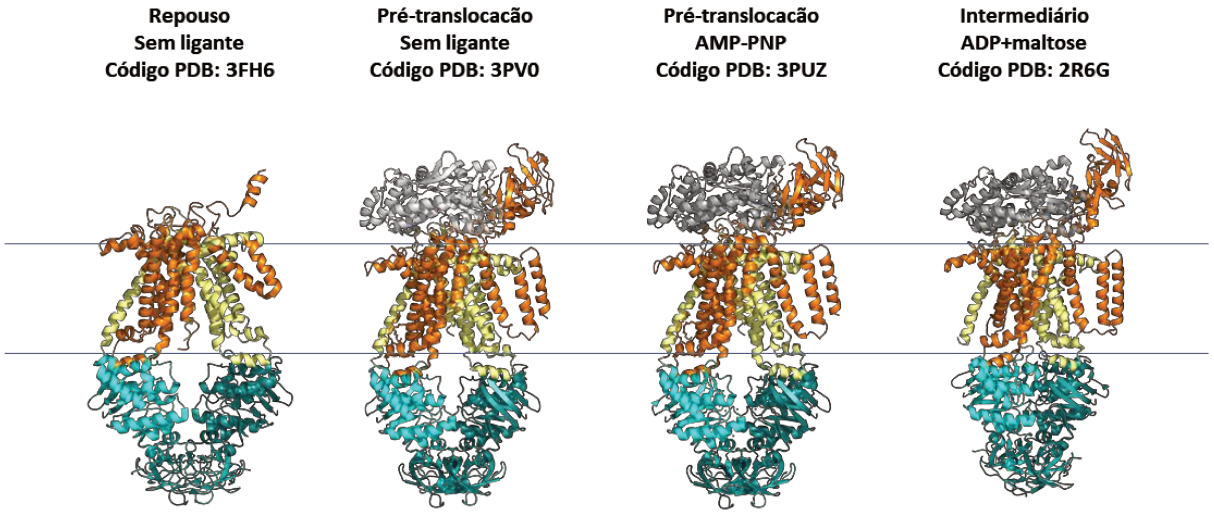


Figura 7. Estados conformacionais do transportador ABC de maltose de *E. coli*. O transportador de maltose foi cristalizado em vários estados conformacionais: voltado para dentro (*inward)* em seu estado de repouso, em um estado de pré-translocação sem ligante, em um estado de pré-translocação em presença de AMP-PNP, e em um estado intermediário voltado para fora da célula, em presença de ADP e maltose.

Inicialmente, este projeto visava a caracterização estrutural/funcional dos componentes dos sistemas de transporte de açúcares de *M. tuberculosis*, assim como o estudo das interações entre os componentes NBDs e TMDs. Como alvos principais, foram consideradas as SBPs LpqY e Rv2041c, cujas estruturas terciárias e funções não são conhecidas. Adicionalmente, essas proteínas são de grande importância no desenvolvimento

de moléculas inibidoras, vacinas e alvos diagnósticos. Infelizmente, após várias tentativas de expressão dos componentes em diferentes células, vetores e condições, não foi possível obtê-las na forma solúvel, o que nos levou ao aprofundamento das abordagens *in silico* dos componentes. Por outro lado, os NBDs eram alvos interessantes devido à sua função essencial no fornecimento de energia dos transportadores e interação com as TMDs. Essa interação, uma vez bloqueada, eliminaria a função do transportador. Similar ao que ocorreu com as periplasmáticas, os NBDs foram de difícil caracterização devido à ausência de expressão ou na forma solúvel. Somente a proteína UgpC foi obtida em pequenas quantidades, porém pouco estável. Como alternativa, trabalhamos com a proteína MalK de *E. coli* e os peptídeos das permeases MalF e MalG, já que os protocolos de produção destas proteínas são amplamente estabelecidos na literatura.

Ainda, focamos na caracterização estrutural dos transportadores de açúcares de *M. tuberculosis*, de forma que pudéssemos inferir informações funcionais e determinantes estruturais importantes, além da relação destes dentro do gênero *Mycobacterium*. Neste objetivo também abordamos as características estruturais das proteínas periplasmáticas em relação a putativos substratos, e a possível promiscuidade na interação das proteínas citoplasmáticas e transmembrana entre os diferentes transportadores. Outro objetivo importante foi o estudo das interações entre os componentes citoplasmáticos e transmembrana, usando peptídeos miméticos das hélices de acoplamento das proteínas transmembrana. O estudo das interações NBD-TMD pode ser de grande importância no desenvolvimento de inibidores, assim como no entendimento do mecanismo dos transportadores ABC, uma vez que o sistema presente em *E. coli* pode ser usado como modelo para outras espécies de bactérias e exportadores de humanos.

No desenvolvimento do projeto foram valiosas e fundamentais as colaborações de diversos pesquisadores: Dr Marko Hyvonen, do Departamento de Bioquímica da Universidade de Cambridge, quem nos deu acesso ao conjunto de vetores de expressão e o sistema RADisplay; Dr. Gabriel Vignoli, Dr. Cristiano Pinto de Oliveira e o Dr. André Luiz Senhem do Instituto de Física da Universidade de São Paulo, em ensaios biofísicos e de SAXS; Dr. André G. Costa-Martins da Escola de Ciências Farmacêuticas da Universidade de São Paulo e José Gabriel Vergara Meza do Instituto de Ciências Biomédicas da Universidade de São Paulo nas análises filogenéticas; e a Dra. Ana Carolina Ramos Moreno do Laboratório de Desenvolvimento de Vacinas do Instituto de Ciências Biomédicas da Universidade de São Paulo na produção de anticorpos.

OBJETIVOS

Objetivo geral

Caracterizar estrutural e funcionalmente os componentes dos transportadores ABC de açúcares presentes em *M. tuberculosis* e usar o transportador de maltose de *E. coli* como modelo para o estudo da interação dos domínios transmembrana com os domínios ligadores de nucleotídeos.

Objetivos específicos

Caracterização estrutural e funcional de ABC de *M. tuberculosis*

1. Genômica comparativa dos transportadores ABC de açúcares de *M. tuberculosis*;
2. Análises filogenéticas dos componentes;
3. Modelagem estrutural e docagem *in silico* dos componentes SBPs;
4. Modelagem estrutural das interações dos componentes NBDs-TMDs;
5. Expressão e purificação de componentes dos sistemas.

Estudos de interação entre TMDs e NBDs utilizando peptídeos sintéticos

1. Desenho de peptídeos, clonagem, expressão e produção da proteína MalK de *E. coli*;
2. Caracterização da interação entre peptídeos e MalK por estudos biofísicos (fluorescência e termoforese);
3. Caracterização de MalK na presença dos peptídeos por espalhamento de raios-X a baixo ângulo (SAXS).

RESULTADOS

Os resultados da presente tese serão apresentados em 2 partes:

No **Capítulo I**, será apresentado o manuscrito intitulado "*Structural and phylogenetic insights of carbohydrate ABC importers from Mycobacterium tuberculosis*", submetido na revista científica *BMC Genomics*, que descreve os componentes dos sistemas de transporte de açúcares do tipo ABC em *M. tuberculosis* a partir de análises genômicas comparativas dentro do gênero *Mycobacterium*, a predição *in silico* das estruturas tridimensionais e sítios de interação dos componentes periplasmáticos, e as interações entre os componentes transmembrana e citoplasmáticos. Essencialmente, cada um dos componentes dos transportadores ABC de açúcares de *M. tuberculosis*, descritos por Braibant (2000), foram submetidos no Protein BLAST para a busca de ortólogos em outras espécies do gênero *Mycobacterium*. Foram analisadas as diferenças em termos de presença e ausência do componente, com base em um corte de cobertura > 90% e identidade > 60%. As estruturas tridimensionais e os sítios de ligação ao substrato das proteínas periplasmáticas LpqY e Rv2041c foram preditas e comparadas com as proteínas periplasmáticas já descritas UspC e UgpB. Por outro lado, a interface de interação entre os componentes transmembrana e citoplasmáticos foi predita; também foi analisada a conservação dos resíduos possivelmente envolvidos na interação. Os resultados foram analisados e discutidos, e sugerem que a exclusividade de pelo menos dois transportadores ABC envolvidos na captação de açúcares em espécies patogênicas do gênero *Mycobacterium* poderia destacar um possível papel deles nos processos de virulência e patogênese. A conservação de resíduos aromáticos nos bolsões de ligação ao substrato das proteínas periplasmáticas mostra que elas têm preferência por açúcares, mas a ausência de conservação nos outros resíduos envolvidos na interação, indica especificidade e afinidades por diferentes substratos. Por outro lado, a conservação entre as interfaces transmembrana-citoplasmática sugere que o transportador UspABC pode usar o componente citoplasmático de qualquer um desses sistemas. No total, esses resultados trazem a primeira comparação desses sistemas e surge a relevância dos sistemas Rv2038c-41c e UgpAEBC para espécies patogênicas do gênero *Mycobacterium*.

Neste capítulo foram incluídos resultados complementares sobre a clonagem, expressão em purificação da proteína UgpC de *M. tuberculosis*, o componente citoplasmático do transportador ABC UgpAEBC. Será descrita também a sua relação filogenética com outras espécies de bactérias, a conservação de motivos no domínio regulatório em comparação com

outras proteínas da mesma família, ensaios de dicroísmo circular, fluorescência intrínseca do triptofano, atividade ATPase, produção de anticorpos, ELISA e *Western Blot*.

No **Capítulo II**, será apresentado o conjunto de dados que serão compilados para o manuscrito "*A new strategy to study interactions between ABC transporter coupling helices and nucleotide-binding domains*". Este trabalho foi desenvolvido a partir do interesse em estudarmos formas de inibição de transportadores ABC, sem a necessidade de expressão do transportador completo. Tentativas prévias de obtenção de proteínas transmembrana para estudos estruturais e funcionais desenvolvidas pelo grupo da Prof. Andrea Balan, foram barradas pela dificuldade de expressão e solubilização dessas proteínas. Como alternativa, tentamos usar apenas peptídeos correspondentes às regiões de interação das TMDs com os NBDs (*coupling helices* ou hélices de acoplamento). Para tal, visando a otimização dos resultados, utilizamos como prova de conceito o transportador ABC de maltose de *E. coli*, um dos transportadores ABC de açúcar mais estudados desde o ponto de vista estrutural e funcional.

CAPÍTULO I

Structural and phylogenetic insights of carbohydrate ABC importers from *Mycobacterium tuberculosis*

Lilia I. De la Torre[1,3,4], José Gabriel V. Meza[2], Sindy Cabarca Barreto[1,3,4], André G. Costa-Martins[5], Andrea Balan[1,3]

[1]Deparment of Microbiology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil

[2]Department of Parasitology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil

[3]Genectics and Molecular Biology Postgraduate Program, Institute of Biology, State University of Campinas, São Paulo, Brazil

[4]Grupo de Investigaciones Biomédicas, Universidad de Sucre, Sucre, Colombia

[5]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

Correspondence:

Prof. Andrea Balan

Laboratory of Applied Structural Biology, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil. Av. Prof. Lineu Prestes, 1374; Cidade Universitária, São Paulo, São Paulo, Brazil; CEP 05508-900.

E-mail: abalan@usp.br

Phone: +55 11 37917745

Abstract

Background

Mycobacterium tuberculosis, the etiological agent of tuberculosis, has at least four ATP-Binding Cassette (ABC) transporters dedicated to carbohydrate uptake: LpqY/SugABC, Rv2038c-41c, UspABC and UgpAEBC. Although the role of these systems in the bacillus is

still unclear, some works have shown that LpqY/SugABC transporter is essential for survival of *M. tuberculosis in vivo* and potentially involved in recycling of cell wall components. The three-dimensional structures of substrate-binding proteins (SBPs) UspC and UgpB were described, however, the affinities for real substrates have been not yet fully determined. Components of these transporters, especially SBPs show high immunogenicity and could be used for development of diagnostic and therapeutic tools. In this work, we used phylogenetic tools and structural bioinformatics to compare the four systems in order to get similarities and differences that can be important for characterization of their roles.

Results

Based on amino acid sequence alignments, phylogenetic analysis, and structural features we showed that the components of the four systems are separated in four different functional groups. A comparison of the putative orthologs of the carbohydrate ABC systems in different species of *Mycobacterium* revealed that Rv2038c-41c and UgpAEBC systems are restricted to pathogenic species of *Mycobacterium*. Using the available crystallographic structures and molecular models of SBPs, we showed that they conserve a characteristic substrate-binding pockets of carbohydrate-binding proteins but do not share similar residues. Moreover, the interface between transmembrane domains (TMDs) and nucleotide-binding domains (NBDs) show conservation in the electrostatic potential.

Conclusions

The exclusivity of at least two different ABC transporters involved in carbohydrate uptake in pathogenic species of *Mycobacterium* genus highlights a possible role of them in virulence and pathogenesis. Despite of the preference for carbohydrates, there is no conservation in the residues that form the substrate-binding pocket of SBPs indicating there are four transporters with different specificities for sugars. The conservation between TMDs and NBDs interfaces suggest a possible promiscuity of the NBD components in *M. tuberculosis* carbohydrate ABC systems. Altogether, these results bring the first comparison of *M. tuberculosis* carbohydrate ABC systems and arise the relevance of Rv2038c-41c and UgpAEBC transporters for pathogenic species of *Mycobacterium*.

Keywords

Carbohydrate uptake, ABC transporters, *Mycobacterium tuberculosis*, Phylogeny, Substrate-Binding Proteins, Protein-Protein Interaction,

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 are infected with *M. tuberculosis* and thus at risk of developing tuberculosis disease [1]. The ability of *M. tuberculosis* to persist inside the host cells under a variety of adverse conditions including oxidative stress, hypoxia and nutrient starvation is 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 [2–4]. Different uptake mechanisms and the broad range of overlapping substrate specificities allow bacteria to quickly adapt to and colonize changing environments, by scavenging essential nutrients from microenvironments to survive within the host. Some mechanisms of nutrient acquisition also are key virulence determinants used by pathogens to mediate disease [5, 6].

Carbohydrates has traditionally been considered an important source of carbon skeleton material and energy supply in bacteria. However, these are not the main carbon and energy sources for *M. tuberculosis*, which prefers host lipids, evidenced in an over-representation of genes in the *M. tuberculosis* genome that encode enzymes for fatty acid metabolism, and upregulation of such genes during macrophage infection [7, 8]. Even though these studies suggest that lipids are the main source of carbon and energy for *M. tuberculosis*, cholesterol import is not required for establishing infection in mice or for growth in resting macrophages [9], suggesting that other yet to be identified carbon sources also have an important role to play [10]. Despite the limited availability of carbohydrates within the macrophage environment, *M. tuberculosis* is equipped with five putative importers of carbohydrate substrates: four members of the ATP-binding cassette (ABC) transporter family and one belonging to the major facilitator superfamily [11].

Bacteria utilize several transport proteins, including ABC transporters to import and expel substrates. ABC transporters, conserved across all organisms, are powered by the energy from ATP to move substrates across cellular membranes. ABC transporters can be classified depending on the direction of translocation of the substrate as importers (into the cell) or exporters (out of the cell). Canonical ABC transporters type importers were identified to date only in prokaryotes and consist of oligoprotein assemblies with two hydrophobic transmembrane domains (TMDs) that form the transport channel, two cytoplasmic nucleotide-

binding domains (NBDs) which in turn are responsible for the breakdown of ATP molecule 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 [12]. ABC transporters probably constitute one of the largest families of paralogous proteins present in *M. tuberculosis*. Genes encoding ABC transporters account for about 2.5% of *M. tuberculosis* genome, being reported 27 complete systems (14 importers and 13 exporters) [11, 13]. The role of ABC importers is well established in the uptake of essential nutrients. The number of ABC importers present in *M. tuberculosis* in comparison with *Escherichia coli*, *Bacillus subtilis* and even *Mycobacterium smegmatis*, revealed that only a few are encoded by the *M. tuberculosis* genome. This is particularly apparent for the transporters involved in carbohydrate uptake [11, 13]. Despite many ABC transporters dedicated to carbohydrate transport have been related to virulence and pathogenesis in pathogenic bacteria, the role of these transporters in *M. tuberculosis* is not clear.

Four operons encode ABC transporters dedicated to carbohydrate uptake in *M. tuberculosis* genome: *lpqY-sugABC, rv2038c-41c, uspABC*, and *ugpAEBC* [13], 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. Several studies demonstrate direct or indirect participation of *M. tuberculosis* ABC importers in survival and virulence of the bacterium within the host. 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 [14]. This transporter also arises interest for the detection of *M. tuberculosis* in sputum sample, since it can probably be the pathway for uptake of a solvatochromic trehalose probe [15]. However, there is a lack of structural and functional information regarding this 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 in apo state (PDB code: 5K2X) and binding studies showed a higher affinity for carbohydrates containing an amino group at the C2 or C3 position, like D-glucosamine-6-phosphate and chitobiose than s*n*-glycero-3-phosphocholine, D-glucose or α,α-D-trehalose [10]. The UgpAEBC ABC transporter is predicted to be involved in scavenge of glycerophospholipids like glycerophosphocholine [16, 17], that 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 glycerophoscholine (PDB code: 6R1B), but functional studies revealed that the protein has no

specificity limited to this substrate [17]. 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) [18]. Immunological studies with 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 as antigen, indicating the capability of this protein for induction of the humoral immune response, and highlighting its potential for development of vaccine candidate against *M. tuberculosis* [19]. The 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 importer systems from *M. tuberculosis* involved in carbohydrate uptake. We evaluated the conservation and phylogenetic relationship in *Mycobacterium* species, and structural features comparisons that might bring light in their role and relevance for the bacillus. We showed that Rv2038c-41c and UgpAEBC systems are exclusive of pathogenic *Mycobacterium* species and the LpqY/SugABC system were the first paralogs to diverge from carbohydrate ABC transporter in *M. tuberculosis*. Also, we showed significant differences in the substrate-binding pockets of SBPs that indicate specificities for different substrates. The absence of NBD in UspABC system might be compensate by other carbohydrate NBDs since the interface with TMDs is conserved.

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

Mycobrowser identifier	Gene name	Possible function (Mycobrowser)	Phenotypes of mutants in different conditions and results of different studies of proteomics, biophysics, and others	References
Rv1235	*lpqY*	Active transport of sugar across the membrane (import)	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.	[20–27]
Rv1236	*sugA*	Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane.		
Rv1237	*sugB*	Active transport of sugar	[a] Rv1238 dimerizes in presence of	

		across the membrane (import). Responsible for the translocation of the substrate across the membrane.	ATP, maximum activity in pH 7.5 and Km of 0.15; presence of regulatory domain as shown for *E. coli* MalK.	
Rv1238 [a]	*sugC*	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.	[20,21,23,25,28]
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] *ex vivo* and *in vivo* models of *M. tuberculosis* infection showed a significantly high level of induced pro-inflammatory cytokines; IFN-γ and TNF-α	[18–21,29]

			secretion increased after stimulation with purified Rv2041c protein to lymphocytes from latent and active TB mice; suggested as a new T-cell antigen that could be a potential vaccine candidate against *M. tuberculosis* infection.	
Rv2316	*uspA*	Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane.	Mutants are essential for in vitro growth and in macrophages.	[20,21,23]
Rv2317	*uspB*	Active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane.		
Rv2318	*uspC*	Active transport of sugar across the membrane (import).	Mutant is essential for in vitro growth and in macrophages; 3D structure available (PDB) shows an acidic carbohydrate-substrate binding cleft with preference for amino sugars; this protein is related to a potential role in recycling components of cell-wall peptidoglycan.	[10,20,21,23]
Rv2832c	*ugpC*	Active transport of Sn-glycerol-3-phosphate across the membrane (import). Responsible for energy coupling to the transport system.	Mutant is essential for in vitro growth.	[20,21,29]
Rv2833c	*ugpB*	Active transport of Sn-glycerol-3-phosphate and glycerophosphoryl diesters across the membrane (import). Sn-glycerol-3-phosphate and	Mutant is essential for in vitro growth, in macrophages and during infection in mice; 3D structure solved at 1.5 Å resolution and affinity for glycerophosphocholine (GPC).	[16,17,20,21,24,29]

		glycerophosphoryl diesters - binding protein interacts with the binding protein-dependent transport system UgpAEC		
Rv2834c	*ugpE*	Active transport of Sn-glycerol-3-phosphate across the membrane (import). Responsible for the translocation of the substrate across the membrane.	Mutant is essential for in vitro growth.	[20,21]
Rv2835c	*ugpA*	Active transport of Sn-glycerol-3-phosphate across the membrane (import). Responsible for the translocation of the substrate across the membrane.	Mutant is essential for in vitro growth and in macrophages; the mutant strain showed slow-down growth or stressed.	[20,21,23,28]

Results

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

The co-occurrence of the four putative carbohydrate ABC transporter components of *M. tuberculosis* in different taxa was analysed using String server (Fig. 1). The operon genes are represented by arrows, whose colours were defined according to their functions. The data show that putative orthologs of the *lpqY/sugABC*, *rv2038c-41c* and *ugpAEBC* operons are prominently present in most of the taxa evaluated, except Eukaryota as expected. The *uspABC* operon may be misrepresented due to the lack of an evident NDB component. It seems clear that in some species of Actinobacteria group (*M. tuberculosis*, *Nocardia brasiliensis* and *Rodococcus fascians*), the operons *lpq/YsugABC*, *rv2038c-41c*, and *ugpAEBC* are conserved (dark red and black bar). However, *Corynebacterium* species has not the same conservation level, as such in other actinobacteria. *C. diphtheriae* shows no conservation for the SBP

component of each operon. The most conserved operon component, as largely described for ABC transporters, is the NBD, such as SugC, Rv2038c and UgpC (blue arrows). The low level of conservation of SBPs, specially LpqY and UspC, is expected since these components usually show low amino acid sequence identity despite the conservation of the three-dimensional folding. Moreover, the completely absence of these two proteins in species like *C. diphtheriae* (Actinobacteria), *Pseudomonas aeruginosa* (Proteobacteria), *Staphylococcus aureus* (Firmicutes), *Treponema pallidum* (Spirochaetales) and protozoan species of Eukarya, suggest no homology with any of the SBPs.

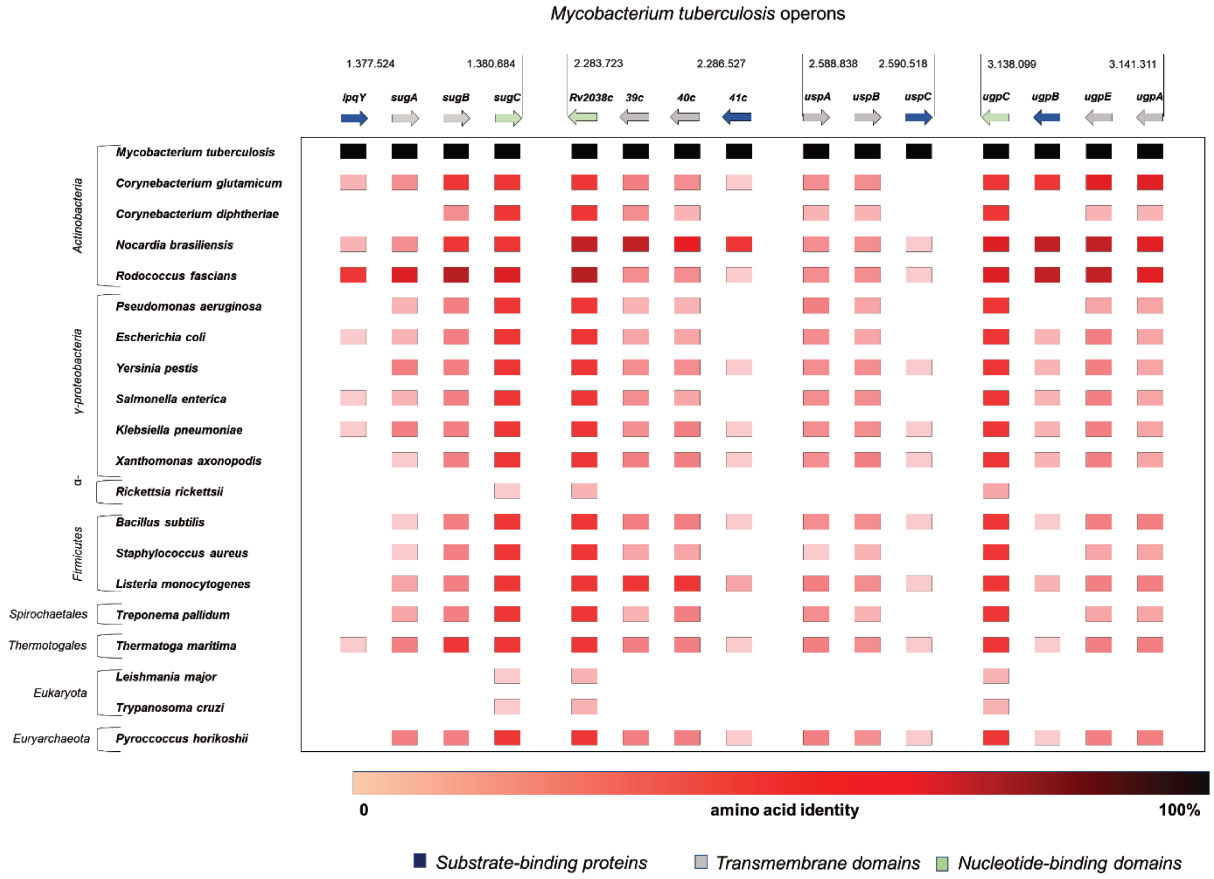


Figure 1. Co-occurrence and genomic proximity of genes encoding for carbohydrate ABC transporters components from *Mycobacterium tuberculosis* across different species. The intensity of red colour reflects a conservation level of the component in the species, from the lightest (least conserved protein) 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 in the left side.

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

In addition to the previous studies, we conducted a comparative genomic analysis of components of putative *M. tuberculosis* carbohydrate ABC transporters in important species

of *Mycobacterium* genus (Table 2). We analysed 14 distinct species: six of these belong to the *M. tuberculosis* complex (MTBC) (three *M. tuberculosis* strains, *M. africanum*, and two *M. bovis* strains), six are classified as pathogenic mycobacteria nonbelonging to the MTBC (*M. avium*, *M. intracellulare*, *M. ulcerans*, *M. marinum*, *M. abscessus* and *M. leprae*) and one 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 *Mycobacterium* genus used in this study conserved orthologs of ABC transporters LpqY/SugABC and UspABC that show high amino acid sequence identities (64 % up to 100%). However, Rv2038-41c is absent in *M. abscessus* and *M. smegmatis*, and UgpAEBC is exclusive of MTBC members and *M. marinum*.

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 homologues 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	***M. tuberculosis* H37Rv**	Rv1235 100%	Rv1236 100%	Rv1237 100%	Rv1238 100%	Rv2038c 100%	Rv2039c 100%	Rv2040c 100%	Rv2041c 100%	Rv2316 100%	Rv2317 100%	Rv2318 100%	Rv2832c 100%	Rv2833c 100%	Rv2834c 100%	Rv2835c 100%
	***M. tuberculosis* H37Ra**	MRA_ RS06560 100%	MRA_ RS06565 100%	MRA_ RS06570 100%	MRA_ RS06575 100%	MRA_ RS10790 100%	MRA_ RS10795 100%	MRA_ RS10800 100%	MRA_ RS10805 100%	MRA_RS12305 100%	MRA_ RS12310 100%	MRA_ RS12315 100%	MRA_ RS15055 100%	MRA_ RS15060 100%	MRA_ RS15065 100%	MRA_ RS15070 100%
	***M. tuberculosis* CDC1551**	MT_ RS06530 100%	MT_ RS06535 100%	MT_ RS06540 100%	MT_ RS06545 100%	MT_ RS10670 100%	MT_ RS10675 99%	MT_ RS10680 100%	MT_ RS10685 100%	MT_ RS12140 99,66%	MT_ RS12145 100%	MT_ RS12150 100%	MT_RS14855 100%	MT_RS14860 100%	MT_ RS14865 100%	MT_ RS14870 100%
	***M. africanum* GM041182**	MAF_ RS06545 100%	MAF_ RS06550 100%	MAF_ RS06555 100%	MAF_ RS06560 100%	MAF_ RS10625 100%	MAF_ RS10625 100%	MAF_ RS10635 100%	MAF_ RS10640 100%	MAF_ RS12095 99,66%	MAF_ RS12100 100%	MAF_ RS12105 99,77%	MAF_ RS14745 100%	MAF_ RS14750 99,77%	MAF_ RS14755 100%	MAF_ RS14760 99,67%
	***M. bovis* AF2122/97**	BQ2027_ MB1267 100%	BQ2027_ MB1268 100%	BQ2027_ MB1269 100%	BQ2027_ MB1270 100%	BQ2027_ MB2064C 100%	BQ2027_ MB2065C 99%	BQ2027_ MB2066C 100%	BQ2027_ MB2067C 99,67%	BQ2027_ MB2343 99,66%	BQ2027_ MB2344 100%	BQ2027_ MB2345 99,77%	BQ2027_ MB2856C 100%	BQ2027_ MB2857C 99,31%	BQ2027_ MB2858C 100%	BQ2027_ MB2860C BQ2027_ MB2859C 100%
	***M. bovis* BCG Pasteur 1173P2**	BCG_ RS06705 100%	BCG_ RS06710 100%	BCG_ RS06715 100%	BCG_ RS06720 100%	BCG_ RS10595 100%	BCG_ RS10600 99,64%	BCG_ RS10605 100%	BCG_ RS10610 100%	BCG_ RS12065 99,66%	BCG_ RS12070 100%	BCG_ RS12075 99,77%	BCG_ RS14680 99,72%	BCG_ RS21385 100%	BCG_ RS14695 100%	BCG_ RS14700 100%
Other pathogenic mycobacteria	***M. avium* 104**	MAV_ RS06590 76,92%	MAV_ RS06595 84,25%	MAV_ RS06600 87,96%	MAV_ RS06605 85,64%	MAV_ RS11815 82,35%	MAV_ RS11810 80,71%	MAV_ RS11805 79,67%	MAV_ RS11800 75,85%	MAV_ RS09985 84,14%	MAV_RS09980 85,77%	MAV_ RS09975 78,86%				
	***M. avium* subsp. paraTBK-10**	MAP_ RS12995 76,71%	MAP_ RS12990 85,59%	MAP_ RS12985 88,32%	MAP_ RS12980 85,90%	MAP_ RS08985 83,15%	MAP_ RS08990 80,71%	MAP_ RS08995 79,67%	MAP_ RS09000 75,40%	MAP_ RS10625 83,79%	MAP_ RS10630 86,13%	MAP_ RS10635 78,41%				
	***M. intracellulare* ATCC 13950**	OCU_ RS312765 78,21%	OCU_ RS31270 84,01%	OCU_ RS31275 87,96%	OCU_ RS31280 85,64%	OCU_ RS37020 82,87%	OCU_ RS37015 81,43%	OCU_ RS37010 81,88%	OCU_ RS37005 77,40%	OCU_ RS35040 82,41%	OCU_ RS35035 86,50%	OCU_ RS35030 79,09%				
	***M. ulcerans* AGY99**	MUL_ RS23385 No sequence	MUL_ RS23380 No sequence	MUL_ RS23375 87,23%	MUL_ RS23370 No sequence	MUL_ RS11890 82,02%	MUL_ RS11895 84,29%	MUL_ RS11900 85,33%	No information	MUL_ RS06750 85,17%	MUL_ RS06745 83,21%	MUL_ RS06740 81,82%				
	***M. marinum* M**	MMAR_ RS21065 76,50%	MMAR_ RS21060 82,99%	MMAR_ RS21055 89,78%	MMAR_ RS21050 84,99%	MMAR_ RS15025 82,30%	MMAR_ RS15030 85,36%	MMAR_ RS15035 85,67%	MMAR_ RS15040 75,57%	MMAR_ RS18015 85,17%	MMAR_ RS18020 85,04%	MMAR_ RS18025 81,82%	MMAR_ RS09415 81.11%	MMAR_ RS09410 75.92%	MMAR_ RS09405 84.00%	MMAR_ RS09400 85.28%
	***M. leprae* TN**	ML_ RS05355 77,25%	ML_ RS05360 79,28%	ML_ RS05365 89,42%	ML_ RS05370 79,74%	ML_ RS07205 79,49%	ML_ RS07210 80,00%	ML_ RS07215 79,67%	ML_ RS07220 78,77%	ML_ RS09020 81,72%	ML_ RS09025 85,04%	ML_ RS09030 75,68%				
	***M. abscessus* 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 64,20%				
	***M. smegmatis* mc2155**	MSMEG_ 5061 69,76%	MSMEG_ 5060 78,03¨%	MSMEG_ 5059 85,04%	MSMEG_ 5058 73,85%					MSMEG_ 4466 81,38%	MSMEG_ 4467 79,17%	MSMEG_ 4468 70,35%				

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

In order to evaluate the protein polymorphism among the orthologs identified in *Mycobacterium* species, amino acid sequences of all the proteins belonging the same functional group (NBD, TMD and SBP) were firstly aligned using Clustal Omega, and then submitted to MEGA-X software for phylogenetics analyses, using a maximum likelihood method. Modern inferences of deep phylogenies using conserved proteins almost exclusively rely on likelihood and Bayesian methods [30]. In parallel, we built structural models for all components of the *M. tuberculosis* and use them to map the insertion and/or deletion regions found in the alignments, in order to describe possible events that occurred after the sequences diverged from a common ancestor. The structures were built using I-TASSER server or Modeller program, and templates, identities and references details are listed in Table S2 (Additional file 1).

The nucleotide-binding domains (NBDs)

The phylogenetic analysis of NBDs showed that they segregate in three main groups each one containing SugC, Rv2038c and UgpC orthologs, respectively (Fig. 2A). All the orthologs that grouped with UgpC belong to the MTBC group and *M. marinum*, and they are closer related to Rv2038c than SugC groups. Similarly, the groups of Rv2038c and SugC NBDs consist of MTBC orthologs but also other pathogenic mycobacteria, and further *M. smegmatis* for SugC (Fig. 2A). We used the NBDs amino acid sequence alignments (Additional file 2) to identify possible polymorphic regions and highlighted them in the structural models (Fig. 2B). This approach could be important to understand the functional differences of these components. The models of NBDs were built from the structural coordinates of the *Thermococcus litoralis* MalK (PDB code: 1G29) [31], which shared amino acid sequence identity higher than 50% (Additional file 1, Table S2). The models for *M. tuberculosis* NBDs analysed in this work, show the conserved catalytic sub-domain, which is similar to the core structure found in many RecA-like motor ATPases [32], but also an additional small C-terminal domain that is unique in some ABC transporters, including that related to carbohydrate and ion uptake [23, 24]. These small domains have regulatory functions with ability to bind the same substrates of the cognate transporters. In these cases, NBDs bound to substrates are maintained in inhibited state, dissociated from each other, with no further transport [34]. Indeed, the alignments revealed that the catalytic domains of the *M. tuberculosis* NBDs are quite similar but

significant differences were found in the regulatory domains (Fig. 2B, green spheres), indicating that high diversification of these components is due to differences in C-terminal regulatory modules, and it could be an important function determinant.

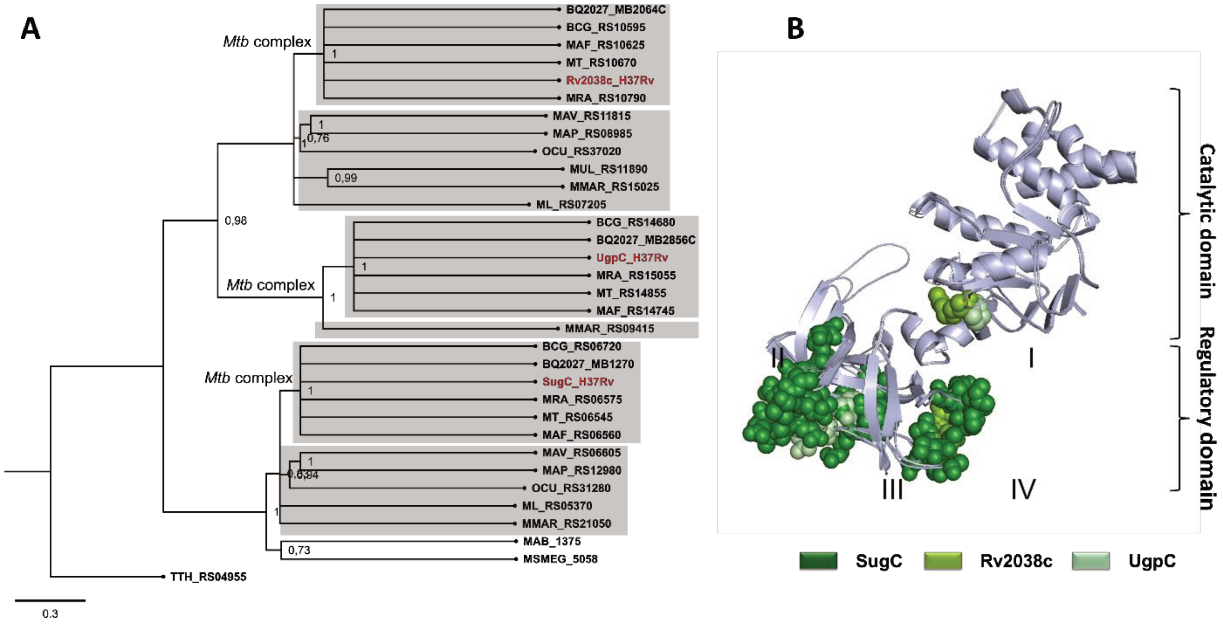


Figure 2. Phylogenetic relationships of NBD components of carbohydrate ABC transporters from species of *Mycobacterium* genus. A Phylogenetic tree of *Mycobacterium* carbohydrate NBD components was built using a Maximum likelihood method and Jones-Taylor-Thornton (JTT) amino acid substitution model in MEGA-X. Proteins are named with the same NCBI locus tag as presented in Table 1. Dark gray: *M. tuberculosis* complex group (MTBC), light gray: other pathogenic mycobacteria, and white: environmental species, represented by *M. smegmatis*. The amino acid sequence of TTH_RS04955 from *Thermus thermophilus*, encoding a putative carbohydrate NBD, was used as outgroup. **B** Structural models of SugC, Rv2038c and UgpC showing the four main representative regions of variation (green coloured spheres) identified in the amino acid sequences alignments. The SugC, Rv2038c and structural models was built based on the structural coordinates of the *Thermococcus litoralis* MalK protein (PDB code: 1G29), which is the NBD component of the trehalose/maltose ABC transporter.

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 analysed 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 was generated in Clustal Omega for each protein group and used as inputs in MEGA-X software for 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. According to the alignments, proteins from group 2 have 5 to 30 additional residues than those from group 1 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. To identify the location of possible regions of polymorphism in the proteins, a structural model of each *M. tuberculosis* TMD component was built, and proteins were compared by pairs (Additional file 3). 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 (Fig. 3B, green spheres). No differences were identified in the coupling helices as well as in the helices that form the pore suggesting that a general and common characteristic is maintained in the translocation pore of carbohydrate transporters (Additional file 4).

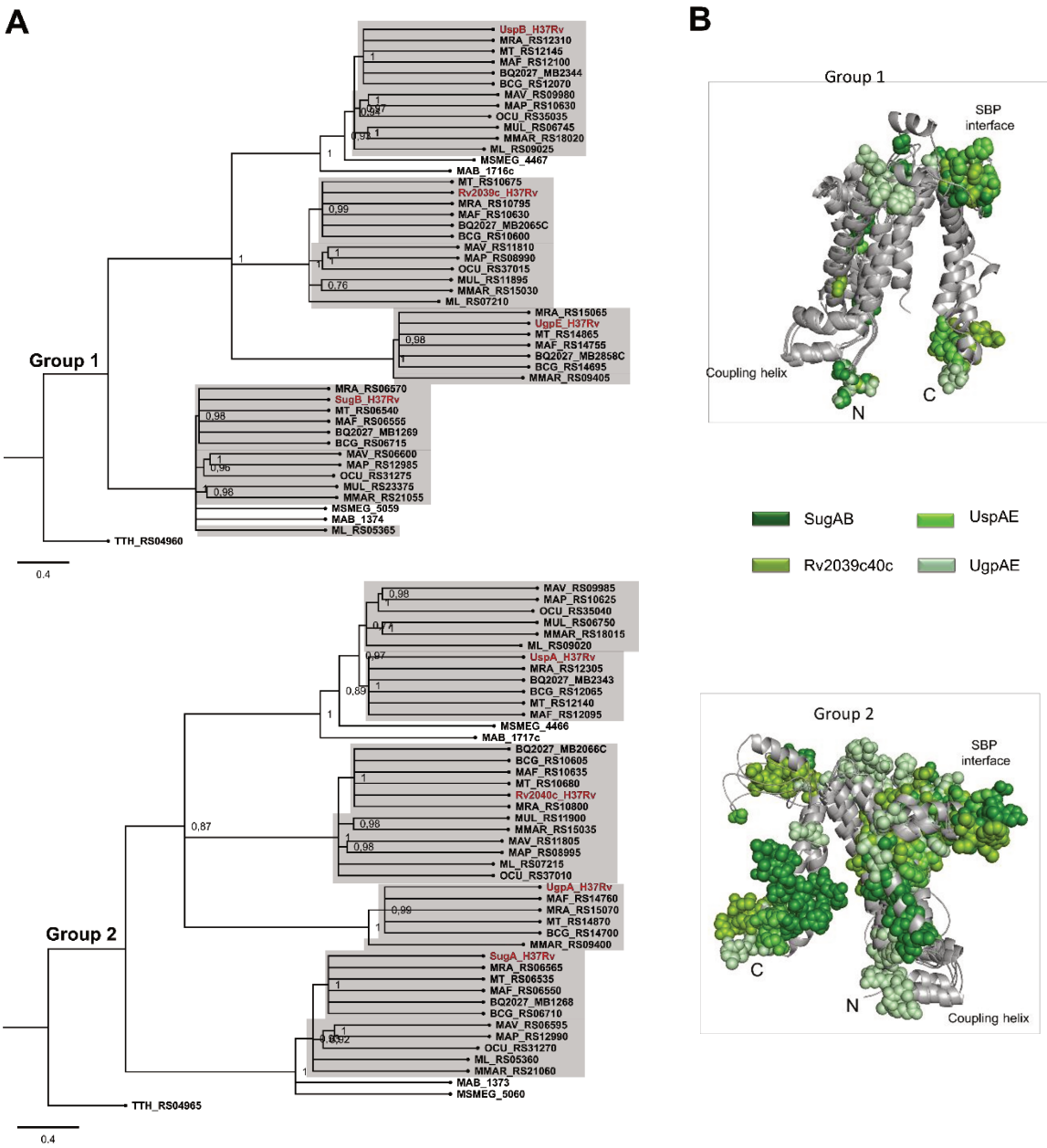


Figure 3. Phylogenetic relationships of TMD components of carbohydrate ABC transporters from species of *Mycobacterium* genus. **A** Phylogenetic tree of *Mycobacterium* carbohydrate TMD components was built using a Maximum likelihood method and Le-Gascuel (LG) amino acid substitution model in MEGA-X. Two trees are showed due the separation of TMDs in two groups, each one with one member of each transporter. Group 1: SugB, Rv2039c, UspB and UgpE, and Group 2: SugA, Rv2040c, UspA and UgpA. Dark gray: *M. tuberculosis* complex group (MTBC), light gray: other pathogenic mycobacteria, and white: environmental species, represented by *M. smegmatis.* The amino acid sequences of TTH_RS04960 and TTH_RS04965 from *Thermus thermophilus*, encoding a putative carbohydrate TMDs, were used as outgroup **B** Structural models of SugB, Rv2039c, UspB, UgpE and SugA, Rv2040c, UspA, and UgpA highlighting the regions that showed significant variation in the amino acid sequences alignments (green coloured spheres). The models were built based on the TMDs structural coordinates of the *E. coli* maltose ABC transporter MalG and MalF, respectively (PDB code: 2R6G).

The substrate-binding proteins (SBPs)

The role of SBPs in ABC importers is of great relevance since they are the components responsible for 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 active state of the transporter [35]. 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 own orthologs (Fig. 4A). UgpB, although almost exclusive of MTBC is closer to LpqY than UspC and Rv2041c. This result is quite different of NBDs or TMDs phylogeny, which SugABC transporter components form a separated clade from other systems. The available structures of *M. tuberculosis* UgpB (PDB code: 4MFI) [16], UspC (PDB code: 5K2X) [10], and molecular models of LpqY and Rv2041c were used for mapping the differences evidenced in the amino acid sequence alignments (Additional file 5). 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 is located the substrate-binding site. A comparison among the protein groups in the alignment allowed us to identify specific regions with amino acids indels, as shown in Fig. 4B. UspC is the shortest protein and differently of the three others, it has a deletion of two set 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 insertion of sequences in domains I and II (Fig. 4B, blue spheres). These regions could be not involved in the carbohydrate-binding site but could be indirectly affect the structure of the binding-pockets. UgpB, that belongs to a transporter almost exclusive of MTBC species, is the protein that presents more sites of variability, in both domains, including regions that directly affect the substrate-binding pocket (Fig. 4B, Additional file 5).

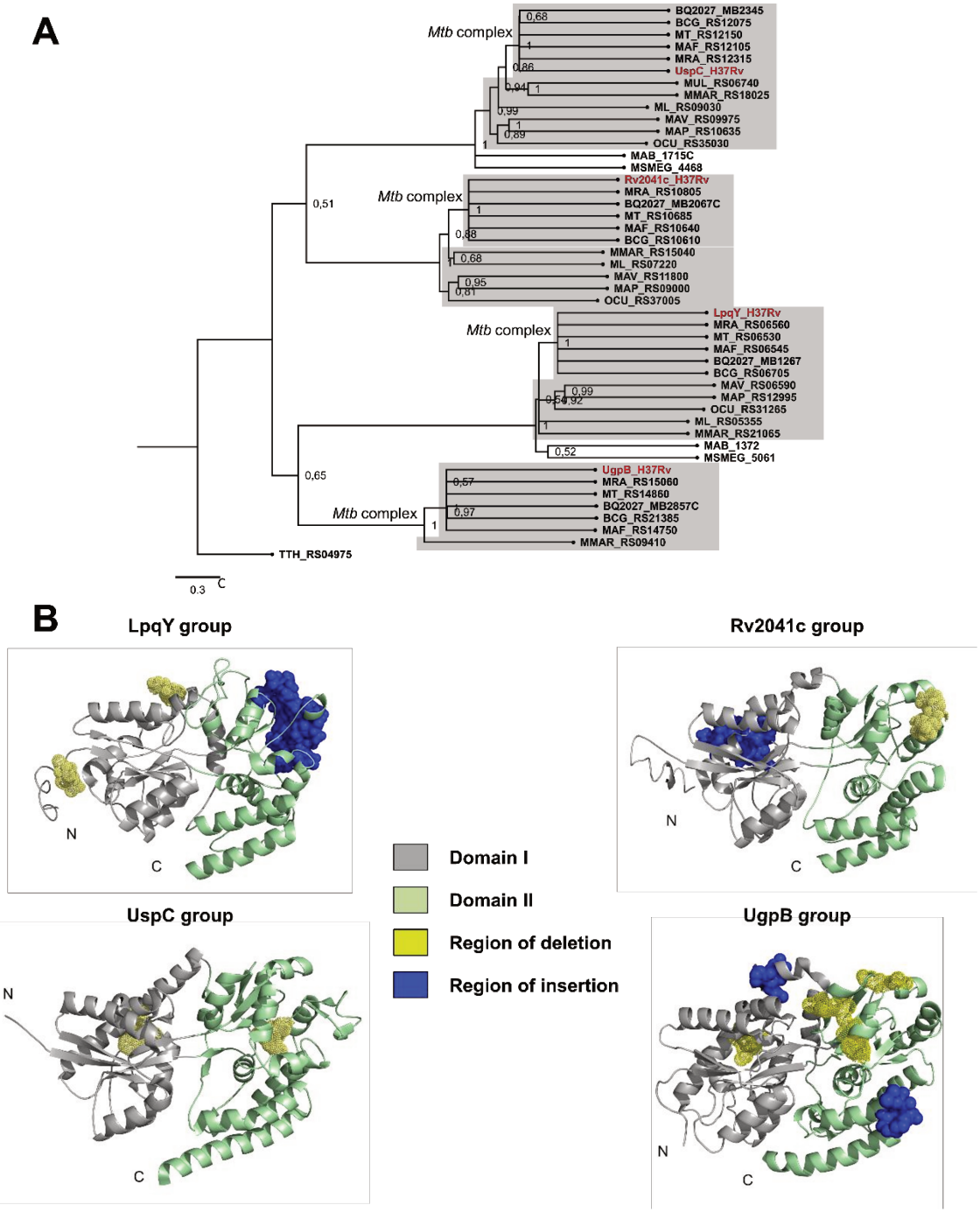


Figure 4. Phylogenetic relationships of SBP components of carbohydrate ABC transporters from species of *Mycobacterium* genus. A Phylogenetic tree of *Mycobacterium* carbohydrate SBP components. The tree was built using a Maximum likelihood method and Whelan and Goldman (WAG) amino acid substitution model in MEGA-X. Dark gray: *M. tuberculosis* complex group (MTBC), light gray: other pathogenic mycobacteria, and white: environmental species, represented by *M. smegmatis*. The amino acid sequence of TTH_RS04975 from *Thermus thermophilus*, encoding a putative carbohydrate SBP was used as outgroup. **B** Crystallographic structure 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 large variation. Regions of amino acid insertion/deletion are represented in blue and yellow spheres, respectively.

Comparison of the substrate-binding pockets among the *M. tuberculosis* H37Rv SBPs

As mentioned previously, the SBPs are the components that define the function of an ABC transporter type importer. In this sense, once the previous results showed that the four proteins separate in different groups, which is suggestive of different functions, we performed comparative analyses of the putative substrate-binding pockets. We used the previous published three-dimensional structures of the UspC and UgpB and the molecular models of LpqY and Rv2041c (Additional file 1, Table S3). Comparison of the general structure of the proteins showed a pattern defined by the two domains (or lobes), where domain I is smaller than domain II and it is more conserved (Fig. 5A, Additional file 6). The structures are quite similar, but in the contrary, the substrate-binding pockets of the four proteins are formed by different residues suggesting an ability to interact with different substrates (Fig. 5B and 5C), which is corroborated by the phylogenetics analyses. The substrate-binding pocket of *M. tuberculosis* UgpB is closely related to the *E. coli* UgpB, but instead to bind sn-glycerol-3-phosphate (G3P) as its ortholog, it showed preferences for glycerophosphocholine (GPC, K_d of 27.3 μM) [16] and a broad range of glycerophosphodiesters [17]. UspC has a cleft between both domains with a set of aromatic side chains, which is assumed to afford potential binding with carbohydrate moieties, and a cluster of acidic residues on N-domain (Asp47, Glu48), inside the pocket (Asp145) and on domain II (Asp216, Asp270, Asp273, Glu410 and Asp414), which also is characteristic of carbohydrate-binding proteins [10]. Thermal shift assays of UspC with different carbohydrates revealed increasing of thermotolerance above 3ºC in presence of D-glucosamine-6-phosphate and chitobiose [10]. Similarly, there is no conservation in the electrostatic potential of the entrance of the pocket and TMDs interface regions (Fig. 5D, black square). Based on the amino acid sequence alignment of the four proteins, it was possible to identify five regions of conservation, as showed in Additional file 6. Four regions are probably involved in the structural folding, but it is interesting to notice that region I is located in domain I, exactly in the interface with the permeases. These results are in agreement with the TMD analyses, which also revealed significant differences in the interface with the SBPs in domain II [36].

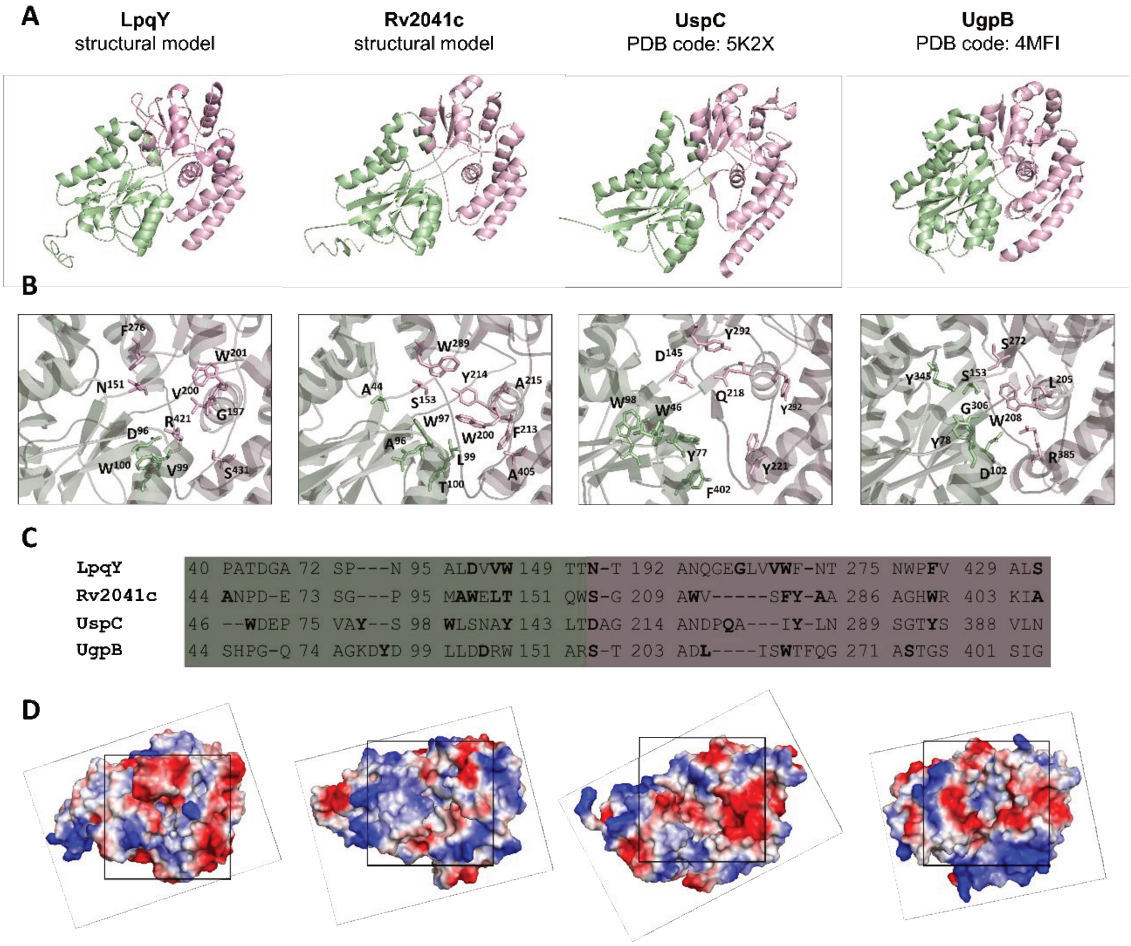


Figure 5. Comparison among the four carbohydrate-binding proteins from *M. tuberculosis*. **A** The crystallographic structures of UgpB (PDB ID: 4MFI) and UspC (PDB ID: 5K2X) were compared to the structural models of Rv2041c and LpqY. Proteins are shown as cartoon with domain I (N-terminal) and II (C-terminal) coloured in green and pink, respectively. **B** Mapping of the residues that form the substrate-binding pockets according to the crystallographic structures and structural models. **C** Highlights of the Clustal Omega amino acid sequence alignment showing the pocket residues of the proteins LpqY, Rv2041c, UspC and UgpB. Residues in bold were identified using ITASSER prediction for LpqY and Rv2041c, and the substrate-binding pockets described for UgpB and UspC [10, 16-17]. **D** Electrostatic potential of the proteins from the pocket entrance perspective (black square). Blue: positive, red: negative, gray: neutral.

Comparison of *M. tuberculosis* H37Rv LpqY and Rv20141c SBPs with putative orthologs

The putative substrate-binding pockets of LpqY and Rv2041c were defined using the structural models and the ligand binding sites predicted using I-TASSER, and further comparisons with other carbohydrate-binding proteins (Fig. 6). The final model of LpqY consists of residues 1 to 468, which corresponds to the mature protein with signal peptide (Additional file 7). The LpqY structural folding as built in the modelling is close related to the trehalose-binding protein of *T. litoralis* (TMBP, PDB code: 1EU8) [31] and the maltose-binding protein of *T. maritima* (MalE3, PDB code: 6DTQ) [37]. Structural alignment of these

structures with LpqY showed root-mean-square deviation (RMSD) of 1.68 Å and 1.31 Å, respectively, and conservation of the residues Asp97, Glu196, Gly197, Glu258, Trp276 and Arg427 (Fig. 6A and B). To simulate the interaction of LpqY with trehalose and maltose, we performed molecular docking analyses using AutoDock Vina, a software that uses a gradient optimization method in executing local optimization to produce the best conformations that have the lowest possible binding energies [38]. The structural model obtained from homology modelling was used as a receptor molecule and the putative ligands trehalose and maltose were obtained from PubChem. Nine substrate conformations of ligands in the substrate-binding site were generated. The best conformations were chosen based on the lowest binding energy, which indicated that maltose has higher binding affinity followed by trehalose (-7.6 kcal/mol and -6.7 kcal/mol, respectively) (Fig. 6B). The structural model of mature protein Rv2041c generated by I-TASSER consists of residues 1 to 439 and showed structural similarity with *Listeria monocytogenes* Lmo0181 (PDB code: 5F7V) [39] with RMSD of 1.66 Å (Fig. 6C). The comparison of the substrate-binding pocket as predicted by I-TASSER showed no conservation of residues and apparently, no possibility for interaction with cycloalternan, the substrate of Lmo0181 (Fig. 6D). Nevertheless, docking of Rv2041c was performed with cycloalternan as ligand but the obtained complexes showed high binding energy and the analysis of the electrostatic potential of both proteins evidenced opposite charges in the substrate-binding pocket (Fig. 6D).

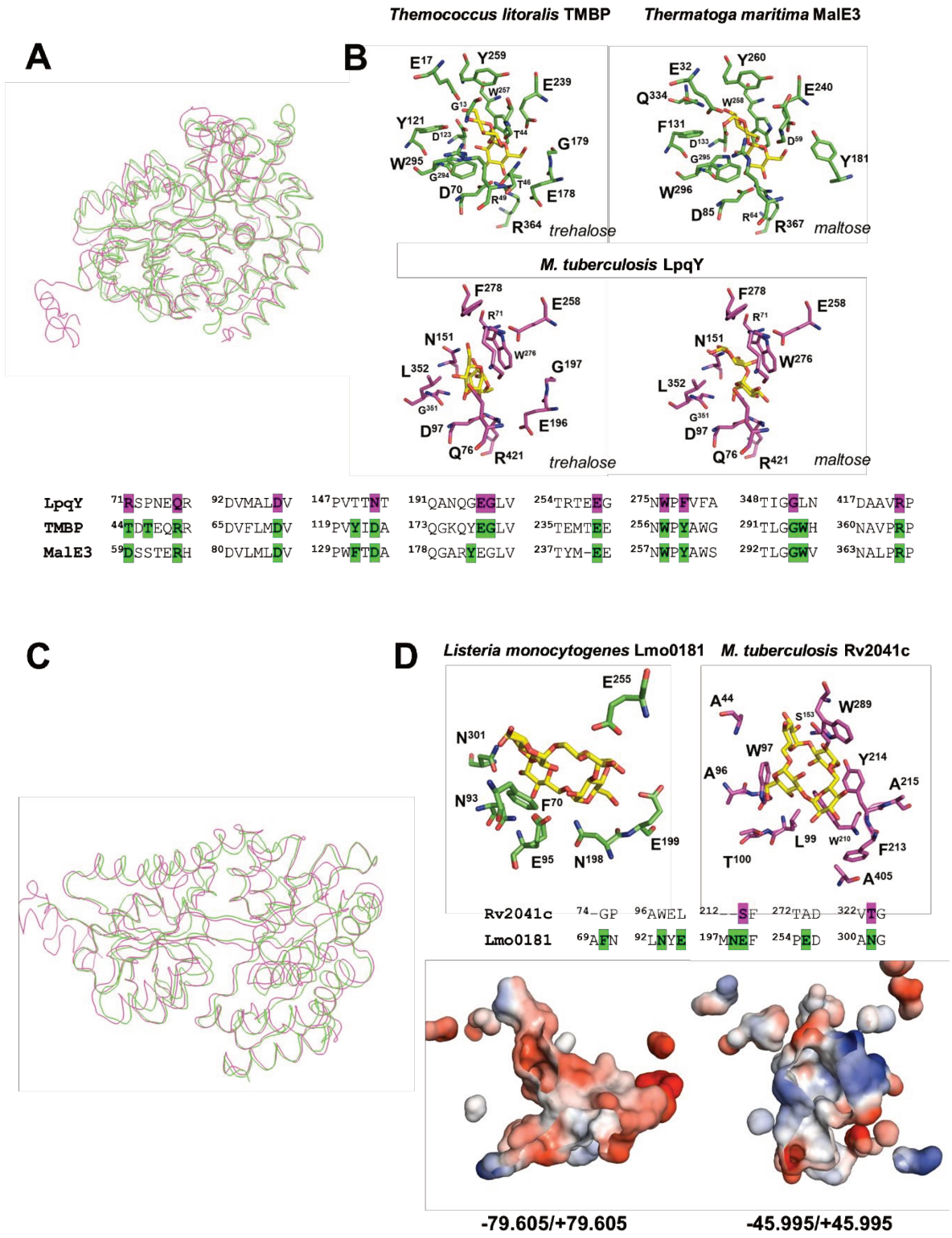


Figure 6. The substrate-binding pockets of *M. tuberculosis* LpqY and Rv2041c in comparison with similar structures. A Structural alignment of *M. tuberculosis* LpqY with *Thermococcus litoralis* trehalose/maltose-binding protein TMBP (PDB code: 1EU8) and *Thermotoga maritima* maltose-binding protein MalE3 (PDB code: 6DTQ). The structures were aligned in Pymol and showed as ribbon. **B** Comparison of binding pocket residues of TMBP and MalE3 structures (green sticks) with LpqY structural model after molecular docking with trehalose and maltose (pink sticks). Residues conserved in the structures are highlighted as in the amino acid sequence alignment. **C** Structural alignment of *M. tuberculosis* Rv2041c and the structure of *Listeria monocytogenes* Lmo0181 bound to cycloalternan (PDB code: 5F7V). The structures were aligned in Pymol and showed as ribbon. **D** Comparison of the residues that coordinate cycloalternan in Lmo0181 and Rv2041c after

docking with the same ligand. Residues are highlighted in bold in the alignment and showed as sticks on the structures. The electrostatic potential of the substrate-binding pockets from both structures was calculated in Pymol and shown as surface. Red: negative charges, blue: positive charges, clear blue: neutral.

Characterization of the interaction between TMD coupling helices and NBDs

TMDs in ABC transporters are responsible for the pore formation but they play an essential role in the activation of the NBDs during the transport. From the four carbohydrate ABC transporters from *M. tuberculosis,* UspABC lacks the ATPase domain and that arise the questioning if could exist some promiscuity between NBDs and TMDs in these *M. tuberculosis* transporters. In this sense, the interfaces of coupling helices and NBDs were comparatively analysed. Firstly, we compared the position of coupling helices in the TMDs structural models, with amino acid sequence alignment and the predictions of TOPCONS web server (Additional file 4). Altogether, these analyses were compared with MalF and MalG sequences of *E. coli* maltose ABC transporter (PDB code: 2R6G) [40]. The results showed that coupling helices of *M. tuberculosis* carbohydrate ABC transporters are located between TM4 and TM5 (Fig. 7A, green and red helices, and Additional file 4). Using GREMLIN complexes server, we showed that TMD components conserve alanine residues in the EAA motif [(E/N/R/K)AA], and charged amino acids (except for UspB that conserve Asn residues) that could have an important role in protein-protein interactions (Fig. 7A-I). SugC, Rv2038c and UgpC conserve a tyrosine and a phenylalanine in positions 88, 89, 89 and 99, 100, 101, respectively. Rv2038c and UgpC proteins seem to have more residues in common (Table S4, Additional file 1). The NBD interfaces also were mapped in the structural models (Fig. 7A, gray region) and had the amino acid sequence aligned. The amino acid conservation of NBD interface regions is higher than the coupling helices, and from six residues identified in positioning of interaction at least four are conserved (Fig. 7A-II). The electrostatic potential of the coupling helices and NBDs interface also was evaluated and revealed a good complementarity. The coupling helices showed a prominent negative interface (Fig. 7B) and the NBDs a complementary set of 8 bunches of positive charges that are spread along the interface generated by the NBD dimers (Fig. 7C). This analysis suggests a maintenance of a profile in these proteins that might be useful for sharing of the NBDs among the different systems (Fig. 7B and C). Indeed, the position of residues of NBDs that interact with the coupling helices in our models, is exactly the same in the dimer structure (Fig. 7C, red

residues in the cyan and deep cyan surfaces, which represent each monomer of NBD in the dimeric structure).

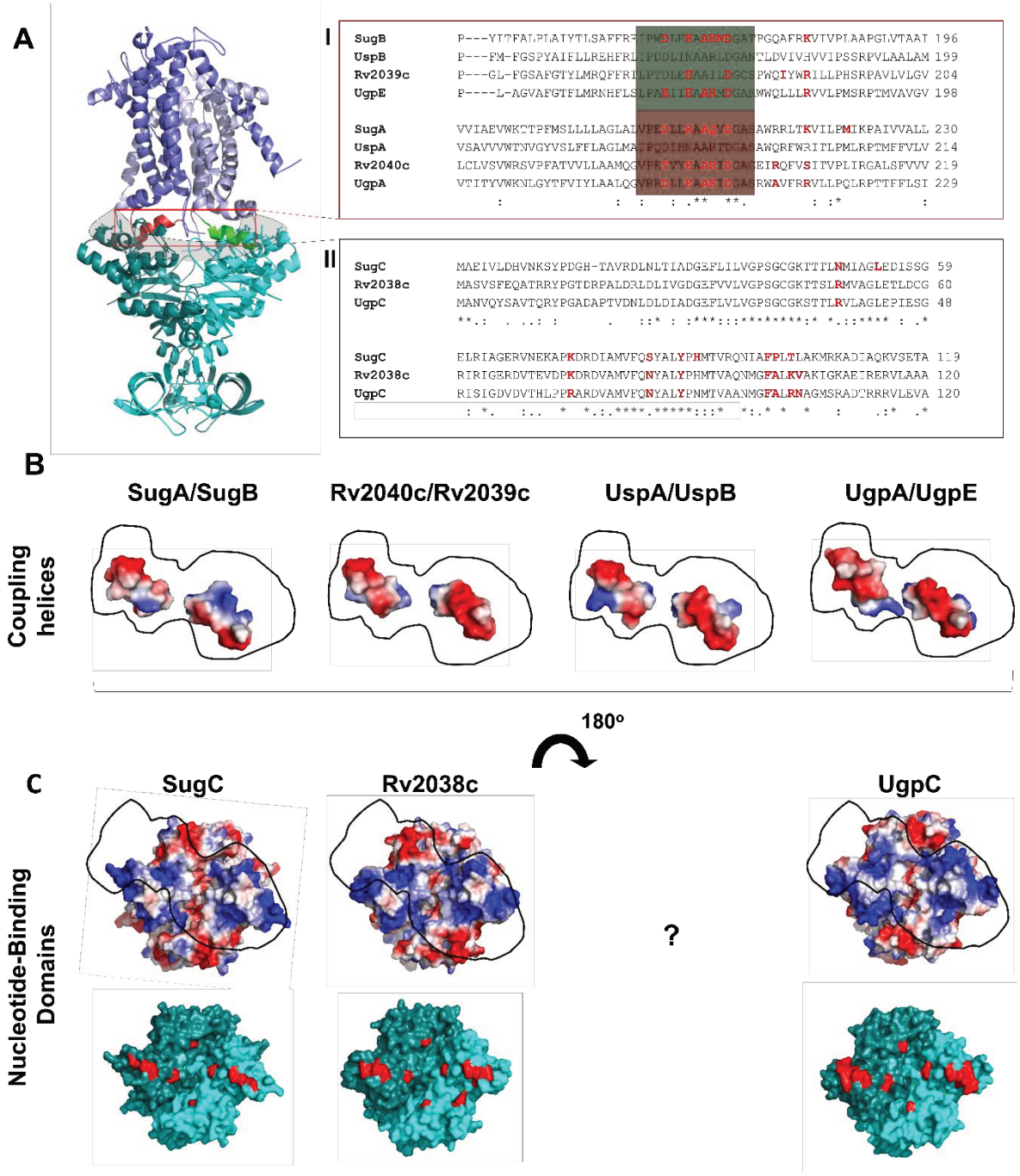


Figure 7. The interface between coupling helices and NBDs of *Mycobacterium tuberculosis* carbohydrate ABC transporters. A Structural model of a carbohydrate transporter showing the dimers of TMDs (pale blue and blue) and NBDs (cyan and deep cyan). The two coupling helices from each monomer are coloured in green and red, respectively, and the NBD region that face the helices is highlighted in a grey box. (I) and (II) Local amino acid sequence alignments of TMDs and NBDs, respectively. Coupling helices are coloured as in the structural model. **B** Electrostatic potential of TMDs coupling helices and NBDs showed as surface. The area in black line highlights the position of the interaction. Blue: positive, red: negative, grey: neutral. **C** Structure of the NBD dimers in surface showing the two monomers and the regions (area in black line) that might interact with coupling helices.

Discussion

The cellular and molecular mechanisms involved in the *M. tuberculosis* nutrition remain largely unknown. Although ABC transporters have a clear contribution to functions that aid colonization of the host environment, such as nutrient scavenging, and evasion or resistance to host defences, these roles rarely are associated to pathogenesis or virulence, in contrast to the role of ABC transporters in eukaryotic diseases [41]. In this way, the study of ABC transporters in carbohydrate uptake across different *Mycobacterium* species (pathogenic and non-pathogenic), and comparisons of these transporters in a functional and structural context could shed light about the relevance of these systems for the pathogenic species. Here, we show that carbohydrate ABC transporters could have a strong role as pathogenesis determinants of *M. tuberculosis*, and the comparison of these systems from an *in silico* structural point of view could reveal the main differences between the components, that also could shed light about their evolution.

According to the co-occurrence of carbohydrate ABC transporter components from *M. tuberculosis* across different species, we show that *Nocardia brasiliensis*, an actinobacteria that also cause a pulmonary disease, shows conservation of Rv2038c-41c and UgpAEBC systems. This fact could reflect the importance of these systems or similar systems for pathogenic species. Nevertheless, *N. brasiliensis* encodes 516 ABC transporter proteins compared to < 100 of *M. tuberculosis*. In this regard, *N. brasiliensis* more resembles a soil bacterium than a pathogenic bacterium [42]. On the other hand, *Rodococcus fascians*, a phytopathogen bacteria, show more conservation of LpqY/SugABC and UgpAEBC. It was showed that infection of planta with *R. fascians* elicits an accumulation of the disaccharide trehalose, after four days post-infection. However, this level strongly decreased infected tissue after 24 days post-infection [43]. Furthermore, it was showed that *R. erythropolys*, an oligotrophic bacterium, that can survive in a completely inorganic medium with no additional carbon source, has a similar trehalose-recycling system as in *Mycobacterium* LpqY/SugABC [44]. The results also evidence that species of *Corynebacterium* has low conservation of carbohydrate ABC systems, especially *C. diphtheriae*, that show complete absence of SBP components. All these together could reflect the diverse lifestyle of Actinobacteria species.

The results revealed that all species of *Mycobacterium* genus used in this study, conserved at least orthologs of two ABC transporters, LpqY/SugABC and UspABC, which is in accordance with previous studies that also demonstrated the presence of these two transporters in *M. smegmatis* [11]. The genomic comparisons between these two species

demonstrated that *M. smegmatis* 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. Unfortunately, most of substrates of the four ABC systems present in *M. tuberculosis* are unknown. These differences reflect the lifestyles of *M. smegmatis* and *M. tuberculosis* in their natural habitats, the soil and human body, respectively [11]. Our results additionally showed that Rv2038c-41c and UgpAEBC are not only restricted to *M. tuberculosis* species, but they also are shared in other pathogenic species of *Mycobacterium* genus. The presence of the UgpAEBC system exclusively in species of *M. tuberculosis* complex group and *M. marinum* could reveal the relevance of this system for the bacilli. The fact could be explained by studies that compare the MTBC and *M. marinum* genomes and suggest genome downsizing through deletion of genes that are dispensable for a pathogenic lifestyle, combined with the acquisition of new genes through horizontal gene transfer (HGT) [45]. With a length of 4.4 Mb, the genome of the MTBC is smaller than the genomes of *M. marinum* (6.6 Mb) [35, 36]. It could be possible that the UgpAEBC transporter was indispensable throughout the evolution of MTBC. UgpB, the SBP component, is a substrate of twin-arginine translocation (Tat) pathway, conserved in different species [37, 38], and upregulated during infection and essential for virulence and survival in several pathogens [39, 40]. Biophysical assays showed its ability to bind GPC, which might be important carbon and phosphate sources inside the host [10, 16]. Indeed, NMR studies of the metabolomic profiling of intact lung tissue at various stages of *M. tuberculosis* infection has revealed a significant increasing of GPC during the start of infection but decreasing with its progress [50]. Similarly, the components of Rv2038-41c system are conserved in *M. tuberculosis* complex group and other pathogenic mycobacteria. In the phylogenetic trees of the three components of the transporters, Rv2038-41c members are always closely related to UgpAEBC. Although the substrate of this transporter is not clear, the SBP Rv2041c role is related to intracellular adaptation within the host and relevant for the pathogen biology and virulence since it is upregulated in phagosome acidic and hypoxic conditions [18]. For all above, we can suggest the Rv2038c-41c and UgpAEBC transporters could be important in the uptake host substrate, and the fact that these transporters are only restricted to pathogenic species, could reflect that they are not dispensable for a pathogenic lifestyle of *M. tuberculosis* and other pathogenic species.

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, which represent 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 interface of permeases. All NBDs analysed conserve a catalytic subdomain, while the main differences are observed in the C-terminal regulatory region. Despite that NBDs segregate mostly according to sequence differences in a region that lies between Walker A and B motifs and includes the helical subdomain, structurally more diverse [51], it could be possible that NBDs containing regulatory subdomains have diverged based on differences in this region. The regulatory subdomain of several carbohydrate NBDs conserved short sequence motifs as FVAxFIGSP, GψRPE, and ExxG, where ψ is an apolar residue and x is any amino acid. This subdomain also conserved C-terminal G and F residues that can serve as signatures [52]. All these motifs and residues are conserved in *Escherichia coli* MalK and UgpC, *Thermococcus litoralis* MalK, and others [52], 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 clusters together depicting the high sequence conservation. Interestingly, these proteins branched out together with homologous proteins from *Pseudomonas syringae* and *Klebsiella pneumoniae,* which are a plant and an animal pathogen respectively [26].

The analysis of the TMDs revealed that differences noticed in both groups are related to regions that perform interaction with SBPs and NBDs suggesting that the permeases, specially that from group 2, could determine the specificity for substrate and/or the communication pathway of SBP to NBD. This is in accordance with Oldham and collaborators (2007) that suggested that the residues responsible for maltose binding in the TM subunits are found exclusively in MalF, not in the second TM subunit, MalG [40]. According to the phylogenetic analysis, it is shown that *M. tuberculosis* carbohydrate TMDs could be diverge from an ancient SugB for group 1, and SugA for group 2. Very little is known about evolutionary relationships about ABC TMDs, however comparative studies of ABC importers TMDs show that TMDs containing six TMs (transmembrane segments), could be originated from a primordial protein containing 3 TMs that suffered an intragenic duplication event [53]. 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 pore, emphaticising the similar character of the substrates.

The periplasmic components showed the highest diversity, as expected. Since they are responsible for the affinity and specificity of the transport, our results suggest that each *M. tuberculosis* carbohydrate transporter has different set of substrates. On the contrary of NBD

and TMD components, the phylogenetic analysis resulted 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 be not necessarily associated to the binding site, but also other regions of the protein that have distinct roles in the transport system and in the bacterium physiology. The structural similarity of the SBPs with carbohydrate-binding proteins and their general characteristics of the pockets that evidence the distribution of solvent-accessible aromatic side chains in the binding cleft and the characteristic acidic molecular surface corroborate their role as carbohydrate transporters. Previous work proposed that *M. tuberculosis* may switches the capacity to acquire carbohydrates and lipids, essential feature of survival of *M. tuberculosis* in macrophages [3, 44]. Carbon and phosphate can be acquired by the action of bacillus phospholipases and glycerophosphodiesterases on host phospholipids. The proximity of the periplasmic proteins to cell-wall carbohydrate could facilitate the acquisition of substrates like GPC, trehalose and chitobiose by UgpB, LpqY and UspC, respectively [10, 14, 17].

Structural details shared by all SBPs components of carbohydrate ABC transporters were analysed. UspC and UgpB proteins exhibit a topology of subcluster DI in the structural classification of substrate-binding proteins [55]. The structural models of LpqY and Rv2041c seem to have the same topology, which is supported by the molecular weights above 40 kDa, also a characteristic of subcluster DI proteins. Curiously, none of the carbohydrate-binding proteins from *M. tuberculosis* seem to belong to the cluster B, which have been classically associated to carbohydrate binding. The substrate-binding sites of LpqY, Rv2041c, UspC and UgpB were compared and despite an apparent conservation in structure of these proteins, the electrostatic potential of substrate cleft does not reflect this conservation. It is evident that these proteins have affinity by different substrates. However, the comparison, in terms of amino acid residues, shows conservation of charge of some residues in substrate cleft of all four structures: a hydrophilic or charged residue (Ser, Asn or Asp) in all proteins, an aromatic residue (Trp) only conserved in Rv2041c and UgpB. On the other hand, LpqY and UspC seem to have other residues in this position (undetermined for LpqY and Gln for UspC), which is a similar arrangement found in maltose-binding proteins [46, 47]. According to our analysis, using SignalP-5.0, we could suggest that LpqY and UspC proteins follow a SEC pathway, in contrast to Rv2041c that follow a TAT pathway. The Tat pathway is important to the virulence and physiology of several bacterial pathogens. Because it is absent in mammals, it has been considered a target for development of new antibacterial agents [58].

The structural prediction of LpqY show that the protein shared several characteristics with other trehalose and/or maltose-binding proteins. Despite the amount of genetic and cellular studies about the role of LpqY/SugABC transporter in the recycling of trehalose, very little is known about the molecular specificity by substrate. We used *T. litoralis* TMBP (PDB code: 1EU8) and *T. maritima* MalE3 (PDB code: 6DTQ), respectively solved in bound forms with trehalose and maltose, to compare the substrate-binding pocket with LpqY. Although the proteins showed low amino acid sequence identity, the three-dimensional structural model of LpqY showed conservation in the residues of the substrate-binding pocket. The comparison showed that at least seven residues involved with trehalose and maltose coordination are conserved in LpqY. The participation of the residues of LpqY in the trehalose coordination was corroborated upon docking analysis. On the contrary to the LpqY binding pocket prediction, little can be inferred about Rv2041c protein, due to an apparent lack of conservation with other proteins.

Finally, our study showed that exist conservation in the coupling helices and NBDs from carbohydrate transporters of *M. tuberculosis.* Although residues of the coupling helices of UgpAE, UspAB and Rv2039/40c are not identical, they have a pattern of negative charges quite similar that fit on the positive residues of the NBDs. Interestingly, the heterodimeric TMDs SugAB and UspAB has a K or R residue in EAA motif instead of the classic glutamic acid. The role of glutamine in EAA was quite discussed due lateral chain form salt bridges. In this way, a lysine or arginine residue also can form salt bridges.

Conclusions

Altogether, the data presented in this work showed a comparative characterization of the four putative carbohydrate-transporters from *M. tuberculosis,* with phylogenetic, structural, and functional analyses that evidenced common features, but mainly important differences that strongly suggest they have different substrates. Moreover, we showed that the ABC transporters Rv2038-41c and UgpABC share high similarity with orthologs present in pathogenic species. The absence of these transporters in environmental species, which has not infection capacity and pathogenesis in human, may also reflect the importance of the carbohydrate uptake for the bacterium. In addition, the results presented here using *M. tuberculosis* and other relevant species consist of an important step for understanding the role of ABC transporters in global carbohydrate metabolism in these bacteria and arises perspectives for further experimental studies.

Methods

Searching for orthologs

The genes coding carbohydrate ABC transporters components of *M. tuberculosis* were obtained from Mycobrowser (https://mycobrowser.epfl.ch/) [59] or KEGG (https://www.genome.jp/kegg/) [60] servers. The co-occurrence analysis across different taxa was evaluated using String server (https://string-db.org/) [61]. The comparative genomic analysis across *Mycobacterium* species was described by Machowski et al. [62]. Briefly, we choose 15 reference genome sequences of *Mycobacterium* strains deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) (Table S1, Additional file 1), and searched for homologues 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/) [63].

Phylogenetic analysis

The alignments and phylogenetic analysis of ABC transporter components (NBD, TMD or SBP) were visualized and performed using MEGA-X [64]. 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 (*lnL*), and the number of parameters (including branch lengths) were calculated. Phylogeny was established from analysis of Maximum Likehood. 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 tree generated were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Molecular modelling

Molecular models of the ABC transporter components were generated using I-TASSER server [65] or Modeller program [66] using as templates the structural coordinates of proteins deposited in the Protein Data Bank (PDB) as described in detail in Table S2 (Additional file

1). A total of five models were generated, and the best models were selected according to the C-score and DOPE value for I-TASSER and Modeller programs, respectively. All models were subsequently validated for their stereo-chemical quality using the program MolProbity [67]. 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) [68].

SBP characterization and binding site prediction

Signal peptide sequences of the substrate-binding proteins were predicted using SignalP-5.0 Server [69]. The characterization of LpqY and Rv2041c substrate-binding sites was identified using the prediction of I-TASSER server in comparison with the structural superimposition of the respective templates (Table S3, Additional file 1). Substrate-binding pockets volume and surfaces of the periplasmic proteins were calculated using the program CASTp [70] with a default probe radius of 1.4 A and MetaPocket [71]. The crystallographic structures of UgpB (PDB code: 4MFI) and UspC (PDB code: 5K2X) were obtained from Protein Data Bank. For prediction of LpqY and Rv2041c substrates affinities, docking analysis were carried out using AutoDock 4.2 [72]. The set of compounds used were collected from PubChem (https://pubchem.ncbi.nlm.nih.gov/). The 2D-to-3D conversion was carried out using the Online SMILES Translator and Structure File Generator (https://cactus.nci.nih.gov/translate/) from The CADD Group's Chemoinformatics Tools and User Services. For docking calculations, Gasteiger charges were added, rotatable bonds were set by AutoDock tools (ADT) and all torsions allowed were chosen to rotate for the ligand. The docking grids were generated with default settings and centred at selected coordinates in the binding pocket. Each grid box was big enough to cover the entire binding pocket. The Genetic Algorithm parameter was selected, and docking runs were performed using Autodock Vina [38].

Characterization of the TMDs and prediction of interfaces with NBDs

To support the TMDs structural models, transmembrane helices, and determination of the helices in contact with NBDs (coupling helices) were predicted using TMHMM Server v. 2.0 [73] and TOPCONS [74]. In addition, amino acid sequence alignments of TMDs and NBDs were performed including as reference the sequences MalG/F and MalK from *E. coli* maltose ABC transporter, respectively. Amino acid sequences of TMDs and NBDs were submitted to Gremlin (http://gremlin.bakerlab.org/) [75] for coevolution-based residue-residue contact predictions. When using Gremlin, for regions >60 residues the e-value threshold was set to E-

06 and the number of iterations with Jackhmmer 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).

List of 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

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data generated and analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and COLCIENCIAS (PhD fellowship for LID and SCB).

Author´s contributions

LID performed all analysis and wrote all sections of the manuscript. JGVM and AGCM performed the phylogenetic analysis. SCB performed molecular modelling. AB supervised, wrote, reviewed, and edited the manuscript.

Acknowledgments

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

References

1. World Health Organization. Global tuberculosis report 2019. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.
2. Timm J, Post FA, Bekker LG, Walther GB, Wainwright HC, Manganelli 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.
3. Niderweis M. Nutrient acquisition by mycobacteria. Microbiology. 2008;154:3.
4. 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.
5. Jeckelmann JM, Erni B. Transporters of glucose and other carbohydrates in bacteria. Pflugers Arch Eur J Physiol. Pflügers Archiv - European Journal of Physiology; 2020;742.
6. 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.
7. 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.
8. 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.
9. Pandey AK, Sassetti CM. Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci U S A. 2008;105:11.
10. 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.
11. 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.
12. Ter Beek J, Guskov A, Slotboom DJ. Structural diversity of ABC transporters. J Gen Physiol. 2014;143:4.
13. Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. [erratum appears in FEMS Microbiol Rev 2002 Mar;26(1):109]. FEMS Microbiol Rev. 2000;24:4.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Kim SY, Shin AR, Lee BS, Kim HJ, Jeon BY, Cho SN, et al. Characterization of immune responses to *Mycobacterium tuberculosis* Rv2041c protein. J Bacteriol Virol. 2009;39:3.
20. 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.
21. 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.
22. 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.
23. Rengarajan J, Bloom BR, Rubin EJ. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. 2005;102:23.
24. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. 2003;100:22.
25. Kruh NA, Troudt J, Izzo A, Prenni J, Dobos KM. Portrait of a pathogen: The *Mycobacterium tuberculosis* proteome In vivo. PLoS One. 2010;5:11.

26. Sabharwal N, Varshney K, Rath PP, Gourinath S, Das U. Biochemical and biophysical characterization of nucleotide binding domain of Trehalose transporter from *Mycobacterium tuberculosis*. Int J Biol Macromol. Elsevier B.V.; 2020;152:1.
27. Wolber JM, Urbanek BL, Meints LM, Piligian BF, Lopez-Casillas IC, Zochowski KM, et al. The trehalose-specific transporter LpqY-SugABC is required for antimicrobial and anti-biofilm activity of trehalose analogues in *Mycobacterium smegmatis*. Carbohydr Res. Elsevier Ltd; 2017;450:10.
28. Parish T, Smith DA, Roberts G, Betts J, Stoker NG, Parish T. The senX3–regX3 two-component regulatory system of Mycobacterium tuberculosis is required for virulence. Microbiology. 2003;149.6.
29. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol. 2003;48:1.
30. Yang Z, Rannala B. Molecular phylogenetics: Principles and practice. Nat Rev Genet. Nature Publishing Group; 2012;13.
31. 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.
32. Vetter IR, Wittinghofer A. Nucleoside triphosphate-binding proteins: Different scaffolds to achieve phosphoryl transfer. Q Rev Biophys. 1999;32:1.
33. 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.
34. Gerber S, Comellas-Bigler M, Goetz BA, Locher KP. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. Science. 2008;321:5886.
35. Thomas C, Tampé R. Structural and Mechanistic Principles of ABC Transporters. Annu Rev Biochem. 2020;89.
36. Chandravanshi M, Samanta R, Kanaujia SP. Conformational Trapping of a β-Glucosides-Binding Protein Unveils the Selective Two-Step Ligand-Binding Mechanism of ABC Importers. J Mol Biol. Elsevier Ltd; 2020;432:20.
37. 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;57:40.
38. Trott,O., Olson AJ. Autodock vina: improving the speed and accuracy of docking. J Comput Chem. 2019;31:2.
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 [Internet]. Nature Publishing Group; 2016;2.
40. Oldham ML, Khare D, Quiocho FA, Davidson AL, Chen J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature. 2007;450.
41. Lewis VG, Ween MP, McDevitt CA. The role of ATP-binding cassette transporters in bacterial pathogenicity. Protoplasma. 2012;249.
42. 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.
43. 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.
44. Matsuoka T, Yoshida N. Functional analysis of putative transporters involved in oligotrophic growth of *Rhodococcus erythropolis* N9T-4. Appl Microbiol Biotechnol. Applied Microbiology and Biotechnology; 2019;103.
45. Veyrier FJ, Dufort A, Behr MA. The rise and fall of the *Mycobacterium tuberculosis* genome. Trends Microbiol. 2011;19:4.
46. Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, Johnson PDR, et al. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. Genome Res. 2008;18:5.
47. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol. Nature Publishing Group; 2018;16.
48. 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.
49. Chandravanshi M, Gogoi P, Kanaujia SP. Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. Gene. Elsevier B.V.; 2016;592:2.
50. Somashekar BS, Amin AG, Rithner CD, Troudt 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.
51. 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.
52. Holland IB, Cole SPC, Kuchler K, Higgins CF. ABC Proteins: From Bacteria to Man. Elsevier Science; 2003; 157-185
53. 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.
54. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A. 2003;100:22.
55. Scheepers GH, Lycklama a Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. FEBS Lett. 2016;590:23.
56. Mächtel R, Narducci A, Griffith DA, Cordes T, Orelle C. An integrated transport mechanism of the maltose ABC importer. Res Microbiol. 2019;170:8.
57. Medrano FJ, De Souza CS, Romero A, Balan A. Structure determination of a sugar-binding protein from the phytopathogenic bacterium *Xanthomonas citri*. Acta Cryst. 2014;70.
58. Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. Annu Rev Microbiol. 2006;60.
59. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis. Elsevier Ltd; 2011;91:1.
60. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:1.
61. 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.
62. Machowski EE, Senzani S, Ealand C, Kana BD. Comparative genomics for *mycobacterial peptidoglycan* remodelling enzymes reveals extensive genetic multiplicity. BMC Microbiol. BMC Microbiology; 2014;14.
63. 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.
64. 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.
65. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9.
66. Sali A, Blundell T. Sali, A. & Blundell, T. L. Comparative modelling by satisfaction of spatial restraints. J. Mol. Biol. 1993;234:3.
67. 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.
68. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0. 2015.
69. 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.
70. 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.
71. Huang B. Metapocket: A meta approach to improve protein ligand binding site prediction. Omi A J Integr Biol. 2009;13:4.
72. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J. Computational Chemistry. 2009;30:16.
73. 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.
74. 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.
75. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. Elife. 2014;3.

RESULTADOS COMPLEMENTARES – CAPÍTULO I: A proteína UgpC de *M. tuberculosis*

Introdução

Os NBDs dos transportadores ABC são um subgrupo da superfamília P-*loop* NTPases (Vetter & Wittinghofer, 1999), que dependem de íons de magnésio para a catálise. Cada NBD possui um núcleo de ~200 aminoácidos e consiste em dois subdomínios principais: o subdomínio maior RecA-*like* também encontrado em outras *P-loop* ATPases, e o subdomínio α-helicoidal estruturalmente mais diverso, exclusivo dos transportadores ABC. No entanto, os NBDs dos transportadores ABC de carboidratos diferem dos NBDs canônicos, possuindo uma extensão C-terminal de aproximadamente 120 a 150 resíduos de aminoácidos. No transportador ABC de maltose de *E. coli* e *Salmonella typhimurium*, o subdomínio C-terminal está envolvido em atividades regulatórias (Shuman, 1998). Vários motivos de sequência curta e resíduos de aminoácidos conservados nesse fragmento, e um motivo conservado na SBP (Schneider, 2001), servem para identificar novos membros da família CUT (*carbohydrate uptake transporter*). Os domínios ABC da família CUT1 são funcionalmente promíscuos e podem atuar em diferentes transportadores funcionalmente relacionados, fortalecendo assim a classificação descrita acima. Por exemplo, foi demonstrado que UgpC de *E. coli* e LacK de *Agrobacterium radiobacter* substituem MalK no transporte de maltose em *E. coli* (Wilken, 1996; Wuttge et al., 2012).

Neste sentido, esta parte do doutorado foi dedicada à caracterização bioquímica e biofísica da proteína UgpC de *M. tuberculosis,* produção de anticorpos anti-UgpC e análise da filogenia e estrutura da proteína. Mostramos que UgpC forma um clado diferente, quando analisada com proteínas homólogas. A UgpC recombinante e produzida em *E. coli* apresentou um espectro de dicroísmo circular, com maior conteúdo de fitas beta e se mostrou funcional, porém com baixa atividade de ATPase. A proteína apresentou baixa imunogenicidade quando inoculada em ratos, evidenciada por um título de anticorpos de 1:6400.

Materiais e métodos

1. Análises filogenéticas e modelagem estrutural

A sequência de aminoácidos da proteína UgpC de *M. tuberculosis* foi obtida na base de dados Mycobrowser (https://mycobrowser.epfl.ch/) (Kapopoulou et al., 2011). As sequências de aminoácidos de possíveis ortólogos da UgpC foram recuperadas da UniProtKB *Reviewed* (Swiss-Prot). Todas as sequências foram alinhadas usando Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). A filogenia foi estabelecida usando o método de Máxima Probabilidade, e o modelo de substituição de aminoácidos Le-Gascuel (LG) (Le & Gascuel, 2001), usando MEGA-X (Kumar et al., 2018). A robustez das árvores inferidas foi testada por análises de *bootstrap* (1000 repetições). As árvores geradas foram visualizadas usando FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Modelos estruturais da proteína UgpC foram gerados usando o programa Modeller (Sali & Blundell, 1994). As coordenadas estruturais da proteína MalK de *Thermococcus litoralis* (Código PDB: 1G29) (Diederichs, 2000) foram usadas como molde. O modelo final foi escolhido com base nos parâmetros geométricos, usando o servidor MolProbity (V. B. Chen et al., 2010).

2. Clonagem do gene *ugpC*

O fragmento correspondente ao gene de interesse foi amplificado a partir do DNA genômico de *M. tuberculosis* H37Rv, *Trudeau Mycobacterial Collection* (TMC) e *American Type Culture Collection* (ATCC). Os oligonucleotídeos utilizados são descritos na tabela 1. As PCRs foram realizadas utilizando-se a DNA polimerase de alta fidelidade Phusion® (New England Biolabs, UK). A confirmação da amplificação foi feita usando eletroforese em gel de agarose.

Tabela 1. Oligonucleotídeos utilizados para a clonagem do gene *ugpC* de *M. tuberculosis*.

Gene*	Oligonucleotídeos	Sequência	Tamanho (pb)
full-ugpC	Forward_Rv2832c_(*Nco*I)	5' TATATCCATGGCTAACGTTCAATACTCTGC 3'	1083
	Reverse_Rv2832c_full_(*Hind*III)	5' ATATAAGCTTAGCGAAGCCGGGTCTCGGTGCGG 3'	
short-ugpC	Forward_Rv2832c_ (*Nco*I)	5' TATATCCATGGCTAACGTTCAATACTCTGC 3'	657
	Reverse_Rv2832c_short_(*Hind*III)	5' ATATAAGCTTACGTGTCAACCTGTTGCAGCACAC 3'	

*Visando melhorar a solubilidade da proteína resultante, foram desenhados dois conjuntos de oligonucleotídeos para o gene *ugpC* que correspondem a variações no tamanho do gene, sendo *full* o gene completo, e *short* o gene sem a região que codifica o domínio regulatório predito a partir do modelo estrutural. Os oligonucleotídeos

possuem sítios de corte para as enzimas de restrição indicadas entre parênteses. Os sítios de corte são compatíveis com o vetor pHAT2 e os vetores da série pOP (Addgene).

Os insertos resultantes foram purificados usando o kit QIAquick PCR Purification (QIAGEN) para remoção de oligonucleotídeos e nucleotídeos excedentes. Os vetores de expressão foram extraídos a partir de células *E. coli* DH5α com o kit QIAprep® Spin Miniprep (QIAGEN). A dupla digestão dos insertos e vetores foi realizada com as respectivas enzimas de restrição (New England Biolabs). Uma vez digeridos os insertos e vetores, estes foram extraídos a partir de gel agarose com o kit QIAquick® GelExtraction Kit (QIAGEN). A ligação foi realizada com o kit Quick Ligation (New England Biolabs). Células *E. coli* DH5α competentes foram transformadas com os produtos de ligação resultantes. A confirmação dos clones foi realizada com auxílio de PCRs a partir de colônias e análises com enzimas de restrição. As construções obtidas foram enviadas para sequenciamento de DNA.

3. Expressão e purificação

Os vetores de expressão contendo os genes *ugpC*, pHAT2-*full ugpC* e pHAT2-*short ugpC* foram transformados na cepa BL21 (DE3) de *E. coli*. A partir das colônias obtidas foi feito um pre-inóculo de 10 ml em meio LB líquido contendo 100 μg/ml de ampicilina, e crescido a 37°C *overnight*. Os pre-inóculos foram inoculados em 1 L de meio LB contendo o mesmo antibiótico para a expressão das proteínas de interesse. As células cresceram a 30°C, 200 rpm até atingir uma OD_{600} entre 0.8 – 1.0. A temperatura foi diminuída para 25°C e as células foram induzidas com 0.1 mM IPTG durante 15 horas. Após a expressão, as culturas foram centrifugadas a 4°C, 6000 rpm durante 20 minutos. O sobrenadante foi descartado e as células foram ressuspensas em 30 ml de tampão de lise (50 mM Tris-HCl pH 8.0 [*full* UgpC] e pH 7.0 [*short* UgpC] 500 mM NaCl, 5 mM $MgCl_2$), 3 mg de lisozima e 3 mM PMSF por litro de cultura. Uma vez que a proteína UgpC não foi expressa na forma solúvel, os passos seguintes consistiram na lavagem do *pellet* obtido na lise celular: primeira lavagem com tampão de lise contendo 1% Triton X-100, segunda lavagem com tampão de lise contendo 1 M NaCl, e terceira lavagem com tampão de lise. Depois das lavagens, os *pellets* foram solubilizados com tampão de lise contendo 8 M ureia, agitando a 4°C durante toda a noite. Em seguida, os *pellets* solubilizados foram centrifugados a 4°C, 16.000 rpm durante 1 hora, e o sobrenadante obtido foi filtrado em membranas de 0.4 μM. A partir do sobrenadante foi feita uma cromatografia de afinidade de níquel em condição desnaturante. As proteínas foram eluídas com 100 mM de imidazol, e as frações resultantes foram reunidas e diluídas em 50 mM Tris-

HCl contendo 0.5 mM de arginina. Depois da diluição, as proteínas foram concentradas em filtros Amicon 30 kDa ou 10 kDa (*Thermo Scientific*), com constante troca de tampão 50 mM Tris-HCl. A proteínas concentradas foram submetidas a dialises extensivas em tampão 50 mM Tris-HCl.

4. Dicroísmo circular (CD) e análise de fluorescência intrínseca dos triptofanos

Para acessar informações sobre a estrutura secundária e verificar o enovelamento correto da proteína *full* UgpC, foram realizados experimentos de CD. Foi usado um espectropolarímetro Jasco J-810 (Jasco International Co., Tóquio, Japão) equipado com um módulo Peltier para o controle de temperatura. As medições foram realizadas em uma cubeta de quartzo com 1 mm de caminho óptico, usando um comprimento de onda de 190-260 nm. A concentração de proteína foi ajustada para 2.5 μM em 5 mM Tris-HCl, pH 8.0. Foram realizadas 20 acumulações de 10 segundos cada. Os dados do CD foram analisados para deconvolução usando a Ferramenta de Análise e Plotagem de CD CAPITO (Wiedemann et al., 2013). A elipticidade é mostrada como a elipticidade residual média [θ] (deg cm^2 $dmol^{-1}$). Os espectros de fluorescência foram obtidos usando um fluorímetro Varian Cary Eclipse, Santa Clara, CA. As amostras foram colocadas em cubetas de quartzo com 4 mm de caminho óptico. O experimento foi realizado à temperatura ambiente (22.5ºC). A concentração de proteína foi ajustada para 5 μM em 50 mM Tris-HCl, pH 8.0. Os espectros de emissão foram obtidos usando um feixe de excitação a 295 nm. Todos os dados foram analisados usando o software de gráficos e análise de dados OriginPro 9.0 (OriginLab Corporation).

5. Ensaio de atividade

A atividade ATPase da proteína *full* UgpC foi determinada usando o kit ATPase/GTPase Activity Assay Kit (Merk), que quantifica a liberação de fosfato inorgânico no ensaio de ATPase. A solução de ensaio foi misturada com uma amostra que continha 1 mg/ml de proteína purificada em 10 μl de tampão 50 mM Tris, pH 8.0. A mistura foi incubada à temperatura ambiente por 20 minutos e a reação foi detida adicionando 200 μl da solução Reagente. O produto colorimétrico foi determinado medindo a absorbância a 620 nm usando um leitor de microplacas Spectramax i3x (Molecular devices Inc). A absorbância do ensaio foi diretamente proporcional à quantidade de Pi liberada e, portanto, à atividade da ATPase.

6. ELISA e *Western Blot*

Os anticorpos específicos anti-UgpC (IgG) presentes nas amostras de soro de camundongos imunizados previamente com a proteína UgpC foram titulados por ensaio imunoabsorbente ligado a enzima (ELISA), conforme descrito anteriormente (Amorim et al., 2010). Resumidamente, uma microplaca de poliestireno Maxisorp Nunc (Sigma-Aldrich, EUA) foi revestida com a proteína UgpC recombinante purificada (2 ng/poço). A placa foi lavada 3 vezes com 1X PBS contendo 0.05% de Tween 20 (PBST) e bloqueada com 1X PBST contendo 3% de leite desnatado e 0.5% de BSA por 2 h a 37°C. Após um novo ciclo de lavagem, as amostras de soro de camundongo foram adicionadas e diluídas em série (log2) a partir de 1:50 na placa, em seguida incubadas em temperatura ambiente por 2 h. Após um novo ciclo de lavagem, o anticorpo IgG anti-*mouse*, conjugado à peroxidase (Sigma-Aldrich), foi adicionado a placa e incubada novamente por 2 h. Após uma lavagem final, as placas foram reveladas com tampão citrato de sódio (pH 5.8) contendo dicloridrato de orto-fenilenodiamina (Sigma-Aldrich) e H_2O_2, e a reação foi interrompida após 20 minutos com a adição de 50 µl de 1M H_2SO_4. A leitura da absorbância foi realizada a 492 nm com um leitor de placas Labsystems Multiscan (ThermoScientific), e usada para calcular os títulos de anticorpos, definidos como a maior diluição do soro capaz de gerar um A_{492nm} de 0.2 acima do soro pré-imune. Para as análises de *Western Blot*, a proteína UgpC foi submetida a SDS-PAGE 12%, transferida para uma membrana PVDF (Merck, EUA) e depois sondada com uma amostra de soro de camundongo (diluição 1: 7.000) usada como anticorpo primário seguindo o protocolo padrão (Sambrook, n.d.). Utilizou-se IgG anti-*mouse*, conjugado a peroxidase (Merk, Alemanha) como anticorpo secundário (diluição 1:15,000), seguido de detecção colorimétrica usando comprimidos prontos para uso NBT/BCIP como substratos (Merck).

7. Predição de epítopos de células B e mapeamento

A previsão do epítopo das células B foi realizada usando o servidor Antibody Epitope Prediction tool (http://tools.iedb.org/bcell/). O método BepiPred-2.0: Sequential B-Cell Epitope Predictor foi selecionado (Jespersen et al., 2017). Para a visualização dos epítopos no modelo estrutural da proteína UgpC, foi utilizado o software PyMOL (The PyMOL Molecular Graphics System, Versão 2.0 Schrodinger, LLC).

8. Declarações de ética

Os procedimentos para manipulação e eutanásia de camundongos foram realizados de acordo com as diretrizes e regulamentos relevantes e foram aprovados pelo Comitê de Ética em Experimentação Animal (CEUA) (protocolo número 050/2014), Instituto de Ciências Biomédicas da Universidade de São Paulo, seguindo normas do Conselho Nacional de Controle de Experimentação Animal (CONCEA, Ministério de Ciência e Tecnologia, Brasil).

Resultados e discussão

Análise filogenética

Uma análise filogenética foi realizada para comparar a proteína UgpC respeito a possíveis proteínas ortólogas em outras espécies de bactéria (Figura 1). A árvore mostra que a UgpC forma um clado com proteínas de espécies da mesma classe (Actinobacteria), diferente de proteínas UgpC de espécies de bactérias Gram-positivas e Gram-negativas. É interessante que a proteína Rv2038c de *M. tuberculosis* é agrupada no mesmo clado, corroborando assim análises prévias (manuscrito 1). Os nossos dados mostram também que os putativos ortólogos de UgpC em Actinobacteria são organizados no mesmo grupo de *M. tuberculosis,* e não com os ortólogos de bactérias gram positivas, contrariando a classificação tradicional de espécies da classe Actinobacteria dentro de Grampositivas. Por outro lado, a proteína SugC foi agrupada com os ortólogos presentes em *Rhodococcus opacus, Treponema palidus, Thermus thermophilus* e *Clostridium acetobutylicum* (Figura 1). O alinhamento de todas as sequências de proteínas identificadas nas diferentes espécies mostrou que as principais diferenças entre elas se encontram no extremo C-terminal da proteína, conhecido como subdomínio regulatório, indicando que a classificação dessas proteínas está relacionada à função das mesmas e possíveis moléculas reguladoras.

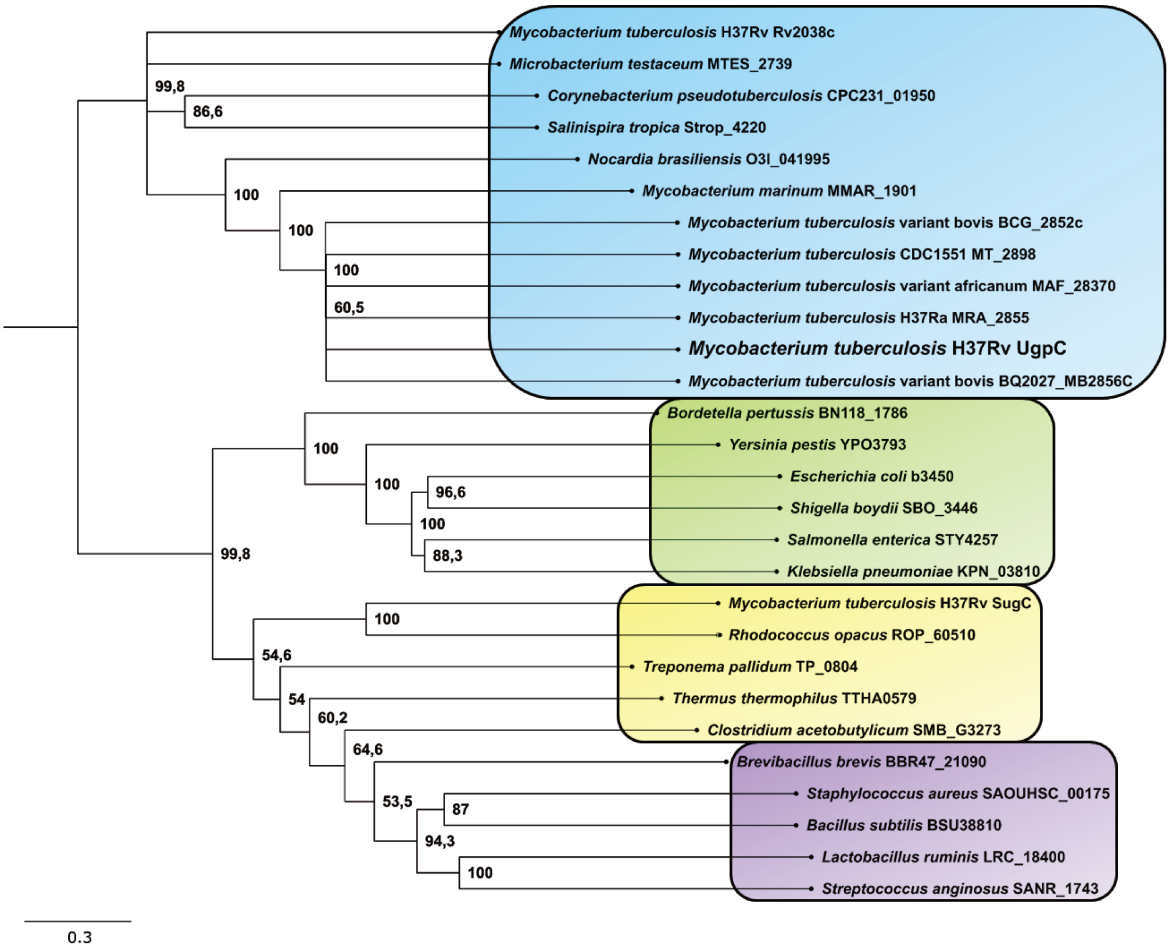


Figura 1. Arvore filogenética das proteínas UgpC de diferentes espécies de bactéria. A arvore inclui espécies da classe Actinobacteria (em azul), grupo das Gram-negativas (verde), grupo de bactéria diverso (amarelo) e grupo das Gram-positivas (roxo).

Para acessar informações sobre a proteína UgpC no contexto de proteínas ABC de açúcares, foi realizado um alinhamento com proteínas ABC já caracterizadas envolvidas no transporte de maltose e sn-glicerol-3-fosfato de *E. coli*, e trealose em *T. litoralis* (Figura 2). É possível observar que a proteína UgpC compartilha motivos já descritos para outras proteínas de açúcares, o que confirma a sua função no transporte dessas moléculas em *M. tuberculosis*. Até o momento não é conhecida a função destes motivos na proteína, porém pelo mapeamento destes motivos no modelo estrutural de UgpC, é possível sugerir participação na organização estrutural, e não envolvida na ligação a uma proteína regulatória o um açúcar repressor.

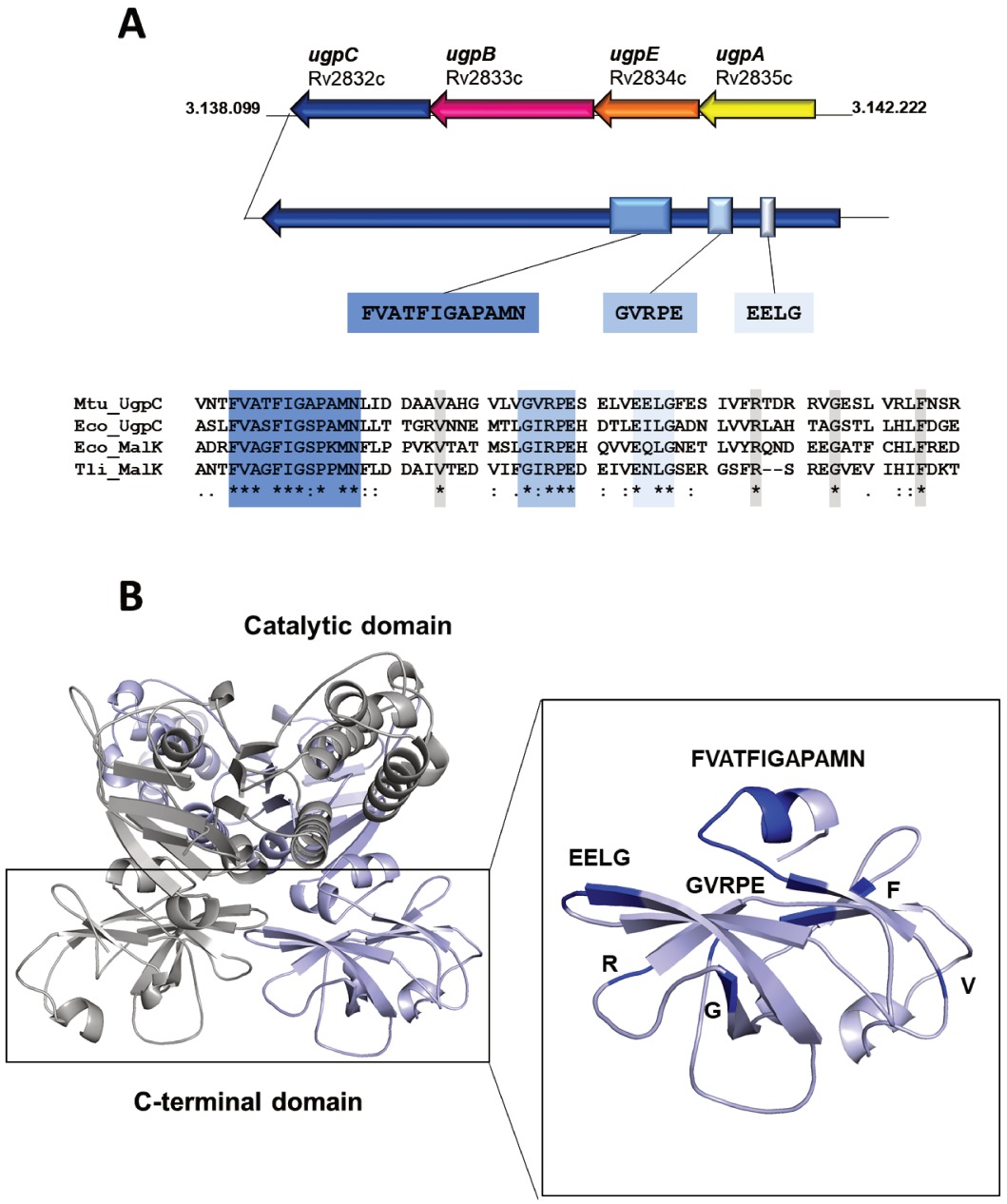


Figura 2. Características funcionais presentes na estrutura da proteína UgpC de *M. tuberculosis*. A Organização genômica do gene *ugpC* e motivos característicos conservados na proteína (em caixas de diferentes tons de azul) em comparação com NBDs de sistemas ABC envolvidos no transporte de açúcares. **B** Modelo da proteína UgpC com dois monômeros (em azul e cinza, respectivamente), cada um com o domínio catalítico maior e regulatório menor. Em detalhe, o domínio regulatório e seus motivos representados em azul escuro.

Testes de expressão e solubilidade

Foram realizadas diferentes construções com diferentes fragmentos do gene, um codificando só o subdomínio catalítico (*short* UgpC), e outro codificando a proteína completa (subdomínio catalítico + subdomínio regulatório, *full* UgpC), além de diferentes vetores de expressão, sem e com proteínas de fusão altamente solúveis, visando obter uma a proteína de

interesse na fase solúvel. Porém, todas as nossas tentativas resultaram em proteínas insolúveis (Tabela 2).

Tabela 2. Testes de expressão e solubilidade da proteína UgpC de *M. tuberculosis*. As diferentes condições testadas são apresentadas, bem como os resultados para cada uma. A maioria das tentativas resultaram em proteína na fase insolúvel.

Construção	Célula	Temperatura (°C)	Horas de indução	Não expresso	Insolúvel	Solúvel
pHAT2-full *ugpC*	BL21 (DE3)	37	3		■	
		25	8		■	
		18	20		■	
	LEMO21 (DE3)	37	3		■	
		25	8		■	
		18	20		■	
	BL21 (DE3) pGRO7	37	3		■	
		25	8		■	
		18	20		■	
	BL21 (DE3) pG-KJE8	37	3		■	
		25	8		■	
		18	20		■	
	ARCTIC EXPRESS (DE3)	12	20		■	
pOP1-full *ugpC*	BL21 (DE3)	37	3	■		
		25	8	■		
		18	20	■		
	LEMO21 (DE3)	37	3	■		
		25	8	■		
		18	20	■		
pHAT2-short *ugpC*	BL21 (DE3)	37	3		■	
		18	20		■	
pOP5GT-full *ugpC*	BL21 (DE3)	37	3		■	
	ARCTIC EXPRESS (DE3)	12	20		■	
pOP5GT-short *ugpC*	BL21 (DE3)	37	3		■	
	ARCTIC EXPRESS (DE3)	12	20		■	
pOP3SU-full *ugpC*	BL21 (DE3)	37	3		■	
		18	20		■	
pOP3SU-short *ugpC*	BL21(DE3)	37	3		■	
		18	20		■	

Expressão de UgpC insolúvel, desnaturação, re-enovelamento e purificação

A proteína foi purificada a partir do *pellet* de culturas da célula *E. coli* BL21 (DE3) portadora do plasmídeo pHAT2-*full* ou *short ugpC* com rendimento de ~1 mg por litro. As

amostras tratadas com 8 M de uréia ou 6M de guanidina foram purificadas por cromatografia de afinidade ao níquel, e eluída com 100 mM de imidazol (Figura 3). As frações foram reunidas e diluídas em tampão de re-enovelamento contendo 50 mM Tris-HCl pH 8.0 (*full* UgpC) ou pH 7.0 (*short* UgpC) e 0.5 M de arginina. Amostras da proteína foram concentradas em filtro Amicon 30 kDa (*full*) ou 10 kDa (*short*), e posteriormente submetidas à diálise extensiva em tampão 50 mM Tris-HCl.

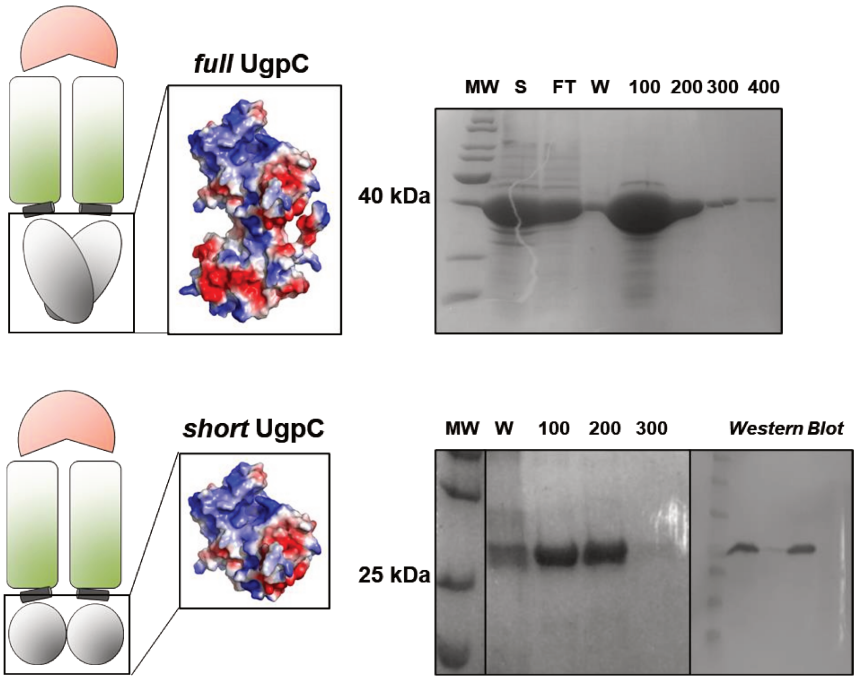


Figura 3. Purificação das proteína *full* UgpC e *short* UgpC. Na parte esquerda, diferentes construções da proteína UgpC: *full* UgpC (domínio catalítico + domínio regulatório) e *short* UgpC (domínio catalítico). Na parte direita é mostrada a eluição das proteínas através de um gradiente de imidazol que ocorreu na fração de 100 mM.

Dicroísmo circular e Fluorescência Intrínseca do Triptofano

O espectro de CD da *full* UgpC apresentou um perfil característico de proteínas α/β, porém com um pico negativo proeminente na região de 218 nm, indicando um maior conteúdo de fitas β, conforme observado nos resultados de deconvolução dos dados (Figura 4). A deconvolução dos dados foi feita usando a ferramenta Capito, e mostrou um conteúdo de 6% hélices α, 54% folhas β, e 49% *ramdon coil*. Estudos sobre domínios ABC de transportadores ABC de açúcares, já mostraram que elas têm um conteúdo maior de hélices α, o oposto do que fora obtido para UgpC, indicando que a mesma pode não ter sido completamente re-enovelada. A proteína também apresentou um espectro de fluorescência com pico máximo de 980 (unidades arbitrárias) no comprimento de onda de 349 nm, revelando que os triptofanos se apresentam em regiões mais expostas ao solvente (Figura 4C e

D). O anterior coincide com o modelo molecular, no qual os dois únicos triptofanos (Trp279 e Trp316) são localizados no extremo C-terminal, um subdomínio da proteína pouco estruturado.

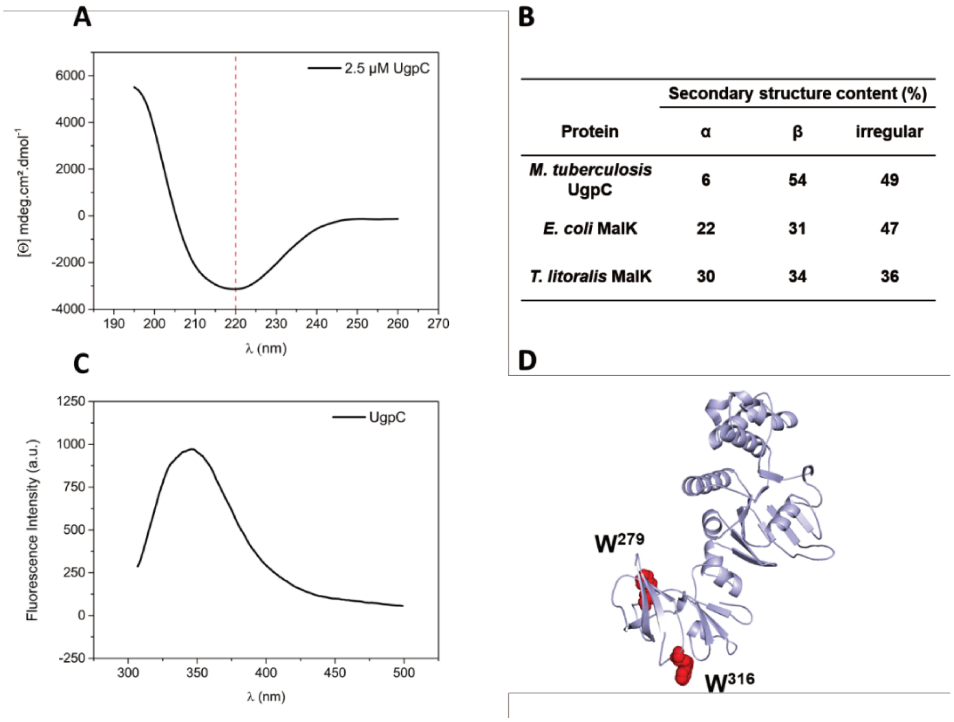


Figura 4. Ensaios biofísicos da proteína *full* UgpC de *M. tuberculosis*. A Dicroísmo circular da proteína UgpC em concentração de 2.5 µM **B** Comparação da estrutura secundaria da proteína UgpC e outras proteínas ABC envolvidas no transporte de açúcares. C Fluorescência intrínseca do triptofano da proteína UgpC. D Mapeamento dos triptofanos na proteína UgpC.

Atividade ATPase

A proteína UgpC *full* de *M. tuberculosis* mostrou uma baixa atividade enzimática quando comparada com outras proteínas relacionadas (Tabela 3), o que pode ser explicado, em parte, pelo não enovelamento completo, conforme evidenciado nos dados de CD.

Tabela 3. Atividade enzimática obtida para *M. tuberculosis* UgpC em comparação com outros NBDs previamente caracterizados na literatura. O reagente verde malaquita interage com o fosfato livre liberado pela enzima resultando em um produto colorimétrico, medido a 620 nm, proporcional à atividade enzimática presente.

Domínio de ligação ao nucleotídeo	Atividade enzimática	Referência
***M. tuberculosis full* UgpC**	0,0083 µmoles/min/mg	Este trabalho
***M. tuberculosis* SugC**	3,78 µmoles/min	(Sabharwal et al., 2020)
***E. coli* MalK**	0,10 µmoles/min/mg	(Chen et al., 2003)

***T. litoralis* MalK**	0,55 µmoles/min/mg	(Diederichs, 2000)

Determinantes da imunogenicidade

Com o objetivo de testar o soro anti-UgpC, um ensaio de ELISA foi realizado para determinar o título de anticorpo, cuja diluição máxima para reconhecimento da proteína foi de 1:6400. Para analisar os possíveis determinantes da imunogenicidade da proteína UgpC, foi realizado um ensaio de Western Blot, com a proteína completa (*full* UgpC), e somente com o domínio catalítico (*short* UgpC). Os resultados mostraram que somente a proteína *full* UgpC foi reconhecida. É interessante ressaltar que o domínio catalítico de proteínas ABC é bastante conservado na arvore evolutiva, lembrando-se que todos os domínios da vida (Bactéria, Archaea e Eukarya) possuem transportadores ABC. Por outro lado, os domínios regulatórios aparecem somente em espécies de bactérias e alguns transportadores ABC específicos. A nossa análise, permite sugerir que os epítopos que estimularam a resposta imune em camundongos, estavam majoritariamente presentes na região regulatória da proteína. Para avaliar a nossa hipótese, fizemos uma predição utilizando a ferramenta BepiPred-2.0, que permite fazer um *screening* de epítopos de células B lineares. Um total de 19 epítopos foram preditos, tanto na região catalítica como na região regulatória, no entanto, epítopos mais longos foram observados na região regulatória. Estudos desenvolvidos por Schneider e colaboradores (2002) mostraram o uso de anticorpos monoclonais como ferramentas para a caracterização funcional de proteínas citoplasmáticas de transportadores ABC de açúcares (Stein et al., 2002). Este resultado é interessante, pois anticorpos contra essas regiões poderiam caracterizar uma nova abordagem na busca de estratégias de bloqueio da atividade de transportadores, como uso dos mesmos como biomarcadores de diagnóstico.

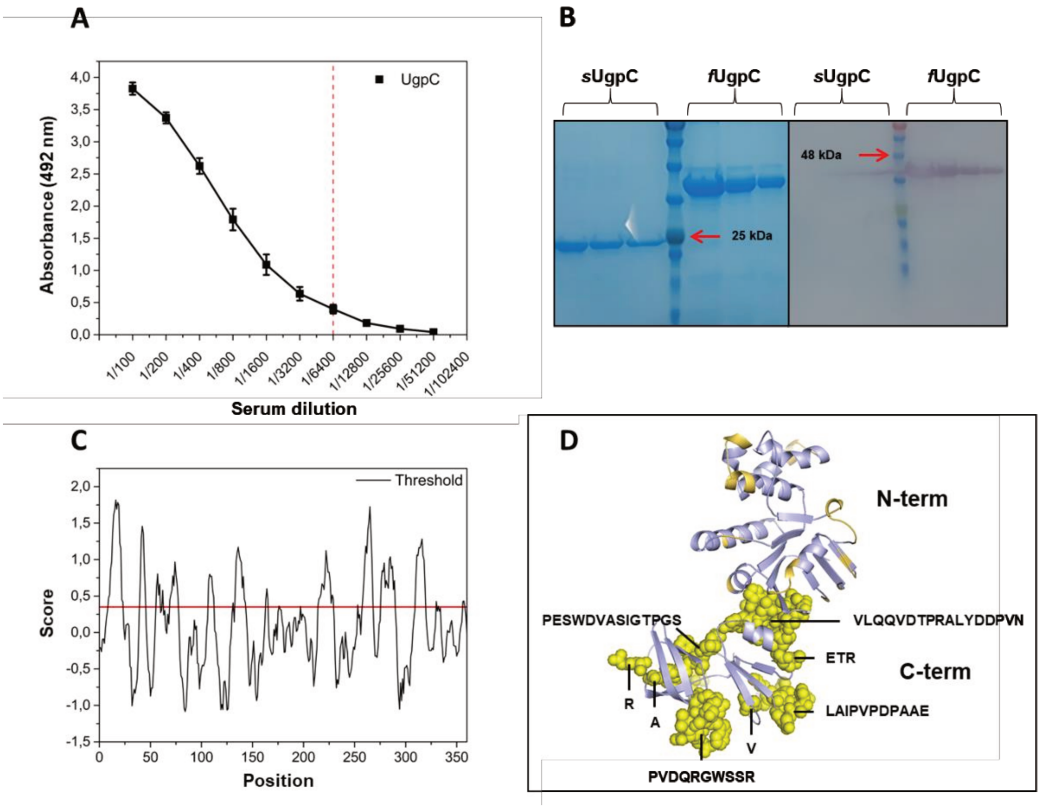


Figura 5. Análises do potencial imunogênico da proteína *full* UgpC. A Reatividade da proteína UgpC recombinante com amostras de soro de ratos previamente inoculados com a proteína. **B** Western Blot das proteínas *full* UgpC e *short* UgpC. Na figura são apresentados o gel resultante da eletroforese de poliacrilamida (SDS-PAGE) e o resultado do *Western Blot* após a transferência das proteínas para a membrana de PVDF. **C** Probabilidade da presença de epítopos lineares na proteína UgpC, em função da sequência de aminoácidos, de acordo com o programa BepiPred-2.0. **D** Mapeamento dos possíveis epítopos preditos para a proteína UgpC de *M. tuberculosis*, realçados em amarelo. Os epítopos que se encontram na região regulatória são apresentados em esferas.

Conclusões

Os nossos resultados permitem concluir que apesar da proteína UgpC de *M. tuberculosis* ser de natureza citoplasmática, ela possui regiões hidrofóbicas que podem interagir com a membrana, e podem ser responsáveis pela insolubilidade da proteína quando expressa de forma recombinante na cepa *E. coli* BL21 (DE3). Além disso, tanto as tentativas de solubilização da proteína completa (*full*) ou sem o domínio regulatório (*short*) foram mal sucedidas, corroborando nossa hipótese e indicando que a região hidrofóbica deve se localizar no domínio catalítico. Infelizmente, mesmo após a obtenção de uma proteína a partir de protocolos de re-enovelamento, não foi possível realizarmos ensaios estruturais ou funcionais, pois a mesma se apresentou bastante instável, polidispersa e/ou sujeita à agregação. Ainda assim, o espectro de dicroísmo circular sugere que a proteína resultante foi parcialmente enovelada o que foi confirmado pelos ensaios de ATPase que mostraram baixa atividade. A

predição de vários epítopos na UgpC foi contrastada com o baixo título encontrado nos ensaios de ELISA. É possível sugerir que os epítopos poderiam ser estruturais e comprometidos pelo fato de a proteína não estar completamente enovelada. Ainda, os resultados de *western blot* mostraram que o anticorpo só reconheceu a construção *full* indicando e que a região regulatória é de fato, importante para a imunogenicidade. Foi importante conseguirmos produzir anticorpos anti-UgpC que serão usados em futuros projetos visando compreender melhor o papel da UgpC em *M. tuberculosis*.

CAPÍTULO II

A new strategy to study interactions between ABC transporter coupling helices and nucleotide-binding domains

Lilia I. De la Torre[1,2], Gabriel S. Vignoli Muniz[3], André L. Sehnem[4], M. Teresa Lamy[3], Cristiano Luís Pinto Oliveira[4], Andrea Balan[1,2]

[1]Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil
[2]Institute of Biology, State University of Campinas, São Paulo, Brazil
[4]General Physics Department, Institute of Physics, University of São Paulo, São Paulo, Brazil
[3]Complex Fluids Group, Experimental Physics Department, Institute of Physics, University of São Paulo, São Paulo, Brazil

Correspondence:
Prof. Andrea Balan
Laboratory of Applied Structural Biology, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil.
E-mail: abalan@usp.br
Phone: +55 11 37917745

Abstract

The ATP-Binding Cassette transporters depend on the Nucleotide-Binding Domains (NBDs or ABC domains) that bind and hydrolyse ATP to supply energy for the transport across the membrane. They are closely associated to the Transmembrane Domains (TMDs) through the so-called coupling helices, a short sequence in the TMDs, containing the conserved EAA motif that are essential for triggering transport. The block of this interaction between the TMDs and NBDs could be a model for inhibition of ABC transporters still not explored. In this sense, the aim of this study was to evaluate the in vitro interaction of mimetic coupling helices peptides with NBDs by biophysical techniques. For such, we used as a proof of concept, the NBD MalK and coupling helices mimetic peptides of MalG and MalF of the well

characterized maltose ABC transporter ($MalEFGK_2$). Using Differential Scanning Fluorimetry (DSF) and Small Angle X-ray Scattering (SAXS) analyses we showed that MalK undergo thermal destabilization and conformational changes in presence of both coupling helices peptides, suggesting an interaction. Moreover, thermophoresis experiments were suitable to measure the NBD and peptides interaction and showed that MalK binds both MalF and MalG coupling helices mimetic peptides with K_d of 47.8 and 20.9 μmol L^{-1}, respectively. On the other hand, the use of intrinsic tryptophan fluorescence technique was not suitable for evaluation of MalK structural changes upon peptides addition and measuring of the interaction between MalK and the peptides. This work opens opportunities for the study of the interaction between TMDs and NBDs in ABC transporters aiming at the development of transport inhibitors and blocking the activity of essential transporters in pathogenic microorganisms.

Introduction

ATP-Binding Cassette (ABC) transporters constitute one of the largest superfamilies of membrane proteins that are responsible for the ATP powered translocation of many substrates, ranging from ions to macromolecules, across cell membranes. The highly conserved ABC domains, also referred to as nucleotide-binding domains (NBDs), provide the nucleotide-dependent engine that drives transport [1]. All ABC transporters share a core modular architecture, which consists of two highly conserved nucleotide-binding domains (NBDs), the unifying hallmark of ABC systems, and two variable transmembrane domains (TMDs) that form the pore translocation pathway [2]. For prokaryotic ABC transporters type importers, substrate translocation is also dependent on another protein component, a substrate-binding protein (SBP) that capture the ligand in the periplasm for delivery to the proper ABC transporter [3]. Prokaryotic canonical ABC importers play important roles in the maintenance of cell integrity, responses to environmental stresses, cell-to-cell communication, and cell differentiation and in pathogenicity [4].

The ABC domains or NBDs bind and hydrolyse ATP to supply energy for the transport across the membrane. The NBDs are a subgroup of the diverse superfamily of P-loop NTPases [5] and depend on magnesium ions for catalysis. Each NBD has a core of approximately 200 amino acids and consists of two subdomains: the larger RecA-type domain, which is also found in other P-loop ATPases, and the structurally more diverse α-helical domain, which is

unique to ABC transporters [2]. The NBDs are characterized by the following sequence motifs and structural elements: the Walker A (or P-loop), Walker B, the A-, D-, and Q-loop, and the H-switch, positioned within the RecA-type ATP-binding core, and the signature motif (LSGGQ), positioned within α-helical domain [6]. The D-loop and Q-loop has a potential role NBD–TMD communication [5, 6]. NBD dimerization is a prerequisite of ATP hydrolysis. Important motifs in transmitting the NBD dimerization state to the TMDs are the so-called coupling helices, short helices in the TMDs that interact with the NBDs [8]. These coupling helices are the only part of the TMD-NBD interface that is structurally conserved between the many different TMD folds. Some coupling helices contain a conserved sequence (EAA motif) [2], mainly in ABC transporters type importers. The coupling helices are embedded in a groove on the NBD surface, similar to a ball-and-socket joint [8], and this groove lies at the ATP-sensitive interface between the RecA-type ATP-binding core and the α-helical domain (near to the Q-loop) [9]. In this way, once the NBDs hydrolyze ATP, each NBD connected to a TMD undergo conformational changes than can be transduced to conformational changes in the TMDs. This process is called alternating access mechanism.

Protein–protein interactions (PPI) are integral to most biological functions. Targeting these interactions with small molecule inhibitors is of increased interest both in academia as well as in the pharmaceutical industry, both for therapeutic purposes and in the search for chemical tools for basic science [10]. NBD-TMD interaction in canonical ABC importers is a protein-protein interaction underexplored, however the role of this could be crucial for the survival of pathogenic microorganisms. Once the interaction triggers conformational changes on the full transporter, the internalization of substrates of great importance for the microorganism survival occurs. Development of strategies with the aim of NBD-TMD inhibition could be of great interest in blocking of ABC transporters of pathogenic microorganism, and also in blocking of human ABC exporters involved in drug resistance and cancer. Although there are cellular, genetic, and structural studies about the role of coupling helices in ABC transporters, approaches to the use of mimetic peptides of coupling helices interacting with NBDs have been little explored. This strategy could be a start point for the use of fragment-based drug discovery and rapid screening of inhibitors of NBD-TMD interactions.

The *Escherichia coli* maltose/maltodextrin ABC transporter type importer $MalEFGK_2$ is one of the best studied ABC transporters to date and could be a great model in the study NBD-TMD and others protein-protein interactions. In this multi-subunit complex, MalE is the SBP, MalF and MalG are the TMDs, and $MalK_2$ is the NBD homodimer. Structures of isolated

MalK NBDs from *E. coli* maltose/maltodextrin ABC transporter ($MalEFGK_2$) have been resolved in different dimeric conformations [6, 7], demonstrating the different conformational states that NBDs adopt in response to different nucleotide binding conditions. Furthermore, the crystallographic structures of the full transporter showed that TMDs interact NBDs using the conserved motif coupling helices, containing the EAA motif [13–17]. In both MalF and MalG TMDs, the glutamate residue in the EAA motif is engaged in a salt-bridge interaction with an arginine residue in MalK NBD, located just after the Walker A motif.

In this work, we use biophysical techniques to track MalK protein changes in presence of MalF and MalG coupling helices peptides. We showed that MalK binds both MalF and MalG coupling helices peptides with a K_d of 47.8 and 20.9 μmol L^{-1}, respectively. Further MalK undergo thermal destabilization in presence of $MgCl_2$ and ATP, and MalF coupling helices peptide. SAXS analysis showed that MalK undergoes conformational changes in presence of both peptides. Intrinsic Fluorescence of Tryptophan analysis showed that the fluorescence of the only two tryptophan in MalK do not change in presence of both coupling helices peptides. Altogether, these results provide a basis for further structural and functional studies of NBDs in interaction with TMD coupling helices, without the need of full transporter analysis. New perspectives were opened by unprecedent investigation of the use of mimetic peptides of TMD coupling helices as elements that could block NBD-TMD interaction, which could contribute to new strategies of blocking protein-protein interaction in these transporters or similar systems, widely distributed in pathogenic bacteria.

1. Materials and methods

1.1. Peptide design and GREMLIN analysis

The amino acid sequence of *E. coli* MalK, MalG and MalG was obtained from the KEGG database (https://www.genome.jp/kegg/genes.html). The MalF and MalG coupling helices were identified using the $MalEFGK_2$ ABC transporter three-dimensional structure (PDB code: 2R6G) [16], and TMHMM server v. 2.0. Gremlin method [18] was used to predict coevolution-based residue-residue contacts. The amino acid sequences of NBD and cytoplasmic regions of TMDs (previously predicted in TMHMM server) were submitted. We accepted interprotein residue pairs with a scaled score ≥1.30 and a probability >0.88 as co-varying pairs; evolutionary couplings (ECs). All figures were generated using the program

PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) [19].

1.2. Chemical synthesis of MalF and MalG coupling helices peptides

The amino acid sequences of coupling helices peptides are shown in Table 1. The peptides were acquired from Biomatik services (Ontario, Canada) and dissolved in acetonitrile for MalFch_peptide or water for MalGch_peptide, according to manufacturer instructions. The purity of the peptides was higher than 95%, as attested by HPLC (Supplementary material).

Table 1. MalF and MalG coupling helices peptides from *E. coli* maltose/maltodextrines ABC transporter MalEFGK2.

Peptide	Amino acid sequence	Molecular mass ($gmol^{-1}$)	Net charge (pH 8.0)
Eco_MalFch_peptide	DDLYEASAMD	1129.10	-4.68
Eco_MalGch_peptide	SSLEEAAALD	1005.00	-3.93

1.3. Protein purification

The *malK* gene was cloned in the pHAT2 vector (kindly donated by Dr. Marko Hyvönen, Addgene catalog number 112583). The MalK fused with N-terminal His-tag was produced as a recombinant protein in *Escherichia coli* BL21 (DE3) and purified as previously describe by Chen et al. [11], with some modifications. Briefly, BL21 (DE3) cells were grown in LB media and 20 μg/ml ampicillin, to log phase and induced at 15 °C by 0.1 mM IPTG for 20 hours. Cells were harvested and broken by sonication in lysis buffer (25 mM Tris pH 8.0, 125 mM NaCl, 5% Glycerol, 5 mM β-mercaptoethanol). After centrifugation (100,000 x g, 60 min), the supernatant was loaded onto a nickel column (HisTrap HP colum, Cytiva) preequilibrated in lysis buffer. After a 10-column volume (CV) wash with lysis buffer followed by 5 CV wash with lysis buffer plus 5 mM imidazole, the protein was eluted using 100 mM imidazole. The fractions containing MalK protein were pooled, concentrated in an Ultra-15 Centrifugal Filter Unit Amicon® (Merck, Mollipore) and further purified on a HiLoad 16/600 Superdex 200 prep grade (Cytiva) size-exclusion in lysis buffer, as a polishing step. Protein purity was tested by 12% SDS-PAGE. Pure protein concentration in work buffer (25 mM Tris-HCl pH 8.0, 125 mM NaCl and 5% glycerol) was quantified using the theoretical extinction coefficient and molecular mass of 40.99 kDa and 20.06 $M^{-1} \cdot cm^{-1}$, respectively.

1.4. Differential scanning fluorescence (DSF)

Interactions of MalF and MalG coupling helices peptides and MalK protein was evaluated using fluorescent thermal denaturation, based on the thermal shifts in melting points of denaturation curves of the MalK protein. The MalK protein samples were mixed with 5X SYPRO™ Orange Protein Gel Stain (ThermoFisher Scientific) for emission fluorescence (excitation and emission at 492 and 610 nm, respectively), which increases when bounded to hydrophobic regions of the protein that are exposed upon thermal denaturation. Increasing concentrations of coupling helices peptides was added to the MalK protein in work buffer (see protein purification, materials and methods) in a final concentration of 100 µM. Each sample was dispensed into 96-well microplates (Hard-Shell® 96-Well PCR plates and Microseal 'B' Adhesive Sealing Films, BIO-RAD). The assays were performed in CFX96 Touch Real-Time PCR Detection System (BIO-RAD). After 1 min of initial incubation at 25 °C, denaturation curves were performed from 25 to 100°C at 1°C min^{-1}. All experiments were repeated at least three times for reproducibility.

1.5. MicroScale Thermophoresis (MST) measurements

The His-tagged MalK protein was labelled with Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation and diluted in work buffer (see protein purification, materials and methods). The coupling helices peptides were added above the concentration of MalK protein in serial dilution. The maximal concentration for coupling helices peptides MalFch and MalGch were 211 and 237 µM, respectively, and the standard concentration for His-tagged MalK protein was 50 nM. Samples were filled into Monolith NT.115 Capillaries (NanoTemper Technologies). The measurements were conducted on a NanoTemper Monolith NT.115 system (NanoTemper Technologies), at 80% power of Nano – RED excitation type for NTA dyed His-tagged MalK protein, and all at medium MST power. Cold Region Start/End were -1s and 0 s, respectively, and Hot Region Start/End were 14s and 15s, respectively. The data were plotted with the K_d fit model (equation 1), that describes a molecular interaction with a 1:1 stoichiometry according to the law of mass action.

Equation 1:

$$F_{norm} = F_1/F_0$$

The F_{norm} is calculated by dividing fluorescence values from the TRIC trace after the laser is turned on (hot region, F_1) by values that are obtained before the laser is turned on (the Initial Fluorescence or cold region, F_0).

For a detailed analysis, a data adjustment with equation 2, with two exponential terms was performed.

Equation 2:

$$F_{norm} = A_0 + A_1 \exp(-t/t_1) + A_2 \exp(-t/t_2)$$

The adjusted parameters are A_0, A_1, A_2, t_1 and t_2. The time constants t_1 and t_2 are inversely proportional to the diffusion constants D_1 and D_2, $t \propto 1/D$. Since D is inversely proportional to the hydrodynamic radius R as in the Stokes-Einstein equation $D = k_B T / 6\pi\eta R$, there is a linearity, $t_i \propto R_i$, $i = \{1, 2\}$. A_1 and A_2 are related to the Soret coefficient of thermophoresis, which still does not have a precise description for its value in complex systems, such as biological molecules dispersed in aqueous electrolytic medium.

1.6. Small-angle X-rays scattering

The experiments were performed in a Bruker-NANOSTAR SAXS equipment, located at the Institute of Physics of the University of São Paulo, equipped with a VÅNTEC-2000 area detector. We performed at least six independent measurements of 1800 s of exposure time for each sample. The samples were filled in 1.5 mm thick cylindrical quartz-glass capillaries supported by homemade stainless-steel cases. X-ray scattering from proteins were obtained after subtraction of contributions of the scattered intensity from the buffer, capillaries, and equipment background noise, using the SuperSAXS package (Oliveira and Pedersen, unpublished). Available for download from: http://stoa.usp.br/crislpo/files/). The results for the intensity I as a function of the momentum transfer, $q = (4\pi/\lambda) \sin \theta$, where λ is the radiation wavelength and 2θ the scattering angle, were obtained. The ATSAS programs package [20] was used to analyse the $I(q)$ results. Briefly, in a first step we verified if the SAXS results are described by the calculated scattering from the known crystallographic structure of the MalK protein (PDB ID: 2AWO). The state of oligomerization was evaluated with the programs CRYSOL [21] and OLIGOMER [22]. The Inverse Fourier Transform method was used to estimate the molecular mass of the protein [23] in each sample by using the program GNOM [24]. The 3D structure was evaluated with CORAL [25] and DAMAVER [26] programs.

1.7. Fluorescence measurements

UV-visible absorption spectroscopy measurements. UV-visible absorption spectra were obtained with an UV-visible spectrophotometer (Varian Cary, Santa Clara, CA).

Steady-State fluorescence measurements. The fluorescence spectra were obtained with fluorimeter (Varian Cary Eclipse, Santa Clara, CA). The experiments were conducted with 1 ml of MalK protein (5 μM) in work buffer at room temperature (22.5 ºC). Increasing concentration of MalFch and MalGch peptides were added to the protein sample from stock solutions (1 mM). Samples were placed in a quartz cuvette with an optical pathway of 4 mm. Emission spectra were obtained using an excitation beam light at 295 nm. No inner filter correction was necessary [27], as absorbance value at 295 nm was found to be smaller than 0.1.

Time-resolved fluorescence spectroscopy

Time-resolved (TR) fluorescence measurements were performed using time-correlated single photon counting technique (TCSPC). The excitation light beam comes from a titanium-sapphire Tsunami 3950 laser from Spectra Physics (Newport Corporation, Irvine, CA, USA), pumped by a solid-state laser Millenia Pro model J80 also from Spectra Physics. The frequency of pulse picker (Spectra Physics model 3980-25) was 8 MHz. The Tsunami was set to give an output of 852 nm and a third harmonic generator BBO crystal (GWN-23PL Spectra Physics) was used to generate the excitation light at 284 nm. The emission was detected at 90 degree from the excitation beam and selected by a monochromator. By using FAST software supplied by Edinburgh Photonics the data fit was used applying the model of exponential decays (Lakowicz 2006) using the following equations:

Equation 3:

$$F(\lambda, t) = \sum_{i=1}^{N} \alpha_i \, e^{-t/\tau_i}$$

Equation 4:

$$f_i = \frac{\alpha_i \tau_i}{\sum_j \alpha_j \tau_j}$$

where F(λ,t) is the number of photons emitted at a given wavelength (λ) and time, t is the time after the excitatory light beam, α_i is the pre-exponential factor, τ_i is the lifetime of the i^{th} component of the decay, and f_i is the fraction contribution of the lifetime τ_I to the intensity decay. We determined these latter factors from the best fitting processes which results from the statistical parameter reduced chi-square (χ^2), $0.95 \leq \chi^2 \leq 1.35$.

Fluorescence Quenching. The experiments were conducted with 1 ml of MalK protein (5 µM) in work buffer at room temperature (22.5 ºC). Increasing concentration of imidazole acrylamide was added to the protein sample from stock solution. It was used an excitation wavelength of 290 nm, and the emission spectrum recording between 300 - 400 nm was acquired. Each emission spectrum was subtracted the dilution. The data was displayed as a Stern-Volmer plot, I_0/I *versus* [acrylamide]. The Stern-Volmer equation was used to calculate the fractional tryptophan accessibility.

2. Results

2.1. The cytoplasmatic region and coupling helices of *E. coli* MalF and MalG TMDs

Once the maltose ABC transporter $MalEFGK_2$ is one of the most characterized ABC transporters until days, it is known that NBD-TMD interface consist of EAA motifs in TMDs and in a region close to Q-loop in NBDs [6]. We mapped these regions to design the mimetic peptides using the crystallographic structure of $MalEFGK_2$ transporter (PDB code: 2R6G) [16]. Also, we used TMHMM server to confirm the location of TMD coupling helices (Fig. 1). We noted that despite the cytoplasmatic region of TMDs is larger than coupling helices, the structured sequence are indeed the coupling helices. This led us to choose them as ligands for our study. The sequence alignment showed that the two amino acid sequence of MalF and MalG proteins conserve residues, however there are residues that are not fully conserved or not conserved at all as Tyr400 in MalF and Glu189 in MalG. These differences may be involved in how each TMD interacts with NBDs.

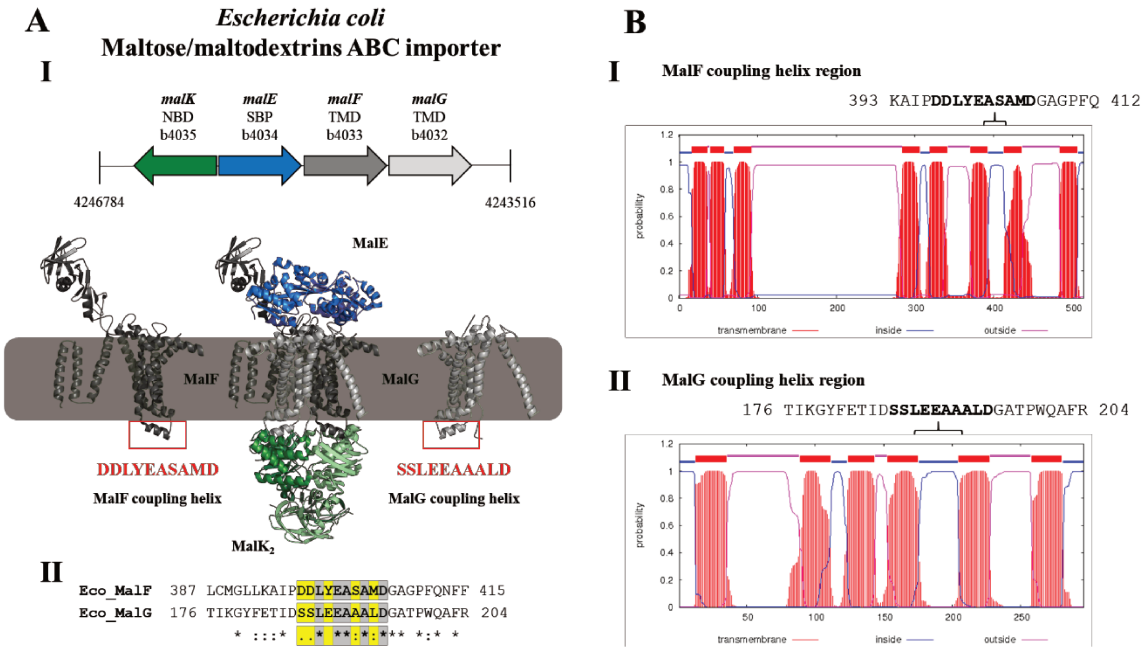
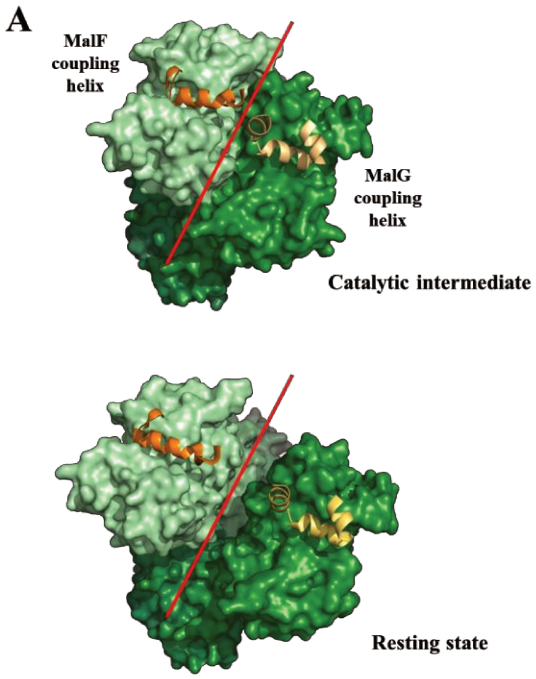


Figure 1. Identification of the coupling helices in *E. coli* MalF and MalG TMDs. A Operon and three-dimensional structure of the maltose ABC transporter from *E. coli* (PDB code: 2R6G), evidencing the positioning and sequence of coupling helices from MalF and MalG TMDs. MalE in blue cartoon, MalK dimer in clear and dark green cartoon, MalF and MalG in dark and clear gray, respectively (I). Alignment of MalF and MalG TMDs amino acid sequences, showing the conservation of cytoplasmatic region and coupling helices (II). **B** TMHMM server prediction of MalF (I) and MalG (II) TMDs, showing the putative cytoplasmatic regions of each TMD, which conserved the coupling helices.

2.2. The residue-residue interface of *E. coli* MalK NBD and MalFG coupling helices

Once the TMD coupling helices were identified, we explored the interaction region on NBD MalK that interacts with coupling helices using Gremlin method (Fig. 2) and crystallographic structures of *E. coli* MalEFGK$_2$ ABC transporter [13, 16]. Gremlin is a robust and accurate method that predict residue-residue interactions across protein interfaces using evolutionary information. We show that the residue-residue interaction on NBD-TMD association is conserved to the two TMD coupling helices, as expected. However, difference in residue-residue interaction of MalK with both coupling helices are shown in Fig. 2. The main differences are in the residues Asp398 and Met405, and Ala192 and Leu194, for MalF and MalG coupling helices, respectively.



B

ATPase	TMDs				
MalK	MalFch	MalGch	Scaled score	Prob	I_prob
Arg47	Glu401	Glu190	3.47/2.40	1.00/1.00	1.00/1.00
Leu52		Leu194	1.59	0.99	0.89
Pro71	Asp398		1.27	0.97	0.83
Ala73		Ala193	1.59	1.00	0.95
Val77	Met405	Leu194	1.24/1.25	0.98/0.96	0.82/0.82
Ser83	Asp398	Ser187	1.77/1.52	1.00/0.99	0.97/0.93
Tyr87	Asp406	Asp195	1.80/2.17	1.00/1.00	0.98/0.99
Phe98		Ala192	2.13	1.00	0.97
Gly99	Asp406	Asp195	1.30/1.22	0.99/0.96	0.86/0.80

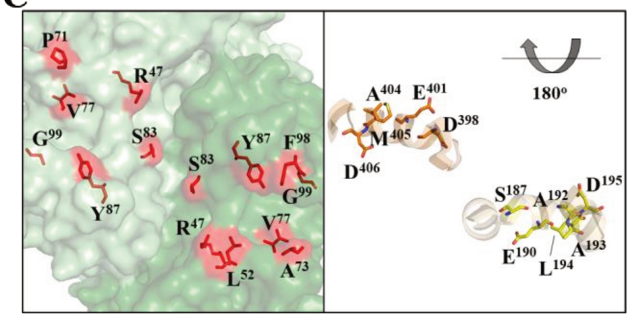


Figure 2. Interaction interface of MalK NBD with MalF and MalG coupling helices. A MalK crystallographic structure in two conformational states. Catalytic intermediate (PDB code: 2R6G), resting (PDB code: 3FH6). In both states the crystallographic structures showed an interaction interface symmetry (red line). **B** Gremlin method prediction of residue-residue interaction of MalK and MalFG coupling helices. Apparently MalK form more bindings with MalG than MalF coupling helix. **C** Mapping of residue-residue interaction of all proteins MalK, MalF and MalG in crystallographic structure (PDB code: 2R6G). MalK is showed in green and coupling helices of MalF and MalG are shown in orange and light orange, respectively. All proteins are shown in the same orientation.

2.3. The mimetic peptides of MalFG coupling helices MalFch and MalGch

The MalFch and MalGch coupling helices peptides were synthetized with a purity > 95%, according to HPLC analysis. The mass spectrometry results showed expected molecular weights of 1129.50 Da and 1005.70 Da for MalFch and MalGch, respectively (Fig 3).

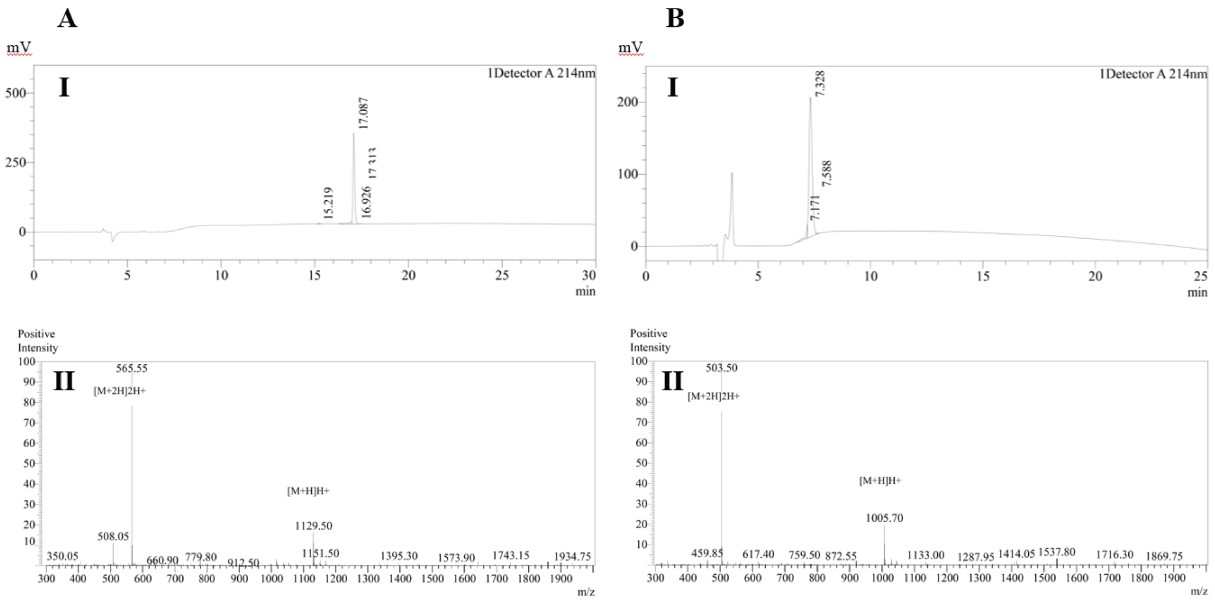


Figure 3. Purity and molecular weight assessment of synthetic coupling helices peptides MalFch and MalGch by RP-HPLC and mass spectrometry, respectively. A. MalFch exhibit a purity > 90% (I) and a M_W of 1129.50 Da (II). **B** MalGch exhibit a purity > 90% (I) and a M_W of 1005.70 Da (II).

2.4. *E. coli* MalK protein was expressed as a soluble and stable protein

To study the MalK NBD – MalF and MalG TMD coupling helices interaction, we produced recombinant MalK protein in *E. coli* cells. After induction of E. coli Bl21 (DE3) cells carrying the pHAT-*malK* plasmid, up to 8 mg/ml of soluble and stable MalK was obtained, with an expected monomeric molecular weight of 40.99 kDa (Fig. 4). The protein was purified using immobilized Nickel affinity chromatography and eluted with 100 mM imidazole. A subsequent step of polishing using size-exclusion chromatography revealed two peaks, the highest one consistent with a dimer ($MalK_2$).

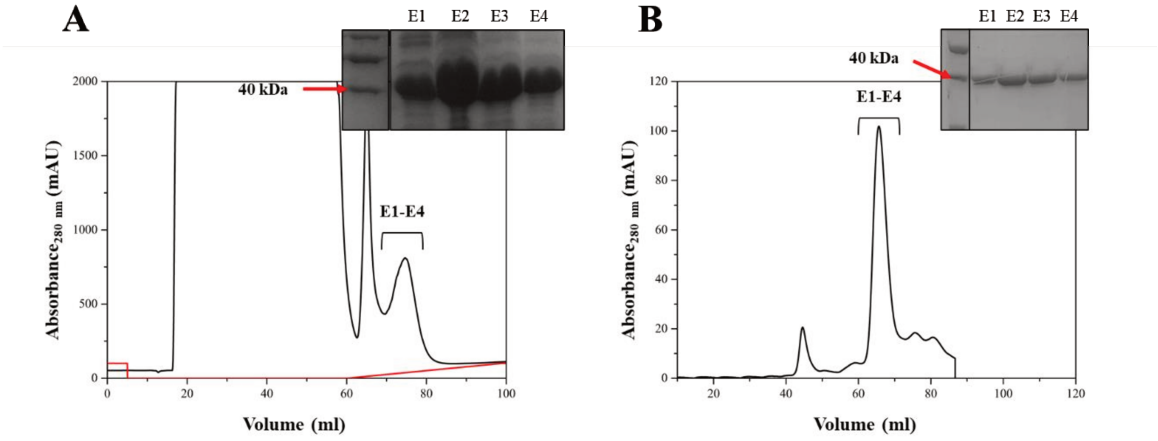


Figure 4. Overexpression and purification steps of *E. coli* MalK. A MalK nickel affinity chromatogram. The inset shows the SDS-PAGE 12% stained with Coomassie brilliant blue. E1-E4 = eluted fractions with 100 mM imidazole. **B** MalK size exclusion chromatogram. The inset shows the SDS-PAGE 12% stained with Coomassie brilliant blue. E1-E4, eluted fractions corresponding to the high peak.

2.5. MalFch and MalGch cause subtle changes on T_m of MalK

The DSF technique allows to track changes in thermal conformational stability of proteins after binding to small molecules (ligands), shifting the midpoint temperatures (T_m) of thermal denaturation curves [28]. The technique is based on the energetic coupling between ligand binding and protein unfolding. The ligand-binding affinity is estimated from the change of the unfolding transition temperature (ΔT_m) obtained in the presence of ligands relative to that obtained in the absence of ligands [29]. Displacements in MalK T_m induced by the TMD coupling helices appeared as feasible approach to identify those binding, allowing us to clarify some aspects of the interaction NBD-TMD using only peptides. Upon addition of $MgCl_2$ and ATP at 5 µM concentration to MalK NBD, 100 µM concentration of the two coupling helices peptides, MalFch and MalGch were added. These peptides caused a MalK-shift to lower values of Tm, ΔT_m of -0.6 and -0.23°C (Fig. 5), respectively, suggesting a coupling helix-induced thermal disturbance, consistent with binding and destabilization of MalK. It is notable that the mixture of MalFch and MalGch peptides induce a positive ΔT_m of 0.4°C, and the addition of $MgCl_2$ and ATP induce a negative ΔT_m of -1.3°C in relation to apo-MalK (data not shown).

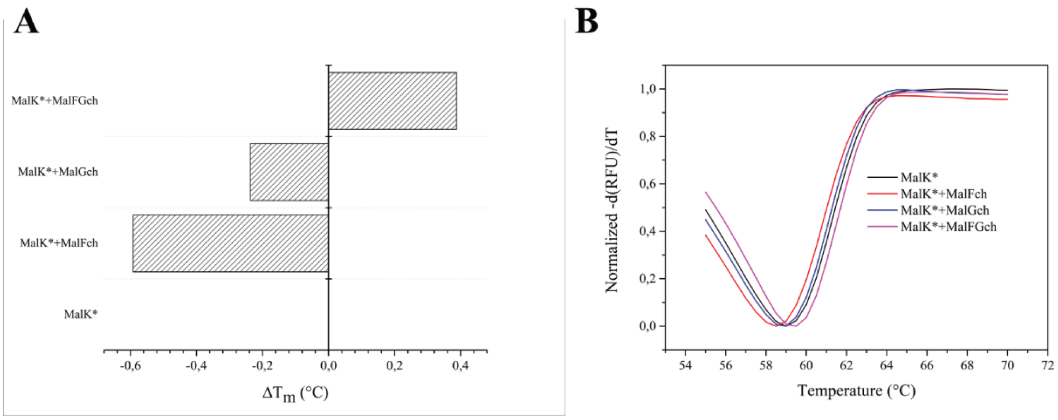


Figure 5. DSF assays of MalK NBD in presence of MalFch and MalGch peptides. A Bar graph illustrating shifts of MalK T_m for a MalFch and MalGch peptides. **B** Thermal denaturation curve for MalK NBD observed by differential scanning fluorescence and T_m shifts observed in the presence of the MalFch and MalGch peptides at 100 µM. Thermal stability curves are plotted against the normalized fluorescence signal. Experiments were carried out in 25 mM Tris-HCl pH 8.0, 125 mM NaCl and 5% glycerol. *MalK protein in presence of $MgCl_2$ and ATP.

2.6. *E. coli* MalK binds both MalFch and MalGch peptides with different K_d

Microscale thermophoresis (MST) allows a quantitative analysis of protein interactions in free solution. Thermophoresis (the directed motion of molecules in temperature gradients) is

highly sensitive to all types of binding-induced changes of molecular properties, be they in size, charge, hydration shell or conformation. In an all-optical approach, an infrared laser is used for local heating, and molecule mobility in the temperature gradient is analysed via fluorescence [30]. So, Microscale Thermophoresis (MST) was used to validate the MalK NBD/MalFch or MalGch peptides interactions and determining the binding affinities. The MalK protein was labelled with the Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation (NanoTemper Technologies). Once MalK protein did not lead to any measurable interaction in work buffer, a serial dilution of each coupling helix peptide, allowing 16-point measurements was performed. The experimental conditions allowed the measurement of the binding affinity for MalK-MalFch and MalGch peptides with a K_d value of 47.8 µmol L^{-1} and 20.9 µmol L^{-1}, respectively, according to K_d Fit Model (equation 1, materials and methods) (Fig. 6).

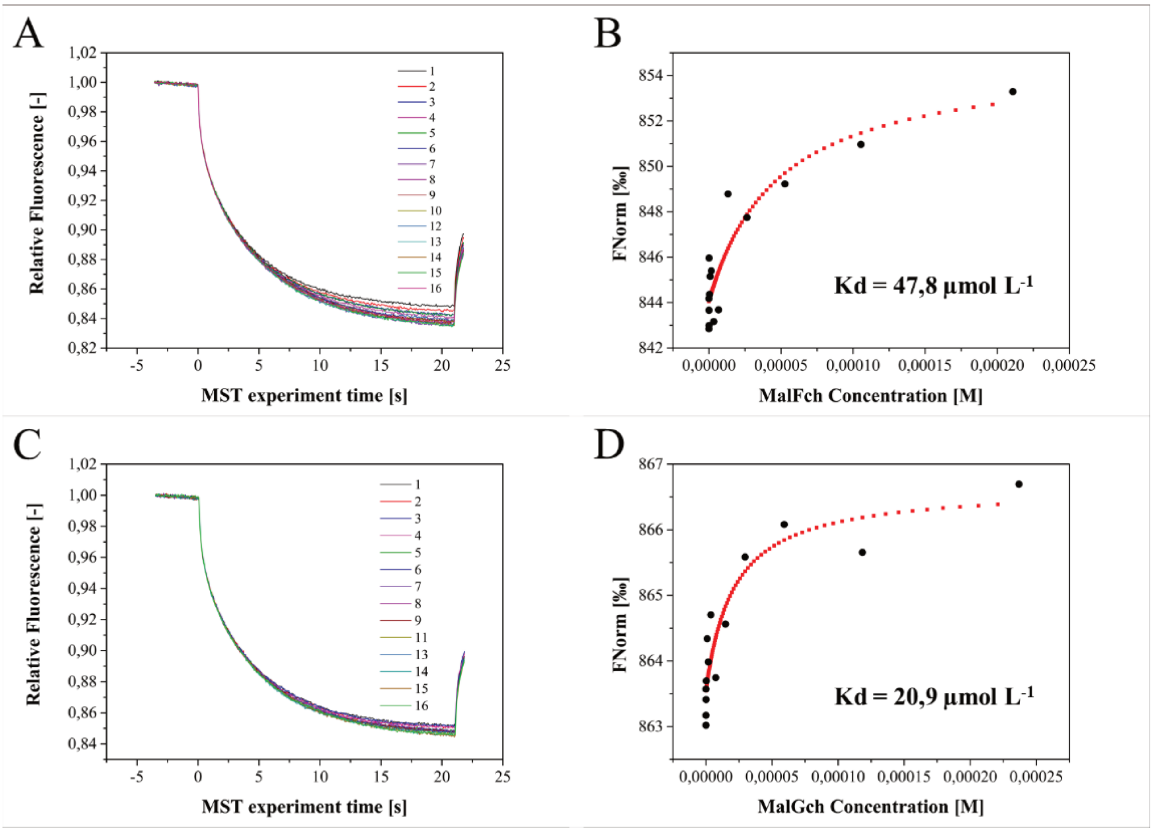


Figure 6. Determination of the affinity of MalFch and MalGch peptides for MalK NBD by microscale thermophoresis (MST). The concentrations of MalK protein, labelled with the fluorescent probe were kept constant at 50 nM, while the concentration of peptides was varied between 0 to 200 µM. **A** MST spectra for MalFch peptide. **B** Dose-response curves for MalFch interacting with MalK protein. **C** MST spectra for MalGch peptide. **D** Dose-response curves for MalGch interacting with MalK protein. Experiments were performed with previous addition of $MgCl_2$ and ATP to MalK protein.

Using the equation 2, described in materials and methods, MalK-MalFch system was analyzed (Fig. 7). The t_1 value shows that diffusion constants remain around constants values

up to certain value of peptide added, and then decreasing up to higher peptide concentration tested. The D_1 value increase around 10% up to the higher peptide concentration, demonstrating a disaggregation with peptide addition. If the amount of peptide continued to increase, the desegregation could continue to occur. The values for the fast term parameters show that the thermophoresis contribution of free fluorescent molecules remains around a constant throughout the experiment, which is reflected in the constant A_2 values. On the other hand, the values of t_2 show that even the rapid response changes with the increasing in peptide concentration. It is interesting to note that the values of t_2 are more than 10 times lower than the values of t_1, that is, the size of the fluorescent molecule must be smaller than MalK-MalFch in this proportion.

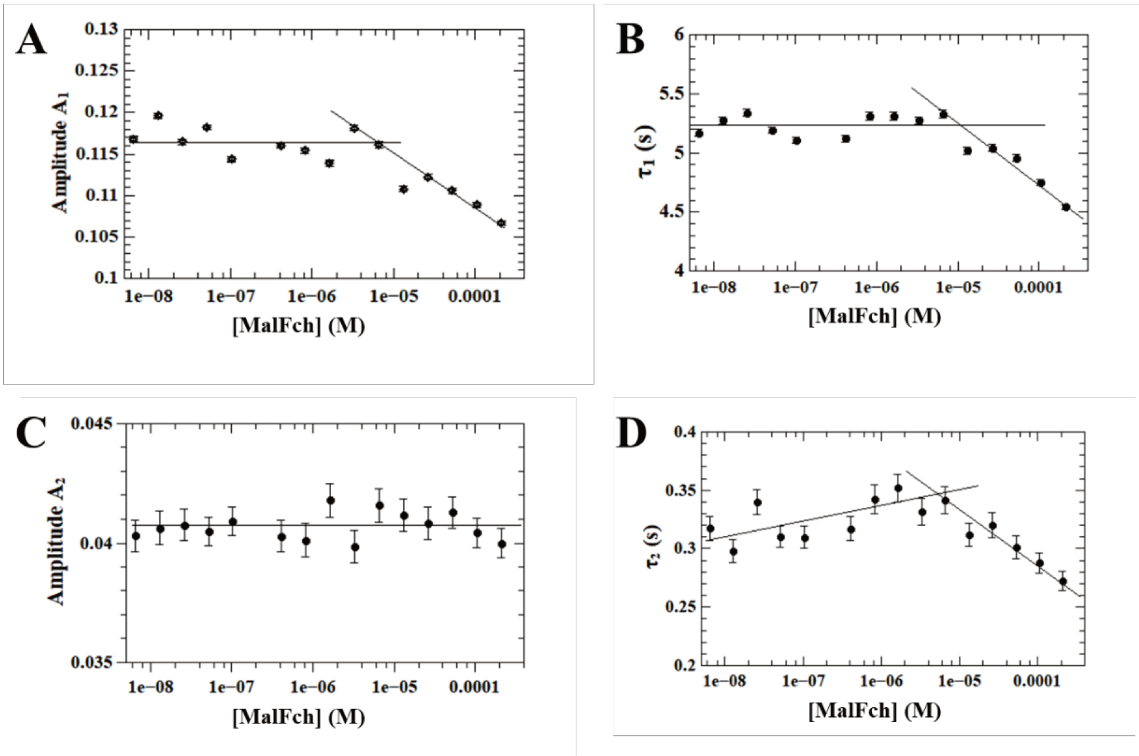


Figure 7. Microscale Thermophoresis parameters of MalK in function of MalFch peptide concentration. Values for the parameters A_1 (**A**) and t_1 (**B**) of the slow response related to the thermophoresis of MalK proteins with the fluorescent molecules bounded as a function of the MalFch peptide concentration. Values for parameters A_2 (**C**) and t_2 (**D**) of the rapid response related to the thermophoresis of unbound fluorescent molecules as a function of the MalFch peptide concentration.

In the case MalK-MalGch system, the interpretation is similarly to the MalK-MalFch system (Fig. 8). Once again it is clear from A_1 and t_1 values that the disaggregation has not yet occurred fully even at the highest concentration of peptide. From A_2 and t_2 it can be concluded that the increase in the concentration of peptides did not influence the thermophoresis of the fluorescent molecules, indicating that there should be no relevant interaction between peptide

and fluorophore. As in the case of MalK-MalGch, t_1 (also D_1) is more than 10 times greater than t_2 (also D_2), as expected.

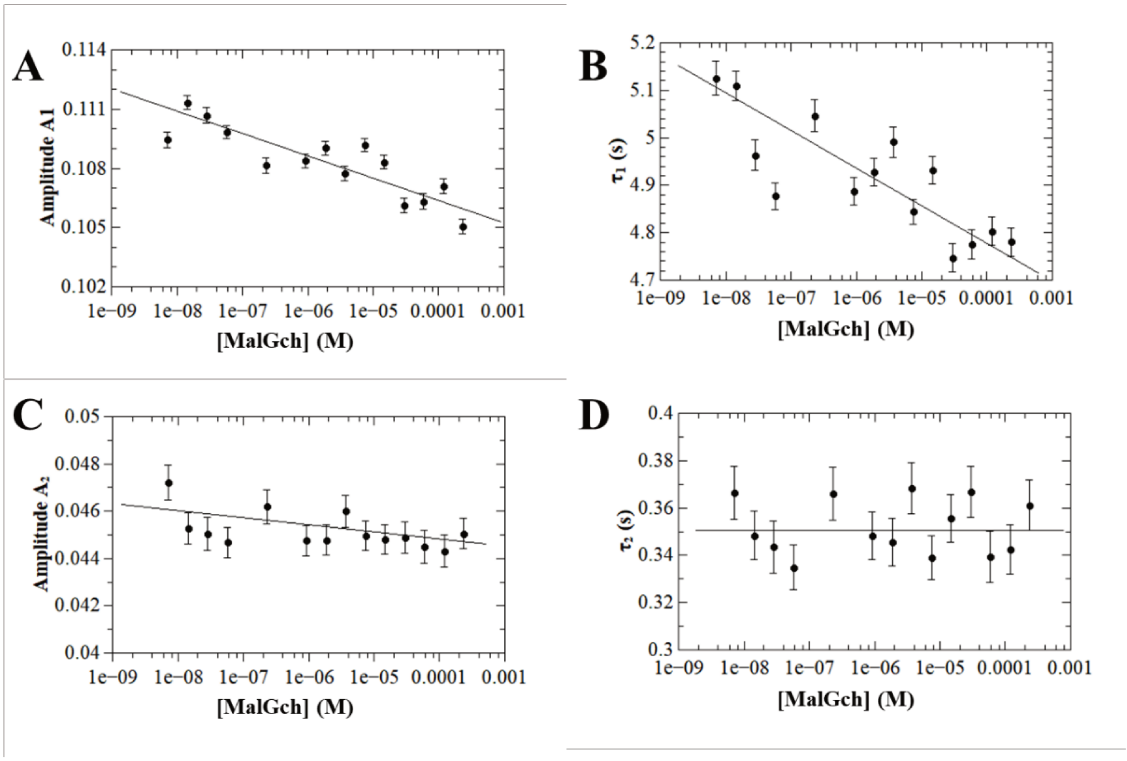


Figure 8. Microscale Thermophoresis parameters of MalK in function of MalGch peptide concentration. Values for the parameters A_1 (**A**) and t_1 (**B**) of the slow response related to the thermophoresis of MalK proteins with the fluorescent molecules bounded as a function of the MalGch peptide concentration. Values for parameters A_2 (**C**) and t_2 (**D**) of the rapid response related to the thermophoresis of unbound fluorescent molecules as a function of the MalGch peptide concentration.

2.7. *E. coli* MalK undergo conformational changes upon MalFch and MalGch peptides addition

A first analysis of the measured data from MalK and peptides was done to verify if I(q) values theoretically calculated using the known crystallographic structure may be similarly to the experimental results. We use the crystallographic structure of MalK (PDB ID: 2AWO). Frequently, there is a need to consider a mixture of different units in solution, like monomers, dimers, and tetramers to fit the calculated I(q) values to experimental result. We performed these analyses using the CRYSOL and OLIGOMER programs, which read the protein crystallographic structure and fit the calculated curve to the experimental results. The experimental result to the MalK protein in solution and the attempt to fit the results by the calculated scattering intensity for monomers, dimers and tetramers are shown in the Fig. 3A and Table 2. A good agreement between experimental and theoretical values is seen only in

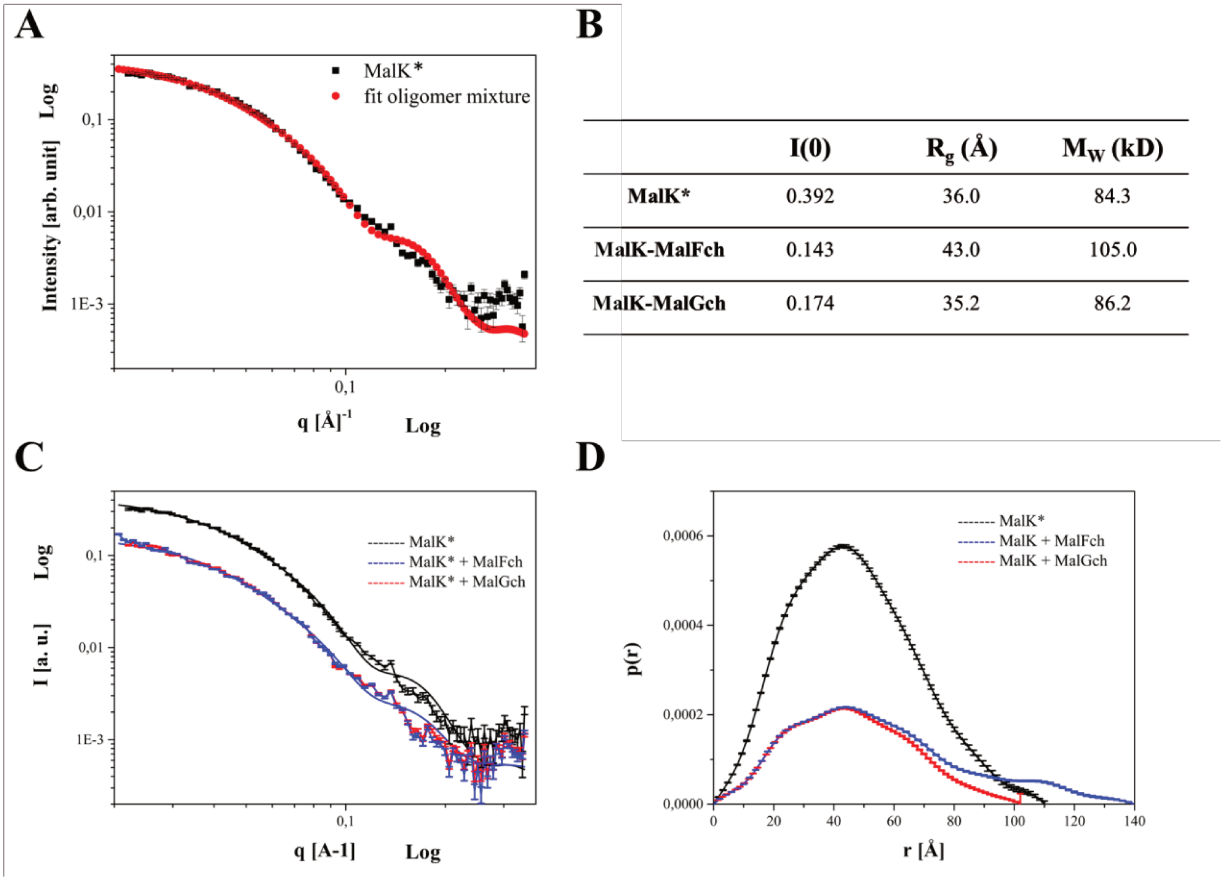
the case we considered a mixture of the three units. By adding the coupling helices peptides to MalK protein, the dimers units are lowered in concentration, as seen in the results (Table 2).

Table 2. Volume fraction of the three oligomeric states in samples of MalK protein in presence of MalFch and MalGch peptides, calculated with OLIGOMER.

Sample	Monomer (%vol)	Dimer (%vol)	Tetramer (%vol)
MalK*	21.4	19.5	59.0
MalK* + MalFch	30.3	5.4	64.2
MalK* + MalGch	34.2	0	65.8

* MalK protein in presence of 5 μM $MgCl_2$ and 5 μM ATP

Another analysis possibility is to consider a dimer in an extended configuration and symmetrically related to the binding site. The OLIGOMER programs are used to find a symmetry between both monomers such that the calculated scattering intensity fits the experimental results. It seems to be not conclusive by these fits if tested conditions have tetramers or the extended dimer. A second analysis of the data was to perform the Inverse Fourier transform to obtain the pair distance function *p(r)* and gyration radius R_g. The program GNOM calculates the function p(r) that best fits its inverse Fourier transform to the I(q) experimental data, by the indirect transform method developed by Glatter [31]. In Fig. 9C we show the experimental results for the scattering intensity and calculated *I(q)* values which is Fourier paired with the *p(r)* values showed in Fig. 9D. From this calculation we obtain the forward intensity $I(0)$, related to the protein M_W through the equation $M_W = I(0)N_A/c(\Delta\rho_M)^2$ [23], with c representing the concentration of proteins in mg/mL and $\Delta\rho_M$ the excess scattering length density per unit mass ($\Delta\rho_M \sim 2\text{x}10^{10}$ cm/g for proteins). The Fig. 9B shows the values for M_W of proteins in all samples. For samples MalK, MalK-MalFch, and MalK-MalGch, the respective increasing in the M_W values is in qualitative agreement with the analysis above, that showed the increasing concentration of tetramers (Table 2). The values in Fig. 9B are averaged from all particles present in solution.



	I(0)	R_g (Å)	M_W (kD)
MalK*	0.392	36.0	84.3
MalK-MalFch	0.143	43.0	105.0
MalK-MalGch	0.174	35.2	86.2

Figure 9. SAXS analysis of *E. coli* MalK and upon addition of MalFch and MalGch peptides. A Experimental scattering curve of MalK and the best intensity fitting of the curves obtained using the program GNOM. **B** $I(0)$, R_g and M_W of three MalK conditions. **C** Experimental scattering curve of MalK in presence of MalFch and MalGch peptides. **D** The normalized pair-distance distribution function $p(r)$ for MalK in presence of MalFch and MalGch peptides. MalFch and MalGch peptides are shown in blue and red, respectively.

2.8. Tryptophan Intrinsic Fluorescence is not a suitable technique to measure MalK-MalFGch peptides interaction

Many biological molecules display fluorescence, such some nucleotides, pyridoxal phosphate, chlorophyl, and proteins [27]. Intrinsic protein fluorescence is due to the aromatic amino acids, mainly tryptophan, considering that phenylalanine has a very low quantum yield, and emission by tyrosine in native proteins is often quenched. Fluorescence due only to tryptophan residues can be preferentially measured by exciting at 295 nm, because at this wavelength there is no absorption by tyrosine. Tryptophan fluorescence is highly sensitive to the environment polarity and shifts in its emission spectrum toward lower wavelengths (blue shift) can be seen as hydrophobicity increases. Changes in emission spectra from tryptophan can be seen in response to protein conformational transitions, subunit association, ligand binding, or denaturation, all of which can affect the local environment surrounding the indole ring [32]. *E. coli* MalK NBD has two tryptophans, including one of them very close to the

cleft for the coupling helices in the NBD (Trp13 and Trp267), that theoretically would allow us to access the MalK intrinsic fluorescence. Fluorescence experiments were performed using 1 ml samples of MalK (5 μM) in 25 mM Tris-HCl pH 8.0, 125 mM NaCl and 5% glycerol and stock solutions of MalFch and MalGch peptides (1 mM) in the same buffer solution. The emission spectra were obtained at 22.5ºC using an excitation beam light at 295 nm. Fig. 11AB displays MalK fluorescence spectra with increasing amounts of peptides. The fluorescence quantum yield is the ratio between the photons emitted and absorbed. Once that interaction of MalK with both peptides did not change its absorption spectra (Fig. 10), it becomes evident that the interaction with both peptides did not result in an increasing or decreasing of fluorescence quantum yield, i.e. an increasing or decreasing of fluorescence emission intensity.

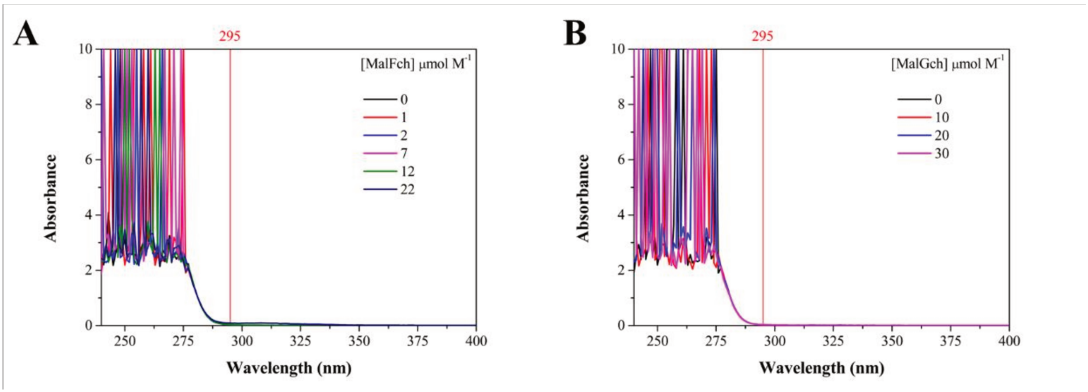


Figure 10. Absorption espectra of MalK as a function of MalFch and MalGch peptides concentration. A Absorption spectra of MalK with increasing concentration of MalFch peptide. **B** MalGch peptide. The experiments were performed with 5 μM of MalK in work buffer. The concentration of the peptides is showed in the figure.

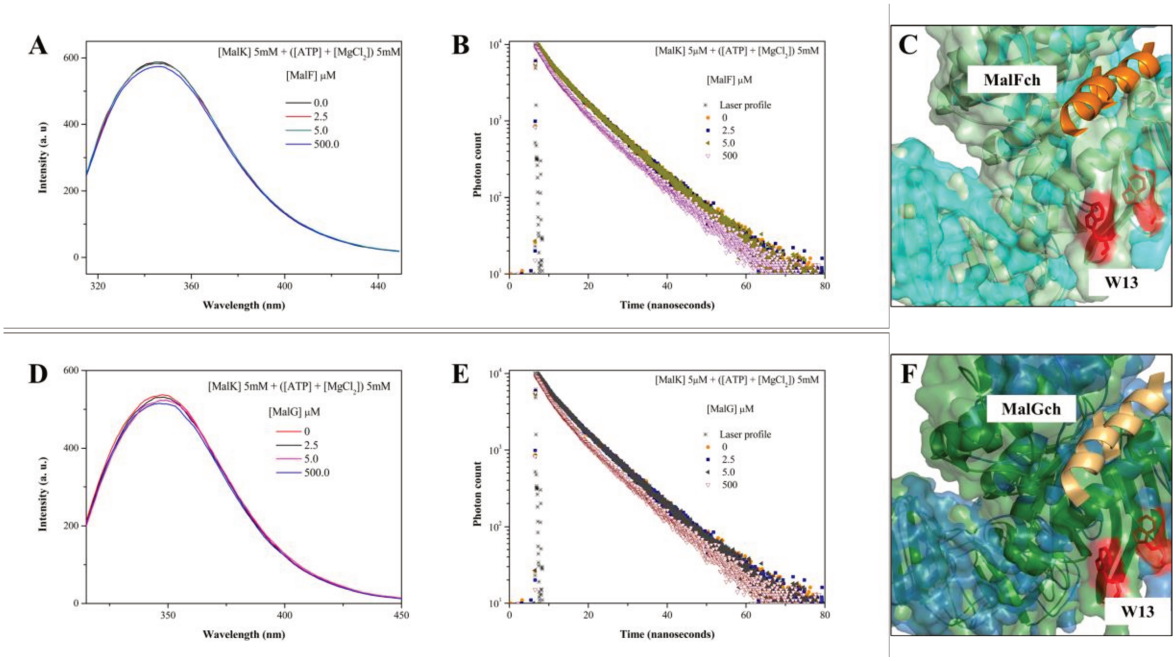


Figure 11. Intrinsic fluorescence measurements of *E. coli* MalK as a function of MalFch and MalGch peptides concentration. A Typical fluorescence spectrum and **B** Time-resolved fluorescence of 5 µM MalK in work buffer in the absence and with increasing concentration of MalFch peptide. **C** Three-dimensional structure of MalK in cartoon/surface showing the position of the Trp13 residue (in red) identified in the protein sequence respect to MalFch peptide. **D** Typical fluorescence spectra and **E** Time-resolved fluorescence of 5 µM MalK in work buffer in the absence and with increasing concentration of MalGch peptide. **F** Three-dimensional structure of MalK in cartoon/surface showing the position of the Trp13 residue (in red) identified in the protein sequence respect to MalGch peptide. Catalytic intermediate (PDB code: 2R6G, green) and resting states (PDB code: 3FH6, blue) are shown. Excitation beam at 295 nm.

To address the accessibility of tryptophan residues, we performed an acrylamide quenching assay (Fig. 12). Acrylamide quenching is sensitive to the degree of tryptophan accessibility to the solvent containing the acrylamide. Since acrylamide can diffuse to the interior of the protein, accessibility to acrylamide may result from tryptophan residues lying at the surface of the protein or from the existence of channels leading to the interior of the protein [33]. Hence one can expect complex Stern-Volmer plots, and even spectral shifts due to selective quenching of exposed versus buried tryptophan residues [27]. We showed that MalK have tryptophan residues that could be accessed by a quenching molecule as acrylamide, as seen in Stern-Volmer plot (Fig. 12B). However, the MalFch and MalGch peptides seem to not provoke excited-state reactions, molecular rearrangements, energy transfer, ground-state complex formation, or collisional quenching in tryptophan residues of MalK, specifically the Trp13, as seen in the Figure11. According to our assumptions this tryptophan should be sensitive to addition of mimetic peptides of TMD coupling helices, but it is not the case.

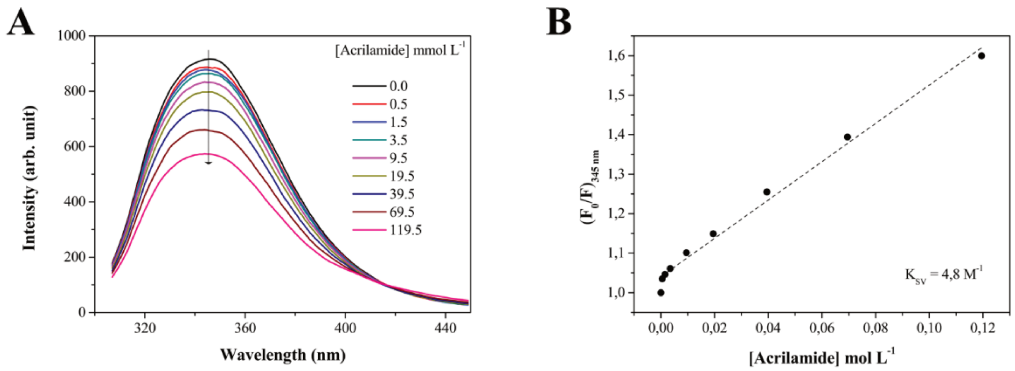


Figure 12. Acrylamide quenching measurements of *E. coli* MalK. A Typical fluorescence spectrum of 5 µM MalK in the absence and with increasing concentration of acrylamide. **B** Stern-Volmer plot of MalK fluorescence at 345 nm as a function of acrylamide concentration.

3. Discussion

The biophysical approach of MalK interacting with MalF and MalG coupling helices presented in this work, could be an important start point to study the interaction among ABC transporter components without the need of transmembrane protein production, that have been shown to be quite challenging. Assuming that ATP hydrolysis trigger conformational changes that switches TMD accessibility, we exposed MalK NBD, with previous addition of $MgCl_2$ and ATP, to MalF and MalG coupling helices peptides, and we measure the MalK responses using biophysical techniques. By analyzing these NBD responses, we showed the relevance of coupling helices in the NBD-TMD interaction.

The coupling helices of TMDs and NBD interaction is a well-known phenomenon. Oldman and collaborators showed that critical contacts are made by a residue on the coupling helix, namely Met405 in MalF and Leu194 in MalG. The side chain of these residues inserts into a hydrophobic pocket formed by MalK residues Ala50, Leu52, Ala73, Val77, Met79 and Phe81. We showed using an evolutionary information that Asp398 and Ser187 (for MalF and MalG, respectively) also could be crucial residues in MalK-MalFG interactions. Curiously, both Met405 and Leu194 are hydrophobic residues, but the main difference among both amino acids is that Leucine is a quite more hydrophobic. In contrast Asp398 and Ser187 are hydrophilic amino acids. Altogether, these differences could be involved in the NBD-TMD interaction mode of $MalEFGK_2$ ABC transporter.

It is known that MalK NBD, differently of other ABC NBDs, remains in dimeric state in Mg^{2+} and ATP free condition [11]. It is explained by the C-term regulatory domains that keep

the two monomers bounded. Using DSF, we showed a MalK single transition around 60°C. After addition of $MgCl_2$ and ATP, the T_m shifted to a lower temperature, suggesting a thermal destabilization. It could be explained by the fact that in presence of ATP and cofactors, the catalytic domain and even the C-term regulatory domain may undergo a conformational change that could destabilize the protein. This result is in contrast with other NBDs of sugar ABC transporters, that shifted to higher temperatures [34]. Both of coupling helices peptides, MalFch and MalGch, provoke a subtle T_m shifts to lower temperatures. This phenomenon could be difficult to understand, but probably the net charge or intrinsic properties of peptides may be involved in the MalK thermal responses. A possible explanation for this phenomenon may be the role described for MalF as the protein that harbors the substrate [16]. If this is the case, the coupling helix of this protein could trigger a destabilization of the MalK protein, that would coincide with a relaxation of the MalK protein, in such a way that the substrate can leave the transmembrane pore. This may also be related to the fact that MalKA85C mutant forms a strong cross-link with MalFS403C in the absence of ATP, but a significantly weaker cross-link in the presence of ATP or after vanadate trapping. In contrast, adding ATP or trapping the complex with vanadate did not affect K-G pairing, although MalK dimers were also formed under these conditions [35]. On the other hand, the presence of both peptides in the MalK samples seems to provoke a subtle T_m shift to higher temperatures.

According to MST results, we show that MalK seems to be an apparent higher affinity for MalGch peptide in contrast to MalFch peptide. First studies showed that substitutions made at the same positions in *malG* or in *malF* coupling helices affect MalG more severely than MalF. At substitution of Ala192 and Ser403 by Asp has a more pronounced effect in MalG (transport defective) than in MalF (reduced transport rate), respectively. At position Gly196 and Gly407, a change to Pro leads to a completely defective MalG protein while MalF is unaffected [36], respectively. The phenomena related to the transport, also could be related to the affinity that MalK bind coupling helices, that in our study shows a higher affinity to MalGch peptide than MalFch peptide. Other analysis of MST data showed that the two systems (MalK-MalFch and MalK-MalGch) undergo an exponential decay, as expected. The MST results indicates that the systems originally have dimers/tetramers that disaggregate as MalFch or MalGch peptides are added. The values of A_1 and as a result of ΔF_{norm} no longer correspond to the peptide binding, but the thermophoresis of the systems is altered by the peptides. This is because the Soret coefficient of thermophoresis is sensitive, etc. [30]. It is not known for sure whether NBD components may have greater or lower affinities for certain

TMD proteins, but the analyzes described above show that NBD MalK can undergo different behaviors according to the coupling helix tested.

SAXS data showed similar behavior for MalK in the presence of both peptides, as also evidenced by MST analysis. Note that the SAXS results are in agreement with the findings of MST suggesting that the MalK proteins disaggregate upon peptide addition. Despite, that both peptides could interact with MalK, it is interesting that P(r) x r results exhibit different behaviors, that coincide with changes in oligomeric states of MalK upon addition of coupling helices peptides. We showed that after MalGch peptide addition the volume of monomers and tetramers increase more than after MalFch addition. Recent studies show that in the presence of the non-hydrolyzable ATP analog (ATPγS) and ADP-phosphate mimics (ADP-VO_4 or ADP-AlF_4), $MalK_2$ was not stabilized in the closed state. Instead, two different semi-closed conformations were obtained, termed asymmetric and symmetric. In the semi-closed asymmetric state (adopted by the majority of the particles), the MalK-MalG interface was closer to the center of the complex, whereas the MalK-MalF interface remained in the open position, similar to the nucleotide-free state. ADP was therefore able to affect the conformation of the MalK-MalG interface [37].

Fluorescence experiments show that both peptides do not change the emission of MalK, increasing or decreasing the fluorescence quantum yield. Following photon absorption, the molecular electronic cloud relaxes returning to the energetic ground state through two different processes: radiative processes which results in light emission, and non-radiative processes which do not implicate in light emission. The presence of both peptides does not interfere with the fluorescence quantum yield indicating that the non-radiative decay processes are not changing. Upon excitation usually tryptophan residues display a large dipole momentum. The tryptophans can lose energy through orientation neighbour solvent molecules. This effect is known as dipolar relaxation and shifts the fluorescent spectrum to higher wavelengths [27]. The interaction between MalK and peptides did not shift the MalK emission spectra, indicating that the region of the tryptophan is not changing in terms of accessibility of the molecules of the solvent.

Altogether, the data presented in this work showed that the use of coupling helices mimetic peptides associated to the thermophoresis assays consist of an interesting alternative for the study of interactions between TMDs and NBDs from ABC transporters. This approach can be

explored for studies of specificities, protein-protein interactions and further development of inhibitors that target the triggering of transport activity in pathogenic bacteria.

References

1. Rees DC, Johnson E, Lewinson O. ABC transporters: The power to change. Nat Rev Mol Cell Biol. 2009;10:218–27.
2. Ter Beek J, Guskov A, Slotboom DJ. Structural diversity of ABC transporters. J Gen Physiol. 2014;143:419–35.
3. Ames GFL. Structure and mechanism of bacterial periplasmic transport systems. J Bioenerg Biomembr. 1988;20:1–18.
4. Eitinger T, Rodionov DA, Grote M, Schneider E. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: Diversity in modular organization and cellular functions. FEMS Microbiol Rev. 2011;35:3–67.
5. Vetter IR, Wittinghofer A. Nucleoside triphosphate-binding proteins: Different scaffolds to achieve phosphoryl transfer. Q Rev Biophys. 1999;32:1–56.
6. Davidson AL, Dassa E, Orelle C, Chen J. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. Microbiol Mol Biol Rev. 2008;72:317–64.
7. Grossmann N, Vakkasoglu AS, Hulpke S, Abele R, Gaudet R, Tampé R. Mechanistic determinants of the directionality and energetics of active export by a heterodimeric ABC transporter. Nat Commun. 2014;5.
8. Dawson RJP, Hollenstein K, Locher KP. Uptake or extrusion: Crystal structures of full ABC transporters suggest a common mechanism. Mol Microbiol. 2007;65:250–7.
9. Thomas C, Tampé R. Structural and Mechanistic Principles of ABC Transporters. Annu Rev Biochem. 2020;89:605–36.
10. Valkov E, Sharpe T, Marsh M, Greive S, Hyvönen M. Targeting protein-protein interactions and fragment-based drug discovery. Top Curr Chem. Germany; 2012;317:145–79.
11. Chen J, Lu G, Lin J, Davidson AL, Quiocho FA. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Mol Cell. 2003;12:651–61.
12. Lu G, Westbrooks JM, Davidson AL, Chen J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. Proc Natl Acad Sci U S A. 2005;102:17969–74.
13. Khare D, Oldham ML, Orelle C, Davidson AL, Chen J. Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations. Mol Cell. Elsevier Ltd; 2009;33:528–36.
14. Oldham ML, Chen J. Snapshots of the maltose transporter during ATP hydrolysis. Proc Natl Acad Sci U S A. 2011;108:15152–6.
15. Oldham ML, Chen S, Chen J. Structural basis for substrate specificity in the *Escherichia coli* maltose transport system. Proc Natl Acad Sci U S A. 2013;110:18132–7.
16. Oldham ML, Khare D, Quiocho FA, Davidson AL, Chen J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature. 2007;450:515–21.
17. Oldham ML, Chen J. Crystal structure of the maltose transporter in a pretranslocation intermediate state. Science (80-). 2011;332:1202–5.
18. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. Elife. 2014;2014:1–21.
19. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version~1.8. 2015.
20. Petoukhov M V., Franke D, Shkumatov A V., Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr. International Union of Crystallography; 2012;45:342–50.
21. Svergun D, Barberato C, Koch MH. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr. International Union of Crystallography; 1995;28:768–73.
22. Konarev P V., Volkov V V., Sokolova A V., Koch MHJ, Svergun DI. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr. 2003;36:1277–82.
23. Pinto Oliveira CL. Investigating Macromolecular Complexes in Solution by Small Angle X-Ray Scattering. Curr Trends X-Ray Crystallogr. 2011.
24. Semenyuk A V., Svergun DI. GNOM. A program package for small-angle scattering data processing. J Appl Crystallogr. International Union of Crystallography; 1991;24:537–40.
25. Petoukhov M V., Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J. 2005;89:1237–50.
26. Volkov V V, Svergun DI. Uniqueness of {\it ab initio} shape determination in small-angle scattering. J Appl Crystallogr. 2003;36:860–4.

27. Lakowicz JR. Principles of fluorescence spectroscopy [Internet]. Second edition. New York : Kluwer Academic/Plenum, [1999] ©1999.
28. Gao K, Oerlemans R, Groves MR. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. Biophys Rev. Biophysical Reviews; 2020;12:85–104.
29. Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem. 2004;332:153–9.
30. Seidel SAI, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, Lazic A, et al. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods. 2013;59:301–15.
31. Pedersen JS. Determination of size distribution from small-angle scattering data for systems with effective hard-sphere interactions. J Appl Crystallogr. 1994;27:595–608.
32. Mö M, Denicola A. Laboratory Exercises Protein Tryptophan Accessibility Studied by Fluorescence Quenching. Biochem Mol Biol Educ Print USA. 2002;30:175–8.
33. Phillips SR, Wilson LJ, Borkman RF. Acrylamide and iodide fluorescence quenching as a structural probe of tryptophan microenvironment in bovine Lens crystallins. Curr Eye Res. 1986;5:611–20.
34. Pretz MG, van der Does C, Albers SV, Schuurman-Wolters G, Driessen AJM. Structural stability of GlcV, the nucleotide binding domain of the glucose ABC transporter of *Sulfolobus solfataricus*. Biochim Biophys Acta - Biomembr. 2008;1778:324–33.
35. Daus ML, Grote M, Müller P, Doebber M, Herrmann A, Steinhoff HJ, et al. ATP-driven MalK dimer closure and reopening and conformational changes of the "EAA" motifs are crucial for function of the maltose ATP-binding cassette transporter ($MalFGK_2$). J Biol Chem. 2007;282:22387–96.
36. Mourez M, Hofnung M, Dassa E. Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. EMBO J. 1997;16:3066–77.
37. Fabre L, Bao H, Innes J, Duong F, Rouiller I. Negative stain single-particle em of the maltose transporter in nanodiscs reveals asymmetric closure of $MalK_2$ and catalytic roles of ATP, MalE, and maltose. J Biol Chem. 2017;292:5457–64.

DISCUSSÃO

A presente discussão abordará os resultados em tópicos. A motivação deste trabalho veio do interesse em responder quais são as funções dos quatro transportadores ABC de açúcares de *M. tuberculosis* e suas características funcionais/estruturais. Adicionalmente, gostaríamos de desenvolver uma maneira de estudar a interação entre TMDs e NBDs sem a necessidade de se trabalhar com os domínios transmembrana, uma vez que os mesmos são de difícil expressão, purificação e reconstituição em proteolipossomos. Esta interação é um ponto chave uma vez que por meio dela ocorre a ativação dos NBDs e consecutiva quebra dos ATPs para liberação da energia necessária ao transporte. Dada a relevância dos transportadores ABC em bactérias, tanto do ponto de vista nutricional (caso dos importadores) como de resistência a drogas (bombas de efluxo, caso dos exportadores), uma forma mais simples de avaliar a interação poderia ser usada em estudos de desenvolvimento de inibidores desses sistemas. A ideia inicial era utilizarmos o transportador UgpAEBC do próprio *M. tuberculosis,* mas as dificuldades de expressão e solubilização dos NBDs e nos levou a utilização do canônico e bem estudado transportador de maltose de *E. coli*. Neste sentido, buscamos abordar cada questionamento usando bioinformática estrutural, ferramentas filogenéticas e técnicas biofísicas.

O crescimento e os requerimentos nutricionais de micobactérias têm sido estudados desde a descoberta de *M. tuberculosis* (Tuberkulose, 1982). A descoberta de que *M. tuberculosis* usa o ciclo do glioxilato para a sobrevivência em camundongos elevou o interesse no metabolismo do carbono desse microrganismo (McKinney et al., 2000; Muñoz-Elías & McKinney, 2005). O fato sugere que o bacilo usa lipídeos como principal fonte de carbono durante a infeção, mas os carboidratos também parecem ter um papel importante. Os primeiros estudos nesse sentido mostraram a relevância de um transportador de dissacarídeos durante a primeira semana de infecção, sugerindo que a bactéria poderia trocar os carboidratos por lipídeos com o desenvolvimento da infecção e início da resposta imune. Estudos desenvolvidos por Titgemeyer e colaboradores (2007), mostraram que *M. tuberculosis* conta com um conjunto de 5 transportadores putativos de carboidratos, 4 deles sendo transportadores do tipo ABC. Interessantemente, esse número é muito baixo em comparações com *M. smegmatis,* o qual possui 28 sistemas para o mesmo propósito, sendo apenas dois compartilhados pelas duas espécies (Titgemeyer et al., 2007). Neste trabalho, mostramos que transportadores como LpqY/SugABC e UspABC além de estar presentes em *M. tuberculosis* e *M. smegmatis*, estão presentes em todas as espécies de *Mycobacterium*

estudadas. Isto permite sugerir que o papel destes transportadores é constitutivo para a fisiologia de *Mycobacterium*, e maiores extrapolações permitiriam sugerir que eles estão envolvidos na reciclagem de componentes da parede celular, funções que serão de igual importância para espécies de vida livre quanto para espécies patogênicas. Em contraste, transportadores como Rv2038c-41c e UgpAEBC, estão presentes em espécies patogênicas, com interesse especial pelo transportador UgpAEBC, exclusivo do MTBC e *M. marinum*. Este fato nos permite sugerir que tais transportadores poderiam estar relacionados aos processos de patogênese e virulência, principalmente dentro do hospedeiro.

As análises filogenéticas mostraram que componentes do transportadores LpqY/SugABC, especificamente o domínio citoplasmático SugC e os domínios transmembrana SugAB, podem ter sido os primeiros parálogos a divergir na evolução, o que destaca o papel destes no gênero *Mycobacterium*. Isto pode ser de grande importância na escolha direcionada de fármacos ou o desenvolvimento de estratégias de diagnóstico, uma vez que proteínas extensivamente conservadas podem não ser de grande interesse. As comparações estruturais mostraram um dos domínios transmembrana como o mais diverso, sendo que essa diversidade pode ser determinada pelas regiões de interação com os domínios periplasmáticos e com o substrato. Isto também foi mostrado para o transportador ABC de maltose de *E. coli*, no qual só uma das proteínas transmembrana, MalF, está envolvida na interação com a proteína periplasmática e o substrato (Oldham et al., 2007). A comparação das proteínas periplasmáticas de todos os transportadores estudados, mostraram características de SBPs do cluster D, segundo a classificação de Scheepers e colaboradores (2016). No entanto, apesar dos bolsões de interação apresentarem uma característica comum que é a presença de resíduos aromáticos, os volumes e os resíduos que formam os bolsões mostraram-se muito diferentes.

A carência de um gene codificando o domínio citoplasmático, dentro do operon *uspABC*, levantou o interesse pela possível promiscuidade destes componentes em relação aos domínios transmembrana. Neste sentido, fizemos uma comparação das interfaces de interação entre estes componentes, e mostramos que uma conservação em resíduos específicos tanto nos domínios citoplasmáticos e transmembrana poderiam permitir dita promiscuidade. Como mencionado, uma vez que essa interação é crucial para o funcionamento dos transportadores ABC, nos propusemos a estudar uma alternativa para avaliar as interações entre estes componentes, utilizando o transportador ABC de maltose de *E. coli*.

A estrutura cristalográfica do transportador ABC $MalEFGK_2$ de *E. coli* em diferentes

conformações tem provido informação crucial sobre a química de ligação e hidrólise do ATP. Hoje é conhecido que o dímero $MalK_2$ livre de nucleotídeos ou no estado de "inativo" apresenta as proteínas transmembrana MalF-MalG na conformação *inward-facing*, enquanto o dímero $MalK_2$ ligado ao ATP está próximo, tendo os TMDs em uma conformação *outward-facing*. A ligação ao substrato, o transporte e a liberação geram passos intermediários adicionais que não têm sido resolvidos pela cristalografia de raios-X (Fabre et al., 2017). Técnicas biofísicas diferentes da cristalografia de raios-X podem ser de grande contribuição no estudo destes passos. Neste trabalho, mostramos por meio de técnicas biofísicas como DSF, MST, SAXS e fluorescência intrínseca do triptofano, que peptídeos miméticos das hélices de acoplamento das proteínas transmembrana podem desencadear respostas na proteína citoplasmática MalK. O ensaio de DSF mostrou que o peptídeo MalFch promove uma desestabilização térmica da proteína MalK, conforme evidenciado pelo valor de ΔT_m. O papel desta desestabilização é pouco compreendido, mas estudos anteriores mostraram que a proteína MalF alberga o substrato (Oldham et al., 2007). Uma possível explicação para o fenômeno, é que a hélice de acoplamento de MalF pode gerar um relaxamento de MalK, de modo que o substrato possa ser liberado e deixar o poro transmembrana. No entanto estudos aprofundados serão necessários para confirmar a nossa hipótese. Os resultados de MST mostraram que MalK se liga aos peptídeos miméticos com uma afinidade na escala micromolar. No entanto a afinidade pelo peptídeo MalG parece ser maior. Isto pode ser explicado por estudos anteriores que demostraram mutações na hélice de acoplamento de MalG afetam severamente o transporte, em comparação com as mesmas mutações na proteína MalF, que não é afetado (Mourez et al., 1997).

Os dados de SAXS mostraram que MalK experimenta mudanças conformacionais na presença de ambos os peptídeos, mas que tais mudanças podem ser assimétricas. Isto também foi demostrado por microscopia eletrônica de partícula única, onde estados assimétricos são observados na presença de ADP, sugerindo que a interface MalK-MalG estava mais fechada ao centro do complexo, enquanto a interface MalK-MalF permanece na mesma posição, quando com o estado livre de nucleotídeos (Fabre et al., 2017).

Os ensaios de fluorescência intrínseca do triptofano, mostraram que não há alteração mensurável na fluorescência de MalK na presença dos peptídeos miméticos, o que pode ser explicado pela localização dos triptofanos na proteína MalK. Ensaios de *quenching* de acrilamida demostraram que os triptofanos estão expostos e indicam que a técnica de fluorescência intrínseca do triptofano não é adequada para medir as interações.

CONCLUSÕES

Os resultados obtidos na presente tese, permitiram evidenciar que os transportadores ABC de açúcares de bactérias patogênicas como *M. tuberculosis*, podem estar envolvidos na patogênese da bactéria, uma vez que dois deles, Rv2038c-41c e UgpAEBC, são conservados somente em espécies patogênicas do gênero *Mycobacterium*. Outros transportadores como LpqY/SugABC e UspABC são conservados tanto em espécies patogênicas e não patogênicas. Além das comparações genômicas mostramos que a filogenia dos componentes destes transportadores pode contribuir no entendimento de características funcionais, com auxílio de análises dos alinhamentos. Neste sentido, mostramos que o transportador LpqY/SugABC, um transportador envolvido na reciclagem de trealose, poderia ser mais antigo na história evolutiva destes sistemas, o que é corroborado também pela presença deste transportador em todas as espécies de *Mycobacterium* avaliadas. As análises filogenéticas também nos permitiram classificar as proteínas TMDs em dois grupos sugestivamente relacionados à função destas proteínas. No caso dos NBDs, mostramos que as regiões C-terminal destas proteínas, ou reguladoras, são as mais diversas, o que indica interações com moléculas diferentes. As análises comparativas dos componentes periplasmáticos SBPs, mostraram que estas proteínas possuem regiões conservadas no domínio I e regiões mais diversas no domínio II, o que pode estar associado a presença de resíduos que permitem a interação com os substratos. Modelagem e *docking* molecular nos permitiram mapear os resíduos no sítio de interação com trealose da proteína LpqY, que pode ser de auxílio em estudos futuros. A proteína Rv2041c, ao ter menor similaridade com proteínas já caracterizadas se mostrou mais desafiantes na predição do sítio de interação e possíveis substratos.

No estudo das interações NBDs – hélices de acoplamento dos TMDs no transportador de maltose de *E. coli*, mostramos que peptídeos sintéticos podem ser usados no estudo destas interações usando técnicas biofísicas. Esta abordagem se mostra interessante, pois a partir do seu desenvolvimento é possível avaliar o papel de diferentes tipos de moléculas na inibição da interação, o que seria uma forma de controle do transporte, e dependendo do transportador, do crescimento do microrganismo. Vale lembrar que a conservação das hélices de acoplamento e o mecanismo em exportadores pode levar ao controle destes, principalmente aqueles envolvidos com a exclusão de drogas, de forma a diminuirmos a resistência bacteriana.

Finalmente, os dados apresentados nessa tese abrem futuras perspectivas de estudos que podem ser direcionados para a caracterização do papel funcional das proteínas estudadas em *M. tuberculosis,* quanto para estudos estruturais e de inibição de transportadores ABC.

REFERÊNCIAS

Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology*, *62*, 53–70.

Amorim, J. H., Porchia, B. F. M. M., Balan, A., Cavalcante, R. C. M., da Costa, S. M., de Barcelos Alves, A. M., & de Souza Ferreira, L. C. (2010). Refolded dengue virus type 2 NS1 protein expressed in *Escherichia coli* preserves structural and immunological properties of the native protein. *Journal of Virological Methods*, *167*(2), 186–192.

Bailo, R., Bhatt, A., & Aínsa, J. A. (2015). Lipid transport in *Mycobacterium tuberculosis* and its implications in virulence and drug development. *Biochemical Pharmacology*, *96*(3), 159–167.

Berntsson, R. P. A., Smits, S. H. J., Schmitt, L., Slotboom, D. J., & Poolman, B. (2010). A structural classification of substrate-binding proteins. *FEBS Letters*, *584*(12), 2606–2617.

Biemans-Oldehinkel, E., Doeven, M. K., & Poolman, B. (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Letters*, *580*(4), 1023–1035.

Borich, S. M., Murray, A., & Gormley, E. (2000). Genomic arrangement of a putative operon involved in maltose transport in the *Mycobacterium tuberculosis* complex and *Mycobacterium leprae*. *Microbios*.

Braibant, M., Gilot, P., & Content, J. (2000). The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiology Reviews*, *24*(4), 449–467.

Brass, J. M., Boos, W., & Hengge, R. (1981). Reconstitution of maltose transport in *malB* mutants of *Escherichia coli* through calcium-induced disruptions of the outer membrane. *Journal of Bacteriology*, *146*(1), 10–17.

Brites, D., & Gagneux, S. (2012). Old and new selective pressures on *Mycobacterium tuberculosis*. *Infection, Genetics and Evolution*, *12*(4), 678–685.

Chen, J. (2013). Molecular mechanism of the *Escherichia coli* maltose transporter. *Current Opinion in Structural Biology*, *23*(4), 492–498.

Chen, J., Lu, G., Lin, J., Davidson, A. L., & Quiocho, F. A. (2003). A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Molecular Cell*, *12*(3), 651–661.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12–21.

Dahl, S. G., Sylte, I., & Ravna, A. W. (2004). Structures and models of transporter proteins. *Journal of Pharmacology and Experimental Therapeutics*, *309*(3), 853–860.

Daniel, T. M. (2006). The history of tuberculosis. *Respiratory Medicine*, *100*(11), 1862–1870.

Davidson, A. L., & Chen, J. (2004). ATP-binding cassette transporters in bacteria. *Annual Review of Biochemistry*, *73*, 241–268.

Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiology and Molecular Biology Reviews*, *72*(2), 317–364.

Dawson, R. J. P., Hollenstein, K., & Locher, K. P. (2007). Uptake or extrusion: Crystal structures of full ABC transporters suggest a common mechanism. *Molecular Microbiology*, *65*(2), 250–257.

Diederichs, K. (2000). Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. *The EMBO Journal*, *19*(22), 5951–5961.

Dippel, R., & Boos, W. (2005). The maltodextrin system of *Escherichia coli*: Metabolism and transport. *Journal of Bacteriology*, *187*(24), 8322–8331.

Dulberger, C. L., Rubin, E. J., & Boutte, C. C. (2020). The mycobacterial cell envelope — a moving target. *Nature Reviews Microbiology*, *18*(1), 47–59.

Ebert, D., & Bull, J. J. (2003). Challenging the trade-off model for the evolution of virulence: Is virulence management feasible? *Trends in Microbiology*, *11*(1), 15–20.

Edson, N. L. (1951). The intermediary metabolism of the mycobacteria. *Bacteriological Reviews*, *15*(3), 147–182.

Fabre, L., Bao, H., Innes, J., Duong, F., & Rouiller, I. (2017). Negative stain single-particle em of the maltose transporter in nanodiscs reveals asymmetric closure of MalK2 and catalytic roles of ATP, MalE, and maltose. *Journal of Biological Chemistry*, *292*(13), 5457–5464.

Fedrizzi, T., Meehan, C. J., Grottola, A., Giacobazzi, E., Fregni Serpini, G., Tagliazucchi, S., Fabio, A., Bettua, C., Bertorelli, R., De Sanctis, V., Rumpianesi, F., Pecorari, M., Jousson, O., Tortoli, E., & Segata, N. (2017). Genomic characterization of Nontuberculous Mycobacteria. *Scientific Reports*, *7*(March), 1–14.

Fenn, J. S., Nepravishta, R., Guy, C. S., Harrison, J., Angulo, J., Cameron, A. D., & Fullam, E. (2019). Structural Basis of Glycerophosphodiester Recognition by the *Mycobacterium tuberculosis* Substrate-Binding Protein UgpB. *ACS Chemical Biology*, *14*(9), 1879–1887.

Fukami-Kobayashi, K., Tateno, Y., & Nishikawa, K. (1999). Domain dislocation: A change of core structure in

periplasmic binding proteins in their evolutionary history. *Journal of Molecular Biology*, *286*(1), 279–290.
Fullam, E., Prokes, I., Fütterer, K., & Besra, G. S. (2016). Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biology*, *6*(6).
Gagneux, S. (2018). Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*, *16*(4), 202–213.
Holland, I. B., Cole, S. P. C., Kuchler, K., & Higgins, C. F. (2003). ABC Proteins: From Bacteria to Man. Elsevier Science.
Jankute, M., Cox, J. A. G., Harrison, J., & Besra, G. S. (2015). Assembly of the Mycobacterial Cell Wall. *Annual Review of Microbiology*, *69*(1), 405–423.
Jeckelmann, J. M., & Erni, B. (2020). Transporters of glucose and other carbohydrates in bacteria. *Pflugers Archiv European Journal of Physiology*.
Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29.
Jiang, D., Zhang, Q., Zheng, Q., Zhou, H., Jin, J., Zhou, W., Bartlam, M., & Rao, Z. (2014). Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS Journal*, *281*(1), 331–341.
Jones, P. M., & George, A. M. (2002). Mechanism of ABC transporters: A molecular dynamics simulation of a well characterized nucleotide-binding subunit. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12639–12644.
Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G. S., & Jacobs, W. R. (2010). Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21761–21766.
Kapopoulou, A., Lew, J. M., & Cole, S. T. (2011). The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*, *91*(1), 8–13.
Kim, S. Y., Shin, A. R., Kim, H. J., Cho, S. N., Park, J. K., & Shin, S. J. (2009). Identification of Rv2041c, a Novel Immunogenic Antigen from *Mycobacterium tuberculosis* with Serodiagnostic Potential. *Scandinavian Journal of Immunology*, *70*(5), 457–464.
Kim, Su Young, Lee, B. S., Sung, J. S., Kim, H. J., & Park, J. K. (2008). Differentially expressed genes in *Mycobacterium tuberculosis* H37Rv under mild acidic and hypoxic conditions. *Journal of Medical Microbiology*, *57*(12), 1473–1480.
Kim, Su Young, Shin, A. R., Lee, B. S., Kim, H. J., Jeon, B. Y., Cho, S. N., Park, J. K., & Shill, S. J. (2009). Characterization of immune responses to *Mycobacterium tuberculosis* Rv2041c protein. *Journal of Bacteriology and Virology*, *39*(3), 183–193.
Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549.
Le Chevalier, F., Cascioferro, A., Frigui, W., Pawlik, A., Boritsch, E. C., Bottai, D., Majlessi, L., Herrmann, J. L., & Brosch, R. (2015). Revisiting the role of phospholipases C in virulence and the lifecycle of *Mycobacterium tuberculosis*. *Scientific Reports*, *5*(November), 1–13.
Le, S. Q., & Gascuel, O. (2001). *An Improved General Amino Acid Replacement Matrix*.
Locher, K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature Structural and Molecular Biology*, *23*(6), 487–493.
Luo, Y., Zhang, T., & Wu, H. (2014). The transport and mediation mechanisms of the common sugars in *Escherichia coli*. *Biotechnology Advances*, *32*(5), 905–919.
Mächtel, R., Narducci, A., Griffith, D. A., Cordes, T., & Orelle, C. (2019). An integrated transport mechanism of the maltose ABC importer. *Research in Microbiology*, *170*(8), 321–337.
McDonough, J. A., McCann, J. R., Tekippe, E. M. E., Silverman, J. S., Rigel, N. W., & Braunstein, M. (2008). Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *Journal of Bacteriology*, *190*(19), 6428–6438.
McKinney, J. D., Höner Zu Bentrup, K., Muñoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., & Russell, D. G. (2000). Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature*, *406*(6797), 735–738.
Mourez, M., Hofnung, M., & Dassa, E. (1997). Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO Journal*, *16*(11), 3066–3077.
Moussatova, A., Kandt, C., O'Mara, M. L., & Tieleman, D. P. (2008). ATP-binding cassette transporters in *Escherichia coli*. *Biochimica et Biophysica Acta - Biomembranes*, *1778*(9), 1757–1771.
Muñoz-Elías, E. J., & McKinney, J. D. (2005). *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nature Medicine*, *11*(6), 638–644.
Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L., & Chen, J. (2007). Crystal structure of a catalytic

intermediate of the maltose transporter. *Nature*, *450*(7169), 515–521.
Pontrelli, S., Chiu, T. Y., Lan, E. I., Chen, F. Y. H., Chang, P., & Liao, J. C. (2018). *Escherichia coli* as a host for metabolic engineering. *Metabolic Engineering*, *50*(February), 16–46.
Rogall, T., Wolters, J., Flohr, T., & Bottger, E. C. (1990). Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *International Journal of Systematic Bacteriology*, *40*(4), 323–330.
Sabharwal, N., Varshney, K., Rath, P. P., Gourinath, S., & Das, U. (2020). Biochemical and biophysical characterization of nucleotide binding domain of Trehalose transporter from *Mycobacterium tuberculosis*. *International Journal of Biological Macromolecules*, *152*, 109–116.
Safi, H., Lingaraju, S., Amin, A., Kim, S., Jones, M., Holmes, M., McNeil, M., Peterson, S. N., Chatterjee, D., Fleischmann, R., & Alland, D. (2013). Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-Arabinose biosynthetic and utilization pathway genes. *Nature Genetics*, *45*(10), 1190–1197.
Saier, M. H. (2000). Families of transmembrane sugar transport proteins. *Molecular Microbiology*, *35*(4), 699–710.
Sali, A., & Blundell, T. (1994). Sali, A. & Blundell, T. L. Comparative modelling by satisfaction of spatial restraints. *Journal of molecular biology*,*234*, 779-815.
Sambrook, J. (n.d.). Molecular cloning : a laboratory manual. Third edition. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2001] ©2001.
Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *100*(22), 12989–12994.
Scheepers, G. H., Lycklama a Nijeholt, J. A., & Poolman, B. (2016). An updated structural classification of substrate-binding proteins. *FEBS Letters*, *590*(23), 4393–4401.
Schneider, E. (2001). ABC transporters catalyzing carbohydrate uptake. *Research in Microbiology*, *152*(3–4), 303–310.
Shuman, H. (1998). Maltose/Maltodextrin System of *Escherichia coli*: Transport, Metabolism, and Regulation. *62*(1), 204–229.
Stein, A., Seifert, M., Volkmer-Engert, R., Siepelmeyer, J., Jahreis, K., & Schneider, E. (2002). Functional characterization of the maltose ATP-binding-cassette transporter of *Salmonella typhimurium* by means of monoclonal antibodies directed against the MalK subunit. *European Journal of Biochemistry*, *269*(16), 4074–4085.
Sutcliffe, I. C., & Harrington, D. J. (2004). Lipoproteins of *Mycobacterium tuberculosis*: An abundant and functionally diverse class of cell envelope components. *FEMS Microbiology Reviews*, *28*(5), 645–659.
Ter Beek, J., Guskov, A., & Slotboom, D. J. (2014). Structural diversity of ABC transporters. *Journal of General Physiology*, *143*(4), 419–435.
Thomas, C., & Tampé, R. (2020). Structural and Mechanistic Principles of ABC Transporters. *Annual Review of Biochemistry*, *89*, 605–636.
Titgemeyer, F., Amon, J., Parche, S., Mahfoud, M., Bail, J., Schlicht, M., Rehm, N., Hillmann, D., Stephan, J., Walter, B., Burkovski, A., & Niederweis, M. (2007). A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Journal of Bacteriology*, *189*(16), 5903–5915.
Tuberkulose, D. Ä. Der. (1982). I. Die Aetiologie der Tuberculose. (Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März er, gehaltenen Vortrage). *Fortschritte Der Medizin*, *100*(12), 539.
Vetter, I. R., & Wittinghofer, A. (1999). Nucleoside triphosphate-binding proteins: Different scaffolds to achieve phosphoryl transfer. *Quarterly Reviews of Biophysics*, *32*(1), 1–56.
Wheeler, P. R., Bulmer, K., & Ratledge, C. (1990). Enzymes for biosynthesis de novo and elongation of fatty acids in mycobacteria grown in host cells: is *Mycobacterium leprae* competent in fatty acid biosynthesis? *Journal of General Microbiology*, *136*(1), 211–217.
Wiedemann, C., Bellstedt, P., & Görlach, M. (2013). CAPITO - A web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics*, *29*(14), 1750–1757.
Wilken, S. (1996). A putative helical domain in the MalK subunit of the ATP-binding-cassette transport system for maltose of *Salmonella typhimurium* ($MalFGK_2$) is crucial for interaction with MalF and MalG . A study using the LacK protein of *Agrobacterium radiobacter* as a tool. *22*, 655–666.
Wuttge, S., Bommer, M., Jäger, F., Martins, B. M., Jacob, S., Licht, A., Scheffel, F., Dobbek, H., & Schneider, E. (2012). Determinants of substrate specificity and biochemical properties of the sn-glycerol-3-phosphate ATP binding cassette transporter (UgpB – AEC_2) of *Escherichia coli*. *86*(September), 908–920.

APÊNDICES

Apêndice A. Arquivos adicionais - CAPÍTULO I

Additional Files

Additional File 1

Table S1. List of *Mycobacterium* species used in this work.

***Mycobacterium* species/strain**	**NCBI accession number**	**Reference link**
M. tuberculosis H37Rv	NC_000962	http://www.ncbi.nlm.nih.gov/pubmed/20980199
M. tuberculosis H37Ra	NC_009525	http://www.ncbi.nlm.nih.gov/pubmed/18584054
M. tuberculosis CDC1551	NC_002755	http://www.ncbi.nlm.nih.gov/pubmed/12218036
M. africanum GM041182	NC_015758	http://www.ncbi.nlm.nih.gov/pubmed/22389744
M. bovis AF2122/97	NC_002945	http://www.ncbi.nlm.nih.gov/pubmed/12788972
M. bovis sp. BCG Pasteur 1173P2	NC_008769	http://www.ncbi.nlm.nih.gov/pubmed/17372194
M. avium 104	NC_008595	Direct Submission
M. avium sub. *paratuberculosis* K-10	NC_002944	http://www.ncbi.nlm.nih.gov/pubmed/16116077
M. intracellulare ATCC 13950	NC_016946	http://www.ncbi.nlm.nih.gov/pubmed/22535933
M. ulcerans AGY99	NC_008611	http://www.ncbi.nlm.nih.gov/pubmed/17210928
M. marinum M	NC_010612	http://www.ncbi.nlm.nih.gov/pubmed/18403782
M. abscessus ATCC19977	NC_010397	http://www.ncbi.nlm.nih.gov/pubmed/23804391
M. smegmatis mc^2155	NC_008596	Direct Submission
M. leprae TN	NC_002677	http://www.ncbi.nlm.nih.gov/pubmed/11234002

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 modelling.

	Templates				
***M. tuberculosis* ABC transporter components**	**Microorganism/ Protein function**	**Query cover (%)**	**Amino acid sequence Identity (%)**	**PDB ID**	**Resolution (Å)**
LpqY Rv1235 (SBP)	*Thermatoga maritima* TMBP - SBP	77	25	6DTQ_A	2.15
SugA Rv1236 (TMD)	*Escherichia coli* MalF - TMD	67	25	2R6G_F	2.8
SugB Rv1237 (TMD)	*Escherichia coli* MalG -TMD	92	31	2R6G_G	2.8
SugC Rv1238 (NBD)	*Thermococcus litoralis* MalK - NBD	97	50	1G29	1.9
Rv2038c (TMD)	*Thermococcus litoralis* MalK - NBD	73	59	1G29	1.9
Rv2039c (TMD)	*Escherichia coli* MalG - TMD	89	26	2R6G_G	2.8
Rv2040c (NBD)	*Escherichia coli* MalF - TMD	79	29	2R6G_F	2.8
Rv2041c (SBP)	*Listeria monocytogenes* Lmo0181 -SBP	91	26	5F7V	1.4
UspA Rv2316 (TMD)	*Escherichia coli* MalF -TMD	80	29	2R6G_F	2.8
UspB Rv2317 (TMD)	*Sphingomonas* sp. AlgM2 - TMD	99	24	4TQU_N	3.2
UspC Rv2318 (SBP)	*Mycobacterium tuberculosis* UspC - SBP	100	100	5K2X	1.5
UgpC Rv2832c (NBD)	*Thermococcus litoralis* Malk - NBD	85	52	1G29	1.9
UgpB Rv2833c (SBP)	*Mycobacterium tuberculosis* UgpB - SBP	100	100	4MFI	1.5
UgpE Rv2834c (TMD)	*Escherichia coli* MalG - TMD	76	30	2R6G_G	2.8
UgpA Rv2835c (TMD)	*Sphingomonas* sp. AlgM1 - TMD	75	26	4TQU_M	3.2

Table S3. **Structures used for prediction of the substrate-binding pocket of LpqY and Rv2041c models.** The program ITASSER was used since it uses different structures for prediction.

PDB Hit	Protein name/substrate	Microorganism	Identity (%)	TM Score	RMSD
		LpqY			
6DTQ	MalE3 maltose	*Thermotoga maritima*	22	0.78814	1.31
6DTR	MalE3 -	*Thermotoga maritima*	22	0.72703	2.77
3K01	GacH -	*Streptomyces glaucescens*	24	0.72774	2.99
6FFL	MalE maltose	*Bdellovibrio bacteriovorus*	18	0.74148	2.00
1EU8	TMBP trehalose	*Thermococcus litoralis*	25	0.79485	1.68
		Rv2041c			
5F7V	Lmo0181 Cycloalternan	*Listeria monocytogenes*	25	0.83902	1.66
4QRZ	Atu4361 Maltotriose	*Agrobacterium fabrum*	22	0.80938	2.11
3K01	GacH -	*Streptomyces glaucescens*	23	0.79341	2.57
4R6K	YesO -	*Bacillus subtilis*	18	0.74237	3.63
4MFI	UgpB -	*Mycobacterium tuberculosis*	23	0.78259	2.84
4QSC	Atu4361 maltose	*Agrobacterium fabrum*	22	0.80831	2.13

Table S4. 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).

NBD	TMD	Scaled Score	Probability	I_Prob
		SugC x SugA		
48_N	200_R	3.27	1.00	1.00
99_F	215_K	3.24	1.00	1.00
102_T	215_K	2.38	1.00	1.00
99_F	202_A	1.77	1.00	0.97
88_Y	205_D	1.68	1.00	0.96
84_S	197_D	1.65	1.00	0.96
140_R	197_D	1.45	0.99	0.92

90_H	220_M	1.44	0.99	0.92
74_K	203_Q	1.38	0.99	0.90
100_P	205_D	1.38	0.99	0.89
SugC x SugB				
48_N	166_K	3.24	1.00	1.00
99_F	181_K	3.16	1.00	1.00
99_F	168_A	2.57	1.00	1.00
102_T	181_K	2.25	1.00	0.99
88_Y	171_D	1.90	1.00	0.98
84_S	163_D	1.89	1.00	0.98
74_K	169_K	1.51	1.00	0.94
53_L	170_M	1.36	0.99	0.89
Rv2038c x Rv2039c				
49_R	174_E	3.17	1.00	1.00
100_F	189_R	2.94	1.00	1.00
100_F	176_A	2.35	1.00	1.00
103_K	189_R	2.32	1.00	1.00
89_Y	179_D	1.84	1.00	0.98
85_N	171_D	1.82	1.00	0.98
75_K	177_I	1.41	0.99	0.91
Rv2038c x Rv2040c				
49_R	189_E	3.62	1.00	1.00
100_F	204_S	3.46	1.00	1.00
100_F	191_A	2.35	1.00	1.00
103_K	204_S	2.33	1.00	1.00
85_N	186_T	1.90	1.00	0.98
89_Y	194_D	1.78	1.00	0.98
101_A	194_D	1.63	1.00	0.96
104_V	200_R	1.49	1.00	0.93
75_K	192_R	1.48	1.00	0.93
UgpC x UgpA				
100_F	214_R	3.50	1.00	1.00
49_R	199_E	3.41	1.00	1.00
100_F	201_A	2.32	1.00	1.00
103_R	214_R	2.25	1.00	0.99
85_N	196_D	1.70	1.00	0.97
89_Y	204_D	1.63	1.00	0.96
101_A	204_D	1.51	1.00	0.94
104_N	210_A	1.49	1.00	0.93
75_R	202_E	1.44	0.99	0.92

UgpC x UgpE				
49_R	168_E	3.23	1.00	1.00
100_F	183_R	3.06	1.00	1.00
100_F	170_A	2.35	1.00	1.00
103_R	183_R	2.32	1.00	1.00
89_Y	173_D	1.89	1.00	0.98
85_N	165_E	1.85	1.00	0.98
75_R	171_R	1.44	1.00	0.91
101_A	173_D	1.36	0.99	0.89

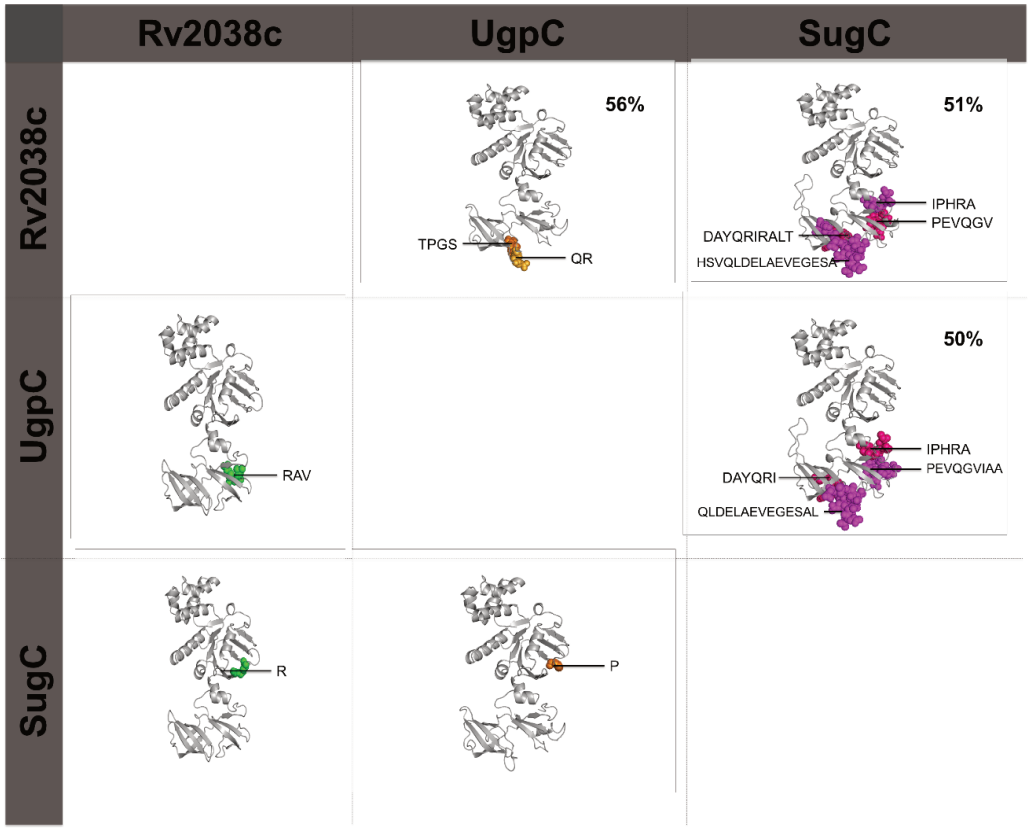
Additional File 2

A

```
  Regulatory region Start …
SugC_H37Rv          SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
MRA_RS06575         SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
MT_RS06545          SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
MAF_RS06560         SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
BQ2027_MB1270       SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
BCG_RS06720         SPAMNFFPARLTAIGLTLPFGEVTLAPEVQGVIAAHPKPENVIVGVRPEHIQDAALIDAY  296
MAV_RS06605         SPAMNFFPATLTPIGLKVPFGEVMLTPEVQQVIAEHPEPDNVIVGARPEHLSDAALIDGY  296
MAP_RS12980         SPAMNFFPATLTPIGLKLPFGEVMLTPEVQQVIAEHPEPDNVIVGARPEHLSDAALIDGY  296
OCU_RS31280         SPAMNFFPATLTPIGLTLPFGEVMLTPDVQEVIAQHPTPGNVIVGVRPEHLSDAALIDGY  296
MMAR_RS21050        SPAMNFFPAVLTPTGLTLPFGEVTLDPQVQQVIAQHPRPANIIVGIRPEQIQDAALIDAY  296
MAB_1375            SPSMNFFPATLTDVGVQLPFGEVTLDAGLYAGITARKPSGDVIVGIRPEQFEDAALVDTY  296
MSMEG_5058          SPAMNFFPATRTDVGVRLPFGEVTLTPHMLDLLDKQARPENIIVGIRPEHIEDSALLDGY  300
Rv2038c_H37Rv       SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
MRA_RS10790         SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
MT_RS10670          SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
MAF_RS10625         SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
BQ2027_MB2064C      SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
BCG_RS10595         SPAMNLFRLSIADSTVSLGDWQILLPRAV------VGTAAEVIIGVRPEHLELGGAGI--  289
MAV_RS11815         SPAMNLFTLPVVDSAVSLGDWPIALPREI------AAAASEVVVGVRPEHFELGGLGV--  289
MAP_RS08985         SPAMNLFTLQVVDSAVSLGDWPIALPREI------AAAASEVVVGVRPEHFELGGLGV--  289
OCU_RS37020         SPAMNLFTLPLVDSAVSLGDWPVAVPREI------AGAAGEVVVGVRPEHFEVGGLGV--  289
MUL_RS11890         SPAMNMFILPIVDSSVLLGDWLIQLPREV------TVPAPEVVVGVRPEHFEVGNLGV--  289
MMAR_RS15025        SPAMNMFILPIVDSSVLLGDWLIQLPREV------TVPAPEVVVGVRPEHFEVGNLGV--  289
ML1424              SPGMNLVTLSIVDSSVLLGDWPIRIPREI------ASAASEVIIGVRPEHFELGSLGV--  289
UgpC_H37Rv          APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
MRA_RS15055         APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
MT_RS14855          APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
MAF_RS14745         APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
BQ2027_MB2856C      APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
BCG_RS14680         APAMNLIDAAVAHGVVRAPDLAIPVPDP---------AAERVLVGVRPESWDVASIGT--  286
MMAR_RS09415        APAMNLIEAPVADGVARLSGLVIPVPRG---------AADRVLIGVRPESWDVVGGDA--  286
                    :*.**:.    .          : :               :::* ***  .

SugC_H37Rv          QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
MRA_RS06575         QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
MT_RS06545          QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
MAF_RS06560         QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
BQ2027_MB1270       QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
BCG_RS06720         QRIRALTFQVKVNLVESLGADKYLYFTTESPAVHSVQLDELAEVEGESALHENQFVARVP  356
MAV_RS06605         QRIRALTFEVKVDMVESLGADKYVYFSTAAWAAHSTQLDELA---AEADAHENQFVARVP  353
MAP_RS12980         QRIRALTFEVKVDMVESLGADKYVYFSTAAWAAHSTQLDELA---AEADAHENQFVARVP  353
OCU_RS31280         QRIRALTFEVKVDLVESLGADKYVYFSTAAWAAHSAQLDDLA---AEGDAHENQFVARVP  353
MMAR_RS21050        QRIRALTFEVNTDLVESLGSDKYVYFSTAGCDVHSAQLDELAGLEGESEVIENRFVARVS  356
MAB_1375            KRITGLTLTVNADVVESLGSDKYVHFTTEGGAAHSDELTELA---QESEVAENEFVARLS  353
MSMEG_5058          ARIRALTFSVRADIVESLGADKYVHFTTEGAGAESAQLAELA---ADSGAGTNQFIARVS  357
Rv2038c_H37Rv       --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
MRA_RS10790         --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
MT_RS10670          --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
MAF_RS10625         --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
BQ2027_MB2064C      --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
BCG_RS10595         --------EMDVDMVEELGADAYLYGRIVSGGC--E--------------MDQSIVARVD  325
MAV_RS11815         --------EMEVDVVEELGADAYLYGRITGSGK--V--------------IDAPIVARVD  325
MAP_RS08985         --------EMEVDVVEELGADAYLYGRITGSGK--V--------------IDAPIVARVD  325
OCU_RS37020         --------EMEVDVVEELGADAYLYGRITGSGK--V--------------IDAPIVARAD  325
MUL_RS11890         --------EMEIDVVEELGADAYLYGRISNGGT--M--------------IDQSVVARAD  325
MMAR_RS15025        --------EMEIDVVEELGADAYLYGRIINGGA--M--------------IDQSVVARAD  325
```

```
ML1424              --------EVEIDMVEELGADAYLYGRIAGASK--V--------------TDQLVVARVD  325
UgpC_H37Rv          ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
MRA_RS15055         ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
MT_RS14855          ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
MAF_RS14745         ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
BQ2027_MB2856C      ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
BCG_RS14680         ----PGSLTVHVELVEELGFESFVYATPVDQRGWSS--------------RAPRIVFRTD  328
MMAR_RS09415        ----AAALGVRVEQVEELGFESFIYATPVAQDGWSS--------------RTRRIVIRSD  328
                            :  : **.** : :::                            .: *

SugC_H37Rv          AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
MRA_RS06575         AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
MT_RS06545          AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
MAF_RS06560         AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
BQ2027_MB1270       AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
BCG_RS06720         AESKVAIGQSVELAFDTARLAVFDADSGANLTIPHRA------------  393
MAV_RS06605         AESKAAIGQTVELALDTTKLMVFDADSGVNLTVAPSDSP----------  392
MAP_RS12980         AESKAAIGQTVELALDTTKLMVFDADSGVNLTVAPSGSP----------  392
OCU_RS31280         GESKAVIGQSIELAFDTTRLVVFDADSGANLTIAPSDAR----------  392
MMAR_RS21050        AESRATLGRPIELAFDTTRLTVFDADTGANLTNPIAAVQ----------  395
MAB_1375            AASKVAEGQPIELIIDTGKLVIFDAESGENLSLAATAE-----------  391
MSMEG_5058          ADSRVRTGEQIELAIDTTKLSIFDAATGLNLTRDITPTDPTEAAGPDAG  406
Rv2038c_H37Rv       GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
MRA_RS10790         GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
MT_RS10670          GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
MAF_RS10625         GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
BQ2027_MB2064C      GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
BCG_RS10595         GRGPPERGSRVRLCPTPGHLHFFAVDGRRIPG-----------------  357
MAV_RS11815         GRNPPEKGSRVRLHPAPGHLHFFGRNGQRIGVRGSW-------------  361
MAP_RS08985         GRNPPEKGSRVRLHPAPGHLHFFGRNGQRIG-------------------  356
OCU_RS37020         GRNPPARGSRVRLHPEPGHVHFFGVDGRRLC-------------------  356
MUL_RS11890         GSNPPERGSRVRLYPQPAQLHFFTVDGRRIA-------------------  356
MMAR_RS15025        GSNPPERGSRVRLYPQPAQLHFFTVDGRRIA-------------------  356
ML1424              GRNPPAKGSRVRLYTEPGNVHFFGVDGHRIS-------------------  356
UgpC_H37Rv          RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
MRA_RS15055         RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
MT_RS14855          RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
MAF_RS14745         RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
BQ2027_MB2856C      RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
BCG_RS14680         RRTAVRVGESLAIVPHSQEVRLFNSRTETRLR-----------------  360
MMAR_RS09415        RHTTVGAGDSLSIAPNPQEVCFFDSRTESRIR-----------------  360
                           *  : :     .: .*
```

B



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 each two proteins can be evidenced by the coloured spheres. The structural models show regions of amino acid insertion/deletion identified when two proteins are compared. The percentage in each box represent the amino acid sequence identity between two orthologues. Structural models of Rv2038c, UgpC and SugC were built from the structural coordinates of *Thermococcus litoralis* MalK (PDB code: 1G29).

Additional File 3

A

```
SugB_H37Rv          ----------------------------------------------M---GARRATYWAVLD  13
MRA_RS06570         ----------------------------------------------M---GARRATYWAVLD  13
MT_RS06540          ----------------------------------------------M---GARRATYWAVLD  13
MAF_RS06555         ----------------------------------------------M---GARRATYWAVLD  13
BQ2027_MB1269       ----------------------------------------------M---GARRATYWAVLD  13
BCG_RS06715         ----------------------------------------------M---GARRATYWAVLD  13
MAV_RS06600         ------------------------------------------MAVNRT---AARRTVLWAVID  18
MAP_RS12985         ------------------------------------------MAVNRT---AARRTVLWAVID  18
OCU_RS35035         ---------------------------------------M----TSPSRAS-TALIYSGL  16
OCU_RS31275         ------------------------------------------MAANRQ---DARRTAIWAVID  18
MUL_RS23375         ---------------------MR--------------AETGTVST---GSGRRTFWVAVD  22
MMAR_RS21055        ---------------------MR--------------AETATVTT---GSGRRTFWVVID  22
MAB_1374            ----------------------------------------MSRP---DSSRTAGWLIAD  16
MSMEG_5059          ---------------------------------------MADRV---DARRATWWSVVN  17
ML_RS05365          ----------------------------------------------M---GARRATIWVIID  13
UspB_H37Rv          ---------------------------------------M----SSPSRVS-NTAVYAVL  16
MRA_RS12310         ---------------------------------------M----SSPSRVS-NTAVYAVL  16
MT_RS12145          ---------------------------------------M----SSPSRVS-NTAVYAVL  16
MAF_RS12100         ---------------------------------------M----SSPSRVS-NTAVYAVL  16
BQ2027_MB2344       ---------------------------------------M----SSPSRVS-NTAVYAVL  16
BCG_RS12070         ---------------------------------------M----SSPSRVS-NTAVYAVL  16
MAV_RS09980         ---------------------------------------M----TSPDRARANIAIYAGL  17
MAP_RS10630         ---------------------------------------M----TSPDRARANIAIYAGL  17
MUL_RS06745         ---------------------MT--------------SLS----RRLSRRLPTRVIYAGL  21
MMAR_RS18020        ---------------------MT--------------SLS----RRLSRRVPTTVIYAGL  21
MAB_1716c           ------------------------------------------------MTWSRNVLVYLLL  13
MSMEG_4467          ----------------------------------------------------MIYAGL  6
ML1769              ---------------------------------------M----SSPSRLS-NAVTYAGL  16
Rv2039c_H37Rv       ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
MRA_RS10795         ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
MT_RS10675          ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
MAF_RS10630         ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
BQ2027_MB2065C      ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
BCG_RS10600         ---------------------MG--------------W---ADRIVHRHFIRGLALYAGL  22
MAV_RS11810         ---------------------MG--------------S---SEAVIKRTVLRAATVYAAL  22
MAP_RS08990         ---------------------MG--------------S---SEAVIKRTVLRAATVYAAL  22
OCU_RS37015         ---------------------MG--------------L---SEGVIRRTALRGGMVYAAL  22
MUL_RS11895         ---------------------MG--------------S---ADRVVQRNIVRAVALYAAL  22
MMAR_RS15030        ---------------------MG--------------S---ADRVVQRKIVRAVALYAAL  22
ML1425              ---------------------MG--------------LAERLNHIVKRSVLRAVVVYIAL  25
UgpE_H37Rv          ----------------------------------------------MTPDRLRSSVGYAAM  15
MRA_RS15065         ----------------------------------------------MTPDRLRSSVGYAAM  15
MT_RS14865          ----------------------------------------------MTPDRLRSSVGYAAM  15
MAF_RS14755         ----------------------------------------------MTPDRLRSSVGYAAM  15
BQ2027_MB2858C      ----------------------------------------------MTPDRLRSSVGYAAM  15
BCG_RS14695         ----------------------------------------------MTPDRLRSSVGYAAM  15
MMAR_RS09405        ------------------------------------MSLRPRGSRAIVRVLGYVAM  20
SugA_H37Rv          ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
MRA_RS06565         ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
MT_RS06535          ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
MAF_RS06550         ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
BQ2027_MB1268       ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
BCG_RS06710         ---------------------MTSVEQRTATAVF-----SRTGSRMAERRLAFMLVAPAA  34
MAV_RS06595         ------------------------------------------MLGRTSEQRLALVLVAPAA  19
MAP_RS12990         ------------------------------------------MLGRTSEQRLALVLVAPAA  19
OCU_RS31270         ------------------------------------------MRTRNRLSEQRLALLLVAPAA  21
MMAR_RS21060        ------------------------------------------MRSRGRRAERRLAFALVAPAT  21
MAB_1373            ----------------------------MTDR-----TSSEGKRAERRLGLLLIAPAA  25
MSMEG_5060          ---------------------MT--AAVTPSASA-----VASDDKKSERRLAFWLIAPAV  32
ML_RS05360          ---------------------MT---AVVGKSWH-----VRASSVQPEQRLAFLLVTPAA  31
UspA_H37Rv          ---------------------------------------MRDAPRRRTALAYALLAPSL  20
MRA_RS12305         ---------------------------------------MRDAPRRRTALAYALLAPSL  20
MT_RS12140          ---------------------------------------MRDAPRRRTALAYALLAPSL  20
MAF_RS12095         ---------------------------------------MRDAPRRRTALAYALLAPSL  20
BQ2027_MB2343       ---------------------------------------MRDAPRRRTALAYALLAPSL  20
```

```
BCG_RS12065          ------------------------------------MRDAPRRRTALAYALLAPSL  20
MAV_RS09985          -----------------------------------M-ASAERRVAPRSTALGYALLAPSL  24
MAP_RS10625          -----------------------------------M-ASAERRVAPRSTALGYALLAPSL  24
OCU_RS35040          -----------------------------------M-AGAARRIRPRSTALGYALLAPSL  24
MUL_RS06750          ----------------------------------------MRATPRRTTALAYALLAPSL  20
MMAR_RS18015         ----------------------------------------MRATPRRTTALAYALLAPSL  20
MAB_1717c            -------------------------------------------MSRSRQTAVAYGLLAPSL  18
MSMEG_4466           ---------------------MAPWRLLT--ANR-F-NEPMATPRVRTTALAYALVAPSL  35
ML1768               MCAHFNTSMTNTGDDTEQSDEVGSISRLRGRKTH-R-TRALTKNHPRSIALAYALLAPSL  58
Rv2040c_H37Rv        ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
MRA_RS10800          ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
MT_RS10680           ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
MAF_RS10635          ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
BQ2027_MB2066C       ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
BCG_RS10605          ----------------------------------------MTRRRGRRAWAGRMFVAPNL  20
MAV_RS11805          ---------------------MTALDTEAG---RRTP--KNPSPWRRHAWAGRLFVAPNM  34
MAP_RS08995          ---------------------MTALDTGAG---RRTP--KNPSPWRRHAWAGRLFVAPNM  34
OCU_RS37010          ------------------------------------------------MFVAPNL   7
MUL_RS11900          ---------------------MTSIDTRSTPEPGAET--GRSRRWQRRSWAGRMFVAPNL  37
MMAR_RS15035         ---------------------MTSIDTRSTPEPAAET--GRSRRWQRRSWAGRMFVAPNM  37
ML1426               ---------------------MTSVETTAVPEPSIAKNHASLPPSRRRAWAGRMFIAPNL  39
UgpA_H37Rv           ---------------------MA-------APQR-A-RLRSSKERVRDYALFVVLVGPNV  30
MRA_RS15070          ---------------------MA-------APQR-A-RLRSSKERVRDYALFVVLVGPNV  30
MT_RS14870           ---------------------MA-------APQR-A-RLRSSKERVRDYALFVVLVGPNV  30
MAF_RS14760          ---------------------MA-------APQR-A-RLRSSKERVRDYALFVVLVGPNV  30
BCG_RS14700          ---------------------MA-------APQR-A-RLRSSKERVRDYALFVVLVGPNV  30
MMAR_RS09400         ---------------------MT------------T-TRPRRNRSWRDYALFVALVGPNL  26

SugB_H37Rv           TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
MRA_RS06570          TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
MT_RS06540           TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
MAF_RS06555          TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
BQ2027_MB1269        TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
BCG_RS06715          TLVVGYALLPVLWIFSLSLKPTST--VKDGKLIPSTVTFDNYRGIFR-GDLFSSALINSI  70
MAV_RS06600          TLVVVYALLPVLWIFSLSLKPTST--VKDGKLIPSAISLENYRGIFR-GDFFSSALINSV  75
MAP_RS12985          TLVVVYALLPVLWIFSLSLKPTST--VKDGKLIPSAISLENYRGIFR-GDFFSSALINSV  75
OCU_RS35035          TLGALITLAPFTLGLMTAFTSAHQFVTGTPLQLPRPPTLANFTDLAG--AGFGRAAAVTA  74
OCU_RS31275          TVVVVYALLPVLWIFSLSLKPTST--VKDGKLIPSSVSLDNYRGIFR-GDIFSSALINSV  75
MUL_RS23375          TMVVVYALLPVLWILSLSPKPTST--VKDGRLIPSSVTLDNYRGVFG-GDLFSSALINSI  79
MMAR_RS21055         TMVVVYALLPVLWILSLSLKPTST--VKDGRLIPSSVTLDNYRGVFG-GDLFSSALINSI  79
MAB_1374             VLVLCYALVPVLWVLSLSLKPTSS--VKDGKFFPWPITLDNYRGIFS-GNVFTSALVNSI  73
MSMEG_5059           ILVIVYALIPVLWILSLSLKPTSS--VKDGKLIPTEITFANYKAIFS-GDAFTSALFNSI  74
ML_RS05365           TFVVGYALLPVLWILSLSLKPAST--VKDGKLIPSLVTFDNYRGIFR-SDLFSSALINSI  70
UspB_H37Rv           TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
MRA_RS12310          TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
MT_RS12145           TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
MAF_RS12100          TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
BQ2027_MB2344        TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
BCG_RS12070          TIGAVITLSPFLLGLLTSFTSAHQFATGTPLQLPRPPTLANYADIAD--AGFRRAAVVTA  74
MAV_RS09980          LVGALITLAPFTLGLLTAFTSAHQFVTGTPLQLPRPPTLSNFADLAG--AGFGRAGAVTA  75
MAP_RS10630          LVGALITLAPFTLGLLTAFTSAHQFVTGTPLQLPRPPTLSNFADLAG--AGFGRAAAVTA  75
MUL_RS06745          VLGALITLLPFALGLLTSFTSAHQFATGTPLQLPHPPTLTNYADLGG--AGFGRAALVTA  79
MMAR_RS18020         VLGALITLLPFALGLLTSFTSAHQFATGTPLQLPHPPTLANYADLGG--AGFGRAALVTA  79
MAB_1716c            TVGAVLTLGPFLLAVSASLKTAKQFATGTPLAPPNPFTLANYTGLAD--AGFGRAALVTV  71
MSMEG_4467           LLGAVITLLPFGLGLLTSFTSAQQFVTESPLSLPRPPTLANYLGLAD--AGFGRAIAVTA  64
ML1769               TLGALITVAPFALGLLTSFTSAHQFTTGTPLQLPRPPTLANYAGLSG--AGFLRATAVTA  74
Rv2039c_H37Rv        IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
MRA_RS10795          IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
MT_RS10675           IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
MAF_RS10630          IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
BQ2027_MB2065C       IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
BCG_RS10600          IGIAWCALFPIIWALSGSLKADGE--VTEPTLFPSHPQWSNYREVFA-LMPFWRMFFNTV  79
MAV_RS11810          LGIAWCALFPIAWAVSGSLKTEGE--VSEPTLVPARPRWSNYTEVFA-LMPIGRMFVNTV  79
MAP_RS08990          LGIAWCALFPIAWAVSGSLKTEGE--VSEPTLVPARPRWSNYTEVFA-LMPIGRMFVNTV  79
OCU_RS37015          LGIAWCALFPIAWAVSGSLKTEGE--VSEPTLLPAHPRWSNYTEVFA-LLPFGRMFFNTV  79
MUL_RS11895          IAIAWCAMFPILWALSGSLKKGGE--VSEPRLLPSDPQWSNYAEVFT-LMPFWRMFFNTV  79
MMAR_RS15030         IAIAWCALFPILWALSGSLKKSGE--VSEPRLLPSDPQWSNYAEVFT-LMPFWRMFFNTV  79
ML1425               TGIAWCALFPIVWALSGSLKKEGE--IREPTLLPAHPQWSNYTEVFD-LIPFWRMFFNTV  82
UgpE_H37Rv           LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
MRA_RS15065          LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
MT_RS14865           LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
MAF_RS14755          LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
BQ2027_MB2858C       LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
BCG_RS14695          LLVVTLIAGPLLFVFFTSFKDQPDIYAQPTSWWPLRWYPQNYRTATE-QIPFWTFLRNSL  74
MMAR_RS09405         LLVMAVVAGPLLFVFFTSFKEQADIYSQPTNWWPPHWYPQNYWTATR-QIPFWTFLRNSV  79
SugA_H37Rv           MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
MRA_RS06565          MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
MT_RS06535           MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
MAF_RS06550          MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
BQ2027_MB1268        MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
BCG_RS06710          MLMVAVTAYPIGYALWLSLQRNNLATPNDT----AFIGLGNYHTILI-DRYWWTALAVTL  89
MAV_RS06595          ILMLAVTAYPIGYAVWLSLQRNNLAAPHDT----AFVGLSNYATILS-DRYWWTALAVTL  74
MAP_RS12990          ILMLAVTAYPIGYAVWLSLQRNNLAAPHDT----AFVGLSNYATILS-DRYWWTALAVTL  74
OCU_RS31270          LLMLTVTAYPIGYAVWLSLQRNNLAVPDDT----AFIGLGNYVTILT-DRYWWTALAVTL  76
MMAR_RS21060         ILMLLVIGYPIGYAMWLSLQRNNLATPNET----RFIGLGNYQTILV-DRYWWTALAVTA  76
MAB_1373             LLMLAVTAYPICYAIWLSVQKYSFASPQDR----KFVWFDNYITVLS-DRYWWSALVVTL  80
MSMEG_5060           LLMLAVTAYPIGYAVWLSLQRYNLAEPHDT----EFIGLANYVTVLT-DGYWWTAFAVTL  87
```

```
ML_RS05360          MLMLVVTAYPIGYAVWLSLQRYSLATPNET----VFIGLGNYQTILT-DPYWWTALAVTL  86
UspA_H37Rv          VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
MRA_RS12305         VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
MT_RS12140          VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
MAF_RS12095         VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
BQ2027_MB2343       VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
BCG_RS12065         VGVVAFLLLPILVVVWLSLHRWDLLGPL------RYVGLTNWRSVLT-DSGFADSLVVTA  73
MAV_RS09985         FGVLAFLLLPILVVIWLSLCRWDLLGPL------RFVGLSNWRSVLT-DAGFGNSLMVTA  77
MAP_RS10625         FGVLAFLLLPILVVIWLSLCRWDLLGPL------RFVGLSNWRSVLT-DAGFGNSLMVTA  77
OCU_RS35040         FGVLAFLLLPILVVIWLSLCRWDLLGPL------RFVGLSNWRSVLT-DGTFGNSLIVTA  77
MUL_RS06750         FGVLAFLLLPILVVIWLSLYRWDLLGPL------RYVGLANWRSVLR-DGDFGNSLVVTA  73
MMAR_RS18015        FGVLAFLLLPILVVIWLSLYRWDLLGPL------RYVGLANWRSVLR-DGDFGNSLVVTA  73
MAB_1717c           FGVAAFLLLPILVVVWLSLCRWDLLGPI------EFVGLDNWRSVLT-DGTFGHSLLVTL  71
MSMEG_4466          FGVVTFLLLPMLVVVWLSLHRWDLLGPI------EFVGLDNWTSVLT-DPAFGKSLVVTL  88
ML1768              FGVVTFLLLPIIVVIWLSLFRWDLLGPL------HYVGLANWRSVLT-DPNFANSLIVTA  111
Rv2040c_H37Rv       AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
MRA_RS10800         AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
MT_RS10680          AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
MAF_RS10635         AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
BQ2027_MB2066C      AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
BCG_RS10605         AAVVVFMLFPLGFSLYMSFQKWDLFTHA------TFVRLDNFRNLFTSDPLFLIAVVNTA  74
MAV_RS11805         VAVAVFMLFPLGFSLYMSFQRWDVFTPP------KFVGLKNFTDLFSSDPLFLIAIRNTV  88
MAP_RS08995         VAVAVFMLFPLGFSLYMSFQRWDVFTPP------KFVGLKNFTDLFSSDPLFLIAIRNTV  88
OCU_RS37010         VAVAVFMLFPLGFSLYMSFQRWDVFTPP------KFVGLKNFGELFTADPLFLIAIRNTV  61
MUL_RS11900         AAVAVFLLFPLGFSLYMSFQNWDLFRAP------TFVGLQNFAKLFTSDPLFLIALRNSV  91
MMAR_RS15035        AAVAVFLLFPLGFSLYMSFQNWDLFRAP------TFVGLQNFAKLFTSDPLFLIALRNSV  91
ML1426              ASVSVFMLFPLGFSLYMSFQKWDMFTPP------VFVGLANFQNLFTSDPLFLIALCNSV  93
UgpA_H37Rv          ALLLLFVYRPLADNIRLSFFDWNVSDPS-A----RFVGLSNYTEWFT-RSDTRQIVFNTA  84
MRA_RS15070         ALLLLFVYRPLADNIRLSFFDWNVSDPS-A----RFVGLSNYTEWFT-RSDTRQIVFNTA  84
MT_RS14870          ALLLLFVYRPLADNIRLSFFDWNVSDPS-A----RFVGLSNYTEWFT-RSDTRQIVFNTA  84
MAF_RS14760         ALLLLFVYRPLADNIRLSFFDWNVSDPS-A----RFVGLSNYTEWFT-RSDTRQIVFNTA  84
BCG_RS14700         ALLLLFVYRPLADNIRLSFFDWNVSDPS-A----RFVGLSNYTEWFT-RSDTRQIVFNTA  84
MMAR_RS09400        ALLLLFVYRPLVDNIRLSFFDWNISDPT-A----EFVGFSNYIEWFS-RPDTRQIVANTA  80
                                *.   .  :                 *:              :

SugB_H37Rv          GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
MRA_RS06570         GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
MT_RS06540          GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
MAF_RS06555         GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
BQ2027_MB1269       GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
BCG_RS06715         GIGLITTVIAVVLGAMAAYAVARLEFPGKRLLIGAALLIT-MFPSISLVTPLFNIERAIG  129
MAV_RS06600         GIGLITTAVAVLLGAMAAYAVARLDFPGKRLLIGATLLIT-MFPAISLVTPLFNIERFLG  134
MAP_RS12985         GIGLITTAVAVLLGAMAAYAVARLDFPGKRLLIGATLLIT-MFPAISLVTPLFNIERFLG  134
OCU_RS35035         LMTAVILVGQLTFSVLAAYAFARLQFPGRDALFWVYIATL-MVPATVTIVPMYLMMAQLG  133
OCU_RS31275         GIGLITTAIAVTLGAMAGYAIARLNFPGKRALVGATLLIT-MFPAISLVTPLFNIERFVG  134
MUL_RS23375         GIGLTTTVIAVLLGAMAACAVARLDFPGKRLLVGVTLLIT-MFPAISLVTPLFNIERRIG  138
MMAR_RS21055        GIGLTTTVIAVLLGAMAAYAVARLDFPGKRLLVGVALLIT-MFPAISLVTPLFNIERRIG  138
MAB_1374            GIGLIATVIAVSVGTMAAYAVARLDFPGKKALIGAALLIA-MFPQISLVTPIFNIERSVG  132
MSMEG_5059          GIGLITTIIAVVIGGMAAYAVARLQFPGKQLLIGVALLIA-MFPHISLVTPIFNMWRGIG  133
ML_RS05365          GIGLTTTVIAVMFGAMAAYAIARLAFPGKRLLVGTTLLVT-MFPAISLVTPLFNIERRVG  129
UspB_H37Rv          LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
MRA_RS12310         LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
MT_RS12145          LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
MAF_RS12100         LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
BQ2027_MB2344       LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
BCG_RS12070         LMTAVILLGQLTFSVLAAYAFARLQFRGRDALFWVYVATL-MVPGTVTVVPLYLMMAQLG  133
MAV_RS09980         LMTAVILVGQLTFSVLAGYAFARLRFPGRDALFWVYIATL-MVPGTVTIVPMYLMMAQLG  134
MAP_RS10630         LMTAVILVGQLTFSVLAGYAFARLRFPGRDALFWVYIATL-MVPGTVTIVPMYLMMAQLG  134
MUL_RS06745         LMTTVILLGQMTFSVLAAYAFARLEFPGRDGLFWVYIATL-MVPATVTVVPLYLMMAQLG  138
MMAR_RS18020        LMTAVILLGQMTFSVLAAYAFARLEFPGRDGLFWVYIATL-MVPATVTVVPLYLMMAQLG  138
MAB_1716c           LMTAIIVIGQLTFSVLAAYAFARLRFPGRDGLFWIYLATL-MIPATVTVVPMYLMLTQVG  130
MSMEG_4467          LMTAVILLGQLVFSVLAAYAFARLEFPGRDVLFWVYLATL-MVPATVTVVPLYLIMAEAG  123
ML1769              LMTAVILLGQLTFSVLAGYAFARLRFPGRDTLFWAYIMTL-MVPATVTVVPMYMMMAQVG  133
Rv2039c_H37Rv       LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQVILMRIVG  138
MRA_RS10795         LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQVILMRIVG  138
MT_RS10675          LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQFILMRIVG  138
MAF_RS10630         LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQFILMRIVG  138
BQ2027_MB2065C      LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQFILMRIVG  138
BCG_RS10600         LYAGCVTAGQVFFCSLAGYAFARLQFRGRDTLFVLYLSTL-MVPLTVTVIPQFILMRIVG  138
MAV_RS11810         LYAGCVTAGHVFFCSLAGYAFARLDFRGRNTLFVVYLGTL-MVPLTVTVIPQFLIMRTLG  138
MAP_RS08990         LYAGCVTAGHVFFCSLAGYAFARLDFRGRNTLFVVYLGTL-MVPLTVTVIPQFLIMRTLG  138
OCU_RS37015         LYAGCVTAGHVFFCSLAGYAFARLDFRGRDALFVVYLGTL-MVPLTVTVIPQFLVMRTLG  138
MUL_RS11895         LYAGCVTAGQIFFCSLAGYAFARLEFRGRDTLFVLYLGTL-MVPLTVTVIPHFILMRTVG  138
MMAR_RS15030        LYAGCVTAGQIFFCSLAGYAFARLEFRGRDTLFVLYLGTL-MVPLTVTVIPQFILMRTVG  138
ML1425              LYAGYVTVGQVFFCSLAGYAFARLQFRGRDALFVLYLGTL-MMPLTVTIIPQFILMRVLG  141
UgpE_H37Rv          IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
MRA_RS15065         IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
MT_RS14865          IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
MAF_RS14755         IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
BQ2027_MB2858C      IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
BCG_RS14695         IITSVLAVVKFTLGVLSAFGLVFVRFPGRTAVFLVIIAAL-MVPNQITVISNYALISHLG  133
MMAR_RS09405        IITSVLAVVKFALGVFSAFGLVFVQFPGKTPVFLLIIAAL-MVPNQITVISNYALIAQLG  138
SugA_H37Rv          AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
MRA_RS06565         AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
MT_RS06535          AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
MAF_RS06550         AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
BQ2027_MB1268       AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
BCG_RS06710         AITAVSVTIEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTVVASYSWYYAWTPGTG  148
```

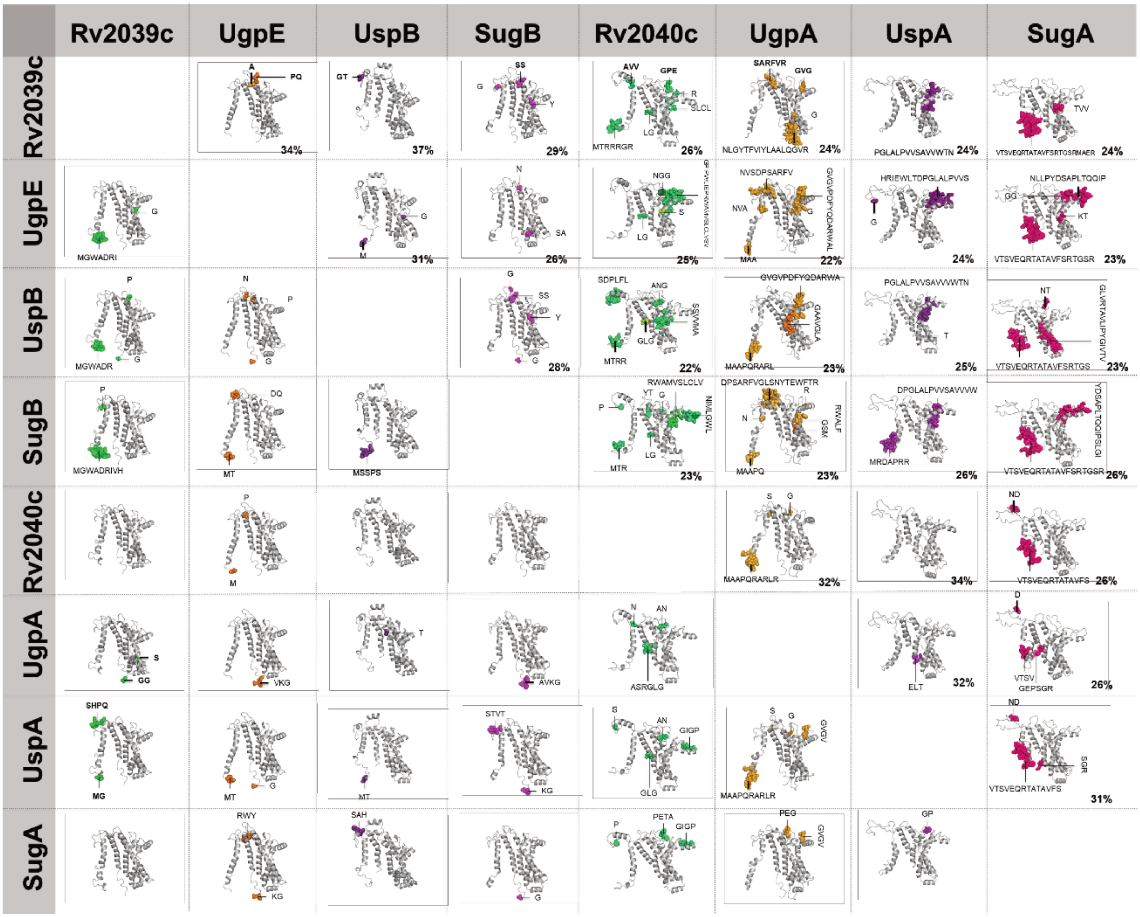
```
MAV_RS06595          GITVVSVSAEFVLGLALALVMHR-TLVGKGLVRTAVLIPYGIVTAVASYSWYYAWTPGTG  133
MAP_RS12990          GITVVSVSAEFVLGLALALVMHR-TLIGKGLVRTAVLIPYGIVTAVASYSWYYAWTPGTG  133
OCU_RS31270          GITVVSVTVEFVLGLTLALVMHR-TLVGKGMVRTAILIPYGIVTVVASYSWYYAWTPGTG  135
MMAR_RS21060         AITVVSVACEFVLGLALALVMHR-SLVAKGLVRTAVLVPYGIVTVVASYSWYYAWTPGTG  135
MAB_1373             VITVISVLIELVLGLALALVMHR-TIFGRGVVRTAVLIPYGIVTVAASYSWYYAWTPGTG  139
MSMEG_5060           GITVVSVAIEFALGLALALVMHR-TIFGKGAVRTAILIPYGIVTVAASYSWYYAWTPGTG  146
ML_RS05360           AITVVSVSIEFILGLMLALVMHR-TLLGKSLVRIAVLIPYSIVTVVASYSWYYAWTPGTG  145
UspA_H37Rv           VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWIVAPTDG  132
MRA_RS12305          VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWIVAPTDG  132
MT_RS12140           VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
MAF_RS12095          VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
BQ2027_MB2343        VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
BCG_RS12065          VFVAIVVPAQTVLGLLAASLLAR-RLPGTGLFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
MAV_RS09985          VFVAMVVPAQTALGLLAATMLAR-RLPGTGLFRTVYVLPWICAPLAIAVLWRWILAPTDG  136
MAP_RS10625          VFVAMVVPAQTALGLLAATMLAR-RLPGTGLFRTVYVLPWICAPLAIAVLWRWILAPTDG  136
OCU_RS35040          IFVAIVVPAQTALGLLAASMLAR-QLRGTGLFRTVFVLPWICAPLAIAVLWRWILAPTDG  136
MUL_RS06750          IFVAIVVPAQTVLGLLAAMMLTR-RLPGTNFFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
MMAR_RS18015         IFVAIVVPAQTVLGLLAAMMLTR-RLPGTNFFRTLYVLPWICAPLAIAVMWRWILAPTDG  132
MAB_1717c            LFVALVIPVQTALGLAAATLLVR-GLPGTVLFRTIYVIPWICAPLAIGVLWHWMLAPTDG  130
MSMEG_4466           LFMALVVPTQTVLGLVAATMLAR-GLPGTGFFRTVYVLPWICAPLAIAVLWRWILAPTDG  147
ML1768               VFVAIVVPTQTMLGLLIATMLAR-QLPGTGVFRTLYVLPWICAPLAIAVLWRWLLAPTDG  170
Rv2040c_H37Rv        VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
MRA_RS10800          VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
MT_RS10680           VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
MAF_RS10635          VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
BQ2027_MB2066C       VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
BCG_RS10605          VYTVGTVVPTVIVSLVVAAFLNR-KIKGISLFRTVVFLPLAISSVVMAVVWQFVFNTDNG  133
MAV_RS11805          IFTLGSVVPTVAISLAVAGVLNQ-KVRGIGIFRTIVFLPLAISSVVMAVVWQFVFNTDNG  147
MAP_RS08995          IFTLGSVVPTVAISLAVAGVLNQ-KVRGIGIFRTIVFLPLAISSVVMAVVWQFVFNTDNG  147
OCU_RS37010          IFTLGTVVPTVVISLVVAGVLNQ-KARGIGIFRTIVFLPLAISSVVMAVVWQFVFNTDNG  120
MUL_RS11900          VYTAGTVVPTVLISLVVAAVLNR-KVPGIGVFRTIVFLPLAISSVVMAVVWQFVFDTNNG  150
MMAR_RS15035         VYTAGTVVPTVLISLVVAAVLNR-KVPGIGVFRTIVFLPLAISSVVMAVVWQFVFDTNNG  150
ML1426               VFTVGTVIPTVLISLVVAGVLNQ-KVKGIGIFRTIVFLPLAISSVVMAVVWQFIFNTHNG  152
UgpA_H37Rv           VFTGAAVVGSMVLGLALAMLLDR-PLRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  143
MRA_RS15070          VFTGAAVVGSMVLGLALAMLLDR-PLRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  143
MT_RS14870           VFTGAAVVGSMVLGLALAMLLDR-PLRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  143
MAF_RS14760          VFTGAAVVGSMVLGLALAMLLDR-PWRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  143
BCG_RS14700          VFTGAAVVGSMVLGLALAMLLDR-PLRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  143
MMAR_RS09400         IFTTAAVAGSMVLGLLLAMLLDQ-PLRGRNLVRSTVFAPFVISGAAVGLAAQFVFDPHFG  139
                                           .    .    .    .    .    .           *

SugB_H37Rv           LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
MRA_RS06570          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
MT_RS06540           LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
MAF_RS06555          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
BQ2027_MB1269        LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
BCG_RS06715          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  170
MAV_RS06600          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKI  175
MAP_RS12985          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKI  175
OCU_RS35035          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRMIPNDLVNAARL  173
OCU_RS31275          LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  175
MUL_RS23375          LFD----------------TWPELILP---YITFALPLAIYTLSAFFREIPWDLEKAAQM  179
MMAR_RS21055         LFD----------------TWPGLILP---YITFALPLAIYTLSAFFREIPWDLEKAAQM  179
MAB_1374             LFD----------------TWPGLIIP---YITFALPLAIYTLSAFFREIPWELEKAAKM  173
MSMEG_5059           LFD----------------TWPGLIIP---YITFALPLAIYTLSAFFREIPWDLEKAAKM  174
ML_RS05365           LFD----------------TWAGLILP---YITFALPLAIYTLSAFFAEIPWDLEKAAKM  170
UspB_H37Rv           LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
MRA_RS12310          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
MT_RS12145           LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
MAF_RS12100          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
BQ2027_MB2344        LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
BCG_RS12070          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRLIPDDLINAARL  173
MAV_RS09980          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRMIPNDLVNAARL  174
MAP_RS10630          LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRMIPNDLVNAARL  174
MUL_RS06745          LRN----------------TFWALVLP---FL-FGSPYAIFLLREHFRIIPNDMINAARL  178
MMAR_RS18020         LRN----------------TFWALVLP---FL-FGSPYAIFLLREHFRIIPNDMINAARL  178
MAB_1716c            LRN----------------TFWALVIP---FM-FGSPYAIFLLREHFRTIPDDLIRAARL  170
MSMEG_4467           LRN----------------TFWALVLP---FV-FGSPYAIFLLREYFRGIPGDLINAARL  163
ML1769               LRN----------------TFWALVLP---FM-FGSPYAIFLLREHFRIIPNDLINAARL  173
Rv2039c_H37Rv        WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MRA_RS10795          WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MT_RS10675           WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MAF_RS10630          WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
BQ2027_MB2065C       WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
BCG_RS10600          WVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MAV_RS11810          WVD----------------TPWAMIVP---GF-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MAP_RS08990          WVD----------------TPWAMIVP---GF-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
OCU_RS37015          WVD----------------TPWAMIVP---GF-FGSAFGTYLMRQFFRTLPTDLEEAAIL  178
MUL_RS11895          LVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPSDLEKAATL  178
MMAR_RS15030         LVD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPSDLEEAATL  178
ML1425               WTD----------------TPWAMIVP---GL-FGSAFGTYLMRQFFRTLPSDLEEAAIL  181
UgpE_H37Rv           LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
MRA_RS15065          LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
MT_RS14865           LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
MAF_RS14755          LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
BQ2027_MB2858C       LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
BCG_RS14695          LRN----------------TFAGIILP----L-AGVAFGTFLMRNHFLSLPAEIIEAARM  172
MMAR_RS09405         LRN----------------TFPGIILP----L-AGVAFGTFLMRNHFLSLPSEVIEAARM  177
```

```
SugA_H37Rv          YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
MRA_RS06565         YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
MT_RS06535          YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
MAF_RS06550         YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
BQ2027_MB1268       YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
BCG_RS06710         YLANLLPYDS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLALVPEDLLRAAQV  204
MAV_RS06595         YLANLLPHGS----APLTAQIPSLAIVVLAEVWKTTPFMSLLLLAGLALVPEDLLKAAQV  189
MAP_RS12990         YLANLLPHGS----APLTAQIPSLAIVVLAEVWKTTPFMSLLLLAGLALVPEDLLKAAQV  189
OCU_RS31270         YLANLLPHGS----APLTSQIPSLAIVVVAEVWKTTPFMSLLLLAGLALVPEELLKAAQV  191
MMAR_RS21060        YLANLLPQGS----APLTQQIPSLGIVVIAEVWKTTPFMSLLLLAGLAMVPEDLLQAAQV  191
MAB_1373            YLANLLPAGS----APLTQQIPSLAVVILAEVWKTTPFMALLLLAGLALVPDDLLKAAQV  195
MSMEG_5060          YLANLLPEGS----APLTDQLPSLAIVVLAEVWKTTPFMALLLLAGLALVPQDLLNAAQV  202
ML_RS05360          YLASLLPQGS----VPLTEQIPSLSIVIIVEVWKTTPFMSLLLLVGLALVPEDLLKAAQM  201
UspA_H37Rv          AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
MRA_RS12305         AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
MT_RS12140          AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
MAF_RS12095         AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
BQ2027_MB2343       AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
BCG_RS12065         AISTVLGHRI----EWLTDPGLALPVVSAVVVWTNVGYVSLFFLAGLMAIPQDIHNAART  188
MAV_RS09985         AVSAVLGHSI----EWLSDPSFALPLVSAVVVWTNVGYVSLSFLAGLLAIPEDIHAAART  192
MAP_RS10625         AVSAVLGHSI----EWLSDPSFALPLVSAVVVWTNVGYVSLSFLAGLLAIPEDIHAAART  192
OCU_RS35040         AVSTVLGHSI----EWLSDPSFALPLVSAVVVWTNVGYVSLSFLAGLLAIPDEVHAAART  192
MUL_RS06750         AISTLLGRRI----EWLTDPDLALPVVAAVVIWTNVGYVSLSFLAGLLAIPDDIHAAART  188
MMAR_RS18015        AISTLLGRRI----EWLTDPDLALPVVAAVVIWTNVGYVSLSFLAGLLAIPDDIHAAART  188
MAB_1717c           AVNALLGQRV----EWLTDPSLALPVVAAVSVWTNVGYVALFFMAGLLAIPGDVHNAARV  186
MSMEG_4466          ALSTVLGTRI----EWLTDPSLALPVVSAVVVWTNVGYVTLFFLAGILNIPADLHNAART  203
ML1768              AFSMVLRHRI----EWLSDPSLALPVVSAVVVWTNVGYVSLSFLAGLLSIPEDINAAART  226
Rv2040c_H37Rv       LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
MRA_RS10800         LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
MT_RS10680          LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
MAF_RS10635         LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
BQ2027_MB2066C      LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
BCG_RS10605         LLNIMLGWLGIGPIPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAARI  193
MAV_RS11805         LLNIMLGWVGLGPVPWLVEPRWAMASLCIVSVWRSVPFATVVLLAAMQGVPETVYEAARI  207
MAP_RS08995         LLNIMLGWVGLGPVPWLVEPRWAMASLCIVSVWRSVPFATVVLLAAMQGVPETVYEAARI  207
OCU_RS37010         LLNIMLGWVGLGPVPWLVEPHWAMVSLCIVSVWRSVPFAAVVLLAAMQGVPGTVYEAAKI  180
MUL_RS11900         LLNIILGWIGVGPVPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAAKI  210
MMAR_RS15035        LLNIILGWIGVGPVPWLIEPRWAMVSLCLVSVWRSVPFATVVLLAAMQGVPETVYEAAKI  210
ML1426              LLNIMLGWIGIGPIPWLVNPGWAMASLCIVSVWRSVPFATVVLLAAMQEVPKTVYEAAKI  212
UgpA_H37Rv          LIQDLLRRIGVGVPDFYQDARWALFMVTITYVWKNLGYTFVIYLAALQGVRRDLLEAAEI  203
MRA_RS15070         LIQDLLRRIGVGVPDFYQDARWALFMVTITYVWKNLGYTFVIYLAALQGVRRDLLEAAEI  203
MT_RS14870          LIQDLLRRIGVGVPDFYQDARWALFMVTITYVWKNLGYTFVIYLAALQGVRRDLLEAAEI  203
MAF_RS14760         LIQDLLRRIGVGVPDFYQDARWALFMVTITYVWKNLGYTFVIYLAALQGVRRDLLEAAEI  203
BCG_RS14700         LIQDLLRRIGVGVPDFYQDARWALFMVTITYVWKNLGYTFVIYLAALQGVRRDLLEAAEI  203
MMAR_RS09400        LVQDLLHRVGVNAPDFYQDTPWALFMVTVTYVWKNLGYTCVIYLAALQGVRRDLLEAAEI  199
                                          :       .               :  :   :  **

SugB_H37Rv          DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
MRA_RS06570         DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
MT_RS06540          DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
MAF_RS06555         DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
BQ2027_MB1269       DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
BCG_RS06715         DGATPGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAA-I-TAPVAIA  226
MAV_RS06600         DGATPAQAFRKVIVPLAAPGVVTAAILVFIFAWNDLLLALTLTA--TKAA-I-TAPVAIV  231
MAP_RS12985         DGATPAQAFRKVIVPLAAPGVVTAAILVFIFAWNDLLLALTLTA--TKAA-I-TAPVAIV  231
OCU_RS35035         DGANTLDEIVHVVIPSSRPVLAALALITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
OCU_RS31275         DGASPAQAFRKVIVPLAAPGVVTAAILVFIFAWNDLLLALTLTA--TKAA-I-TAPVAIV  231
MUL_RS23375         DGATPYQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAS-I-TAPIAIA  235
MMAR_RS21055        DGATPYQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTA--TKAS-I-TAPVAIA  235
MAB_1374            DGATPAQAFWKVIAPLATPGIVTSAILVFIFAWNDLLLAISLTA--TDRS-I-TAPVAIA  229
MSMEG_5059          DGATPAQAFRKVIAPLAAPGIVTAAILVFIFAWNDLLLALSLTA--TQRA-I-TAPVAIA  230
ML_RS05365          DGATSGQAFRKVIVPLAAPGLVTAAILVFIFAWNDLLLALSLTS--TKAA-I-TAPVAIT  226
UspB_H37Rv          DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
MRA_RS12310         DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
MT_RS12145          DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
MAF_RS12100         DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
BQ2027_MB2344       DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
BCG_RS12070         DGANTLDVIVHVVIPSSRPVLAALAMITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  229
MAV_RS09980         DGAHTLDVIVHVVIPSSRPVLAALALITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  230
MAP_RS10630         DGAHTLDVIVHVVIPSSRPVLAALALITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  230
MUL_RS06745         DGANTLDVLVHVVIPSNRPVLAAMTLITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  234
MMAR_RS18020        DGANTLDLLVHVVIPSSRPVLAAMTLITVVSQWNNFMWPLVITS--GHKW-R-VLTVATA  234
MAB_1716c           DGANTLDILWHVVVPVSRPILITLTLITVVSQWNNFMWPLVITS--GGNW-R-VLTVATA  226
MSMEG_4467          DGANTLDIITHVVVPASRPILVTLAMITVVSQWNNFLWPLVISS--GDTW-R-VLTVATA  219
ML1769              DGANTLDVIVHVVIPSNRSVLAALTLITVVSQWNNFMWPLVITS--GNKW-R-VLTVATA  229
Rv2039c_H37Rv       DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
MRA_RS10795         DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
MT_RS10675          DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
MAF_RS10630         DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
BQ2027_MB2065C      DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
BCG_RS10600         DGCSPWQIYWRILLPHSRPAVLVLGVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
MAV_RS11810         DGCTPWQVYWRILLPHARPAVMVLAVLTWVNVWNDFLWPLLMIQ--RSSL-A-TLTLGLV  234
MAP_RS08990         DGCTPWQVYWRILLPHARPAVMVLAVLTWVNVWNDFLWPLLMIQ--RSSL-A-TLTLGLV  234
OCU_RS37015         DGCSPWQVYWRILLPHARPAVMVLAVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  234
MUL_RS11895         DGCSPWQIYWRILLPHARPAVMVLGVLTWVNVWNDFLWPLLMVQ--RDSI-A-TLTLGLV  234
MMAR_RS15030        DGCSPWQIYWRILLPHARPAVMVLGVLTWVNVWNDFLWPLLMVQ--RDSI-A-TLTLGLV  234
ML1425              DGCSPWQIYWRVLLPHAKPAVGVLAVLTWVNVWNDFLWPLLMIQ--RNSL-A-TLTLGLV  237
UgpE_H37Rv          DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
```

```
MRA_RS15065          DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
MT_RS14865           DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
MAF_RS14755          DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
BQ2027_MB2858C       DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
BCG_RS14695          DGARWWQLLLRVVLPMSRPTMVAVGVITVVNEWNEYLWPFLMSD--DESV-A-PLPIGLT  228
MMAR_RS09405         DGARWWQLLFRVVLPMSGPTMVAFGIITVVNEWNEYLWPFLMSD--DESV-A-PLPVGLT  233
SugA_H37Rv           DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
MRA_RS06565          DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
MT_RS06535           DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
MAF_RS06550          DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
BQ2027_MB1268        DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
BCG_RS06710          DGASAWRRLTKVILPMIKPAIVVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  259
MAV_RS06595          DGAGAWRRLTRVTLPIIKPAVVVALLFRTLDAFRIFDNIYVLTN--GANN---TGSVSML  244
MAP_RS12990          DGAGAWRRLTRVTLPIIKPAVVVALLFRTLDAFRIFDNIYVLTN--GANN---TGSVSML  244
OCU_RS31270          DGAGPWRRLTGVILPIIKPAVVVALLFRTLDAFRIFDNIYVLTN--GANN---TDSVSIL  246
MMAR_RS21060         DGAGAWRRLTKIILPLIKPAVLVALLFRTLDAFRIFDNIYVLTG--GDND---TASVSIL  246
MAB_1373             DGANGWSRLIRVTVPIMKPAILVALLFRTLDAFRIFDNIYVLTG--GSNN---TGSVSIL  250
MSMEG_5060           DGAGPWKRLTKVILPMIKPAILVALLFRTLDAFRIFDNIYILTG--GSND---TGSVSIL  257
ML_RS05360           DGAGAWRRLTKIILPIIKPAVMVALLFRTLDAFRIFDNIYVLTR--GVNN---TDSVSIL  256
UspA_H37Rv           DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
MRA_RS12305          DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
MT_RS12140           DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
MAF_RS12095          DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
BQ2027_MB2343        DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
BCG_RS12065          DGASAWQRFWRITLPMLRPTMFFVLVTGIISAAQVFDTVYALTG--GGPQGS-TDLVAHR  245
MAV_RS09985          DGANAWQRFWRITMPMLRPTTFFVLVTGIVSSAQVFDTVYALTG--GGPAGS-TDLVAHR  249
MAP_RS10625          DGANAWQRFWRITMPMLRPTTFFVLVTGIVSSAQVFDTVYALTG--GGPAGR-TDLVAHR  249
OCU_RS35040          DGANAWQRFWRITIPMLRPTTFFVLVTGIVSTAQVFDTVYALTG--GGPAGS-TDLVAHR  249
MUL_RS06750          DGANAWQRFWRITLPMLRPTTFFVLVTGIVSAAQIFDIVYALTG--GGPEGS-TELVAHQ  245
MMAR_RS18015         DGANAWQRFWRITLPMLRPTTFFVLVTGIVSAAQIFDIVYALTG--GGPEGS-TELVAHQ  245
MAB_1717c            DGATAWQRFRRITLPMLRPTFFFVLVTGVVSAVQTFDTVYALTS--GGPQNR-TDLAAHR  243
MSMEG_4466           DGADAWQRFRHITLPMLRPTMFFVLVTGIVSAAQVFDTVYALTA--GGPQGR-TDLIAHR  260
ML1768               DGADAWQRFWRITLPMLRPTMFFVLVTGIVSTAQVFDTVYALTD--GGPKSR-TDLVAHR  283
Rv2040c_H37Rv        DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
MRA_RS10800          DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
MT_RS10680           DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
MAF_RS10635          DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
BQ2027_MB2066C       DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
BCG_RS10605          DGAGEIRQFVSITVPLIRGALSFVVVISIIHAFQAFDLVYVLTGANGGPETA-TYVLGIM  252
MAV_RS11805          DGAGEIRQFFAITVPLIRGSISFVVIISIIHAFQAFDMVYVLTGPSGGPESA-TYVLGIM  266
MAP_RS08995          DGAGEIRQFFAITVPLIRGSISFVVIISIIHAFQAFDMVYVLTGPSGGPESA-TYVLGIM  266
OCU_RS37010          DGASEIRQFVSITVPLIRGSISFVVIISVIHAFQAFDMVYVLTGANGGPESA-TYVLGIM  239
MUL_RS11900          DGAGEIRQFVSITVPLIGGVMSFVVVISIIHAFQAFDLVYVLTGPNGGPETG-TYVLGIM  269
MMAR_RS15035         DGAGEIRQFVSITVPLIRGAMSFVVVISIIHAFQAFDLVYVLTGPNGGPETG-TYVLGIM  269
ML1426               DGAGEIRQFISITVPFIRGAISFVVVISFIHAFQTFDLVYVLNGPNGGPELA-TYVLGIM  271
UgpA_H37Rv           DGASRWAVFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  261
MRA_RS15070          DGASRWAVFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  261
MT_RS14870           DGASRWAVFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  261
MAF_RS14760          DGASRWAVFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  261
BCG_RS14700          DGASRWAVFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  261
MMAR_RS09400         DGASRWTSFRRVLLPQLRPTTFFLSITVLINSLQVFDVINVMTR--GGPEGTGTTTMVYQ  257
                     **.       :  *         :   :   .       :

SugB_H37Rv           NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
MRA_RS06570          NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
MT_RS06540           NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
MAF_RS06555          NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
BQ2027_MB1269        NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
BCG_RS06715          NFTG-SSQFEEPTGSIAAGAIVITIPIIVF--VLIFQRRIVAGLTSGAVKG-         274
MAV_RS06600          SFSG-SSQFEEPTGSIAAGAVVITVPVILF--VLIFQRRIVAGLTSGAVKG-         279
MAP_RS12985          SFSG-SSQFEEPTGSIAAGAVVITVPIILF--VLIFQRRIVAGLTSGAVKG-         279
OCU_RS35035          DLQ---SRFNSQWTLVMAATTIAIAPLIVL--FVAFQRHIVASIVVSGLK--         274
OCU_RS31275          NFSG-SSQFEEPTGSIAAGAVVITVPIIAF--VLIFQRRIVAGLTSGAVKG-         279
MUL_RS23375          NFAG-SSQFEEPTGSIAAGAMVITVPIIVF--VLIFQLRIVAGLTSGAVKG-         283
MMAR_RS21055         NFAG-SSQFEEPTGSIAAGAMVITVPIIVF--VLIFQRRIVAGLTSGAVKG-         283
MAB_1374             NFTG-SSQFEEPTGSIAAGAIVITVPIIIF--VLIFQRRIVAGLTSGAVKG-         277
MSMEG_5059           NFTG-SSQFEEPTGSIAAGAMVITIPIIIF--VLIFQRRIVAGLTSGAVKG-         278
ML_RS05365           NFTG-SSQFEEPTGSIAAGAIVITAPIIVF--VWIFQRRIVAGLTSGAVKG-         274
UspB_H37Rv           DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
MRA_RS12310          DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
MT_RS12145           DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
MAF_RS12100          DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
BQ2027_MB2344        DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
BCG_RS12070          DLQ---SRFNDQWTLVMAATTVAIVPLIAL--FVTFQRHIVASIVVSGLK--         274
MAV_RS09980          DLQ---TRFNAQWTLVMAATTIAIVPLVTL--FVIFQRHIVASIVVSGLK--         275
MAP_RS10630          DLQ---TRFNAQWTLVMAATTIAIVPLVTL--FVIFQRHIVASIVVSGLK--         275
MUL_RS06745          DLQ---SRFNAQWTLVMAATTVAIVPLIVL--FVAAQRHIVSSIAVSGLK--         279
MMAR_RS18020         DLQ---SRFNAQWTLVMAATTVAIVPLIVL--FVAAQRHIVSSIVVSGLK--         279
MAB_1716c            GLQ---SQFNAQWPIVMAATTVAIVPLVIL--FVAFQKYIVRSITLSGLK--         271
MSMEG_4467           GLQ---TQYNAQWTLVMAATTVAIVPLIVV--FATLSRHIVRSIVVTGIK--         264
ML1769               DLQ---SRFNAQWTLVMAATTVAIVPLIVL--FVGFQRHIIPSIVVSGLK--         274
Rv2039c_H37Rv        RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
MRA_RS10795          RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
MT_RS10675           RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
MAF_RS10630          RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
BQ2027_MB2065C       RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
BCG_RS10600          RLR---GEYVARWPVLMAASMLMLVPLVIL--YAVAQRSFVRGIAVTGLGG-         280
MAV_RS11810          RMK---GEYVARWPVLMAASMLIMLPLVII--YAIAQRSFVRGIAVTGMGG-         280
```

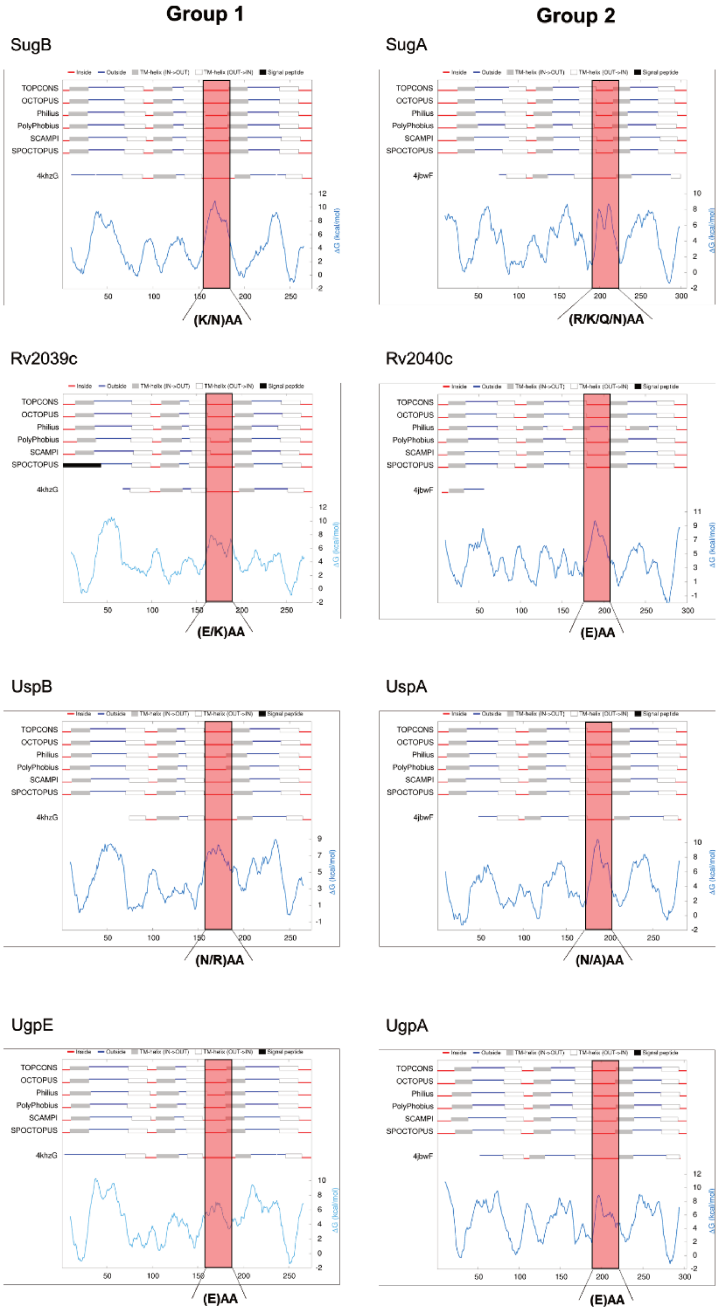
```
MAP_RS08990          RMK---GEYVARWPVLMAASMLIMLPLVII--YAIAQRSFVRGIAVTGMGG-          280
OCU_RS37015          RMK---GEYVARWPVLMATSILIMLPLVVI--YAFAQRSFVRGIAVTGMGG-          280
MUL_RS11895          RMQ---GEYVARWPVLMATSMLILLPLVAV--YAVAQRAFIRGIAVTGLGG-          280
MMAR_RS15030         RMQ---GEYVARWPVLMATSMLILLPLVVV--YAVAQRAFIRGIAVTGLGG-          280
ML1425               RMR---GEYGTCWPVIMATSMLIILPLVII--YTIAQRAFVRGITVTRIGG-          283
UgpE_H37Rv           FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
MRA_RS15065          FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
MT_RS14865           FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
MAF_RS14755          FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
BQ2027_MB2858C       FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
BCG_RS14695          FLQ--QAEGVTNWGPVMAVTLLAMLPILLV--FIALQRQMIKGLTSGAVKG-          275
MMAR_RS09405         FLQ--QAEGVTNWGPVMAVTLLAMLPILVI--FIGLQRRMIKGLTSGAVKG-          280
SugA_H37Rv           GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
MRA_RS06565          GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
MT_RS06535           GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
MAF_RS06550          GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
BQ2027_MB1268        GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
BCG_RS06710          GYDNLFKGFNVGLGSAISVLIFGCVAVIAFIFIKLFGAAA----PGGEPSGR          307
MAV_RS06595          GYDNLFKGFNVGLGSAISVLIFGCVGLIALVFVKVFGAAA----PGGDVDGR          292
MAP_RS12990          GYDNLFKGFNVGLGSAISVLIFGCVGLIALVFVKVFGAAA----PGGDVDGR          292
OCU_RS31270          GYDNLFKGFNVGLGSAISVLIFVCVGLIALLFIKVFGAAA----PGGDVDGR          294
MMAR_RS21060         GYDNLFKGFNVGLGSAISVLIFGCVTLIAVVFITLFGAHQFHSAPGGVTDAR          298
MAB_1373             GYDNLFKAFNVGLGSAISVLIFLCVAIIAVIFIKGFGASAPTT--DGEEARR          300
MSMEG_5060           GYDNLFKAFNVGLGSAISVLIFLSVAIIAFIYIKIFGAAA----PGSDEEVR          305
ML_RS05360           GYDNLFKGFNVGLGSAISVLIFGCVALIAVIFIKVFGVVA----PGGDSNGY          304
UspA_H37Rv           IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
MRA_RS12305          IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
MT_RS12140           IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
MAF_RS12095          IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
BQ2027_MB2343        IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
BCG_RS12065          IYAEAFGAAAIGRASVMAVVLFVILVGATVVQHLYFRRRISYELT-------          290
MAV_RS09985          IYAEAFGSAAIGRASVMAVVLFVILIGVTLVQHLYFRRRISYDLT-------          294
MAP_RS10625          IYAEAFGSAAIGRASVMAVVLFVILIGVTLVQHLYFRRRISYDLT-------          294
OCU_RS35040          IYAEAFGSAAIGRASVMAVVLFVILIGVTLLQHLYFRRRISYDLT-------          294
MUL_RS06750          IYAEAFGSAAIGRASVMAVVLFVILIGVTFVQHMYFRRRISYDLT-------          290
MMAR_RS18015         IYAEAFGSAAIGRASVMAVVLFVILIGVTFVQHMYFRRRISYDLT-------          290
MAB_1717c            VYAEAFEAAHVGRSAAMSLILFLILVTITVIQHRYFRARVSYDLV-------          288
MSMEG_4466           IYAEAFGAAAVGRAAVMALVLFVLLVGVTIVQHLYFRRRISYELT-------          305
ML1768               IYAEAFGSAAIGRASVMAVLLFVVLFGVTVVQHLYFQRRISYELT-------          328
Rv2040c_H37Rv        LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
MRA_RS10800          LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
MT_RS10680           LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
MAF_RS10635          LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
BQ2027_MB2066C       LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
BCG_RS10605          LFQHAFSFLEFGYASALAWVMFAILLVLTVLQLRITHRRSW-EASRGLG---          300
MAV_RS11805          LFQHAFSFLEFGYASALAWVMFAVLLVLTVVQLRVSHRRSL-ETSRGLK---          314
MAP_RS08995          LFQHAFSFLEFGYASALAWVMFAVLLVLTVVQLRVSHRRSL-ETSRGLK---          314
OCU_RS37010          LFQHAFSFLEFGYASALAWVMFAVLLVLTVVQLKVSHRRSL-ETSRGLK---          287
MUL_RS11900          LFQHAFSFLEFGYASALAWVIFAILLVLTVLQLRITRRRSW-EASSGLS---          317
MMAR_RS15035         LFQHAFSFLEFGYASALAWVIFAILLVLTVLQLRITRRRSW-EASSGLS---          317
ML1426               LFQHAFSFLEFGYASALALVTFAILLVLIVLQLLLNRRHSW-EVSGGLS---          319
UgpA_H37Rv           VYVETFRNFRAGYGATVATIMFLVLLAVTYYQVRVMDRGQRQ----------          303
MRA_RS15070          VYVETFRNFRAGYGATVATIMFLVLLAVTYYQVRVMDRGQRQ----------          303
MT_RS14870           VYVETFRNFRAGYGATVATIMFLVLLAVTYYQVRVMDRGQRQ----------          303
MAF_RS14760          VYVETFRNFRAGYGATVATIMFLVLLAVTYYQVRVMDRGQRQ----------          303
BCG_RS14700          VYVETFRNFRAGYGATVATIMFLVLLAVTYYQVRVMDRGQRQ----------          303
MMAR_RS09400         VYLETFRNFRAGYGATIATMMFLFLLVVTYYQVRVMDRGQQQ----------          299
                                          :   .
```

B



Additional file 3. Structural and amino acid sequence differences found in the *M. tuberculosis* carbohydrate TMD components. **A** Amino acid sequence alignment of TMDs performed with Clustal Omega. Proteins can be divided in group 1 (SugB, Rv2039c, UspB and UgpE) and group 2 (SugA, Rv2040c, UspA and UgpA). **B** Structural comparison of TMDs. Comparisons were performed for each two models and the coloured residues represent the position of amino acid insertion/deletion between two proteins. The percentage in each box represents the amino acid sequence identity between two orthologues. Models of all proteins were built using Modeller program.

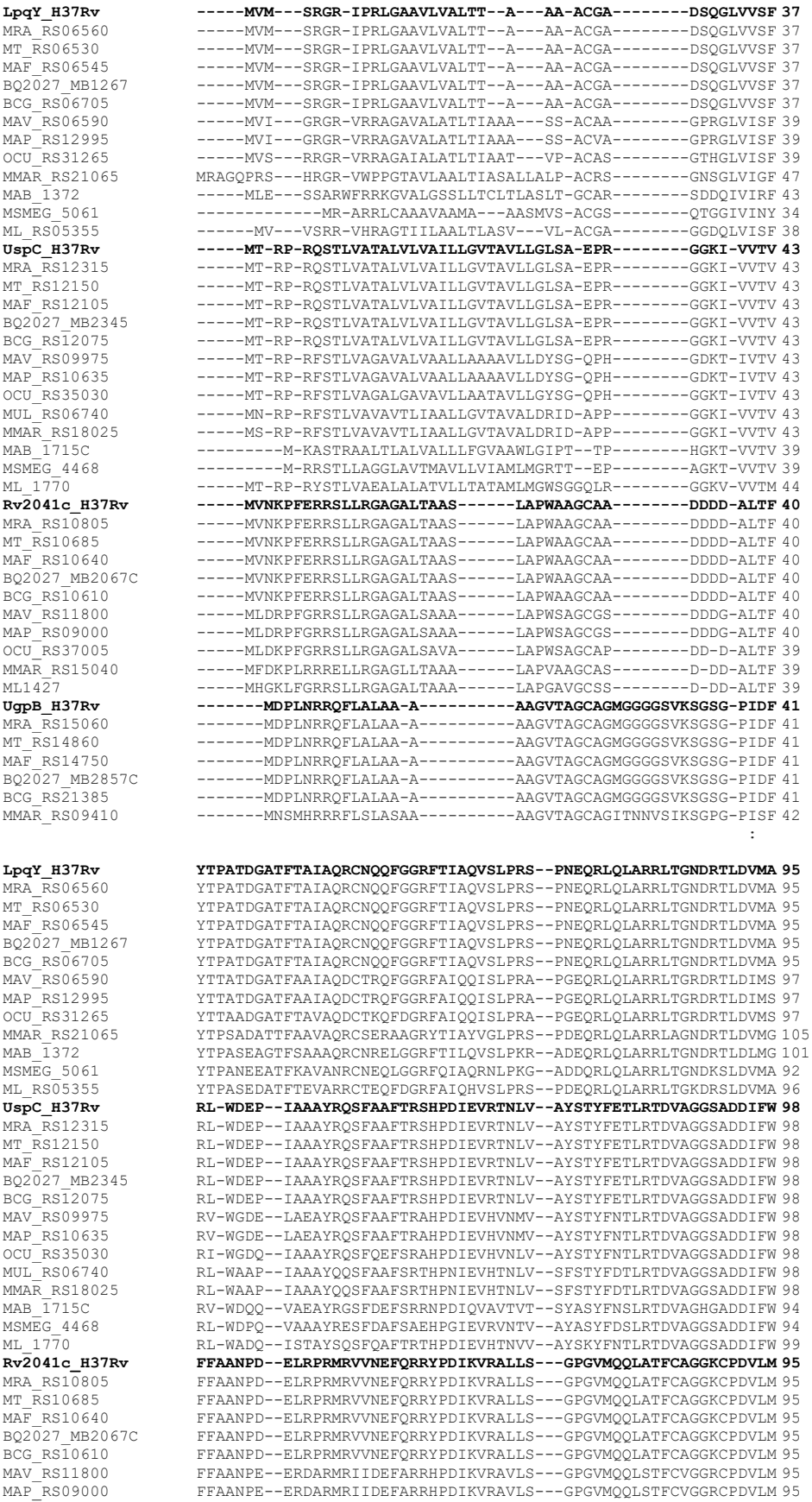
Additional File 4



Additional file 4. Prediction of topology of the TMDs components from *M. tuberculosis* H37Rv carbohydrate ABC transporters. Amino acid sequences of the proteins were submitted to TOPCONS program. The red bars highlight the position of coupling helices.

Additional File 5

A



```
OCU_RS37005          FFAANPD--EREARMRIIDEFARRHPDIKVRAVLS---GPGVMQQLSTFCVGGKCPDVLM 94
MMAR_RS15040         FFAANPE--EADARIRVVDAFGRSHPDIKVRTLLS---GPGALQQISTFCAGGKCPDVLM 94
ML1427               FFAANPE--ETNARMRIVGEFQRDHPDIKVRAVLS---GPGVMQQLSTFCAGGKCPDVLM 94
UgpB_H37Rv           WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
MRA_RS15060          WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
MT_RS14860           WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
MAF_RS14750          WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
BQ2027_MB2857C       WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
BCG_RS21385          WSSHPGQ--SSAAERELIGRFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVL 99
MMAR_RS09410         WSNHPGQ--SSGVEKVLIDRFQQQFPGLSVKLVDAGMDYDEVAEKFNAALIGDDVPDVVV 100
                                                             :           *:.

LpqY_H37Rv           LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
MRA_RS06560          LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
MT_RS06530           LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
MAF_RS06545          LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
BQ2027_MB1267        LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
BCG_RS06705          LDVVWTAEFAEAGWALPLSDDPA--GLAENDAVADTLPGPLATAGWNHKLYAAPVTT-NT 152
MAV_RS06590          LDVVWTAEFAEAGWALPLSDDPA--ARAEADATVDTLPGPLSTARWHDRLFAAPVTT-NT 154
MAP_RS12995          LDVVWTAEFAEAGWALPLSDDPA--GRAEADATVDTLPGPLSTARWHDRLFAAPVTT-NT 154
OCU_RS31265          LDVVWTAEFAEAGWALPLSDDPA--GRAEPDATADVLPGPLSTARWEGKLFAAPVTT-NT 154
MMAR_RS21065         LDVMWTAEFAEAGWALPLSEDPA--GLAETDATIDTLPGPLKTATWKHRLYAAPVTT-NT 162
MAB_1372             MDVVWTAEFAEAGWALPLSDDPA--GVTEAAAQRDALAGPLESARWKGKLYAAPLST-NT 158
MSMEG_5061           LDVVWTAEFAEAGWAVPLSEDPA--GLAEADATENTLPGPLETARWQDELYAAPITT-NT 149
ML_RS05355           MDVVWTAEFAEAGWTLPLSEDPA--GLAEPDAIVDTLPGPLATATWKRKLYAAPVTT-NT 153
UspC_H37Rv           LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
MRA_RS12315          LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
MT_RS12150           LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
MAF_RS12105          LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
BQ2027_MB2345        LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
BCG_RS12075          LSNAYFAAYADSGRLMKIQT--D---------AADWEPAVVDQFTRSGVLWGVPQLTDAG 147
MAV_RS09975          LSNAYLAAYADSGRLLNILDTLGTNA------AADWERPVVEQFTRHGQLWGVPQLTDAG 152
MAP_RS10635          LSNAYLAAYADSGRLLNILDTLGTNA------AADWERPVVEQFTRHGQLWGVPQLTDAG 152
OCU_RS35030          LSNAYLAAYADSGRLLDIGAALGPTA------ASDWERPVVEQFTRNGTLWGVPQLTDAG 152
MUL_RS06740          LSNAYLAAYADSGRLMKIDTAVD---------PGEWEPAVVDQFTRNGVLWGVPQLTDAG 149
MMAR_RS18025         LSNAYLAAYADSGRLMKIDTAVD---------PGEWEPAVVDQFTRNGVLWGVPQLTDAG 149
MAB_1715C            LSNAYLSDYADTGNLVPVEP------------RADWDPSVVAQFTRDGKLWGVPQLSDAG 142
MSMEG_4468           ISNGYFAGYADNGHLLDIADLLGPDA------ATAWEPSVVEQFTRNGALWGVPQLTDAG 148
ML_1770              LSSAYLAAYADNGRLINISNSLGQRA------TSDWEPAVVDQFTRAGALWGVPQLTDAG 153
Rv2041c_H37Rv        AWELTYAELADRGVLLDLNTLLARDQAFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
MRA_RS10805          AWELTYAELADRGVLLDLNTLLARDQAFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
MT_RS10685           AWELTYAELADRGVLLDLNTLLARDQAFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
MAF_RS10640          AWELTYAELADRGVLLDLNTLLARDQAFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
BQ2027_MB2067C       AWELTYAELADRGVLLDLNTLLARDQVFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
BCG_RS10610          AWELTYAELADRGVLLDLNTLLARDQAFAAELKSDSIGALYETFTFNGGQYAFPEQW-SG 154
MAV_RS11800          AWELSYAELADRGVLLDLGPLLARDKAFAQQLQADSIPALYETFTFNGKQYALPEQW-SG 154
MAP_RS09000          AWELSYAELADRGVLLDLGPLLARDKAFAQQLQADSIPALYETFTFNGKQYALPEQW-SG 154
OCU_RS37005          TWELSYAELADRGVLLDLNPLLARDKAFAQQLKADSIPALYETFAFNGSQFALPEQW-SG 153
MMAR_RS15040         AWELSYAELADRGVFLDLNTMLARDQKFADELAADSIGALYDTFAYNGGQYAFPEQW-SG 153
ML1427               AWDLTYAELADRGVLLDLNTLLGQDKAFAAELKSDSIEPLYETFTFNGGQYAFPEQW-SG 153
UgpB_H37Rv           LDDRWWFHFALSGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
MRA_RS15060          LDDRWWFHFALSGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
MT_RS14860           LDDRWWFHFALSGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
MAF_RS14750          LDDRWWFHFALSGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
BQ2027_MB2857C       LDDRWWFHFALIGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
BCG_RS21385          LDDRWWFHFALIGVLTALDDLFGQVG----VDTTDYVDSLLADYEFNGRHYAVPYAR-ST 154
MMAR_RS09410         LDDIWWFHFALSGVISPLDKLFRQIG----VDTSDYVDTLLADYEFNGRHYALPYAR-ST 155
                             *  *    :                                  :. *

LpqY_H37Rv           QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
MRA_RS06560          QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
MT_RS06530           QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
MAF_RS06545          QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
BQ2027_MB1267        QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
BCG_RS06705          QLLWYRPDLVNS-----P-PTDWNAMIA------EA----ARL--------------HAA 182
MAV_RS06590          QLLWYRPDLVPQ-----P-PRTWDAVVT------EA----ARL--------------HAA 184
MAP_RS12995          QLLWYRPDLVLQ-----P-PRTWDAVVT------EA----ARL--------------HAA 184
OCU_RS31265          QLLWYRPDLVRQ-----P-PRTWDGMVR------EA----TRL--------------HAA 184
MMAR_RS21065         QLLWYRPDLVAQ-----P-PETWNAVVA------EA----GRL--------------RAA 192
MAB_1372             QLLWYRKDLVPE-----P-PATWDQTVR------EA----ENF--------------ARS 188
MSMEG_5061           QLLWYRADLMPA-----P-PTTWDGMLD------EA----NRL--------------YRE 179
ML_RS05355           QLLWYRTDLVDQ-----P-PGDWNGMVA------EA----ARL--------------HAA 183
UspC_H37Rv           IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
MRA_RS12315          IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
MT_RS12150           IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
MAF_RS12105          IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
BQ2027_MB2345        IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
BCG_RS12075          IAVFYNADLLAAAGVDPT-QVD-NLRWSRGDD-DTLRPMLARLTVDADGRTANTPGFDAR 204
MAV_RS09975          IALYYNADLLGAAGIDPA-QLN-SLRWNPAGG-DTLRPLLARLTVDADGNRGDTRGFDPG 209
MAP_RS10635          IALYYNADLLGAAGIDPA-QLN-GLRWDPAGG-DTLRPLLARLTVDADGNRGDTRGFDPG 209
OCU_RS35030          IALYYNADLLGAAGVDPA-QLS-TLRWGPDEP-DTLRPVLARLTVDADGNRGDSKGFDAG 209
MUL_RS06740          IAVFYNADLLAAAGVDPA-ELD-GLRWSPGPD-DTLRALLARLTVDADGHVGGTPGFDPG 206
MMAR_RS18025         IAVFYNADLLAAAGVDPA-ELD-GLRWSPGPD-DTLRALLARLTVDADGHVGGTPGFDPG 206
MAB_1715C            IALYYNKNLLDAAQVDPA-ELA-ELRWDPDPEVDTLRPMLHRLTAP-------------- 186
MSMEG_4468           IAVYYNADLLEKAGVSPA-DLS-TLRWSNGPD-DTLRPLLARLTVEES-----------G 194
ML_1770              IAVFYNADLLTAAGIDPV-QLN-RMQWTSNDD-DTLRPLLTQLTLDTNGHVAKTPGFDSR 210
Rv2041c_H37Rv        NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ-------------GRSG 199
MRA_RS10805          NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ-------------GRSG 199
```

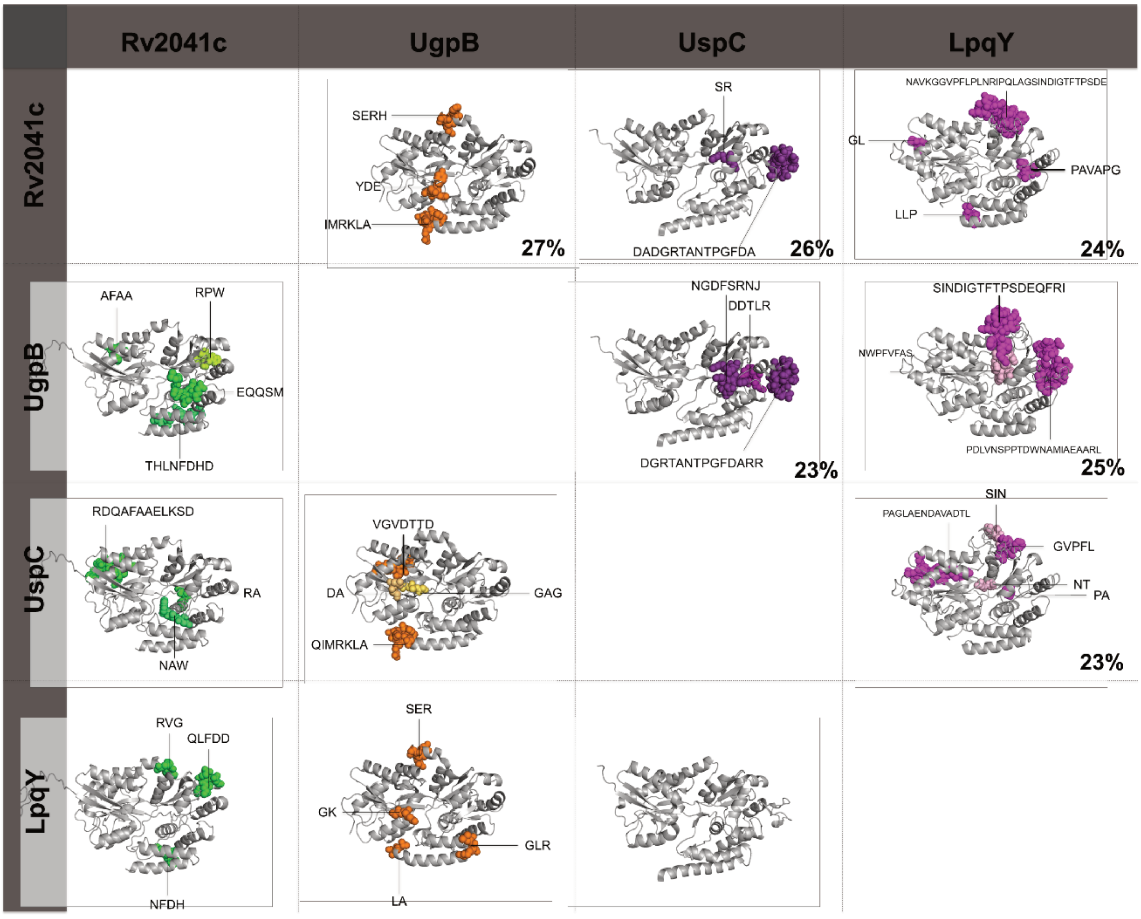
```
MT_RS10685          NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ--------------GRSG 199
MAF_RS10640         NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ--------------GRSG 199
BQ2027_MB2067C      NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ--------------GRSG 199
BCG_RS10610         NFLFYNKQLFDDAGVPPP-PGSWERPWSFAEFLDAAQALTKQ--------------GRSG 199
MAV_RS11800         NYLFYNKRLFDEAGVPSP-PTAWERPWDFSEFLDTARALTKR--------------DASG 199
MAP_RS09000         NYLFYNKRLFDEAGVPSP-PAAWEHPWGFSEFLNAARALTKR--------------DASG 199
OCU_RS37005         NYLFYNKRLFADAGVPAP-PTTWQRPWDFSGFLDTARALTKR--------------DTSG 198
MMAR_RS15040        NFLFYNRGLFADGGVPPP-PTSWDQPWSFAEFLDTARALTKR--------------DGAG 198
ML1427              NYLFYNKQLFTNAGVQPP-PCTWEQPWSFTEFLDTARALTKR--------------DSSG 198
UgpB_H37Rv          PLFYYNKAAWQQAGLPDRGPQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
MRA_RS15060         PLFYYNKAAWQQAGLPDRGPQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
MT_RS14860          PLFYYNKAAWQQAGLPDRGPQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
MAF_RS14750         PLFYYNKAAWQQAGLPDRGPQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
BQ2027_MB2857C      PLFYYNKAAWQQAGLPDRGLQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
BCG_RS21385         PLFYYNKAAWQQAGLPDRGLQSWS------EF-DEWGPELQRVVGA----GRSAHGWANA 203
MMAR_RS09410        PLFYYNKAVWQRAGLPDRGPNSWQ------EF-DDWGPRLQRVVDE----RQWAHGWANA 204
                      .:*.                                .

LpqY_H37Rv          GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
MRA_RS06560         GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
MT_RS06530          GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
MAF_RS06545         GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
BQ2027_MB1267       GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
BCG_RS06705         GEPSWIAVQANQGE---GLVVWFNTLLVSAGGSVLSEDGRHVTLTDTPAHRAATVSALQI 239
MAV_RS06590         GQPSWIAVQANEGE---GLVVWFNTLLASGGGRVLSEDGRRVTLTDTPAHRAATVNALRI 241
MAP_RS12995         GRPSWIAVQANEGE---GLVVWFNTLLASGGGRVLSEDGRRVTLTDTPAHRAATVNALRI 241
OCU_RS31265         GQPSWIAVQANEGE---GLVVWFNTLLASGGGQVLSEDGRRVTLTDTPAHRAATVDALRT 241
MMAR_RS21065        GQPTWIAVQANQGE---GLMVWFNTVLSSVGGQVLSDDGTRVTLTDTPAHRAATVAALRV 249
MAB_1372            EGPSWIALQGKQYE---GLVVLFNTLLSSAGGSVLAGDGKTVTLTDTPAHRQATVTALRT 245
MSMEG_5061          GGPSWIAVQGKQYE---GMVVWFNTLLQSAGGQVLSDDGQRVTLTDTPEHRAATVKALRI 236
ML_RS05355          GEPSWIAVQANQGE---GLVVWFNTLLASAGGQVLSEDGRHVTLTDTPEHRAATVRALRI 240
UspC_H37Rv          RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
MRA_RS12315         RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
MT_RS12150          RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
MAF_RS12105         RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
BQ2027_MB2345       RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
BCG_RS12075         RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--DGKFAFDN-----PGAIEAFRY 251
MAV_RS09975         RVRQWGYNAANDPQ---GIYLNYI--G-SAGGVFQR--GDEFAFDN-----PAAVSAFRY 256
MAP_RS10635         RVRQWGYNAANDPQ---GIYLNYI--G-SAGGVFQR--GDEFAFDN-----PAAVSAFRY 256
OCU_RS35030         RVRQWGYNAANDPQ---GIYLNYI--G-SAGGVFQR--GDEFAFDN-----PGAVTAFQY 256
MUL_RS06740         RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFMR--DNEFAFDN-----PPAIDAFRY 253
MMAR_RS18025        RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFMR--DNEFAFDN-----PPAIDAFRY 253
MAB_1715C           --GHWGYNAANDLQ---GIYLNYL--G-SAGAVFQA--DDKFAFAK-----PRAEMAFTY 231
MSMEG_4468          RTRQWGYNAANDLQ---GIYLNFI--G-SAGGTFSE--GDRFTFDN-----PQAVEAFEY 241
ML_1770             RVRQWGYNAANDPQ---AIYLNYI--G-SAGGVFQR--GDEFAFDN-----PSAVEAFRY 257
Rv2041c_H37Rv       RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
MRA_RS10805         RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
MT_RS10685          RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
MAF_RS10640         RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
BQ2027_MB2067C      RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
BCG_RS10610         RDRQWGFVNAWVSFYAAGLFAMNN--GVPWSVPRMN--PTHLNFDH-----DGFLEAVQF 250
MAV_RS11800         RAAQYGFVNTWGSYYSAGLFAMNN--GVPWSDPRLN--PTHFNFDN-----AAFQEAVQF 250
MAP_RS09000         RAAQYGFVNTWGSYYSAGLFAMNN--GVPWSDPRLN--PTHFNFDN-----AAFQEAVQF 250
OCU_RS37005         RAAQYGFVNTWVSTYSAGLFAMNN--GVPWSNPRLN--PTHLNFDD-----AAFQEAVQF 249
MMAR_RS15040        RVTQWGFVDTFVPPLSAGLFALNN--GVPWSTPRMN--PTHLNFDD-----DAFIEAVQF 249
ML1427              RVTQWGFVNTWLSYYTAGLFALNN--GVPWSNPRMN--PTHLNFDD-----DAFIEAVQF 249
UgpB_H37Rv          DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
MRA_RS15060         DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
MT_RS14860          DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
MAF_RS14750         DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
BQ2027_MB2857C      DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
BCG_RS21385         DLISWTFQGPNW----------------AFGGAYSD--KWTLTLTE-----PATIAAGNF 240
MMAR_RS09410        ELISWTFQGVNW----------------TFDGSYSD--LWTLRFTD-----PDTIAAGNF 241
                        :                           .          . : .          *

LpqY_H37Rv          LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
MRA_RS06560         LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
MT_RS06530          LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
MAF_RS06545         LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
BQ2027_MB1267       LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
BCG_RS06705         LKSVATTPGADPSITRTEEGS--ARLAFEQGKAALEVNWPFVFASMLENAVKGGVPFLPL 297
MAV_RS06590         LKSVATAPGADPSITRTDEGT--ARLAVEQGRAALAVNWPYALASMLDNAVKGGVPFLPL 299
MAP_RS12995         LKSVATAPGADPSITRTDEGT--ARLAVEQGRAALAVNWPYALASMLDNAVKGGVPFLPL 299
OCU_RS31265         LKSVATAPGADPSITRTDEGT--ARLAVEQGRAALAVNWPYALASMLENAVKGGVPVLPL 299
MMAR_RS21065        LKSVATAPGADPSITRAEAGT--ARLAFEQGKAALELNWPYVFASMLENAVKGGVPFLPL 307
MAB_1372            MKSIATADGADPSITQTDEGT--ARLAFEQGKAAFEINWPFVMASMMENAIKGGVPFLRL 303
MSMEG_5061          IKSVATAPGADPSITQTDENT--ARLALEQGKAALEVNWPYVLPSLLENAVKGGVSFLPL 294
ML_RS05355          IKAVATVPGADPSITRTDEAT--ARLAVEQGKAALEVNWPFVLASMLENAVKGGVAFLPL 298
UspC_H37Rv          LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
MRA_RS12315         LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
MT_RS12150          LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
MAF_RS12105         LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
BQ2027_MB2345       LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
BCG_RS12075         LVGLINDDHVAPPASDTNDNGDFSRNQFLAGKMALFQSGTYSLAPVARD----------- 300
MAV_RS09975         LVDLINRDHVAPPAADTNDNGDFSRNQFLAGRMALFQSGTYNLAPVARD----------- 305
MAP_RS10635         LVDLINRDHVAPSAADTNDNGDFSRNQFLAGRMALFQSGTYNLAPVARD----------- 305
OCU_RS35030         LVDLINRDHVAPPAADTNDNGDFSRNQFLAGRMALFQSGTYNLAPVARD----------- 305
MUL_RS06740         LVGLINNDHVAPPASDTNDNGDFSRNQFLAGRMALFQSGTYSLAPVARD----------- 302
```

```
MMAR_RS18025        LVGLINNDHVAPPASDTNDNGDFSRNQFLAGRMALFQSGTYSLAPVARD---------- 302
MAB_1715C           LVDLINVDRVAPSAADTNDNNDFSRNQFLQGRMALFQSGTYNLAQIQAN---------- 280
MSMEG_4468          LVRLINTDRVAPPASDTNDNGDFSRNAFLAGRMALFQSGTYNLAAIADQ---------- 290
ML_1770             LVGLINNDHVAPPASETNNNGDFSRNQFLSGRMALFQSGTYNLALIARE---------- 306
Rv2041c_H37Rv       YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
MRA_RS10805         YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
MT_RS10685          YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
MAF_RS10640         YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
BQ2027_MB2067C      YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
BCG_RS10610         YADLTNKHKVAPSAAEQQSMS--TADLFSVGKAGIALAGHWRYQTFDRA---------- 297
MAV_RS11800         YADLANKYRVAPNASETQSMS--TPNLFAVGRAAMALGGHWRYQTYLRA---------- 297
MAP_RS09000         YADLANKYRVAPNGSETQSMS--TPNLFAVGRAAMALGGHWRYQTYLRA---------- 297
OCU_RS37005         YADLANEHRVAPNASETQSMS--TPNLFAVGRAAIALGGHWRYQTYLRA---------- 296
MMAR_RS15040        YADLSAKHQVAPTASELQSIA--TTDLFSLGKAAMALGGHWRYQTFDRA---------- 296
ML1427              YCDLTNKYQVAPDASEQQWMA--TADLFSLGKAAIALGGHWRYQTFMRA---------- 296
UgpB_H37Rv          YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLAGITAS---------- 282
MRA_RS15060         YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLAGITAS---------- 282
MT_RS14860          YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLAGITAS---------- 282
MAF_RS14750         YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLAGITAS---------- 282
BQ2027_MB2857C      YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLPGITAS---------- 282
BCG_RS21385         YRNSIHGKGYAAVAND-------IANEFATGILASAVASTGSLPGITAS---------- 282
MMAR_RS09410        YRDSIHTKRYAAVANN-------IANEFATGIMASAVASTGSLTGITQM---------- 283
                                                 .  *  .

LpqY_H37Rv          NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
MRA_RS06560         NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
MT_RS06530          NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
MAF_RS06545         NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
BQ2027_MB1267       NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
BCG_RS06705         NRIPQLAGSINDIGTFTPSDEQFRIAYDASQQVFGFAPYPA-------VAPGQPAKVTIG 350
MAV_RS06590         NRDPRLAGSINDVGIFVPTDEQYRIAYQASQKVFGFAPYPG-------AAPGLPAKVTIG 352
MAP_RS12995         NRDPRLAGSINDVGIFVPTDEQYRIAYQASQKVFDFAPYPG-------AAPGLPAKVTIG 352
OCU_RS31265         NQDPRLAGSINDVGTFVPSDEQFRIAYQASQKVFGFAPYPG-------VAPGHPAKVTIG 352
MMAR_RS21065        DRLPELAGSINDVGTFVPTDEQFRIAYLASQKVFGFAPYPA-------VLPGRPARVTIG 360
MAB_1372            DRRPDLTGAIGDAGRFAPSEEQFAAAYEAASAALGFAPFPS-------VVAGQPARVTIG 356
MSMEG_5061          DGDPALQGSINDVGTFSPTDEQFDIAFDASKKVFGFAPYPG-------VNPDEPARVTLG 347
ML_RS05355          DRISELAGSINNVGTFVPNDEQFRIAYQATRNVFGFAPYPS-------VSRSEPAKVTIG 351
UspC_H37Rv          ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
MRA_RS12315         ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
MT_RS12150          ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
MAF_RS12105         ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
BQ2027_MB2345       ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
BCG_RS12075         ----------------------------ALFHWGVAMLPAGPAGRVSVTNGIAAA---- 327
MAV_RS09975         ----------------------------ARFRWGVAMMPAGPVGRVSVTNGIAAA---- 332
MAP_RS10635         ----------------------------ARFRWGVAMMPAGPVGRVSVTNGIAAA---- 332
OCU_RS35030         ----------------------------ARFRWGVALMPSGPAGRVSVTNGIAAA---- 332
MUL_RS06740         ----------------------------ATFRWGVAMLPIGPAGRVSVTNGIAAA---- 329
MMAR_RS18025        ----------------------------ATFRWGVAMLPIGPAGRVSVTNGIAAA---- 329
MAB_1715C           ----------------------------ATFPWDVAMMPAGPQGRVSVTNGIVAA---- 307
MSMEG_4468          ----------------------------APFPWGVAMLPIGPKGRVSVTNGIAAA---- 317
ML_1770             ----------------------------ARFHWGIAMMPTGPQGRVSVTNGIAVA---- 333
Rv2041c_H37Rv       ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
MRA_RS10805         ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
MT_RS10685          ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
MAF_RS10640         ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
BQ2027_MB2067C      ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
BCG_RS10610         ----------------------------DGLDFDVAPLPIGPRG-------RAACSDIG 321
MAV_RS11800         ----------------------------EGLDFDVAPLPVGPAVGK----GQPACSDIG 324
MAP_RS09000         ----------------------------EGLDFDVAPLPVGPAVGK----GQPACSDIG 324
OCU_RS37005         ----------------------------EDLDFDVAPLPVGPALAN----GRAACSDIG 323
MMAR_RS15040        ----------------------------DGLEFDVTALPLGPNG-------RAACSNIG 320
ML1427              ----------------------------EGLDFDVTSLPIGPSAGTVPATRSGACSDIG 327
UgpB_H37Rv          ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
MRA_RS15060         ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
MT_RS14860          ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
MAF_RS14750         ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
BQ2027_MB2857C      ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
BCG_RS21385         ----------------------------ARFDFGAAPLPTGPDAAPACPTGG------- 306
MMAR_RS09410        ----------------------------AKFDFGAAPVPTGPGGAPGCPTGG------- 307
                                                  . :  *

LpqY_H37Rv          GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
MRA_RS06560         GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
MT_RS06530          GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
MAF_RS06545         GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
BQ2027_MB1267       GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
BCG_RS06705         GLNLAVA--KTTRHRAEAFEAVRCLRDQHNQRYVSLEGGLPAVRASLYSDPQ---FQAKY 405
MAV_RS06590         GANLAVA--STTRHRAEAFEAIRCVRSLQHQKYVAIQGGLPPVRTSLYSDPQ---FQTKY 407
MAP_RS12995         GANLAVA--STTRHRAEAFEAIRCVRSLQHQKYVAIQGGLPPVRTSLYSDPQ---FQTKY 407
OCU_RS31265         GANLAVA--STTRHRAEAFEAIRCLRSLQHQKYVSIAGGLPPVRTSLYSDPQ---FQAKY 407
MMAR_RS21065        GLNLAVA--STSRHKAEAFEAVRCLRDPQSQKRLAIEGGLPAVRTSLYADPQ---FQAKY 415
MAB_1372            GLNIAVA--KTTRYKAQAFEALNCLRSRENQKATAIEGGLPAVDSSLYDDPE---FQAKY 411
MSMEG_5061          GLNLAVA--STSQHKAEAFEAIRCLRNVENQRYTSIEGGLPAVRTSLYDDPA---FQKKY 402
ML_RS05355          GLNLAVA--STSRHKAEAFEAVRCLRSEQSQQYLSIEGGLPAVRASLYSDPA---FQDKY 406
UspC_H37Rv          --GN-----SASKHPDAVRQVLAWMGSTEGNSYLGRHGAAIPAVLSAQPVYF--DYWSA- 377
MRA_RS12315         --GN-----SASKHPDAVRQVLAWMGSTEGNSYLGRHGAAIPAVLSAQPVYF--DYWSA- 377
MT_RS12150          --GN-----SASKHPDAVRQVLAWMGSTEGNSYLGRHGAAIPAVLSAQPVYF--DYWSA- 377
MAF_RS12105         --GN-----SASKHPDAVRQVLAWMGSTEGNSYLGRHGAAIPAVLSAQPVYF--DYWSA- 377
```

```
BQ2027_MB2345       --GN-----SASKHPDAVRQVLAWMGSTEGNSYVGRHGAAIPAVLSAQPVYF--DYWSA- 377
BCG_RS12075         --GN-----SASKHPDAVRQVLAWMGSTEGNSYVGRHGAAIPAVLSAQPVYF--DYWSA- 377
MAV_RS09975         --GN-----AATKHPGAVRQVLAWMGSRQGNEYLGRYGAAIPAVTSAQPVYF--GYWAA- 382
MAP_RS10635         --GN-----AATKHPAAVRQVLAWMGSRQGNEYLGRYGAAIPAVTSAQPVYF--RYWAS- 382
OCU_RS35030         --GN-----AASRHPDAVRRVLAWMGSRQGNEYLGRYGAAIPAVSSAQPVYF--HYWAT- 382
MUL_RS06740         --GN-----SASQHPDAVRQVLAWMGSSAGNEYLGRDGVAIPAVRSAQPSYF--AFWEA- 379
MMAR_RS18025        --GN-----SASQHPDAVRQVLAWMGSSAGNEYLGRDGVAIPAVRSAQPSYF--AFWEA- 379
MAB_1715C           --AN-----SSSPHPDAVHKVLAWMGSTDGNSFLGRSGSAIPAVLSARAPYF--QYWAD- 357
MSMEG_4468          --GN-----AATRHPEAVREVLAWMGSRRGNEFVGRRGAAIPAVLAAQPVYH--EYWAS- 367
ML_1770             --GN-----SATKHPDAVRQVLAWMGSREGNAYLGRHGAAVPAVRSAQSVYF--DYWAA- 383
Rv2041c_H37Rv       VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
MRA_RS10805         VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
MT_RS10685          VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
MAF_RS10640         VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
BQ2027_MB2067C      VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
BCG_RS10610         VTGLAIA--ATSRRKDQAWEFVKFATGPVGQALIGESRLFVPVLRSAINSHG--FANAH- 376
MAV_RS11800         ATGLAIS--SSSPRKEQAWEFVKFATGPVGQALIGESCLFVPVLRSALKSDG--FARAH- 379
MAP_RS09000         ATGLAIS--SSSPRKEQAWEFVKFATGPVGQALIGESCLFVPVLRSALKSDG--FARAH- 379
OCU_RS37005         ATGLGIS--STSPRKDQAWEFVKFATGPVGQALIGESCLFVPVLRSALNSEG--FANAH- 378
MMAR_RS15040        ATGLAIA--ASSARKEQAWEFVKFATGPAGQALIGETDLFVPVLGSALNSAG--FAKAH- 375
ML1427              ATGLAIA--ASSSRKEQAWEFVKFATGPAGQALIGESCLFVPVLQSAIYSTG--FAKAH- 382
UgpB_H37Rv          -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLADN 365
MRA_RS15060         -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLADN 365
MT_RS14860          -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLADN 365
MAF_RS14750         -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLVDN 365
BQ2027_MB2857C      -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLADN 365
BCG_RS21385         -AGLAIPAKLSEERKVNALKFIAFVTNPTNTAYFSQQTGYLPVRKSAVDDASERHYLADN 365
MMAR_RS09410        -AGLAIPTKLSDERKRNALRFIEYITNPVNTAYFSQRTGYLAVRKSAADVPSEQKYLADH 366
                                 :      . . :    .       .       .  :

LpqY_H37Rv          --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
MRA_RS06560         --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
MT_RS06530          --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
MAF_RS06545         --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
BQ2027_MB1267       --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
BCG_RS06705         --------PMHAIIRQQLTDAAVRPATPVYQALSIRLAAVLSPITEID--PESTA----- 450
MAV_RS06590         --------PMYTIIRRQLTDAAVRPATPVYQTVSIRLASTLSPITGID--PERTA----- 452
MAP_RS12995         --------PMYTIIRRQLTDAAVRPATPVYQTVSIRLAATLSPITGID--PERTA----- 452
OCU_RS31265         --------PMYTIIRRQLTDAAVRPATPVYQAVAIRLAATLSPITEID--PERTA----- 452
MMAR_RS21065        --------PMYRIIRQQLTEAAVRPATPVYQAVSLRLSAELSPVTEID--PERTA----- 460
MAB_1372            --------PAYAIIRDQLANAAVRPASPYYQAISTRVSAVLAPITGID--PERTA----- 456
MSMEG_5061          --------PQYEIIRQQLTNAAVRPATPVYQAVSTRMSATLAPISDID--PERTA----- 447
ML_RS05355          --------PMYTIIRQQLTDAAVRPVTPAYQAVSIRLSAALNPITDID--PEPMA----- 451
UspC_H37Rv          --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
MRA_RS12315         --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
MT_RS12150          --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
MAF_RS12105         --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
BQ2027_MB2345       --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
BCG_RS12075         --RGVDVTPFFAVLNGPRIA---APGGAGFAAGQQALEPYFDEMFLGRGDVTTTLRQAQA 432
MAV_RS09975         --RGVDVTPFFAVLNGPRIA---APGGAGFAAGNDALRPYFDEMFSGRGDVATTLRRAQA 437
MAP_RS10635         --RGVDVTPFFAVLNGPRIA---APGGAGFAAGNDALRPYFDEMFSGRGDVATTLRRAQA 437
OCU_RS35030         --RGVDVTPFFTVLNGPRIA---APGGAGFAAGNDALQPYFDEMFLGRGDVASTLRRAQA 437
MUL_RS06740         --KGVNVNPFFAVLSGSRIP---APGGAGFAAGNDALKPYFDEMFLGRGQVATILREAQA 434
MMAR_RS18025        --KGVNVEPFFAVLSGSRIP---APGGAGFAAGNDALKPYFDEMFLGRGQVATILREAQA 434
MAB_1715C           --KGVDVSPFFEVLRGQQIA---APGGQGFGAGFAALKPYFAEMFLGRLDVREALQQAQR 412
MSMEG_4468          --RGVDVSPFFRVLQGPRIA---APGGAGFPAGFEALTPYFAEMFLGRRDVAGTLAEAQR 422
ML_1770             --KGIDVTPFFSVLDGPHIP---APGGAGFPAGDDALQSYFDEMFLGHGDVEKILCQAQA 438
Rv2041c_H37Rv       --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
MRA_RS10805         --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
MT_RS10685          --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
MAF_RS10640         --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
BQ2027_MB2067C      --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
BCG_RS10610         --RRV--GNLAVLSEGPAYS-EGLPVTPAWEKIAALMDRYFGPVLRGSRPATSLTGLS-Q 430
MAV_RS11800         --RRV--GNLGVLTDGPAFS-QGLPITPAWEKVNALMDRNFGPILRGSRPATSLAGLS-R 433
MAP_RS09000         --RRV--GNLGVLTDGPAFS-QGLPITPAWEKVNALMDRNFGPILRGSRPATSLAGLS-R 433
OCU_RS37005         --RRI--GNLSVLTEGPAYS-QGLPITPAWEKVNALIDRNFGPVLRGRQPATSLAGLS-R 432
MMAR_RS15040        --RRV--DNLAVLTAGPGHS-QGLPITPAWPKVHALMDRNFGPVLRGSRPATSLIGMS-R 429
ML1427              --NRV--ANLAVLTGGPVHS-AGLPITPAWEKINALMDRNFGPVLRGVRPATSLAGLA-R 436
UgpB_H37Rv          PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
MRA_RS15060         PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
MT_RS14860          PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
MAF_RS14750         PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
BQ2027_MB2857C      PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
BCG_RS21385         PRARVALDQLPHT-RTQDYARVFLPGGD------RIISAGLESIGLRGADVTKTFTNIQK 418
MMAR_RS09410        PRARVAINQLPHT-RPQDYARVFLPGAD------RIISAGLESIGLRGTNVAKTFASIER 419
                                          *           :   :  :

LpqY_H37Rv          ---DELAAQAQKAIDGMGLLP       468
MRA_RS06560         ---DELAAQAQKAIDGMGLLP       468
MT_RS06530          ---DELAAQAQKAIDGMGLLP       468
MAF_RS06545         ---DELAAQAQKAIDGMGLLP       468
BQ2027_MB1267       ---DELAAQAQKAIDGMGLLP       468
BCG_RS06705         ---DELAAQAQKAIDGMGLLP       468
MAV_RS06590         ---DQLSAEVQKAIDGKGLLP       470
MAP_RS12995         ---DQLSVEVQKAVDGKGLLP       470
OCU_RS31265         ---DELSAEVQKAIDGKGLLP       470
MMAR_RS21065        ---DKLAAQVQQAIDGKGLLP       478
MAB_1372            ---DLLNDQVQKAVDGRGLIP       474
```

```
MSMEG_5061          ---DELTEAVQKAIDGKGLIP      465
ML_RS05355          ---DRLAAQVQKAIDGKGLLP      469
UspC_H37Rv          AANAATQR-------------      440
MRA_RS12315         AANAATQR-------------      440
MT_RS12150          AANAATQR-------------      440
MAF_RS12105         AANAATQR-------------      440
BQ2027_MB2345       AANAATQR-------------      440
BCG_RS12075         AANAATQR-------------      440
MAV_RS09975         AANAAAARR------------      446
MAP_RS10635         AANAAARR-------------      445
OCU_RS35030         AANAAAARQ------------      446
MUL_RS06740         AANAAARR-------------      442
MMAR_RS18025        AANAAARR-------------      442
MAB_1715C           AANRALER-------------      420
MSMEG_4468          AANAAASR-------------      430
ML_1770             AANTAAHR-------------      446
Rv2041c_H37Rv       AVDEVLRNP------------      439
MRA_RS10805         AVDEVLRNP------------      439
MT_RS10685          AVDEVLRNP------------      439
MAF_RS10640         AVDEVLRNP------------      439
BQ2027_MB2067C      AVDEVLRNP------------      439
BCG_RS10610         AVDEVLRNP------------      439
MAV_RS11800         SVDEVLRSP------------      442
MAP_RS09000         SVDEVLRSP------------      442
OCU_RS37005         SVDEVLRSP------------      441
MMAR_RS15040        ALDEVLRSP------------      438
ML1427              AVDEVLNSP------------      445
UgpB_H37Rv          RLQVILDRQIMRKLAGHG---      436
MRA_RS15060         RLQVILDRQIMRKLAGHG---      436
MT_RS14860          RLQVILDRQIMRKLAGHG---      436
MAF_RS14750         RLQVILDRQIMRKLAGHG---      436
BQ2027_MB2857C      RLQVILDRQIMRKLAGHG---      436
BCG_RS21385         RLQVILDRQIMRKLAGHG---      436
MMAR_RS09410        QLQIILDRQIVRKLRQHG---      437
```

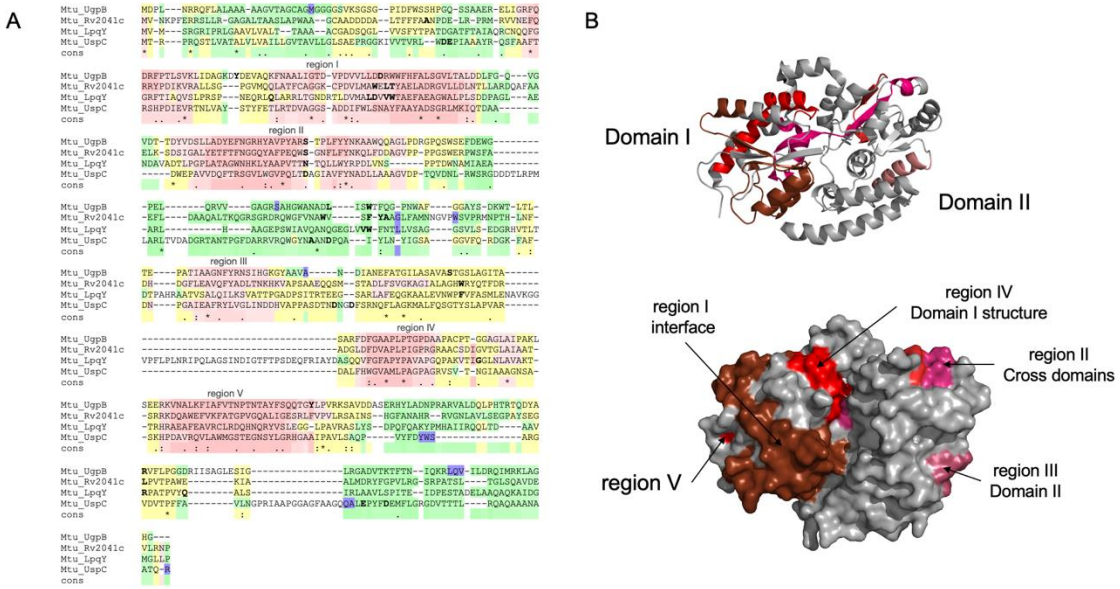
B



Additional file 5. Structural and amino acid sequence differences found in the *M. tuberculosis* carbohydrate SBP components. A Amino acid sequence alignment of the SBPs showing the four groups (coloured in gray, orange, salmon, and green). The alignment was made using Clustal Omega. **B** Proteins in

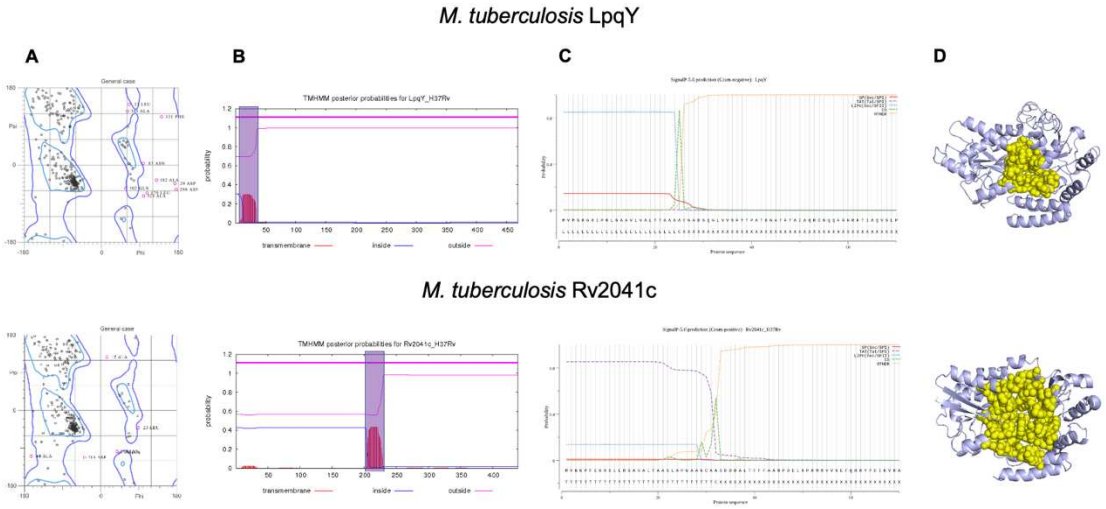
cartoon represent the modelled structures of all SBPs, except for UgpB and UspC that have crystallographic structures (PDB codes: 4MFI and 5K2X, respectively). After the amino acid sequence alignment of each two structures variable regions were showed by coloured spheres.

Additional File 6



Additional file 6. Amino acid sequence alignment of the *M. tuberculosis* carbohydrate-binding proteins using structural information. A Amino acid sequence alignment of LpqY, Rv2041c, UspC and UgpB was obtained with Expresso program in the T-coffee server (http://tcoffee.crg.cat/apps/tcoffee/do:expresso). The colours follow the program pattern where the highest to lowest conservation is shown by red, yellow, green, and blue, respectively. Amino acids that form the substrate-binding pocket are highlighted in bold. **B** Three-dimensional structure of UgpB showing the five regions with highest conservation. The putative function of each region is pointed in the structure represented in surface.

Additional file 7



Analysis of the Rv2041c and LpqY structural models generated by I-TASSER program. A Ramachandran plot for the final selected structural models showing more than 96% of its amino acid residues in the favoured region. **B** Prediction of transmembrane domains (in magenta area) in the proteins obtained with TMHMM program. **C** Signal P-5.0 prediction showing the possible translocation pathway. **D** Substrate-binding pocket prediction of Rv2041c and LpqY proteins using CASTp3.0 server.

ANEXOS

Anexo 1 - Manuscrito sobre a proteína PotF de *X. citri*

Manuscrito "*The citrus plant pathogen Xanthomonas citri has a dual polyamine-binding protein*", submetido ao periódico Journal Molecular Biology (JMB), no qual participei como colaboradora na obtenção de dados de Espalhamento de Raios-X a baixo ângulo, ensaios de *Thermal Shift* e Fluorescência Intrínseca do Triptofano, visando a obtenção das constantes de associação da proteína em presença de diversas poliaminas (espermidina, espermina e putrescina). Este trabalho foi desenvolvido como um estudo paralelo dentro laboratório da Dra. Andrea Balan. A proteína PotF faz parte do transportador PotFGHI de *X. citri*, e similar aos transportadores de açúcares de *M. tuberculosis* e a transportador de maltose de *E. coli*, este é um importador ABC do tipo I. A proteína PotF faz parte do cluster D na classificação de proteínas periplasmáticas de ligação ao substrato. Curiosamente, as proteínas periplamásticas de *M. tuberculosis* envolvidas na captação de açúcares e a proteína MalE de *E. coli*, são agrupadas dentro do mesmo cluster.

Journal of Molecular Biology

The citrus plant pathogen Xanthomonas citri has a dual polyamine-binding protein

--Manuscript Draft--

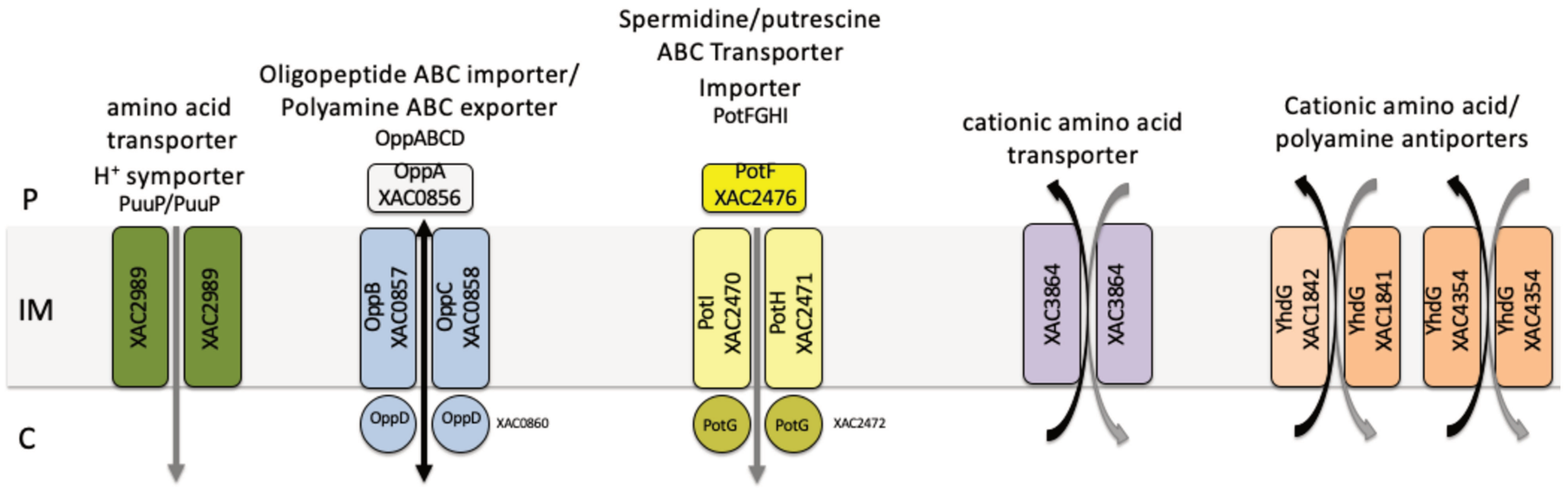
Manuscript Number:	
Article Type:	Full Length Article
Section/Category:	Structure, chemistry, processing and function of biologically important macromolecules and complexes
Keywords:	Polyamine-binding protein; PotF; Xanthomonas citri; ligand-induced conformational changes
Corresponding Author:	Andrea Balan, Ph.D. Universidade de São Paulo SAO PAULO, SÃO PAULO BRAZIL
First Author:	Aline S Cremonesi
Order of Authors:	Aline S Cremonesi
	Lilia I De La Torre
	Maximillia F Souza
	Gabriel S V Muniz
	Maria T Lamy
	Cristiano L P Oliveira
	Andrea Balan, Ph.D.
Abstract:	ATP-Binding Cassette transporters (ABC transporters) are membrane proteins involved in import and export of different molecules, including ions, sugars, peptides, drugs, and others. Due the diversity of substrates, they have large relevance in physiological processes such as virulence, pathogenesis, and antimicrobial resistance. In Xanthomonas citri subsp. citri, the phytopathogen responsible for the citrus canker disease, it was shown that 20% of the ABC transporters are expressed under infection conditions, highlighting components of the putative putrescine/polyamine ABC transporter PotFGHI. Polyamines are ubiquitous molecules that mediate the cell growth and proliferation and play important role in the infection. In this work, we studied the periplasmic-binding protein PotF (Xac2476) and showed that the protein is highly conserved in Xanthomonas sp. genus and that X. citri has a set of proteins, which functions are directly related to the import or assimilation of polyamines. Ligand-protein binding constant values (K_b) were determined through fluorescence spectroscopy. For putrescine, it was determined a $K_b = 1.7 \times 10^5 M^{-1}$. Spermidine binds PotF much more strongly with a $K_b \geq 5 \times 10^8 M^{-1}$. Molecular modelling analysis associated with experiments of circular dichroism, intrinsic tryptophan fluorescence and small angle X-ray scattering (SAXS) revealed that the interactions performed by the two polyamines induced an increasing of the protein thermal stability and significant structural changes that coincides with closing of the domains. We also explored the conserved aspects of the ligand-binding site as well as differences for ligand binding specificity.
Suggested Reviewers:	Ana Carolina Zeri ana.zeri@lnls.br
	Luis S Netto nettoles@ib.usp.br
	Kazue Igarashi iga16077@faculty.chiba-u.jp
	Karsten Niehaus kniehaus@genetik.uni-bielefeld.de
	Jorgelina Ottado ottado@ibr-conicet.gov.ar
	Maira Goytia mgoytia@emory.edu

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

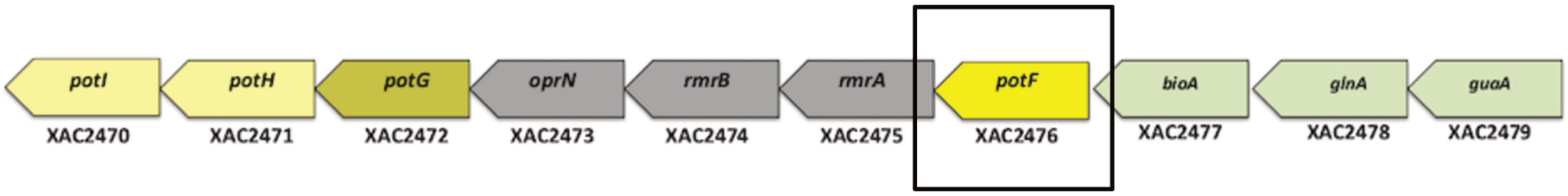
Opposed Reviewers:	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

All transporters for polyamine-related molecules

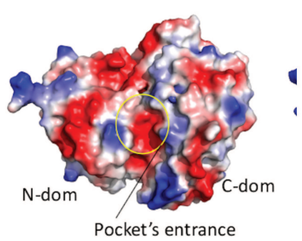


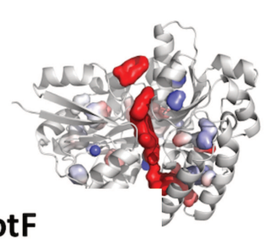
Putrescine-spermidine operon in *X. citri*



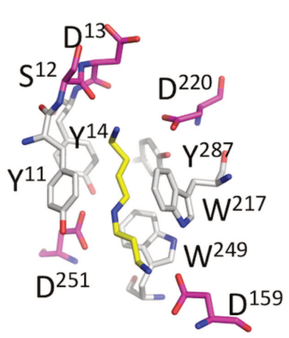
Electrostatic potential, pore
And ligand-binding site

***X. citri* PotF**

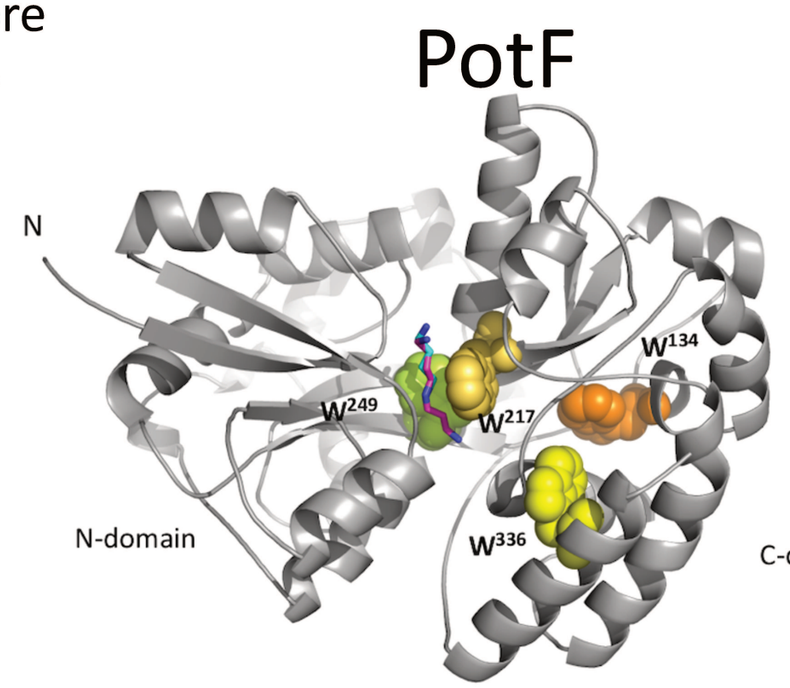


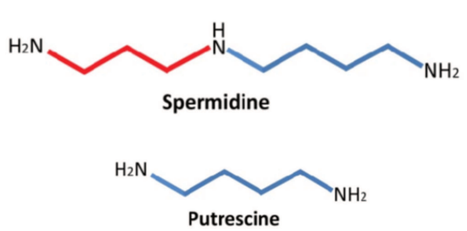


***X. citri* PotF**

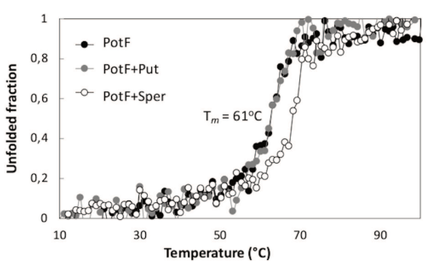


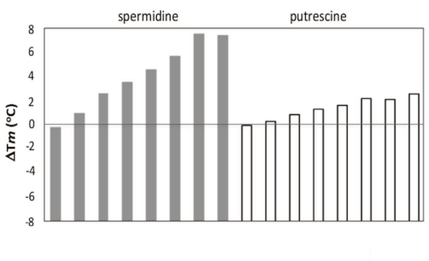
PotF

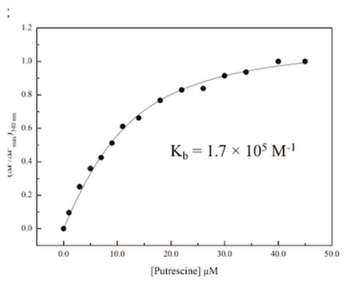


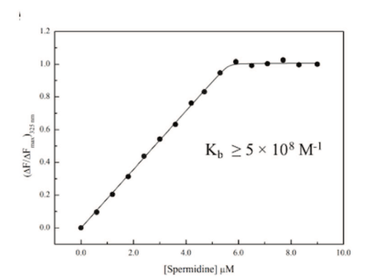


PotF binds putrescine/spermidine

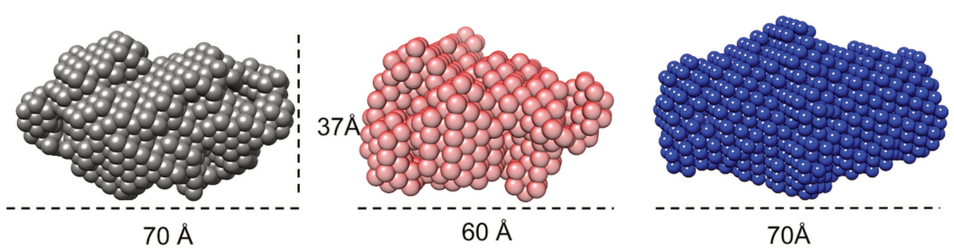








Structural changes



The citrus plant pathogen *Xanthomonas citri* has a dual polyamine-binding protein

Aline Sampaio Cremonesi, Lilia I. De la Torre, Maximillia Frazão de Souza, Gabriel S. Vignoli Muniz, Maria Teresa Lamy, Cristiano Luis Pinto Oliveira and Andrea Balan

Highlights

- Characterization of polyamine-related operons in *X. citri;*
- PotF is the putrescine/spermidine-binding protein of *X. citri;*
- PotF has preference for spermidine (higher K_b);
- Ligands induce the increasing in the thermal stability of PotF;
- SAXS analysis of PotF with ligands showed conformational changes.

Manuscript (docx or tex file) Click here to view linked References

The citrus plant pathogen *Xanthomonas citri* has a dual polyamine-binding protein

Aline Sampaio Cremonesi[1, 2], Lilia I. De la Torre[2, 3], Maximillia Frazão de Souza[4], Gabriel S. Vignoli Muniz[5], Maria Teresa Lamy[5], Cristiano Luis Pinto Oliveira[4] and Andrea Balan[2#]

1 - Programa de Pós-graduação Interunidades em Biotecnologia, Universidade de São Paulo, SP, Brazil;
2 - Laboratório de Biologia Estrutural Aplicada LBEA, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, SP, Brazil.
3 - Programa de Pós-graduação em Genética, Universidade Estadual de Campinas, Campinas, SP, Brazil;
4 - Grupo de Fluidos Complexos, Departamento de Física Experimental, Instituto de Física, Universidade de São Paulo, São Paulo, Brazil;
5 - Instituto de Física, Universidade de São Paulo, São Paulo, Brazil.

#Correspondence to Andrea Balan: abalan@usp.br
Applied Structural Biology Laboratory (LBEA)
Department of Microbiology
Institute of Biomedical Science II – University of São Paulo
Cidade Universitária - São Paulo - SP
CEP 05508-000
+ 55 (11) 3091-7745
+ 55 (11) 983471717

Running title: *Xanthomonas citri* has a dual putrescine/spermidine-binding protein PotD

Keywords:

Polyamine-binding protein; PotF; *Xanthomonas citri*; ligand-induced conformational changes

Abstract

ATP-Binding Cassette transporters (ABC transporters) are membrane proteins involved in import and export of different molecules, including ions, sugars, peptides, drugs, and others. Due the diversity of substrates, they have large relevance in physiological processes such as virulence, pathogenesis, and antimicrobial resistance. In *Xanthomonas citri* subsp. *citri,* the phytopathogen responsible for the citrus canker disease, it was shown that 20% of the ABC transporters are expressed under infection conditions, highlighting components of the putative putrescine/polyamine ABC transporter PotFGHI. Polyamines are ubiquitous molecules that mediate the cell growth and proliferation and play important role in the infection. In this work, we studied the periplasmic-binding protein PotF (Xac2476) and showed that the protein is highly conserved in *Xanthomonas* sp. genus and that *X. citri* has a set of proteins, which functions are directly related to the import or assimilation of polyamines. Ligand-protein binding constant values (K_b) were determined through fluorescence spectroscopy. For putrescine, it was determined a $K_b = 1.7 \times 10^5$ M^{-1}. Spermidine binds PotF much more strongly with a $K_b \geq 5 \times 10^8$ M^{-1}. Molecular modelling analysis associated with experiments of circular dichroism, intrinsic tryptophan fluorescence and small angle X-ray scattering (SAXS) revealed that the interactions performed by the two polyamines induced an increasing of the protein thermal stability and significant structural changes that coincides with closing of the domains. We also explored the conserved aspects of the ligand-binding site as well as differences for ligand binding specificity.

Introduction

Xanthomonas citri (*X. citri*) is a gram-negative phytopathogenic bacterium responsible for the citrus canker, disease that causes significant losses of citrus fruits affecting the Economy in Brazil and in the World [1, 2]. Production of xanthan gum polysaccharide [3, 4] induction of adhesins [5, 6] and type III and IV effector proteins are one

of the many mechanisms involved in the infection and pathogenesis [7, 8]. The relationship between ATP-Binding Cassette systems (ABC transporters) and physiological processes has been largely demonstrated in bacteria and also in the *Xanthomonas* genus [9–13]. Proteomics analyses made from leaves extracts from plants infected with *X. citri* revealed that more than 20% of the ABC transporters from *X. citri* are expressed under infection and pathogenesis. Among the highly expressed transporters there are components of the putative putrescine/polyamine ABC transporter PotFGHI [14, 15], encoded by *potFGHI* operon. Polyamines (putrescine, spermidine and spermine) are ubiquitous in almost all prokaryotic and eukaryotic cells. In eukaryotes, they play roles in synthesis and structure of nucleic acids and proteins, protection of oxidative damage, apoptosis among others [16]. In prokaryotes they are capable to bind to nucleic acids and through these interactions they mediate the cell growth and proliferation [17–19], gene regulation and differentiation [20], with active role in the infection caused by different microorganisms [21–23]. Polyamine signals from mammalian host are responsible for induction of the type III secretion system in *Pseudomonas aeruginosa* [24], one of the key determinants for virulence for this bacterium. In *E. coli*, three systems are related to polyamine import: the ABC transporters PotFGHI (specific for putrescine) and PotABCD (spermidine-preferential uptake system), responsible for maintenance of the levels of polyamines in presence of glucose, and the Puu system, which imports putrescine for its utilization as an energy source in glucose depletion [18, 25]. Structures of PotF and PotD, the periplasmic-binding proteins of the two ABC systems, are available and revealed that the binding site is characterized by four acidic residues that recognize the positive charged nitrogen and five aromatic side chains that anchor the methylene backbone by van der Walls interactions [18].

In the PotFGHI system of *X. citri,* PotF is the periplasmic binding protein but its ligand is not known. In this work we focused on the functional and structural characterization of this protein, based on molecular modelling, biophysical assays, and small angle X-ray scattering (SAXS). We produced the recombinant purified protein and showed that *X. citri* PotF binds both putrescine and spermidine. For putrescine it was determined a K_b = 1.7 $\times 10^5$ M^{-1},

whereas for spermidine a much higher association was observed, with an apparent binding constant $K_b \geq 5 \times 10^8\ M^{-1}$. SAXS and fluorescence spectroscopy showed that upon binding the protein suffers significant structural changes. Based on the structural data from orthologs, we showed the similarities and differences identified in the ligand-binding pocket and vicinity that are responsible for the dual specificity. Altogether, the results presented in this work provide the basis for further structural and functional studies of polyamine signal importation system in *X. citri* and other species of *Xanthomonas* genus.

Results

X. citri has a set of proteins related to uptake and assimilation of polyamines conserved in the *Xanthomonas* genus

X. citri potF gene (Xac2476) is located in a cluster of 10 genes, separated from the putative putrescine ABC transporter components (PotIHG or Xac2470-2472) by three genes that respectively, encode an outer membrane protein (OprN) and two components of multi-drug efflux pumps (RmrB and RmrA) (Fig. 1a). Also, upstream of *potF* sequence we identified genes encoding enzymes related to putrescine biosynthesis, such as *bioA* (putrescine aminotransferase, Xac2477), *glnA* (glutamine synthetase, Xac2478) and *guaA* (glutamine amidotransferase, Xac2479). *X. campestris* Xcc2346 protein, the putative ortholog of *X. citri* GlnA, it was identified by proteomic analysis in extracellular extracts of the bacterium cultivated in minimum medium [26]. In *P. aeruginosa* organization, the multi-drug efflux pump genes are absent but there are two genes encoding the periplasmic-binding proteins SpuD and SpuE, which bind respectively, putrescine and spermidine [27]. These two proteins use the same transport system in the inner membrane, formed by proteins SpuFGH. Differently, in *E. coli*, the genes encoding the components of putrescine (*potFGHI* genes) and polyamine (*potABCD*) transporters are located in two separated operons that encode their own specific transporters (Fig. 1a). Apparently, in *X. citri,* PotGHI

system is the corresponding ABC transporter for PotF, which specificity for different polyamines is not known. A search for additional polyamine transporters in the inner membrane using *E. coli* orthologs revealed at least five other systems that might transport polyamines: Xac2989 (PuuA/PuuP), a putative amino acid transporter (H^+) symporter, the proteins Xac3863/3864, Xac1841/1842 (YhdG) and Xac4354 (YhdG) forming three distinct cationic amino acid/polyamine antiporters, and the proteins Xac0857/Xac0858/Xac0860. The latter were previously described as the components of an oligopeptide ABC transporter type importer [28], but in our search, the proteins appear also as *E. coli* orthologs corresponding to components of an ABC transporter type exporter for polyamines, the SapBCDE system [29] (Fig. 1b) (Table S1, Supplementary material). Besides the transporters in the inner and outer membranes, we also identified in *X. citri* a set of proteins putatively involved in polyamines biosynthesis and catalysis, such as described in *E. coli* and *P. aeruginosa* [30–32]. Essentially, the enzymes belong to the catabolism of L-arginine to spermidine or putrescine and putrescine to succinate (Fig. 2). In KEGG database, proteins Xac3923 (SpeA, arginine decarboxylase), Xac2302 (hypothetical) and Xac3002 (synthetase/amidase) are related to polyamine biosynthesis but their role in the bacterium is not clear.

The ligand-binding pocket of *X. citri* PotF conserves functional and structural characteristics for interaction with both putrescine and spermidine

To identify proteins that had structural similarity to *X. citri* PotF, its amino sequence was submitted to a Blastp against the Protein Data Bank. The search resulted in four spermidine/putrescine-binding proteins that shared more than 45% of amino acid sequence similarity with *X. citri* PotF: *E. coli* PotF (bound to spermidine and putrescine, PDB codes 4JDF and 1A99, respectively) [33], and PotD (monomer and dimer conformations, PDB codes 1POT and 1POY) [34], *P. aeruginosa* SpuD (apo and bound to putrescine, PDB codes 3TTK and 3TTM) and SpuE (apo and bound to spermidine, PDB codes 3TTL and 3TTN)

[27] (Table S2, Supplementary material). For comparison with the putative orthologs, a three-dimensional model of *X. citri* PotF was built using the structural coordinates of *P. aeruginosa* SpuD (PDB code 3TTM), which shared 56% of amino acid sequence identity. To identify the putative ligand-binding pocket of *X. citri* protein, the model of PotF was superimposed with the structures of *E. coli* PotF and PotD and *P. aeruginosa* SpuD and SpuE and analysed together with the structural alignment performed with Expresso from T-Coffee Multiple Sequence Alignment program [35]. The structural alignment of the proteins showed that the residues that form the binding site are conserved (Fig. 3a). These residues were selected in each protein and mapped in *X. citri* PotF model (Fig. 3b). Four residues that promote polar interactions with the polyamines in these structures are identical in *X. citri* PotF: S^{12}, D^{13}, D^{220} and D^{251} (Fig. 3b, purple sticks). Similarly, the presence of a set of aromatic residues that might help with the stacking of the ligand is highly conserved (gray sticks), strongly suggesting that *X. citri* PotF could bind putrescine. When *E. coli* PotD and *P. aeruginosa* SpuE bound to spermidine were used for comparative analyses, we observed the relevance of polar and aromatic residues, similarly to what was identified for putrescine binding, but due the increased size of spermidine and occupation in the channel pocket, additional interactions are performed with E^{181} (SpuE) and Y^{85} and N^{327} (PotD), respectively. In *X. citri* protein, D^{159} is correspondent to E^{181} and N^{327} but no aromatic residue is localized in similar position to PotD Y^{85} (Fig. 3b). We also compared the electrostatic potential of the proteins looking at the entrance of the pocket (formed between N- and C-domains of PotF) aiming to determine specific features that might differentiate between putrescine- and spermidine-binding proteins (Fig. 4a). Proteins are shown as surface colored in red, blue, and light gray, respectively for negative, positive, and neutral charges. In general, the proteins presented similar charge pattern with the pocket's entrance (yellow circle) quite negative. On the other hand, significant differences can be evidenced in the pocket's volume and *X. citri* PotF shows a long channel (Fig. 4b, red surface). The high level of conservation in the ligand-binding residues, similar charges, and shape of the pocket, show that *X. citri*

PotF has features that suggested it would be capable for binding of both putrescine and spermidine.

X. citri PotF was expressed as a soluble and stable protein and suffered structural changes in presence of both putrescine and spermidine

To study the capability of *X. citri* PotF to bind spermidine and putrescine, a recombinant protein was produced in *E. coli* cells. After induction of *E. coli* Artic Express cells carrying the pET28-*potF* plasmid, 2.5 mg/L of soluble and stable PotF was expressed with expected molecular mass of 38.2 kDa (Fig. S1, Supplementary material). The protein was purified using immobilized metal affinity chromatography (IMAC) and eluted with 150 mL imidazole linear gradient. The purified samples were still submitted to the size-exclusion chromatography revealing one unique peak. Dynamic light scattering analysis of PotF in absence or presence of spermidine and putrescine showed that samples were monodisperse (0.2 polydispersity) consisting of a protein of estimated molecular weight of 37 kDa, which was compatible with its monomeric state (38.2 kDa). The hydrodynamic radius of the apo PotF was 3.37 nm and suffered a slight increasing in presence of putrescine (3.43 nm) and spermidine (3.53 nm) (Fig. S1, Supplementary Material).

The samples were submitted to circular dichroism analysis (CD) (Fig. 5). The CD spectrum of apo PotF was characteristic of an alpha/beta protein, as expected from the bioinformatics analyses and structural model, with minima signals at 208 nm and 222 nm. The presence of putrescine or spermidine in the samples did not significantly affect the secondary structure content and profile of the protein (Fig. 5a). In addition, the thermal stability of the protein was evaluated in apo state and in presence of the putative ligands, before (20°C) and after incubation of the protein sample at 100°C. The spectrum of the protein submitted to high temperature revealed the loss of the peak at 222 nm and decreasing in the α-helices content (Fig. 5b, dashed line) when compared to the protein at 20°C (solid line). A slight increasing of this peak is observed in the profile of the protein

submitted to decreasing of the temperature (100ºC to 20ºC) (Fig. 5b, dot line). Although PotF showed loss of secondary structure content with increasing of the temperature, it did not suffer complete denaturation, as observed by the CD values. To determinate the temperature of melting (T_m) and influence of the polyamines, PotF was submitted to the thermal denaturation analysis in absence and presence of 50 μM of putrescine and spermidine. The results revealed that the T_m of PotF, which was 61ºC, was not significantly changed in presence of putrescine and remained around 63°C ± 2°C but suffered significant increasing of 5 degrees after spermidine addition (T_m = 68°C) (Fig. 5c). Thermal shift analysis of PotF with increasing concentrations of the ligands corroborated the previous results and indicated that spermidine significantly induces the protein thermal stability (Fig. 5d).

PotF binds both putrescine and spermidine

X. citri PotF is fluorescent due to the presence of aromatic amino acid residues. In particular, PotF has four tryptophans (Fig. 6a), including two very close to the ligand-binding site (W^{217} and W^{249}) that allowed us to monitor the variations in the intrinsic fluorescence upon titration with putrescine and spermidine (Fig. 6b and 6d). Fluorescence experiments were performed using 1 ml samples of PotF (9 μM) in 50 mM Tris-HCl pH 8.0 and a stock solution of putrescine or spermidine (1 mM) in the same buffer solution. The emission spectra were obtained at 22.5ºC using an excitation beam light at 295 nm. Fig. 6b displays PotF fluorescence spectra in buffer with increasing amounts of putrescine. The fluorescence quantum yield is the ratio between the number of photons emitted and absorbed. Once that interaction of PotF with putrescine did not change its absorption spectra, it becomes evident that the interaction with putrescine resulted in an increasing of the fluorescence quantum yield, *i.e.* an increasing of the fluorescence emission intensity.

As there is no change of the position of the fluorescent band (see Fig. 6b), the fraction of the fluorescence intensity change at its maximum, 340 nm, ($\Delta F/\Delta F_{max} = (F-F_0)/(F-$

F_{max}), is plotted as a function of the putrescine concentration (Fig. 6c). F_0, F, and F_{max} are the fluorescence intensities of PotF at 340 nm in the absence of ligands, at a given ligand concentration, and at saturating concentration, respectively. By using eq. 01 (Material and Methods) to fit the experimental results the binding constant (K_b) was determined as 1.7×10^5 M^{-1} (Fig. 6c).

Similarly, the behavior of PotF in the presence of spermidine was evaluated by the changes in the PotF intrinsic fluorescence (Fig. 6d and 6e). Spermidine also did not change significantly PotF absorption spectrum. However, different from putrescine, increasing the concentration of spermidine not only increases the protein fluorescence intensity but also causes a blue shift in its emission spectrum (Fig. 6d). Under saturated conditions, the position of the maximum emission shifts 15 nm, from 340 to 325 nm. It is interesting to point out that the change in PotF fluorescence spectrum saturates when spermidine concentration (5.9 μM) is smaller than that of PotF (9 μM). This could indicate that during the process of PotF production some molecules ended up by not presenting free spermidine binding site. Considering the shift in the fluorescent spectrum, the graphic of the fraction of the fluorescence change, $\Delta F/\Delta F_{max}$, versus the spermidine concentration (Fig. 6e), used the fluorescence at 325 nm, which is the maximum of the fluorescent band at saturated spermidine concentration. The fitting with equation 01, yielded a K_b value of 5.0×10^8 M^{-1}.

Small angle X-ray scattering (SAXS) analysis of PotF in presence of putrescine and spermidine

In order to investigate PotF structural conformation in the apo and in the presence of 300 μM of putrescine or spermidine, we performed SAXS measurements. The SAXS data fit were computed by GNOM program [37] assuming a monodisperse system (Fig. 7a), and the evaluated pair distribution functions, p(r), are shown in Fig. 7b for each analyzed sample. The p(r) function shows similar behavior in the three SAXS profiles, where a bell-shape was

obtained with a maximum at 25 Å and a maximum dimension within the particle (D_{max}) approximately 73 Å. Low-resolution models were obtained using the program DAMMIN [38]. From the Kratky Plot ($I.q^2$ vs. q) analysis we obtained information related to the protein flexibility (Fig. 7c). In this plot, compact samples may present a bell-shape with the curve-approaching zero at high q values. Any degree of internal flexibility may cause an increasing of the final region. As it can be seen, the samples present similar behavior, indicating a globular compact folded particle. The average model from 10 independent runs to each SAXS data shows a prolate shape where it is observed a slightly more compact arrangement in case of PotF in the presence of spermidine (Fig. 7d). The samples also present similar structural parameters of radius of gyration (R_g) (PotF: R_g = 22.0 ± 0.1 Å; PotF + Putrescine: R_g = 21.9 ± 0.1 Å; PotF + Spermidine: R_g = 21.8 ± 0.1 Å, derived from GNOM [37].

In attempt to compare experimental SAXS data collected for PotF and theoretical SAXS intensity calculated from the high-resolution 3D models, we used CRYSOL program [39] and the structural coordinates of *P. aeruginosa* SpuD (PDB 3TTM) as entry (Table S2, Supplementary material). The superposition of the calculated (3TTM) and experimental SAXS profiles (PotF-Apo and in the presence of ligands) are shown in Fig. 8a as well as the determined *ab initio* structural models in comparison with the 3TTM structure Fig. 8b. The results indicated that the crystal model 3TTM was able to describe the SAXS data for the native protein and for the sample with ligands, with a slightly better agreement for the protein in the presence of putrescine (χ^2= 4.0).

Discussion

Polyamines are important molecules found in all living organisms, including plants, where they play several physiological functions from embryogenesis to flowering [50]. *X. citri* is a phytopathogenic bacterium that infects citrus plants. The citrus canker disease has no treatment and the knowledge of the essential mechanisms that the bacterium uses for infection, virulence and pathogenesis are important steps for the development of forms of

disease control. In this work, we characterized functional and structural aspects of the *X. citri* PotF, showing consequent conformational changes and capability of the protein to interact with not only one, but two polyamines, putrescine and spermidine. The comparison of the ligand-binding site of *X. citri* PotF with *P. aeruginosa* SpuD (spermidine-binding protein) and *E. coli* PotF (putrescine-binding protein) revealed the presence of W273 in a key position. Wu and co-workers [27] showed that the presence of this amino acid is essential for differentiation between spermidine or putrescine binding. This comparison suggested that *X. citri* PotF could bind spermidine and not putrescine. Curiously, our further experiments evidenced that, in fact, PotF has preference for spermidine, but also binds putrescine. The binding affinities and the effects of the ligand interaction were also evaluated in biophysical assays that showed more significant changes of the protein in the presence spermidine than putrescine, including a higher thermal stability. That might be explained by the fact that spermidine is twice the size of putrescine, which would lead to a higher occupancy of the ligand binding site and interactions with residues. The significant quenching evidenced in the fluorescence assays also was in accordance with the structural model that showed the presence of two tryptophan (W^{217} and W^{249}) in the ligand-binding site. Fluorescence experiments showed that both polyamines changed the emission of PotF, increasing the fluorescence quantum yield. The presence of both polyamines increases the fluorescence quantum yield, indicating that the non-radiative decay processes decrease in the ligand-bound protein. It is interesting that this could be interpreted as an increase of the rigidity of the environment of the tryptophan, since it is already well established that the increase in the rigidity can enhance the fluorescence quantum yield [42]. The interaction with putrescine did not shift the PotF emission spectrum (Fig. 6b), indicating that the region of the tryptophan residues is not changing in terms of accessibility to the molecules of the solvent. In contrast, the interaction between PotF and spermidine shifts the emission spectra to about -15 nm (Fig 6d). This is an indication that tryptophan dipolar relaxation is decreasing, suggesting a change in PotF conformation where tryptophan residues would sense a more hydrophobic environment. This result is directly related to the larger size of

spermidine in comparison with putrescine and its occupancy in the ligand-binding site, facing W217 and W249 (see Fig. 6a).

SAXS data showed similar behaviour for PotF in the apo and in the presence of both ligands and suggested that PotF in the presence of spermidine acquired a slightly more compact arrangement than the apo PotF. The crystal model 3TTM was able to describe experimental SAXS profile in a reasonable agreement in all cases, with indicated that the overall shape of the high-resolution structure for the protein, especially in the presence of putrescine where a slightly better fit was obtained.

The physiological relevance of this found is important, since both putrescine and spermidine are abundant polyamines present in citrus plants. The ability of the phytopathogen to capture and use host polyamines in its own metabolism should affect plant performance and increase the virulence and pathogenesis of *X. citri.* Moreover, we compared the genomic organization of polyamine-related genes in *Xanthomonas* species and other gamma-proteobacteria. Despite the differences found in the genetic organization, PotF is highly conserved in all the species of *Xanthomonas* genus, with amino acid sequence identities higher than 96% from 100% of query coverage, indicating the functional conservation and relevance of these proteins for bacteria that infect plants during the evolution. Moreover, the presence of one bivalent operon instead two, reveals a greater use of the genome in *X. citri* than in *E. coli* or *P. aeruginosa*.

Altogether, the data presented in this work showed that *X. citri* has a dual spermidine/putrescine-binding protein. Many phytopathogenic fungi, fungi-related organisms and proteobacteria from gamma subclass present genes for their import and metabolism. Due the high conservation of PotF in *Xanthomonas* genus, the data obtained in this work might help to the understanding of the physiological role of PotF in distinct pathogenic processes induced by different *Xanthomonas* species.

Material and Methods

Bioinformatics analysis

The sequences of genes and proteins from *X. citri, X. campestris, P. aeruginosa* and *E. coli* used in this study were obtained from Kyoto Encyclopedia of Genes and Genomes database (Kegg - http://www.genome.jp/kegg/). Table S1 (Supplementary material) shows the KEGG references and functions of all genes. Search for putative promoter regions and transcriptional units were performed using B-prom predictor (Softberry - http://www.softberry.com) and BioCyc site (http://biocyc.org/), respectively. Thre-dimensional models of *X. citri* PotF were built based on the structural coordinates of the *P. aeruginosa* SpuD in apo form (PDB code 3TTK) and in complex with putrescine (PDB-ID code 3TTM) [27] using the Modeller program [43].

DNA amplification and plasmid construction

A DNA fragment of 350 kb containing the *potF* gene (XAC2476, GI 1156547) without the first 21 nucleotides of the signal peptide was amplified by PCR from genomic DNA of *Xanthomonas citri* 306. The forward 5′ GGGCCCCCGCAACCTTTG 3′ and reverse 5′ CCATGGAACCACCGAGCAAC 3′ oligonucleotides, were used for the fragment amplification and insertion of the *Apa*I and *Nco*I restriction enzymes sites, respectively, in the start and final of the gene. The PCR product was cloned in pGEM®-T Easy (Promega) to generate the pGEM_potF, which was subsequently digested with *Apa*I and *Nco*I. The digested fragment was sub-cloned into the pET28a generating the pET28a_*potF* vector. The mature PotF expressed from this construct presented a N-terminal His_6-tagged. Standard molecular biology techniques were performed as described elsewhere [44]. DNA sequencing was used for construct verification.

Expression of the recombinant putrescine-binding protein PotF

The pET28a_potF vector was used to transform *E. coli* Artic Express cells using heat-shock method [45]. Pre-cultures of transformed cells were grown overnight at 37°C and 200 r.p.m. in LB broth containing the appropriate antibiotics (50 µg/mL kanamycin and 20 µg/mL gentamicin). Two percent of the pre-cultures were used to inoculate 1 liter of LB broth that was incubated at 37°C until the optical density at 600 nm reached 0.5 – 0.6, when the recombinant protein expression was induced with 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside), at 12°C, 200 r.p.m. for 24 hours. Cells were harvested by centrifugation at 4000×*g* for 10 min. at 4°C and stored at −20°C for at least 4 hours before the protein extraction procedure.

Protein extracts and purification of the His_6-tagged PotF protein

Induced bacterial cells were resuspended in affinity chromatography buffer A (50 mM sodium phosphate, pH 7.4, 150 mM NaCl, 5% glycerol, 0.1% Triton X-100, 1 mM PMSF, 5 mM Benzamidine and 1 mM β-mercaptoethanol) and incubated on ice with lysozyme (250 µg/mL) for 45 min. Cells were disrupted by 6 cycles of sonication (60% amplitude, 10 second pulses and interval of 20 seconds between the pulses) and the lysate was centrifuged at 30,000×*g* for 60 min at 4°C. The supernatants were purified by conventional chromatographic techniques using an ÄKTA FPLC system (Amersham Biosciences) and the HiTrap Chelating HP 5 mL (GE) column carried with 100 mM $NiSO_4$ and pre-equilibrated with buffer A. The column was washed with 20 mL buffer A and proteins were eluted with a 150 mL linear gradient (0 to 100%) of buffer B (50 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 5% glycerol, 1 mM PMSF, 5 mM Benzamidine, 1 mM β-mercapethanol and 500 mM imidazole). Fractions containing PotF were pooled, concentrated to 2 mL using a centrifugal filter device and loaded onto a Superdex 75 16/60 gel filtration column. Isocratic elution was performed with gel-filtration buffer (50 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 5% glycerol and 1mM DTT) at a flow rate of 0.3 mL/min.

Circular dichroism (CD) and Dynamic Light Scattring (DLS)

Far UV CD spectra were recorded on a Jasco-810 spectropolarimeter using a Peltier system PFD 425S for temperature control. 3 µM protein samples were prepared in 10 mM sodium phosphate buffer pH 7.4 and 15 mM NaCl. CD spectra were acquired at 20°C using a 1 mm path length cell at 0.5 nm intervals over the wavelength range from 190 to 260 nm. Ellipticity is reported as the mean residual ellipticity [θ] (deg cm^2 $dmol^{-1}$). Samples were subjected to thermal unfolding from 10°C to 100°C with spectra collected at 1°C intervals. The loss of secondary structure was followed by measuring the ellipticity at 222 nm using 0.5°C intervals. Midpoint transition temperatures were calculated as the center of the Gaussian fit of the first derivative of the denaturation curves. Refolding assays were started at 100°C and the temperature lowered to 10°C with concomitant acquisition of the ellipticity at 222 nm using 1.0°C intervals. DLS data were collected at 18°C using samples of recombinant PotF at 0.7 ml to 0.1 mg/mL diluted in buffer containing 50 mM sodium phosphate buffer pH 7.4, 150 mM NaCl and 5% glycerol. It was performed 300 accumulations of 10 seconds each.

Fluorescence measurements

UV-visible absorption spectroscopy measurements. UV-visible absorption spectra were obtained with an UV-visible spectrophotometer (Varian Cary, Santa Clara, CA).

Steady-State fluorescence measurements. The fluorescence spectra were obtained with a fluorimeter (Varian Cary Eclipse, Santa Clara, CA). Samples were placed in a quartz cuvette with an optical pathway of 4 mm. The experiments were conducted at room temperature (22.5ºC), with 1 ml samples of PotF (9 µM) in 50 mM Tris-HCl pH 8.0 and stock solutions of putrescine or spermidine (1 mM) in the same buffer solution added to the desired concentrations. Emission spectra were obtained using an excitation beam light at 295 nm. No inner filter correction was necessary [42], as Absorbance values at the excitation light,

295 nm, were found to be smaller than 0.05. From the emission spectra, an apparent binding constant was obtained (K_b) by using the nonlinear least squares method to fit the experimental data with the expression for the one-site binding model [36], equation 01.

$eq.01$:

$$F - F_0 = \frac{F_{max} - F_0}{2}\left\{\left(1 + \frac{C_L}{C_{PotF}} + \frac{1}{K_b C_{PotF}}\right) - \sqrt{(1 + \frac{C_L}{C_{PotF}} + \frac{1}{K_b C_{PotF}})^2 - 4\frac{C_L}{C_{PotF}}}\right\}$$

Where C_L and C_{PotF} are the ligand (putrescine or spermidine) and PotF molar concentrations, respectively; F_0, F, and F_{max} are the fluorescence intensities of PotF at a given wavelength in the absence of ligands, at a given ligand concentration, and at saturating concentration, respectively.

Small-angle X-ray scattering (SAXS)

SAXS analysis. The samples of apo PotF were placed at reusable quartz capillaries in a sample holder with controlled temperature (20°C). PotF sample at 5 mg/mL were diluted in 50 mM sodium phosphate buffer pH 7.4, and 150 mM NaCl. SAXS data was recorded at the Beamline SAXS1 at Synchrotron National Laboratory (LNLS) for three PotF samples with protein concentration of 130 μM diluted in 50 mM Tris HCl buffer pH 8.0, 100 mM NaCl and 3 mM DTT (Dithiothreitol). Ligands were equally added at final concentration of 300 μM. SAXS images were recorded in a 2D photon counting detector PILATUS for individual sample exposure time of 10s. A total of 10 images was collected for each sample and respectively for each sample buffer. The integration of the data was performed using Fit2D [46], resulting in a q range of $0.013 < q < 0.47$. The samples profiles were after submitted to data treatment using SUPERSAXS package (Oliveira, C.L.P and Pedersen, J.S; available at http://stoa.usp.br/crislpo/files/), where the scattering of buffer is subtracted from the corresponding protein scattering. The final intensity data, as a function of the modulus of the reciprocal space momentum transfer q, described by $4\pi \sin\theta/\lambda$ where θ is the

scattering angle and λ is the radiation wavelength, were analysed with GNOM program [37] applying Indirect Fourier Transformation (IFT) for a monodisperse system.

Data analysis. The pair distribution function p(r) were computed by experimental SAXS profile best fitting using the GNOM program [37] assuming a monodisperse system, providing the maximum diameter (D_{max}) and the radius of gyration (Rg).

Fitting of SAXS and structure modeling. To perform the modeling of the PotF structure, a molecular model was built based on the structural coordinates of the PotF protein from *P. aeruginosa*, solved at 2.0 Å resolution (PDB code 3TTM) [27] using the Modeller program [43]. Amino acid sequence alignment was performed using ClustalW [47]. The theoretical scattering curve of the modelled structure was calculated and compared with the experimental SAXS curve using the program CRYSOL [40] with entry the crystal model PotD (PDB 3TTM) from *P. aeruginosa*. The modelled structure was optimized against the SAXS data by using the program CORAL [41]. *Ab initio* models where performed using DAMMIN program [38] where simulated annealing method is applied to optimized dammy atom arrangement drive by D_{max} from p(r) function and χ^2 criterion. When necessary, the alignment of the *ab initio* models was performed using SUPCOMB [48] and DAMAVER [49], a set of programs based on alignment, selection of the most typical and building of the average model.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grant numbers 2011/22386-2 (PhD fellowship for ASP), 2018/20162-9 (Regular Research Project); Conselho Nacional de Desenvolvimento Científico (CNPq), grant number 401505/2016-2 (Universal Research Project); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); COLCIENCIAS (PhD fellowship for LILT). MTL is recipient of CNPq research fellowships. GSVM and MTL are part of the

National Institute of Science and Technology Complex Fluids (INCT-FCx), financed by CNPq (465259/2014-6) and FAPESP (2014/50983-3). We thank the infrastructure offered by the Brazilian National Laboratory of Synchrotron Light (LNLS), Campinas, Brazil; and the Department of Experimental Physics, University of São Paulo.

Author contributions

Aline Sampaio: Investigation, writing-original draft preparation. **Lilia Iriarte De La Torre:** Investigation and fluorescence data. **Maximillia Frazão de Souza**: writing and SAXS data analysis. **Sun Yang**: Investigation, SAXS data. **Gabriel Vignoli:** Investigation, fluorescence data analysis. **Maria Teresa Lamy:** Conceptualization, fluorescence analysis. **Cristiano L. P. Oliveira:** SAXS experiments, conceptualization and data analysis. **Andrea Balan:** Supervision, Writing-Reviewing and Editing.

Declaration of Competing Interests

All authors declare no competing interests.

References

1. Brunings, A. M. & Gabriel, D. W. (2003). *Xanthomonas citri*: breaking the surface. *Mol Plant Pathol*. 4, 141–157.
2. Yan, Q. & Wang, N. (2012). High-throughput screening and analysis of genes of Xanthomonas citri subsp. citri involved in citrus canker symptom development. *Mol Plant-Microbe Interact*. 25, 69–84.
3. Vojnov, A. A., Zorreguieta, A., Dow, J. M., Daniels, M. J., Dankert, M. A. (1998). Evidence for a role for the gumB and gumC gene products in the formation of xanthan from its pentasaccharide repeating unit by *Xanthomonas campestris*. *Microbiology*.

144, 1487–1493.

4. Nitschke, M. & Rodrigues, V. (2000). Effect of virulence and serial transfers of *Xanthomonas campestris* on xanthan gum production. *Brazilian J Microbiol*. 31, 58–60.

5. Mhedbi-Hajri, N., Darrasse, A., Pigné, S., Durand, K., Fouteau, S., Barbe, V., Manceau, C., Lemaire, C., Jacques, M. A. (2011). Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evol Biol*. 11, 1-12.

6. Dunger, G., Guzzo, C. R., Andrade, M. O., Jones, J. B., Farah, C. S. (2014). *Xanthomonas citri* subsp. citri type IV pilus is required for twitching motility, biofilm development, and adherence. *Mol Plant-Microbe Interact*. 27, 1132–1147.

7. Ryan, R. P., Vorhölter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., Dow, J. M. (2011). Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. *Nat Rev Microbiol*. 9, 344–355.

8. Souza, D. P., Oka, G. U., Alvarez-Martinez, C. E., et al. (2015). Bacterial killing via a type IV secretion system. *Nat Commun*. 6, 1–9.

9. Oshiro, E. E., Tavares, M. B., Suzuki, C. F., Pimenta, D. C., Angeli, C. B., de Oliveira, J. C. F., Ferro, M. I. T., Ferreira, L. C. S., Ferreira, R. C. C. (2010). Distribution and biological role of the oligopeptide-binding protein (OppA) in *Xanthomonas* species. *Genet Mol Biol*. 33, 341–347.

10. Casabuono, A., Petrocelli, S., Ottado, J., Orellano, E. G., Couto, A. S. (2011). Structural analysis and involvement in plant innate immunity of *Xanthomonas axonopodis* pv. *citri* lipopolysaccharide. *J Biol Chem*. 286, 25628–25643.

11. Tófoli De Araújo, F., Bolanos-Garcia, V. M., Pereira, C. T., et al. (2013). Structural and physiological analyses of the alkanesulphonate-binding protein (SsuA) of the citrus pathogen *Xanthomonas citri. PLoS One*. 8, 1-14.

12. Pegos, V. R., Nascimento, J. F., Sobreira, T. J. P., Pauletti, B. A., Paes-Leme, A., Balan, A. (2014). Phosphate regulated proteins of *Xanthomonas citri* subsp. *citri*: A proteomic approach. *J Proteomics*. 108, 78–88.

13. Moreira, L. M., Facincani, A. P., Ferreira, C. B., Ferreira, R. M., Ferro, M. I. T., Gozzo, F. C., de Oliveira, J. C. F., Ferro, J. A., Soares, M. R. (2015). Chemotactic signal transduction and phosphate metabolism as adaptive strategies during citrus canker induction by *Xanthomonas citri*. *Funct Integr Genomics*. 15, 197–210.

14. Bachmann, B. J. & Low, K. B. (1980). Linkage map of *Escherichia coli* K-12, edition 6. *Microbiol Rev*. 44, 1–56.

15. Higgins, C. F. (2001). ABC transporters: Physiology, structure and mechanism - An overview. *Res Microbiol*. 152, 205–210.

16. Liu, T., Huang, B., Chen, L., Xian, Z., Song, S., Chen, R., Hao, Y. (2018). Genome-wide identification, phylogenetic analysis, and expression profiling of polyamine synthesis gene family members in tomato. *Gene*. 661, 1–10.

17. Kusano, T., Yamaguchi, K., Berberich, T., Takahashi, Y. (2007). The polyamine spermine rescues Arabidopsis from salinity and drought stresses. *Plant Signal Behav*. 2, 251–252.

18. Igarashi, K. & Kashiwagi, K. (2000). Polyamines: Mysterious modulators of cellular functions. *Biochem Biophys Res Commun*. 271, 559–564.

19. Tabor, C. W. & Tabor, H. (1985). Polyamines in microorganisms. *Microbiol Rev*. 49, 81–99.

20. Bae, D. H., Lane, D. J. R., Jansson, P. J., Richardson, D. R. (2018). The old and new biochemistry of polyamines. *Biochim Biophys Acta - Gen Subj*. 1862, 2053–2068.

21. Ilari, A., Fiorillo, A., Baiocco, P., Poser, E., Angiulli, G., Colotti, G. (2015). Targeting polyamine metabolism for finding new drugs against leishmaniasis: a review. *Mini Rev Med Chem*. 15, 243–252.

22. Martini, C., Michaux, C., Bugli, F., et al. (2015). The polyamine N-acetyltransferase-like enzyme PmvE plays a role in the virulence of *Enterococcus faecalis*. *Infect Immun*. 83, 364–371.

23. Wojtasik, W., Kulma, A., Namysł, K., Preisner, M., Szopa, J. (2015). Polyamine metabolism in flax in response to treatment with pathogenic and non–pathogenic

Fusarium strains. *Front Plant Sci*. 6, 1–12.

24. Zhou, L., Wang, J., Zhang, L. H. (2007). Modulation of bacterial type III secretion system by a spermidine transporter dependent signaling pathway. *PLoS One*. https://doi.org/10.1371/journal.pone.0001291

25. Terui, Y., Saroj, S. D., Sakamoto, A., Yoshida, T., Higashi, K., Kurihara, S., Suzuki, H., Toida, T., Kashiwagi, K., Igarashi, K. (2014). Properties of putrescine uptake by PotFGHI and PuuP and their physiological significance in *Escherichia coli*. *Amino Acids*. 46, 661–670.

26. Watt, S. A., Wilke, A., Patschkowski, T., Niehaus, K. (2005). Comprehensive analysis of the extracellular proteins from *Xanthomonas campestris* pv. *campestris* B100. *Proteomics*. 5, 153—167.

27. Wu, D., Lim, S. C., Dong, Y., Wu, J., Tao, F., Zhou, L., Zhang, L. H., Song, H. (2012). Structural basis of substrate binding specificity revealed by the crystal structures of polyamine receptors SpuD and SpuE from *Pseudomonas aeruginosa*. *J Mol Biol*. 416, 697–712.

28. Moutran, A., Quaggio, R. B., Balan, A., De Souza Ferreira, L. C., De Cassia Cafe Ferreira, R. (2004). The Oligopeptide Permease (Opp) of the Plant Pathogen *Xanthomonas axonopodis* pv. *citri*. *Curr Microbiol*. 48, 354–359.

29. Sugiyama, Y., Nakamura, A., Matsumoto, M., Kanbe, A., Sakanaka, M., Higashi, K., Igarashi, K., Katayama, T,. Suzuki, H., Kurihara, S. (2016). A novel putrescine exporter SapBCDF of *Escherichia coli*. *J Biol Chem*. 291, 26343–26351.

30. Kurihara, S., Oda, S., Tsuboi, Y., Hyeon, G. K., Oshida, M., Kumagai, H., Suzuki, H. (2008). γ-glutamylputrescine synthetase in the putrescine utilization pathway of *Escherichia coli* K-12. *J Biol Chem*. 283, 19981–19990.

31. Kurihara, S., Tsuboi, Y., Oda, S., Kim, H. G., Kumagai, H., Suzuki, H. (2009). The putrescine importer puuP of *Escherichia coli* K-12. *J Bacteriol*. 191, 2776–2782.

32. Lu, C. D., Itoh, Y., Nakada, Y., Jiang, Y. (2002). Functional analysis and regulation of the divergent *spuABCDEFGH-spuI* operons for polyamine uptake and utilization in

Pseudomonas aeruginosa PAO1. *J Bacteriol*. 184, 3765–3773.

33. Vassylyev, D. G., Tomitori, H., Kashiwagi, K., Morikawa, K., Igarashi, K. (1998). Crystal structure and mutational analysis of the *Escherichia coli* putrescine receptor. Structural basis for substrate specificity. *J Biol Chem*. 273, 17604—17609.

34. Sugiyama, S., Matsu, I. Y. O., Maenaka, K., Vassylyev, D. G. (1996). PotD protein complexed with spermidine and the mechanism of polyamine binding. *Protein Sci*. 5, 1984–1990.

35. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 39, 13–17.

36. Valeur, B. & Berberan-Santos, M. N. (2012). In Molecular Fluorescence: principles and applications. Second edition. Wiley-VCH; Wiley-VCH Verlag GmbH & Co. KgaA. Weinheim, Germany.

37. Semenyuk, A. V. & Svergun, D. I. (1991). GNOM. A program package for small-angle scattering data processing. *J Appl Crystallogr*. 24, 537–540.

38. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*. 76, 2879–2886.

39. Pinto Oliveira, C. L. (2011). Investigating Macromolecular Complexes in Solution by Small Angle X-Ray Scattering. Curr Trends X-Ray Crystallogr. https://doi.org/10.5772/30730

40. Svergun, D., Barberato, C., Koch, M. H.. (1995). CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr*. 28, 768–773.

41. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria. G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr*. 45, 342–350.

42. Lakowicz, J. R. (2006). In Principles of fluorescence spectroscopy. Third Edition. Springer US. New York, NY, USA

43. Sali, A. & Blundell, T. (1994). Comparative modelling by satisfaction of spatial restraints. *J Mol Biol*. 234, 779–815.

44. Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D. Seidman, J. G., et al. (1998). In Current protocols in molecular biology. John Wiley & Sons, Inc., New York, USA.

45. Sambrook, J. (2001). Molecular cloning : a laboratory manual. Third edition. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press.

46. Hammersley, A. P. (2016). FIT2D: A multi-purpose data reduction, analysis and visualization program. *J Appl Crystallogr*. 49, 646–652.

47. Larkin, M.A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. 23, 2947-2948.

48. Kozin, M. B., Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J Appl Crystallogr*. 34, 33–41.

49. Volkov, V. V., Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr*. 36, 860–864.

50. Killiny, N., Nehela, Y. (2020). Citrus polyamines: structure, biosynthesis, and physiological functions. *Plants.* 9, 426, 1-28.

Figure Captions

Fig. 1. Genetic organization of *pot* genes of *Xanthomonas citri* and related in comparison with *Xanthomonas campestris, Escherichia coli and Pseudomonas aeruginosa*. (a) Cluster organization of Pot systems in *E. coli, P. aeruginosa, X. campestris* and *X. citri.* Yellow arrows represent genes from putative ABC transporter for putrescine/polyamine. KEGG codes and amino acid sequence identity is shown for each gene in comparison with *X. citri* ortholog. (b) Additional putrescine and polyamine transporter systems as described in *E coli* and their putative orthologs identified in *X. citri.* (*) *sapEDCB* operon is referred as a

putrescine ABC transporter type exporter, but *sapA* (b1294) gene encodes a periplasmic binding protein that usually belongs to importers. List of all genes/proteins and their functions is presented in Table S1 (Supplementary material).

Fig. 2. General view of the putative systems for polyamines transport and related enzymes found in *Xanthomonas citri*. Proteins were identified using *E. coli* and *P. aeruginosa* orthologs described in the literature and search for key words [polyamines], [spermine], [spermidine] and [putrescine] in *X. citri* databank (NCBI Reference sequence NC_003919.1). The pathways are shown as described for orthologs and in the Uniprot databank (https://www.uniprot.org). (?) enzymes missing or not identified, (P) periplasm, (IM) inner membrane and (C) cytoplasm. *X. citri* enzymes are showed according the KEGG reference. In grey boxes we show the intermediates compounds for each pathway. The list of all proteins, their putative functions, amino acid sequence identity and orthologs for the search are described in Table S1 (Supplementary material).

Fig. 3. Comparison between the ligand-binding site of *Xanthomonas citri* PotF and its orthologs that have three-dimensional structures available. (a) Structure-based amino acid sequence alignment showing residues that are involved with putrescine or spermidine coordination in the periplasmic binding proteins and their conservation in *X. citri* PotF. The alignment was performed using *Expresso* program from T-Coffee server [35] and highlights the residues involved with polyamines binding (in bold). The numbers are described according to the protein' structures. Conservation of residues is shown as: (*) identical; (:) similar; (.) different. PotF_Xac_2476_apo: *X. citri* PotF; PotF_Eco_1A99_put: *E. coli* putrescine-binding protein PotF bound to putrescine (PDB code 1A99) [33], SpuD_Pae_3TTM_put: *P. aeruginosa* putrescine-binding protein SpuD bound to putrescine (PDB code 3TTM) [27], SpuE_Pae_3TTN_spd: *P. aeruginosa* spermidine-binding protein SpuE bound to spermidine (PDB code 3TTN) [27], and PotD_Eco_1POT_spd: *E. coli* spermidine/putrescine-binding protein PotD bound to spermidine (PDB code 1POT) [34]. (b)

Structural comparison of the ligand-binding sites. The residues that interact with putrescine in *E. coli* PotF and *P. aeruginosa* SpuD and spermidine in *E. coli* PotD and *P. aeruginosa* SpuE were identified and used for mapping of the ligand-binding pocket in *X. citri* PotF. Polar and aromatic residues are shown in purple and gray sticks, respectively. Putrescine and spermidine are shown in yellow stick.

Fig. 4. Electrostatic potential and volume of the pockets in polyamine binding proteins in comparison with *Xanthomonas citri* PotF. (a) Comparison of the surface electrostatic potential from *X. citri* PotF and the polyamine binding proteins. The figure shows the proteins from the pocket (black circle) entrance perspective, detaching N- and C-domains. Electrostatic potential is shown as red, blue and gray for negative, positive and neutral charges, respectively. (b) Differences in the pocket of the five proteins. The pocket inner volumes are detached as red surfaces in the cartoon representation of the proteins. PDBs used for calculation are showed above the name of the proteins.

Fig. 5. Circular dichroism analysis of *Xanthomonas citri* PotF. (a) CD spectra of PotF (solid line) in comparison with PotF + 30 μM putrescine (dashed line) and PotF + 30 μM spermidine (dotted line). (b) CD spectra of the recombinant PotF (solid line) at 20ºC, after the increasing of the temperature up to 100ºC (dashed line) and after the decreasing of the temperature back to 20ºC (dotted line). (c) Thermal-induced unfolding of PotF (●), PotF + 50 μM of putrescine (●) and PotF + 50 μM of spermidine (○). Unfolding was followed by measuring the ellipticity at 222 nm during increasing of the temperature from 10°C to 100°C, with 0.5°C of intervals between each spectrum acquisition. CD spectra were acquired with 3 μM of protein in 10 mM sodium phosphate buffer pH 7.4 and 10 mM NaCl as described in Material and Methods.

Fig. 6. Intrinsic fluorescence measurements of *Xanthomonas citri* PotF. (a) Three-dimensional structure model of PotF in cartoon showing the position of the four residues of

tryptophans (showed as balls) identified in the protein sequence. Putrescine and spermidine (cyan and pink sticks, respectively) were docked in the ligand-binding pocket for to show the proximity with the tryptophans. Domains N- and C are pointed. (b) Typical fluorescence spectra of 9 μM PotF (in buffer 50 mM Tris pH 8.0 and 50 mM NaCl) in the absence and with increasing concentration of putrescine. (c) Normalized change of PotF fluorescence at 340 nm as a function of putrescine concentration. (d) Typical fluorescence spectra 9 μM PotF (in buffer 50 mM Tris pH 8.0 and 50 mM NaCl) in the absence and with increasing concentration of spermidine. (e) Normalized change of PotF fluorescence at 325 nm as a function of spermidine concentration. Excitation at 295 nm.

Fig. 7. SAXS data analysis for *X. citri* apo PotF and in presence of putrescine and spermidine. (a) Experimental SAXS profile of PotF, PotF + putrescine and PotF + spermidine and the fitting obtained using the GNOM program with IFT method assuming a monodisperse system. (b) The pair-distance distribution function p(r) respectively for each analyzed SAXS profile in (a). (c) Kratky plot obtained by $I(q)xq^2$ vesus q for apo PotF and PotF in presence of putrescine and spermidine. (d) *Ab initio* models using DAMMIN program for apo PotF (gray) and PotF in presence of spermidine (red) and putrescine (blue), the average envelops for ten individually runs, respectively for each sample, are shown in sphere representation. The models show a slightly compact conformation for PotF in the presence of spermidine

Fig. 8. Comparison of the SAXS experimental data to the crystal model 3TTM. (a) Superposition of the experimental intensity of SAXS measured for apo PotF (up), PotF + spermidine(middle) and PotF + putrescine(down)with the theoretical SASX profile for 3TTM model (inset plot as green *cartoon* representation), resulting in a χ^2 of 5.8, 4.5 and 3.7, respectively. (b) Comparison between Ab initio models and 3TTM for apo PotF (up), PotF + spermidine(middle) and PotF + putrescine(down).

Figures

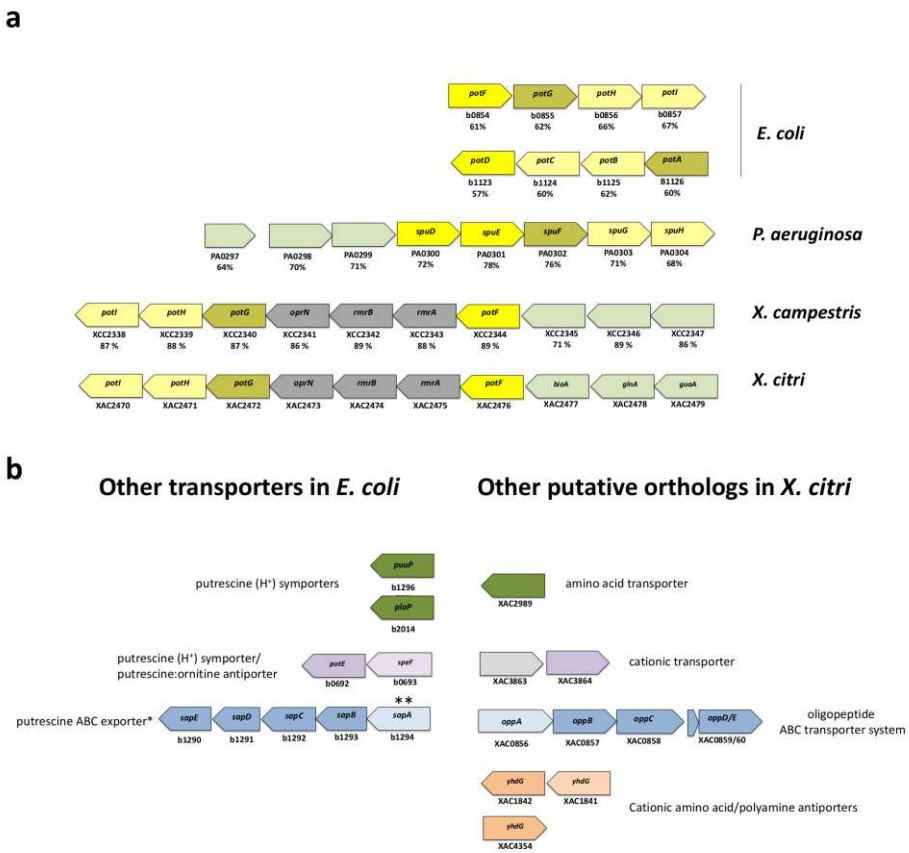
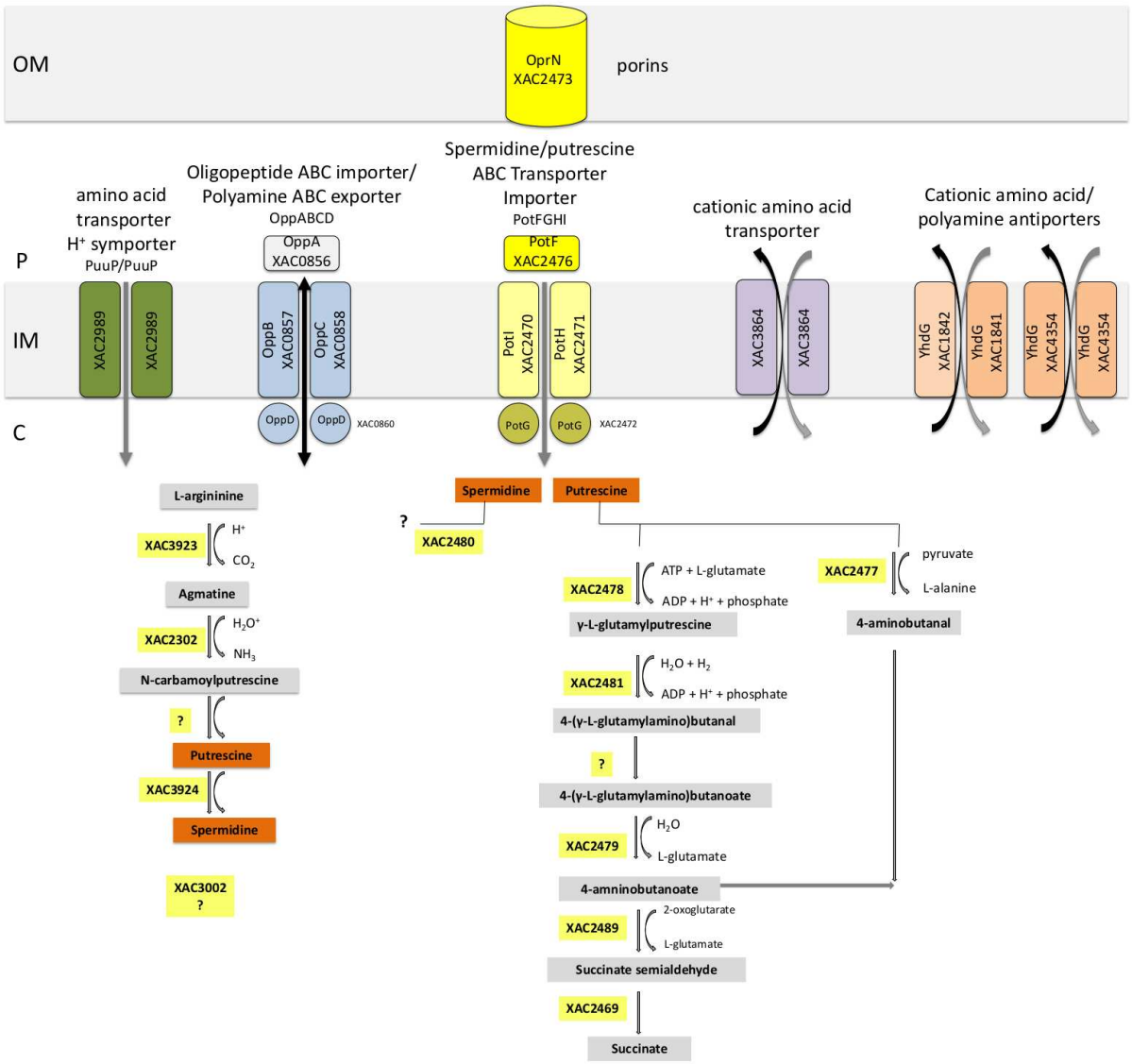


Fig. 1



743
744 **Fig. 2**

745

a

```
PotF_Xac_2476_apo VYNYSDYIA16.....157PSDMI161.....215VGWSGDI221.....244EGAPQWFDML253
PotF_Eco_1A99_put IYNWSDYIA42.....157PEEVF161.....242IGWAGDV248.....270EGAMAFFDVF280
SpuD_Pae_3TTM_put VYNWSDYIA40.....181PTEIL185.....239IGYSGDI245.....267EGAGSFFDMV277
SpuE_Pae_3TTN_spd IYNWTDYIA39.....179GDEML183.....237FGYSGDV243.....266EGANLWFDLM275
PotD_Eco_1POT_spd FYNWTEYVP39.....169AREVF173.....227MIWNGSA233.....250EGGIFWMDSL259
                  .**: :*:      :::        . : *.       **:  ::*
```

b

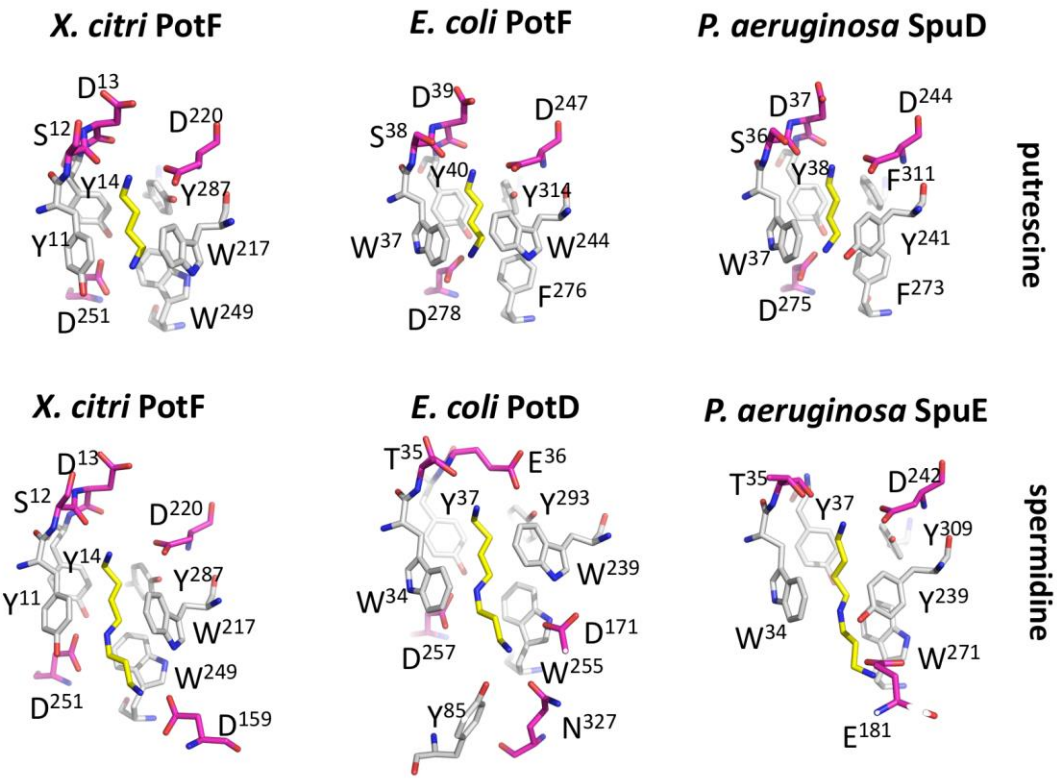


Fig. 3

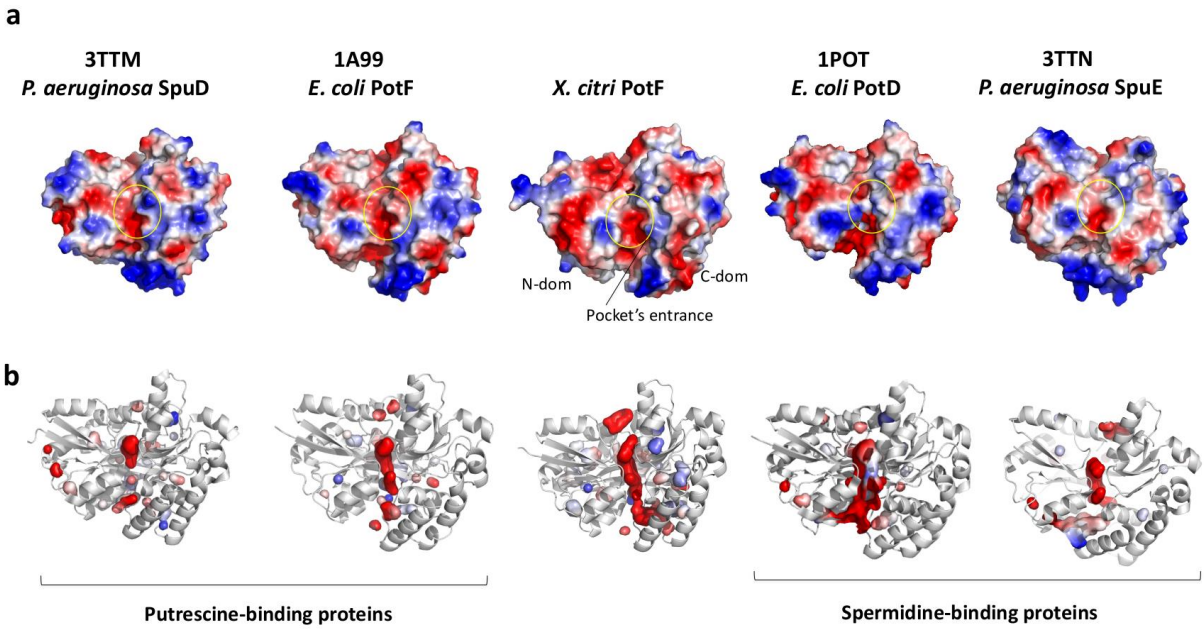

a
3TTM
P. aeruginosa SpuD
1A99
E. coli PotF
X. citri PotF
1POT
E. coli PotD
3TTN
P. aeruginosa SpuE
N-dom
C-dom
Pocket's entrance
b
Putrescine-binding proteins
Spermidine-binding proteins

Fig. 4

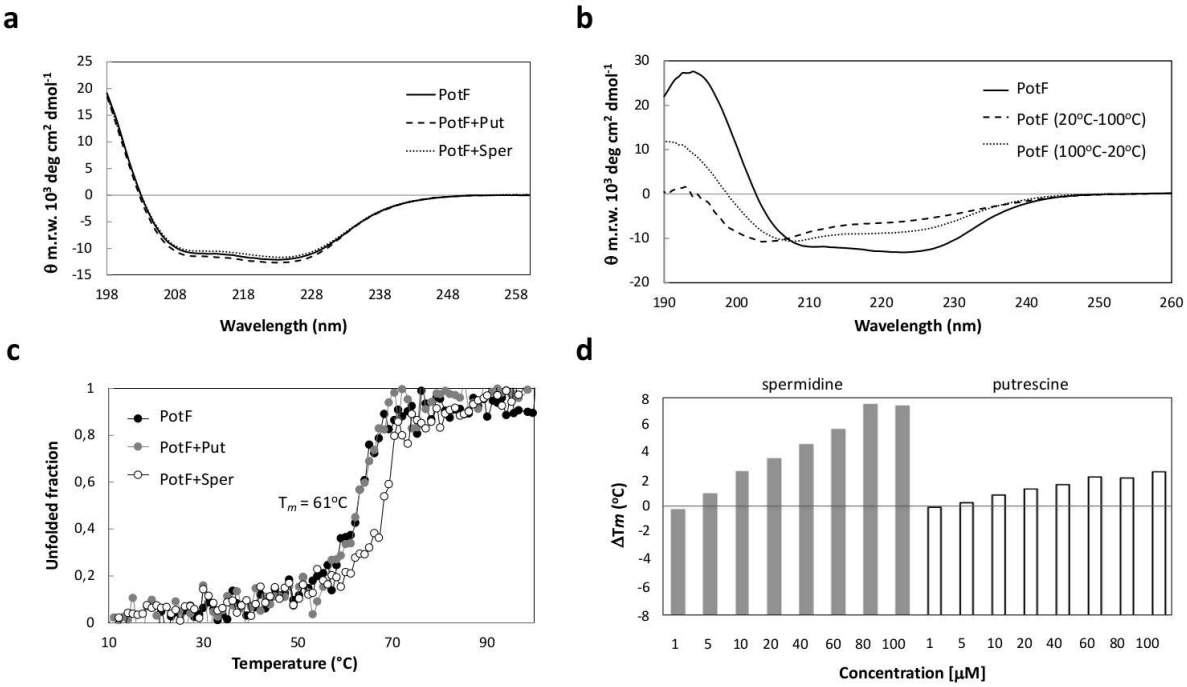

a
θ m.r.w. 10³ deg cm² dmol⁻¹
PotF
PotF+Put
PotF+Sper
Wavelength (nm)
b
θ m.r.w. 10³ deg cm² dmol⁻¹
PotF
PotF (20°C-100°C)
PotF (100°C-20°C)
Wavelength (nm)
c
Unfolded fraction
PotF
PotF+Put
PotF+Sper
T_m = 61°C
Temperature (°C)
d
spermidine
putrescine
ΔTm (°C)
Concentration [μM]

Fig. 5

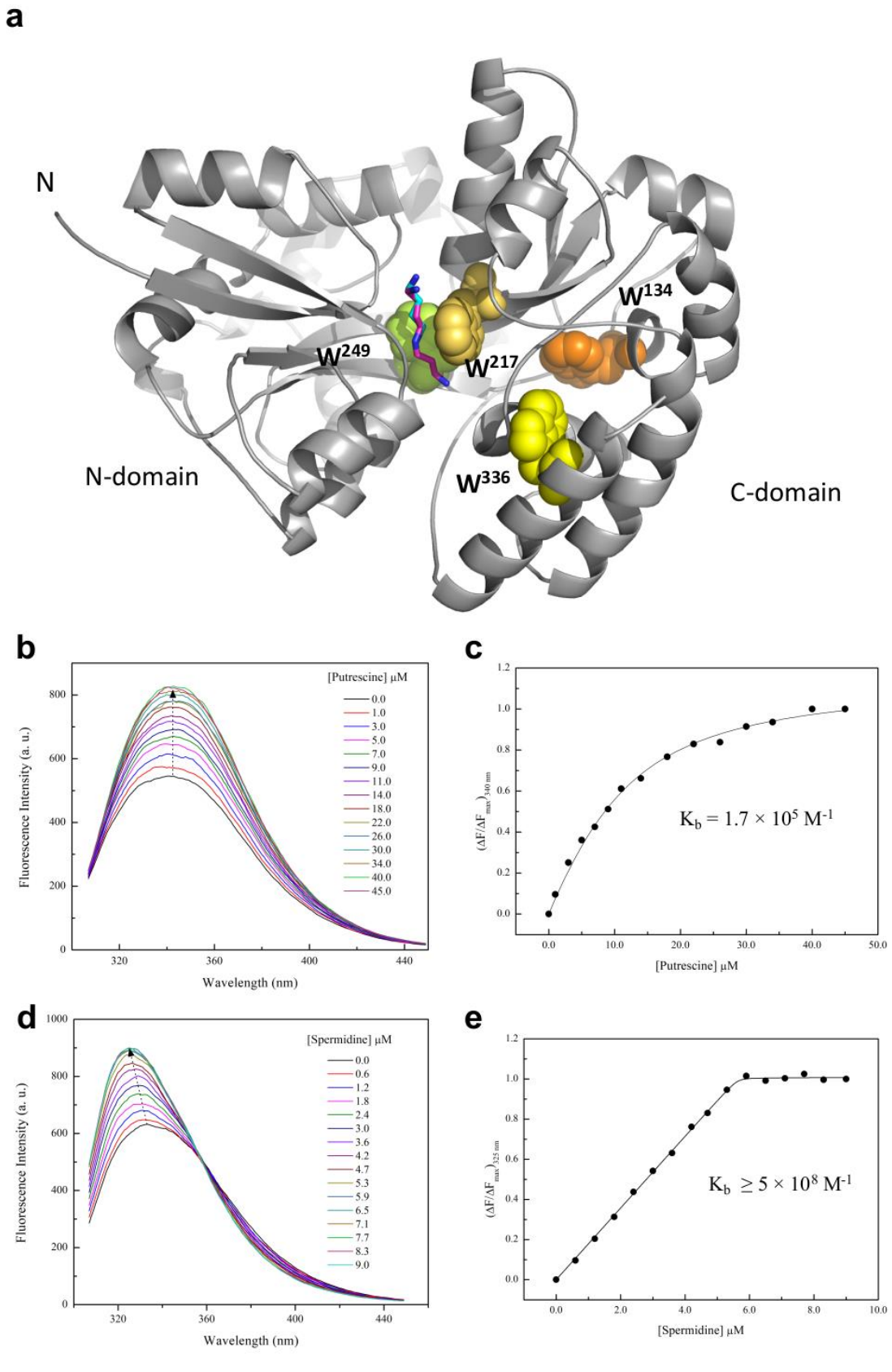

a
N
W134
W249
W217
W336
N-domain
C-domain
b
[Putrescine] μM
0.0
1.0
3.0
5.0
7.0
9.0
11.0
14.0
18.0
22.0
26.0
30.0
34.0
40.0
45.0
Fluorescence Intensity (a. u.)
Wavelength (nm)
c
Kb = 1.7 × 10^5 M^-1
[Putrescine] μM
d
[Spermidine] μM
0.0
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.7
5.3
5.9
6.5
7.1
7.7
8.3
9.0
Fluorescence Intensity (a. u.)
Wavelength (nm)
e
Kb ≥ 5 × 10^8 M^-1
[Spermidine] μM

Fig. 6

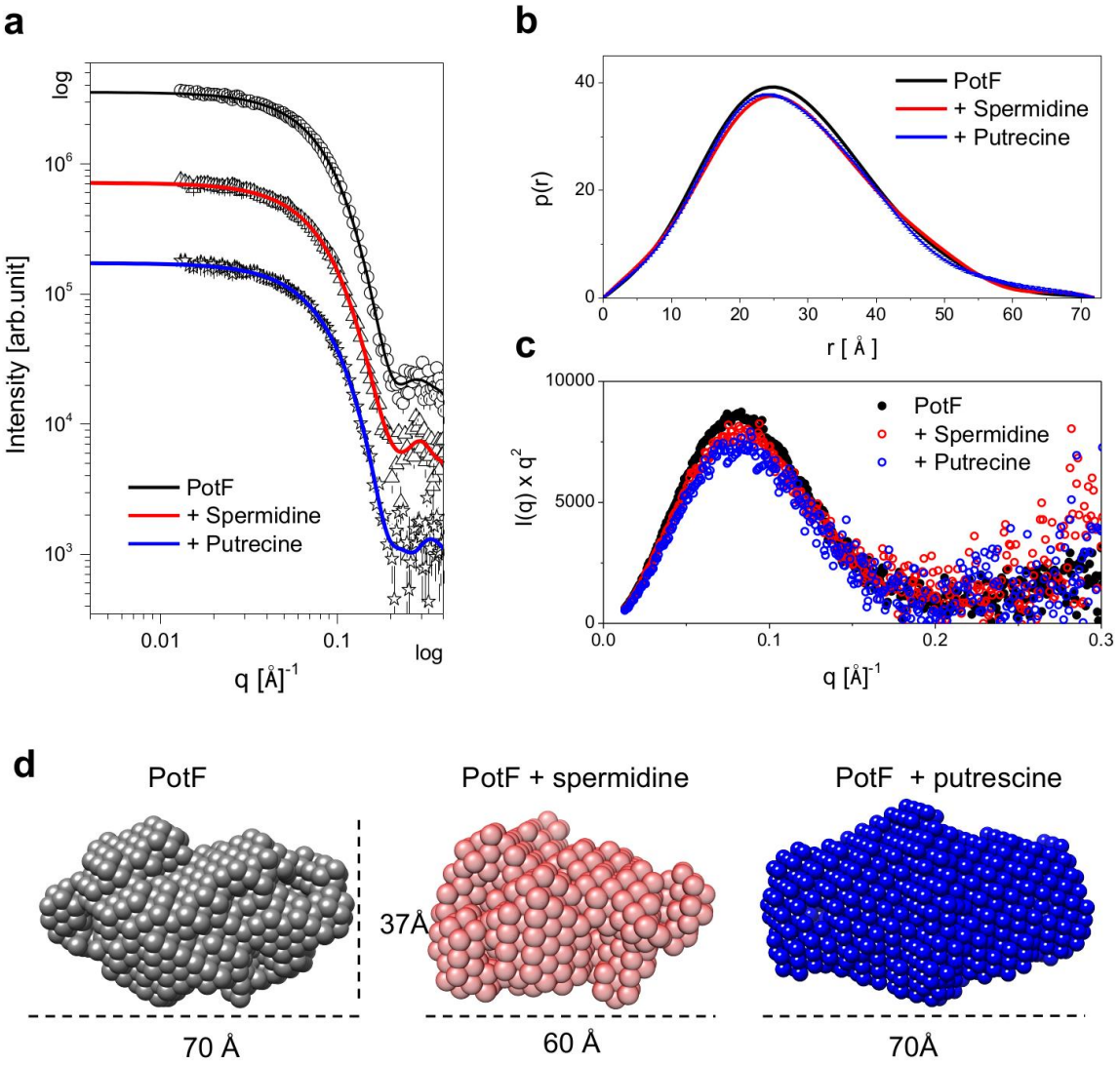

a
log
10^6
10^5
10^4
10^3
Intensity [arb.unit]
PotF
+ Spermidine
+ Putrecine
0.01
0.1
log
q [Å]$^{-1}$
b
40
20
0
p(r)
PotF
+ Spermidine
+ Putrecine
0
10
20
30
40
50
60
70
r [Å]
c
10000
5000
0
I(q) x q^2
PotF
+ Spermidine
+ Putrecine
0.0
0.1
0.2
0.3
q [Å]$^{-1}$
d
PotF
PotF + spermidine
PotF + putrescine
37Å
70 Å
60 Å
70Å

Fig. 7

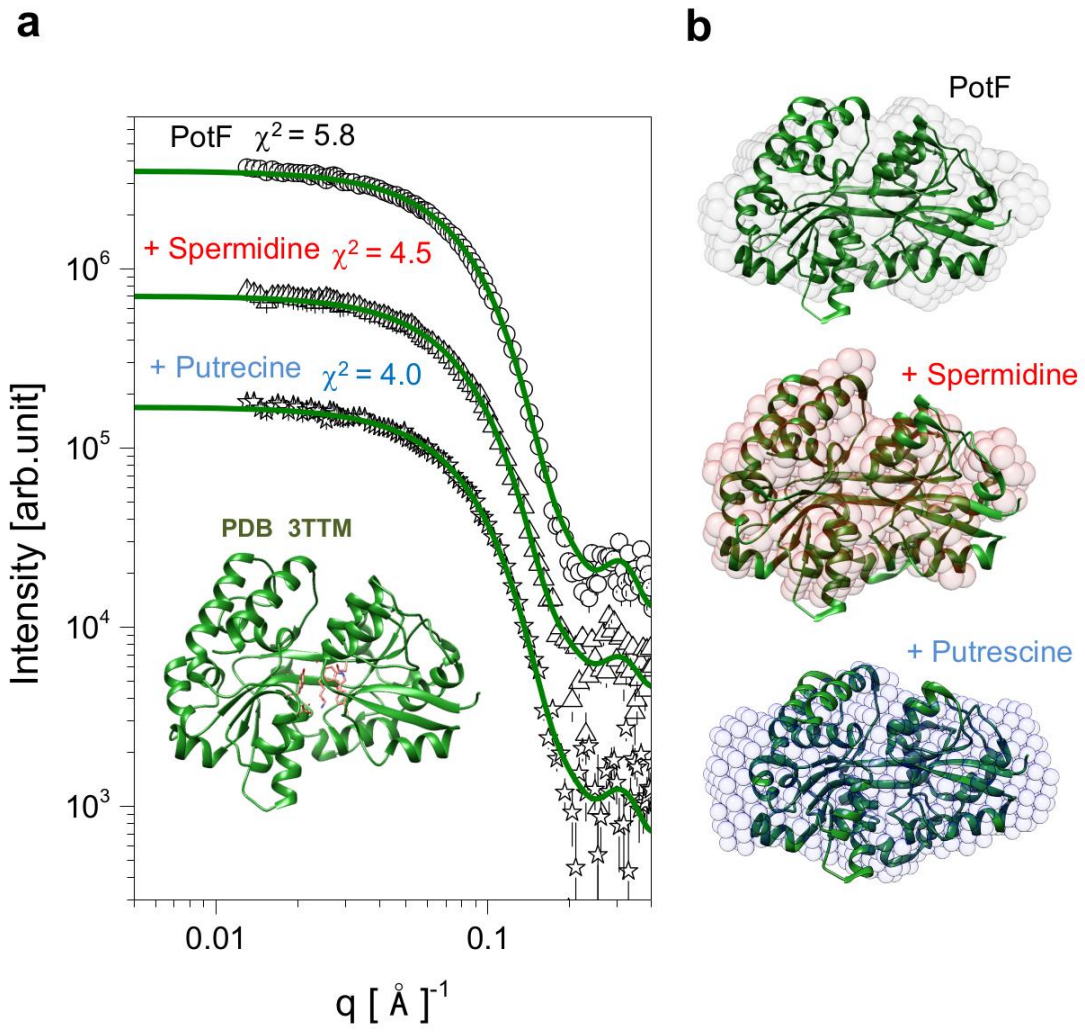

a
b
PotF χ^2 = 5.8
+ Spermidine χ^2 = 4.5
+ Putrecine χ^2 = 4.0
PDB 3TTM
Intensity [arb.unit]
10^6
10^5
10^4
10^3
0.01
0.1
q [Å]$^{-1}$
PotF
+ Spermidine
+ Putrescine

Fig. 8

Click here to access/download
Supplementary Material (To be Published)
Cremonesi_et_al_Suppl_Material.pdf

Anexo 2 - Artigo sobre peptídeos antimicrobianos sintéticos

Artigo "*Interaction of synthetic antimicrobial peptides of the Hylin a1 family with models of eukaryotic structures: zwitterionic membranes and DNA*", publicado no periódico Biochemistry and Biophysics Report (BBR). O trabalho foi desenvolvido pelo grupo de Biomembranas do Instituto de Física da USP, e eu participei como colaboradora na obtenção de dados de mobilidade electroforética dos AMPs em presença de DNA plasmidial.

Biochemistry and Biophysics Reports 24 (2020) 100827

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

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

Interaction of synthetic antimicrobial peptides of the Hylin a1 family with models of eukaryotic structures: Zwitterionic membranes and DNA

Gabriel S. Vignoli Muniz [a,*], Lilia I. De la Torre [b], Evandro L. Duarte [a], Esteban N. Lorenzón [c], Eduardo M. Cilli [d], Andrea Balan [b], M. Teresa Lamy [a,**]

[a] Instituto de Física, Universidade de São Paulo, SP, Brazil
[b] Instituto de Ciências Biomédicas, Universidade de São Paulo, SP, Brazil
[c] Unidade Acadêmica Especial Ciências da Saúde, Universidade Federal de Jataí, Jataí, GO, Brazil
[d] Instituto de Química, Universidade Estadual Paulista, Araraquara, SP, Brazil

ARTICLE INFO

Keywords:
Antimicrobial peptides
DPPC vesicles
DNA
DSC
Fluorescence spectroscopy
Electrophoresis

ABSTRACT

Antimicrobial peptides (AMPs) have been appointed as a possible alternative to traditional antibiotics in face of pathogens increasing resistance to conventional drugs. Hylin a1 (IFGAILPLALGALKNLIK), an AMP extracted from the skin secretion of a South American frog, *Hypsiboas albopunctatus*, was found to show a strong cytotoxicity against bacteria and fungus, but also a considerable hemolytic action. Considering the toxicity of the peptide in eukaryotic cells, this work focuses on investigating the effects of the interaction of the Hylin a1 analogues W^6Hya1, D^0W^6Hya1 and K^0W^6Hya1 with models of eukaryotic structures, namely zwitterionic liposomes of dipalmitoyl phosphatidylcholine (DPPC) and calf-thymus DNA (CT DNA). Through intrinsic Trp fluorescence we determined that the peptide affinity for fluid DPPC bilayers follows the decreasing order: D^0W^6Hya1 (+2) > W^6Hya1 (+3) » K^0W^6Hya1 (+4). Fluorescence data also indicate that the Trp residue in the more positively charged peptide, K^0W^6Hya1, is less deep in the bilayer than the residue in the other two peptides. This finding is supported by differential scanning calorimetry (DSC) data, which shows that both D^0W^6Hya1 and W^6Hya1 disturb DPPC gel-fluid transition slightly more effectively than K^0W^6Hya1. DPPC DSC profiles are homogeneously disturbed by the three peptides, probably related to peptide-membrane diffusion. Surprisingly, the peptide that displays the lowest affinity for PC membranes and is located at the more superficial position in the bilayer, K^0W^6Hya1, is the most efficient in causing formation of pores on the membrane, as attested by carboxyfluorescein leakage assays. The three peptides were found to interact with CT DNA, with a deep penetration of the Trp residue into hydrophobic pockets of the double helix, as indicated by the significant blue shift on the Trp fluorescence, and the displacement of DNA-bound ethidium bromide by the peptides. The experiments of DNA electrophoresis confirm that Hylin peptides bind DNA in a concentration-dependent manner, inducing complete DNA retardation at the relative AMP/plasmid DNA weight ratio of ~17. These findings could help to better understand the AMPs toxic effects on eukaryotic cells, thus contributing to the design of healthier therapeutic agents.

1. Introduction

Due to the increasing resistance of pathogens against traditional antibiotics, as well as the emergence of new diseases, a severe and relevant threat to public health on a global scale, the scientific community has been making efforts in search for new therapeutic drugs [1–3]. Among different classes of molecules with activity against pathogenic microorganism, antimicrobial peptides (AMPs) have been proposed as a possible next generation of therapeutic agents [4–6]. Furthermore, AMPs can affect different pathogens or parasites ranging from bacteria, protozoa, and even virus.

AMPs consist of a group of short molecules, generally containing between 4 and 50 amino acids. More than 3000 AMPs have already been discovered in different living beings, from prokaryotic to unicellular eukaryotes, as well as multicellular organisms [7–9]. AMPs are rich in residues of arginine and/or lysine, which give them positive net charge

* Corresponding author.
** Corresponding author.
E-mail addresses: gvignoli@if.usp.br (G.S. Vignoli Muniz), mtlamy@usp.br (M.T. Lamy).

https://doi.org/10.1016/j.bbrep.2020.100827
Received 17 August 2020; Received in revised form 26 September 2020; Accepted 27 September 2020
Available online 3 November 2020
2405-5808/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

when in physiologic pH. Moreover, due to the presence of hydrophobic and hydrophilic residues, AMPs present an amphiphilic character and many of these molecules interact strongly with amphiphilic aggregates and lipid bilayers [7]. Considering that AMPs are generally cationic, most of the studies emphasize their interaction with negative structures, particularly negative domains in lipid bilayers [10].

Nonetheless, the precise antibiotic mechanism of the AMPs is still a matter of debate [11–13]. For instance, it has been reported that some AMPs may act as destabilizing lipid bilayers, inducing membrane disruption. On the other hand, they could also act by either inducing changes in a cellular enzyme, and/or inhibiting nucleic acid synthesis [14,15]. As an example, the AMP coprisin (TCDVLSFEAKGIAVNHSACALHCIALRKKGGSCQNGVCVCRN) has the ability to cross the membrane of *Escherichia coli* without inducing membrane permeabilization, and once it is in the cytosol, this peptide induces the cell's death through apoptosis [16]. Moreover, some AMPs, such as Hecate (FALALKALKKALKKLKKALKKAL), and its analogue GA-Hecate, present a dual action: at low concentrations they induce apoptosis whereas at high concentration they provoke membrane disruption [15].

Like most drugs, AMPs can also cause damaging effects on mammalian cells. Hence, synthetic AMPs, based on the primary sequence of native AMPs, have been designed trying to magnify the cytotoxicity against pathogens and minimize any possible secondary effects on healthy cells [17].

The three peptides used in this work (Fig. 1) are based on the sequence of the native Hylin a1 (IFGAILPLALGALKNLIK) extracted from the skin secretion of the frog *Hypsiboas albopunctatus*. Hylin a1 (Hya1) presents a considerable hemolytic and a strong antimicrobial action [18]. The changes in the native sequence of Hya1 consist in the substitution of a leucine by a tryptophan (Trp) residue at the sixth position of the peptide chain (W^6Hya1), and a modification at the N-terminus group, with the insertion of an amino acid residue, either an aspartate residue (D^0W^6Hya1) or a lysine one (K^0W^6Hya1).

The changes in the natural sequence of the native Hya1 modulated the peptides toxicity against cultures of bacteria and fungus [19]. The introduction of a tryptophan residue (W^6Hya1), without modification of the peptide's net charge, resulted in an increase of activity against Gram-positive bacteria and fungus cultures, as well as a higher hemolytic action. Similarly, the addition of a negative charge (D^0W^6Hya1) enhanced the peptide antibiotic activity against Gram-positive bacteria, but a decrease against Gram-negative bacteria was observed [19]. The extra positive charge in the peptide chain (K^0W^6Hya1) expanded its antimicrobial spectrum, with activity against cultures of Gram-positive and Gram-negative bacteria and fungus, but an increase in its hemolytic action [19]. In addition, K^0W^6Hya1 also presents activity against planktonic and biofilm growth of oral bacteria [20], and it has been used in the control of bacterial diseases of citrus [21].

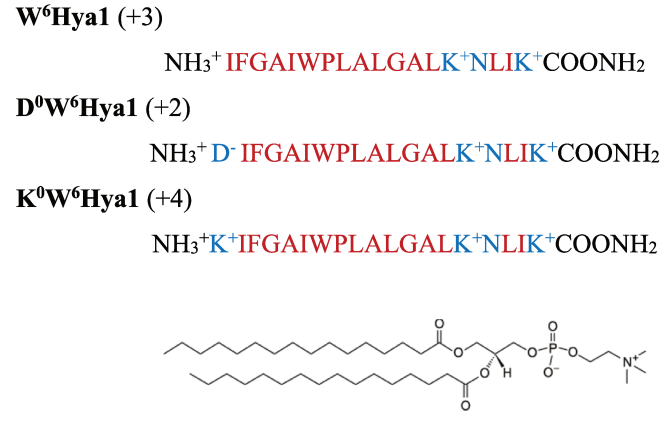


Fig. 1. Primary structures of W^6Hya1, D^0W^6Hya1 and K^0W^6Hya1, and the chemical structure of the lipid DPPC. Polar amino acid residues are drawn in blue and non-polar in red. The peptides net charges are indicated in parentheses. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Furthermore, Hylin peptides are unstructured in water, whereas in the presence of zwitterionic micelles of LPC (dodecylphosphocholine) the peptides display an ordered secondary structure composed mostly by α-helix [19]. Previously, it has been shown that Hylin peptides interact with anionic amphiphilic aggregates [19,22]. The comparative interaction of K^0W^6Hya1 with zwitterionic and anionic membranes, mimicking mammalian and bacterial membranes, respectively, was investigated [23], showing that though the peptide interacts with DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine), its interaction with DPPG (1,2-dipalmitoyl-*sn*-glycero-3-phospho-(1′-*rac*-glycerol)) was much stronger.

The focus of the present work is a comparative study of the effects of the interaction of the three Hylin analogues, W^6Hya1 (+3), D^0W^6Hya1 (+2), and K^0W^6Hya1 (+4) with models of eukaryotic structures, namely DPPC, a zwitterionic lipid used here to mimic healthy mammalian membranes [24,25], and calf-thymus DNA (CT DNA), aiming at the design of healthier therapeutic agents. Different methodologies were applied, such as intrinsic Trp fluorescence, as the three peptides have a Trp residue, differential scanning calorimetry of DPPC membranes disturbed by the peptides, and the ability of the peptides to cause pore formation in PC membranes, through the measurement of the leakage of entrapped carboxyfluorescein (CF), a fluorescent dye, in PC large unilamellar vesicles. The interaction of the peptides with CT DNA was studied via both Trp fluorescence and by the competitive studies with CT DNA previously bound to ethidium bromide. Furthermore, the interaction of Hylin peptides with plasmid DNA was evaluated by electrophoresis experiments.

2. Materials & methods

2.1. Chemicals and reagents

1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) and 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) were acquired from Avanti Polar Lipids. Calf-thymus DNA (CT DNA), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), sephadex-G25 medium column, glucose, chloroform, ethylenediamine tetraacetic acid (EDTA), 3,8-Diamino-5-ethyl-6-phenylphenanthridinium bromide (EB), sodium hydroxide (NaOH), hydrochloric acid (HCl), and sodium chloride (NaCl) were purchased from Sigma Aldrich (St Louis, MO). The plasmid pOP3BP (pDNA) was kindly gifted from Marko Hyvönen (Dep. of Biochemistry, University of Cambridge). All solutions or dispersions were prepared with Milli-Q water or chloroform.

2.2. Peptide synthesis

The peptides were synthesized manually using the N-9-fluorenylmethoxycarbonyl (Fmoc) chemistry, purified, and characterized according to the experimental protocol, as described somewhere else [19]. The purity of the Hylin peptides was found to be higher than 98%, as attested by Reversed-phase high-performance liquid chromatography (RP-HPLC) (Fig. SM1).

2.3. Large unilamellar vesicle preparations

The desired amount of lipids was solubilized in chloroform. By using a flux of gaseous nitrogen, the solvent was evaporated and thus a thin film of lipids formed at the bottom of the glass tube. Then, the lipid film was kept under low pressure conditions for a minimum of 3 h. Aqueous dispersions were prepared by the addition of buffer (10 mmol L^{-1} HEPES, 3 mmol L^{-1} NaCl, pH 7.4) to the lipid film, followed by vortexing for 2 min at 50 °C. Finally, lipid dispersions were extruded through polycarbonate filters (mini-extruder by Avanti Polar Lipids, 19

mm membranes with 100 nm pores, 31 times) above the lipid gel−fluid transition temperature (≥50 °C), for the formation of large unilamellar vesicles (LUVs). All lipid dispersions used in this work were freshly prepared on the same day of the experiments. Through inorganic phosphate assay [26] we determined the lipid concentration before and after the extrusion process: the difference was smaller than 5%.

2.4. CT DNA stock solution

CT DNA was diluted into buffer (10 mmol L^{-1} HEPES, 3 mmol L^{-1} NaCl, pH 7.4), followed by intensive stirring for three days, and kept at 4 °C for no longer than a week.

The CT DNA concentration was determined by the UV absorbance at 260 nm after 1:20 dilution in water, using an extinction coefficient of the DNA per molar nucleotide concentration equals to 6600 L mol^{-1} cm^{-1} [27,28]. The stock solution of CT DNA gave a ratio of UV absorbance at 260 and 280 nm higher than 1.8, indicating that the CT DNA was sufficiently free of protein contamination.

2.5. Absorption spectroscopy

Optical absorption spectra were obtained with an UV–Vis spectrophotometer (VarianCary, Santa Clara, CA). Samples were placed in a quartz cuvette (0.4 × 1.0 cm), with the absorption optical pathway of 0.4 cm. The temperature was controlled with a Carry Peltier thermostat, and measurements were performed at 25 °C or 50 °C.

2.6. Fluorescence spectroscopy

Steady state fluorescence measurements were performed using a Fluorimeter (VarianCary, Santa Clara, CA) with slits for excitation and emission of 5 nm and a bandpass of 2 nm. Temperatures were controlled by a Carry Peltier thermostat. Fluorescence experiments were performed with 1 ml solutions of AMP (20 μmol L^{-1}) in buffer (HEPES 10 mmol L^{-1}, 3 mmol L^{-1} NaCl, pH 7.4), upon titration with DPPC vesicles or CT DNA. The membrane stock dispersion consisted of 10 mmol L^{-1} of extruded DPPC (100 nm), and CT DNA stock solution was approximately 3 mmol L^{-1}.

The experiments with membranes were conducted at DPPC gel (25 °C) and fluid phases (50 °C), with the excitation beam light at 280 nm. The experiments with CT DNA were performed at 25 °C. To avoid any absorption by DNA nucleobases, DNA experiments were performed with the excitation beam light at 295 nm. The fluorescence spectra were corrected by the appropriate dilution due to the addition of lipids or DNA. Moreover, the inner filter correction [29] was applied to all the fluorescent emission spectra by using Equation (1):

$$F_{corr}(\lambda) = F_{obs}(\lambda) 10^{(A_{exc}\, l + A_{ems}\, l')} \quad (1)$$

where F_{corr} (λ) and F_{obs} (λ) are the corrected and observed fluorescence intensity at a given λ, A_{exc} and A_{ems} are the absorbance per unit of pathway at the excitation and emission wavelengths, respectively. l and l′ are the optical pathways for excitation (0.2 cm), and for emission (0.5 cm), respectively. When necessary (Fig. SM2), the fluorescence spectra were transformed from wavelength to energy, and the intensity multiplied by λ^2. This procedure is necessary given that the emission spectrum is recorded with a constant wavelength bandpass, not energy [30,31].

2.7. Entrapment of carboxyfluorescein (CF) in LUVs and leakage assay

CF solutions were prepared in buffer pH 8.5. After CF solubilization, the sample pH was readjusted to 7.4 with HCl. Lipid films were hydrated with buffer (10 mmol L^{-1} HEPES, 3 mmol L^{-1} NaCl, 1 mmol L^{-1} EDTA, pH 7.4 solution) containing 50 mmol L^{-1} carboxyfluorescein (CF). The lipid dispersion (~6 mmol L^{-1}) was extruded, as previously described. In order to remove non-encapsulated CF, the lipid dispersion was eluted through a Sephadex-G25 medium column with 10 mmol L^{-1} HEPES, pH 7.4 with 1 $mmol^{-1}$ EDTA, 3 mmol L^{-1} NaCl, and 150 mmol L^{-1} glucose, the latter was added to the buffer to adjust the osmolarity inside and outside of the liposomes. Vesicles with encapsulated CF were collected in the void volume of the column. Lipid concentration was determined by inorganic phosphate assay [26].

Lipid dispersion (100 μmol L^{-1}) was placed in quartz cuvettes (1.0 × 1.0 cm, 2.0 mL) and the fluorescent emission measured with a Fluorescence Spectrometer (Varian Cary Eclipse, Santa Clara, CA), and the temperatures were controlled with a Carry Peltier thermostat. The CF release measurements were performed under constant stirring. CF encapsulating in LUVs was used as a model to evaluate the abilities of Hylin peptides to induce pore formation in zwitterionic bilayers. At 50 mmol L^{-1} the encapsulated CF is self-quenched, hence virtually non-fluorescent. Due to AMP or detergent action, CF might be released from the liposomes into the bulk, diluting CF and increasing the CF fluorescence intensity. CF emission was continuously recorded in time (one measurement per second), at 25 °C, λ_{exc} = 490 nm and λ_{em} = 512 nm. In all experiments, Hylin peptides (0.05 μmol L^{-1}) were added to lipid dispersion (100 μmol L^{-1}) at the 100th second and, at the end of the experiment, 2000th second, Triton X-100 (12 μL of 10% w/v) was added to promote complete CF leakage.

The percentage of CF leakage, (%) Leakage, was determined according to Equation (2):

$$(\%)\ Leakage(t) = 100 \times \frac{(I(t) - I_0)}{(I_{total} - I_0)} \quad (2)$$

where I(t) is the fluorescence intensity at time t, I_0 is the initial fluorescence, before peptide addition, and I_{total} is the maximum fluorescence obtained after the addition of Triton X-100. The kinetics were performed using zwitterionic liposomes of DPPC in the gel phase at 25 °C. As the experimental procedure with fluid DPPC (50 °C) was found to be quite unreliable [23], to mimic the fluid phase of the dipalmitoyl membranes, similarly prepared vesicles of POPC were used at 25 °C.

2.8. Differential scanning calorimetry (DSC)

DSC profiles were obtained with a microcalorimeter (Microcal VP-DSC, Northampton, MA). Samples were heated from 15 °C to 60 °C at a scan rate of 20 °C per hour. The sample cell (500 μL) was filled with a 3 mmol L^{-1} lipid dispersion with or without the addition of the desired AMP concentration. In this work, we will refer to the concentration of AMP as the percentage of the [AMP] with respect to the molar concentration of lipid (% [AMP] = 100 [AMP]/[L]), where [L] is the lipid concentration. We corrected the DSC traces taking into consideration the dilution due to the addition of the peptides. Baseline subtractions and peak integrals were performed using the MicroCal Origin software with the additional module for DSC data analysis provided by MicroCal.

2.9. Competitive studies with ethidium bromide (EB)

Ethidium bromide (EB) was solubilized in buffer (10 mmol L^{-1} HEPES, 3 mmol L^{-1} NaCl, pH 7.4) to make a stock solution (5 mmol L^{-1}), and submitted to ultrasound for 3 min to assure complete solubilization. The samples consist of 1 ml solution of EB (15 μmol L^{-1}) previously incubated with CT DNA (10 μmol L^{-1}) for 15 min to assure thermal equilibrium and the formation of the complex (EB-DNA; 15:10 in moles). Then, samples were placed in a quartz cuvette (0.4 × 1.0 cm, 1 ml). Finally, samples were titrated with Hylin peptides, and EB-CTDNA fluorescence spectra were measured. Experiments were performed with an excitation beam light at 545 nm.

2.10. DNA-binding assay

For Hylin analogues, W^6Hya1, D^0W^6Hya1, and K^0W^6Hya1, binding

reaction with plasmid DNA (plasmid pOP3BP, 4691 bp) was monitored using agarose gel electrophoresis. The nucleic acid binding efficiency was estimated by determining the degree of delayed mobility of the plasmid DNA (pDNA) bands, which is reflected in an up-shift of the DNA to higher molecular weight, indicating changes in the ratio of charge mass of the DNA-peptide complex [32]. Reactions containing 600 ng of pDNA and increasing amounts of each AMP in water for a final volume of 20 μL, were incubated at a constant temperature of 21 °C for 45 min. After incubation, 8 μl of each reaction was mixed with 2 μL of GelPilot DNA Loading Dye, 5x (QIAGEN), and submitted to electrophoresis on agarose gels (1% w/v), which contained 1X UniSafe Dye (UNISCIENCE), in TAE buffer (40 mmol L^{-1} Tris, 20 mmol L^{-1} acetic acid, and 1 mmol L^{-1} EDTA Sodium salt dihydrate), at 60 V during 90 min. Agarose gel electrophoresis was performed in a horizontal gel apparatus Mini-Sub Cell GT (BIORAD). The migration of pDNA was visualized after staining with the fluorescent intercalated UniSafe Dye under a UV illuminator. All experiments were repeated at least three times for reproducibility.

3. Results

3.1. Peptides interaction with zwitterionic membranes

Most of the bilayers composed of one saturated lipid species display two different thermal phases: a gel and a fluid phase. In the gel phase the lipid molecules are more organized and packed when compared to those in the fluid phase. In both phases, the lipid molecules are constrained to the two-dimensional plane of the membrane, but in the fluid phase the lipids are looser and can diffuse faster within the plane [33]. For DPPC membranes, the gel-fluid transition temperature is about 40 °C. The interaction of exogenous molecules with lipid membranes might be profoundly affected by the lipid phase [33]. Hence, we investigated the interaction of the three AMPs, W^6Hya1, D^0W^6Hya1, and K^0W^6Hya1, with DPPC membranes in their gel (25 °C) and fluid (50 °C) phases. That would somehow mimic both more packed and less packed lipid domains in biological membranes, respectively.

3.1.1. Fluorescence spectroscopy

The Hylin peptides studied herein are fluorescent due to the replacement of the leucine residue at the 6th position of the peptide chain by the aromatic tryptophan (Trp). To comparatively analyze Hylin peptides binding properties to eukaryotic-like membranes, we monitored the changes of AMP emission due to presence of DPPC LUVs.

Trp fluorescence spectrum is very sensitive to its environment [30], making it an excellent fluorescent probe to investigate changes in its vicinity. Accordingly, when a Trp residue moves to a more hydrophobic environment, its emission spectrum shifts to smaller wavelengths (higher energies), in comparison with the spectrum of the fluorophore in aqueous environment, as the dipolar relaxation decreases considerably [30]. Usually, parallel to that, there is an increase in the fluorescence intensity due to the reduction in non-radiative deactivation processes related with interactions with solvent molecules and/or a decrease of molecular mobility.

Let us compare the interaction of each peptide with zwitterionic LUVs of DPPC in the gel (25 °C) and fluid phases (50 °C). Fig. 2 exhibits the evolution of the AMPs intrinsic fluorescence emission spectra with increasing amounts of DPPC. It is evident that the three peptides bind to the zwitterionic membranes, at both gel (25 °C) and fluid (50 °C) bilayer phases, as the Trp fluorescence spectrum changes in the presence of DPPC. However, it is also evident that DPPC fluid membranes (Fig. 2, right column) induce stronger modifications on the Trp fluorescence spectra as compared to gel membranes (Fig. 2, left column).

To analyze and compare the changes caused by DPPC membranes on the Trp spectrum of the three peptides, the shifts on the position of the maximum emission of the three peptides in the presence of increasing amounts of lipids are shown in Fig. 3. For gel DPPC bilayers (Fig. 3a), the

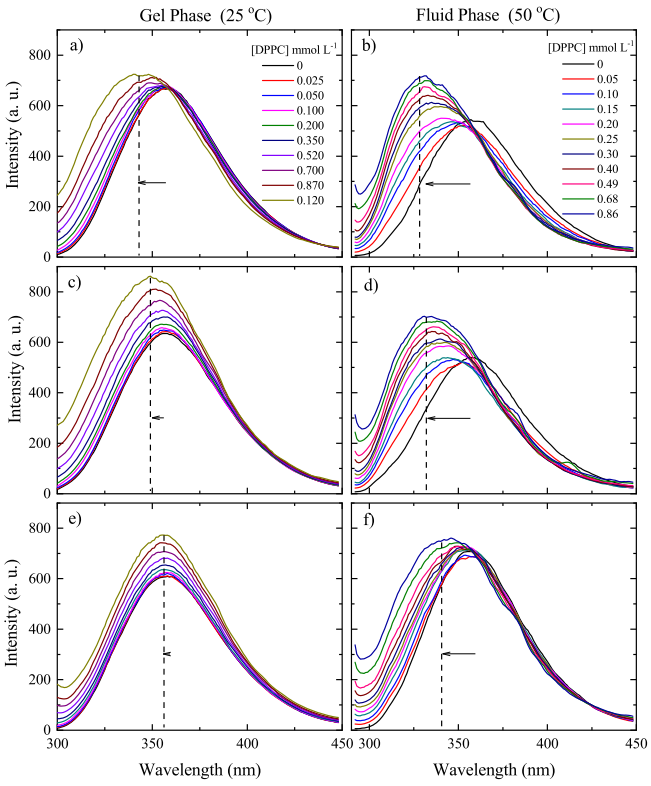


Fig. 2. Typical fluorescence spectra of W^6Hya1 (a, b), D^0W^6Hya1 (c, d), and K^0W^6Hya1 (e, f), obtained from titrating the peptide solution with gel (25 °C) (left column), and fluid (50 °C) (right column) vesicles of DPPC. $\lambda_{exc} = 280$ nm. AMPs concentration = 20 μmol L^{-1}. Dashed lines indicate the center (maximum) of the emission bands at the maximum lipid concentration used. The fluorescence intensities at these positions were used to calculate the apparent dissociation constant for each sample (Fig. 4 and Eq. (2)). Arrows indicate spectra shifts at the maximum lipid concentration.

Trp blue shifts are significantly smaller than those observed in the presence of fluid membranes (Fig. 3b). That indicates that either the peptides (or, at least, the Trp residue in the peptides) are deeper inside fluid DPPC bilayers as compared with gel membranes, and/or that the peptides exhibit a small partition into gel DPPC bilayers and stay mostly in the water medium.

For fluid DPPC membranes, the maximum lipid concentration used here is close to a lipid saturation concentration, where most of the peptides are bound to the membrane, as indicated by the flattening of the curves in Figs. 3b and 4b, for the highest DPPC concentrations used (around 1 mmol L^{-1}). Hence, it is possible to conclude that in fluid DPPC bilayers, Trp in the more charged peptide, K^0W^6Hya1 (net charge +4; ▼in Fig. 3b), is in a shallower position in the bilayer (smaller blue shift) as compared with Trp in the two other peptides, W^6Hya1 and D^0W^6Hya1 (net charges +3 (▱) and +2 (⊗), respectively, in Fig. 3b). Moreover, Trp in W^6Hya1 (▱) seems to be in a somewhat deeper position in the membrane (larger blue shift) than the residue in D^0W^6Hya1 (⊗).

For the three peptides, in both gel and fluid membranes, the blue shifts (Fig. 3) in the presence of the highest lipid concentrations used here are listed in Table 1. Higher lipid concentrations could not be used due to the significant light scattering yielded by them, making the inner filter corrections used here unreliable (see Materials and methods). Hence, even though the effect on the Trp fluorescence spectrum due to the peptides binding to gel DPPC is far from saturated, it follows the same trend observed for the peptides in fluid DPPC, with K^0W^6Hya1 (▼) displaying the least effect (Fig. 3a).

Apart from the blue shift caused by DPPC vesicles on the Trp spectrum, there is a clear increase in the spectrum intensity, mainly observed

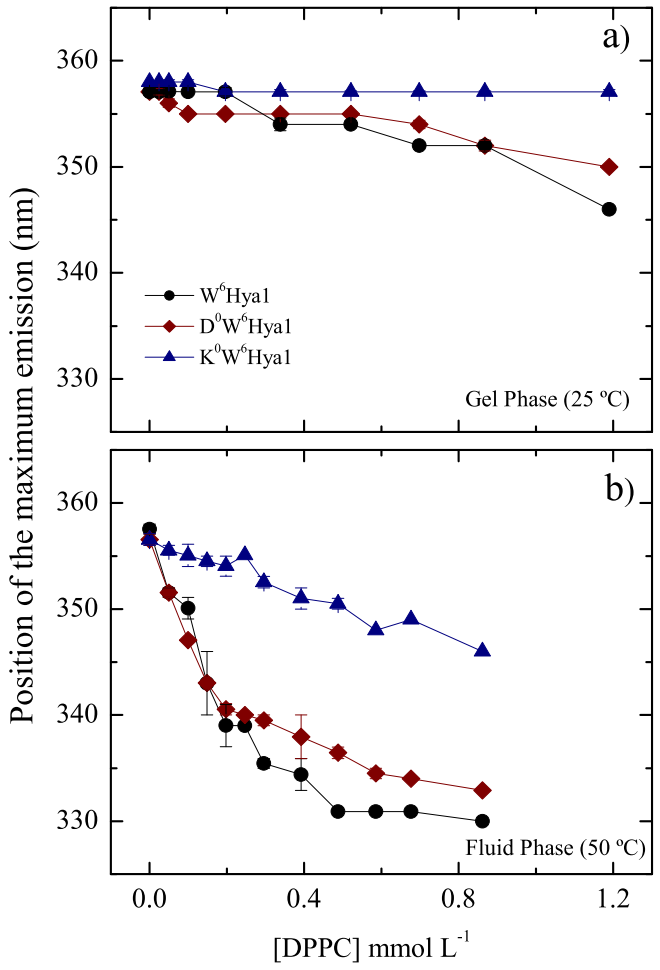


Fig. 3. Position of the maximum emission of W^6Hya1 (black circle), D^0W^6Hya1 (wine diamond), and K^0W^6Hya1 (navy triangle) as a function of DPPC concentration: (a) in the lipid gel phase (25 °C) and (b) in the fluid phase (50 °C). The AMPs concentration was 20 μmol L^{-1} λ_{exc} = 280 nm. Error bar indicates standard deviation of at least three experiments with different samples. If not shown, it was found to be smaller than the symbol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

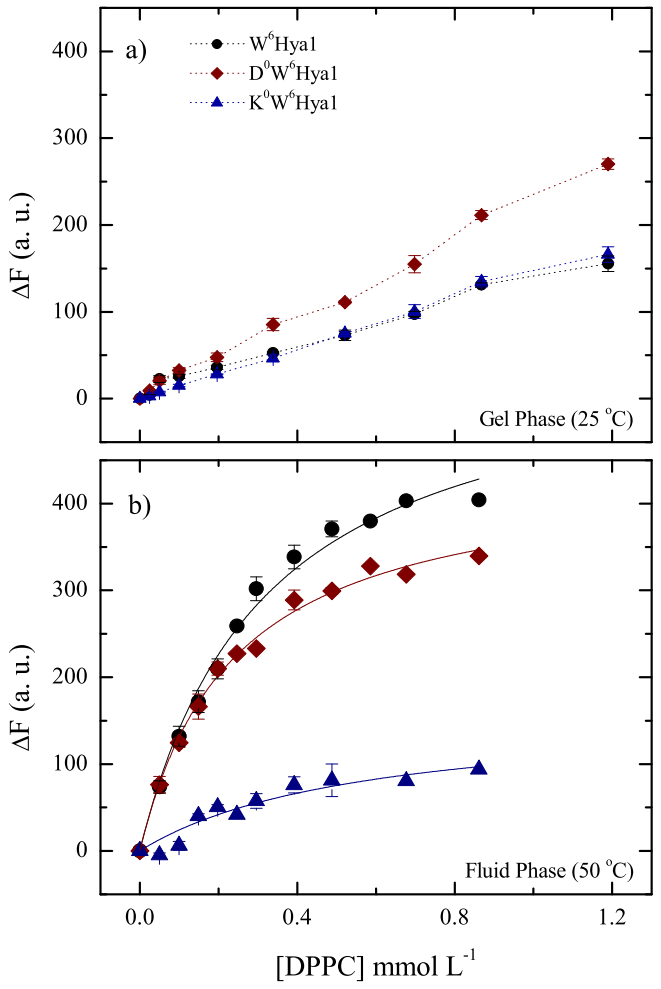


Fig. 4. Increase of the Trp fluorescence intensity (ΔF = F–F_0, F_0 being the fluorescence intensity in the absence of lipids), at the wavelength positions indicated in Fig. 2, of W^6Hya1 (black circle), D^0W^6Hya1 (wine diamond) and K^0W^6Hya1 (navy triangle), in the presence of DPPC vesicles in the (a) gel (25 °C) and (b) fluid (50 °C) lipid phases. λ_{exc} = 280 nm. The dotted lines are guides for eyes and full lines correspond to the fitting of the data with Eq. (3). AMPs concentration = 20 μmol L^{-1} λ_{exc} = 280 nm. Error bar indicates standard deviation of at least three experiments with different samples. If not shown, it was found to be smaller than the symbol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the presence of fluid DPPC membranes (Fig. 2, right column). This is also an indication that the Trp residue, in the peptides, is inserted into the membrane, in a microenvironment of lower polarity and/or more vibrational restriction. Hence, the variation of the fluorescence intensity (ΔF = F–F_0, F_0 being the fluorescence intensity in the absence of lipids) was plotted as a function of the lipid concentration, for the three peptides, in the membranes in both gel and fluid phases (Fig. 4). The fluorescence intensities were measured at the wavelength where the emission intensity is at its maximum in the presence of the highest lipid concentration used (see dashed lines in Fig. 2).

Similar to what was observed for the blue shift (Fig. 3a), the Trp fluorescence intensity variations in the presence of gel DPPC are not very intense (Fig. 4a), they do not saturate with the amount of lipids used here (around 1 mmol L^{-1}). That strongly suggests that at the DPPC gel phase most of the AMPs remain in the aqueous phase. Hence, it was not possible to address an apparent dissociation constant (K_d) for the peptides in the presence of gel DPPC vesicles.

However, the plots ΔF x [DPPC] obtained in the presence of fluid DPPC (Fig. 4b) could be well fitted with the conventional binding isotherm (Eq. (3), full lines in Fig. 4b):

Table 1
AMPs Trp blue-shifts due to the binding to DPPC vesicles at the highest lipid concentration used here, at gel (25 °C) and fluid (50 °C) lipid phases (data in Fig. 3). AMPs net charge are shown in parentheses.

AMP	$\Delta\lambda_{max}$ (nm) gel (25 °C)	$\Delta\lambda_{max}$ (nm) fluid (50 °C)
W^6Hya1 (+3)	(11.0 ± 0.3)	(27.5 ± 0.8)
D^0W^6Hya1 (+2)	(7.1 ± 0.2)	(24.6 ± 0.5)
K^0W^6Hya1 (+4)	(3.1 ± 0.8)	(11.5 ± 0.6)

$$\Delta F = F - F_0 = \frac{(F_\infty - F_0)[L]}{(K_d + [L])} \quad (3)$$

where F_∞ is the fluorescence intensity of the AMPs at lipid saturating concentration. The apparent dissociation constant (K_d) is the concentration of lipids which induces 50% of changes in the fluorescence intensity, under the experimental conditions employed. Here, we use K_d values to compare the affinities of the three AMPs to fluid DPPC

membranes, not for the actual measurement of a true partition constant. Hence, it is important to keep in mind that the apparent K_d calculated here does not take into consideration the increase of the membrane charge due to cationic binding peptides [34].

Table 2 displays K_d, ΔF_∞, and χ^2 values obtained through the best fitting processes. Note that the dissociation constant obtained for K^0W^6Hya1 is about two times higher than those obtained for the other two AMPs. However, the nonlinear least squares fit for K^0W^6Hya1 yielded a χ^2 value equal to 0.90, indicating that this K_d value has a high associated error. This occurs due to the low affinity of this peptide to DPPC membranes, rendering the determination of F_∞ rather inaccurate. It is interesting to observe that K_d values follow the crescent order D^0W^6Hya1 (+2) < W^6Hya1 (+3) < K^0W^6Hya1 (+4). It is important to keep in mind that a higher K_d value reflects a lower affinity to DPPC membrane. Therefore, the net charge dependence is clearly indicating that the increase of a positive charge at the N-terminus of the peptide hampers the AMP association to zwitterionic membranes.

To use the variations on the Trp fluorescence intensity to discuss the penetration of the peptides into the lipid bilayer, we should look at the values of ΔF_∞ (Eq. (3) and Table 2), as, theoretically, this is the value obtained when all peptides are bound to the vesicles. In accord with the blue shift (Fig. 3 and Table 1), they indicate that Trp in W^6Hya1, is somehow deeper in the membrane ($\Delta F_\infty = 59$) than the residue in the other two peptides. Moreover, the significant lower ΔF_∞ value obtained for the more cationic peptide (K^0W^6Hya1) ($\Delta F_\infty = 17$) strongly indicates that the double positive charge at its N-terminal keeps this peptide less deep in the membrane as compared with D^0W^6Hya1 ($\Delta F_\infty = 44$).

Though there is no theoretical argument for associating a binding isotherm, like Eq. (3), with the shift of the maximum of the emission band, it was interesting to find that the plots of ΔF x [DPPC] (Fig. 4, and Table 2) and ΔE x [DPPC] behave similarly, as shown in Fig. SM2 and Table SM1, displaying the same trend. ΔE is $E - E_0$, where E is the energy of the emission band corresponding to the maximum position of the band, and E_0 is the value in the absence of lipids.

3.1.2. *Differential scanning calorimetry (DSC)*

Saturated lipid bilayers often display a very narrow peak of heat capacity which is characteristic of a cooperative process, being very dependent on lipid-lipid interaction. As the presence of an exogenous molecule may interfere with the phase transition process, DSC thermograms can provide important information about the interaction of exogenous molecules with lipid systems [35].

In particular, multilamellar DPPC membranes (non-extruded) display a very characteristic pre-transition peak centered at about 33 °C and a thin main transition peak around 40 °C. Whereas, 100 nm extruded lipid dispersions of DPPC (Large Unilamellar Vesicles, LUVs) exhibit a subtle pre-transition peak and a less cooperative main transition peak about the same temperatures observed for multilamellar DPPC dispersions [23,33].

Fig. 5 displays DSC profiles of DPPC LUVs in the absence and presence of increasing amounts of W^6Hya1 (Fig. 5a), D^0W^6Hya1 (Fig. 5b), and K^0W^6Hya1 (Fig. 5c). Increasing amounts of the three peptides cause similar effects on the DPPC thermogram, and similar to that reported before for K^0W^6Hya1 [23]. The broadening of the DPPC bilayer transition observed in the presence of the three peptides suggest a homogeneous perturbation of the lipids due to peptide binding, decreasing the cooperativity of the gel-fluid transition. That is, there is no indication of the coexistence of peptide-bound and peptide-free regions (bulk lipid) in the bilayer, as observed with K^0W^6Hya1 in the presence of anionic DPPG membranes [23]. The enthalpy of the main transition remains roughly the same for all samples studied, $\Delta H \approx 8$ kcal mol^{-1}, in accord with previous studies with pure DPPC [36]. Though this was not investigated

Table 2
Apparent dissociation constants of the three AMPs for fluid (50 °C) DPPC membranes, obtained from steady-state fluorescence spectroscopy (data in Fig. 4). AMPs net charge are shown in parentheses.

AMP	K_d [× 10^{-4} mol L^{-1}]	ΔF_∞[× 10 a. u.]	χ^2
W^6Hya1 (+3)	(3.2 ± 0.3)	(59 ± 2)	0.99
D^0W^6Hya1 (+2)	(2.4 ± 0.2)	(44 ± 1)	0.99
K^0W^6Hya1 (+4)	(6 ± 3)	(17 ± 4)	0.90

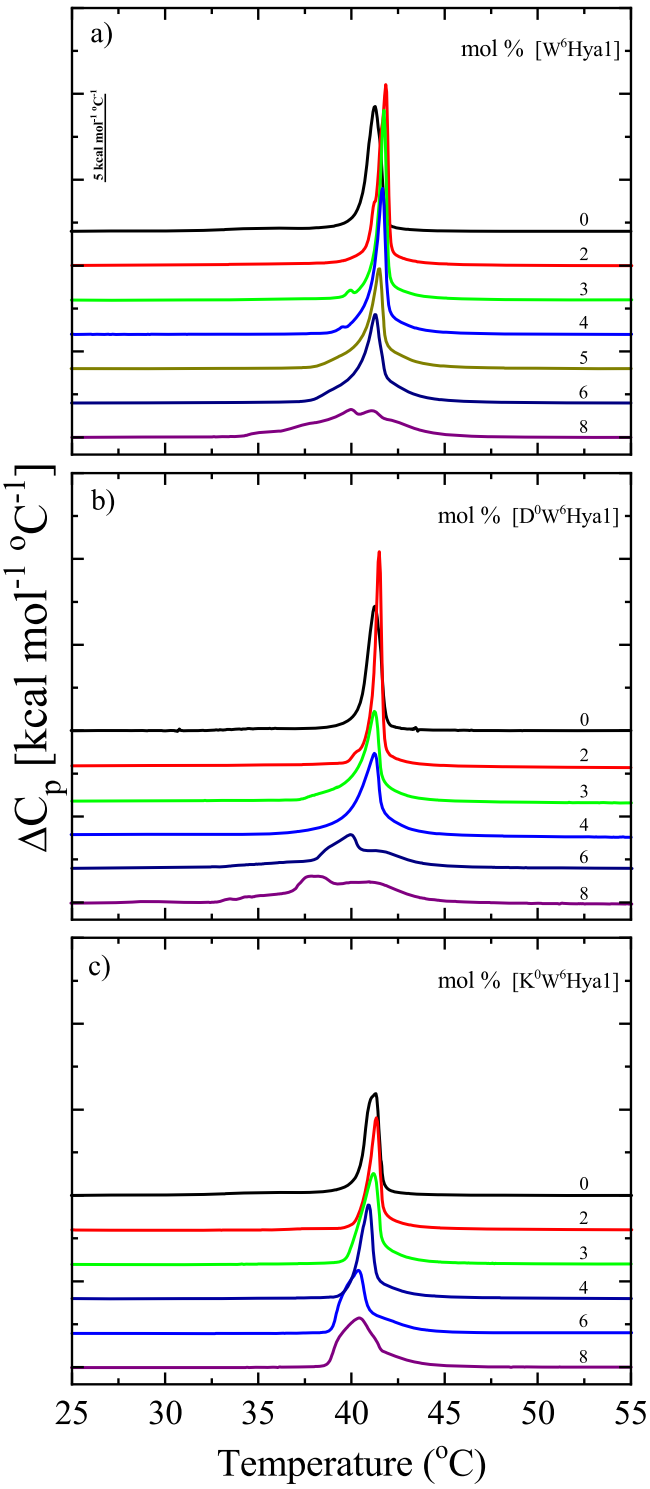


Fig. 5. DSC thermograms of 100 nm extruded lipid dispersions composed of 3 mmol L^{-1} DPPC with increasing peptide–lipid molar ratio, from 2% to 8% of (a) W^6Hya1 or (b) D^0W^6Hya1 and (c) K^0W^6Hya1. Scans were obtained using a scan rate of +20 °C/h, and they are shifted for clarity. Duplicated samples showed similar results.

in the present work, it is interesting to note that the lowest peptide concentration used here (2 mol%) causes a slightly narrowing of the thermal transition (Fig. 5).

All thermograms were found to be not only reproducible, in a second scan, but also reversible (Fig. SM3), supporting the assumption that the three peptides cause a homogeneous effect on the bilayer. Furthermore, DSC thermograms seem to be in agreement with the fluorescence findings which shows that the most cationic peptide displays a lower affinity and a shallower interaction with DPPC membranes (Tables 1 and 2). Namely, at 8 mol%, K^0W^6Hya1 is less effective in disturbing DPPC thermograms, as it is still possible to observe a well resolved thermal peak, whereas for the two other peptides, at this same concentration, a very broad band is observed.

3.1.3. Carboxyfluorescein (CF) leakage assay

It was previously shown that the cationic peptide K^0W^6Hya1 could not only disturb anionic membranes, but also causes a significant leakage of carboxyfluorescein (CF) through zwitterionic PC membranes [23], at rather low peptide/lipid relative concentration. Hence, we compare the effect of the three peptides studied here, with different net charges, on the leakage of CF through PC membranes. Similar to previous results [23], Fig. 6 shows that a very small peptide/lipid relative concentration (0.05 mol%) is enough to cause a significant leakage in fluid POPC membranes, and has almost no effect on gel DPPC bilayers, both at 25 °C. Very interesting to note that the more cationic peptide, K^0W^6Hya1, is the one that causes the major leaking at the end of the process, after 2000 s, in fluid PC vesicles. The spontaneous CF release, in the absence of peptides, was found negligible (less than 2% in 2000 s), as the control consists of liposomes containing CF in the absence of peptides. As found for many CF-leakage experiments [37–39] the kinects were well fitted by two exponentials, indicating two kinect processes, with different decay times (Fig. SM4). Indeed, this finding is in agreement with existing data for the interaction of K^0W^6Hya1 with PC vesicles [23]. Important to have in mind that at the concentrations used here the peptides do not significantly change the size of the vesicles, as attested by dynamic lighting scattering experiments (not shown).

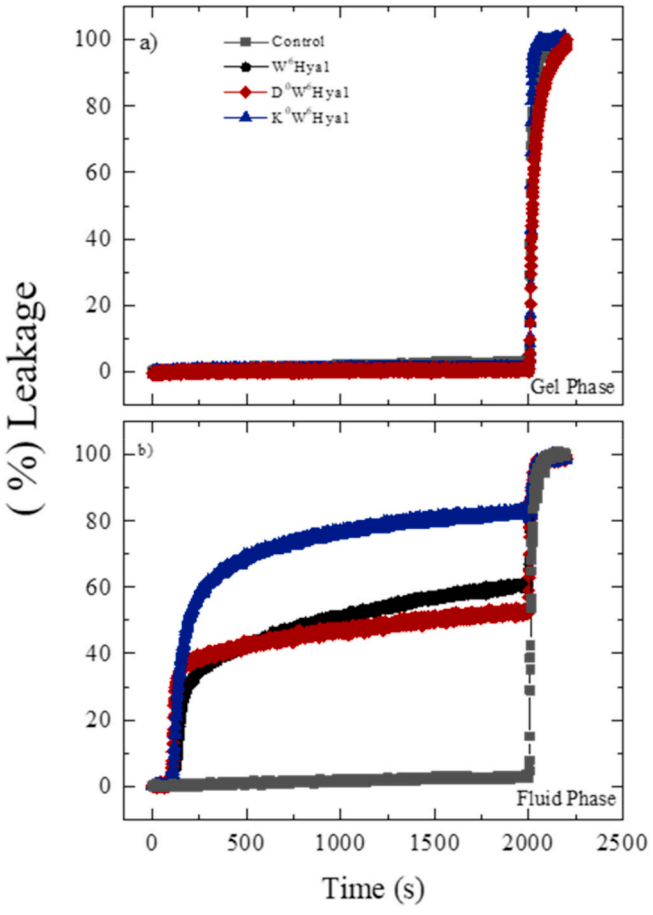


Fig. 6. Typical kinetics of CF leakage through LUVs composed of gel membranes of DPPC (a), and fluid membranes of POPC (b), at 25 °C, in the presence of 0.05 μmol L^{-1} of W^6Hya1 (black circle), D^0W^6Hya1 (wine diamond) and K^0W^6Hya1 (navy triangle). Lipid concentration used was 100 μmol L^{-1}. The control (gray square) consists of pure PC LUVs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Hylin peptides interaction with DNA

Since Hylin peptides form pores in zwitterionic lipid bilayers, we can presume that they will invade cell cytosol. Therefore, we found important to inquire about Hylin peptides interaction with structures present in the cellular cytosol, such as nucleic acids. Indeed, several peptides show nucleic acid binding abilities [40–42]. Due to electrostatic interaction, it was expected that cationic AMPs would display DNA-binding activity.

3.2.1. Trp fluorescence spectroscopy with CT DNA

We comparatively evaluated the interaction of Hylin peptides with CT DNA, through intrinsic AMPs fluorescence spectroscopy. Fig. 7 displays the normalized spectra of Hylin peptides, W^6Hya1 (Fig. 7a), D^0W^6Hya1 (Fig. 7b), K^0W^6Hya1 (Fig. 7c), with increasing amounts of CT DNA. Note that in the presence of CT DNA, the peptides exhibit a significant blue shift: Fig. 7d is the plot of the position of the Trp maximum emission in each peptide versus CT DNA concentration (in pair base molar concentration). Upon CT DNA titration, the peptides absorption spectra display significant light scattering (Fig. SM5), mainly for CT DNA concentrations above 40 μmol L^{-1}, rendering quite inaccurate the values of the measured fluorescence intensities, even after the corrections used here (see Material and methods). Hence it was not possible to calculate an apparent binding constant for the peptides with CT DNA, as shown in Eq. (3).

The maximum blue shifts ($\Delta\lambda_{max}$) observed at the highest CT DNA concentration used here (80 μmol L^{-1}) for the Hylin peptides are listed in Table 3. We observe significant blue shifts due to the presence of CT DNA: W^6Hya1 (+3) and K^0W^6Hya1 (+4) display a maximum blue shift of about 17 nm. In contrast, the interaction of the peptide D^0W^6Hya1 (+2) with CT DNA shifts the emission peak 10 nm only. The observed blue shifts indicate that the polarity in the vicinity of the Trp residues has changed, this could be correlated with DNA binding and/or conformational changes in the Hylin peptides.

3.2.2. Competitive CT DNA-binding between ethidium bromide and hylin peptides

The interaction of exogenous molecules and DNA can be driven by covalent and non-covalent interactions, the later are notably divided in two binding modes: intercalative and groove. In the intercalative binding mode, the molecules allocate themselves into DNA double helix adjacent to the nucleobases, whereas in the groove binding mode the molecules interact with the negative sugar-phosphate DNA backbone, mediated by electrostatic and hydrogen bond interactions. Usually small molecules can bind to DNA by more than only one mode [43,44].

The observed Trp blue shifts for Hylins in the presence of DNA show that these peptides interact with DNA. To verify if Hylin peptides bind to DNA via the intercalative mode, we evaluated the ability of the peptides to displace ethidium bromide (EB), previously bound to CT DNA. EB is a dye that intercalates into DNA double helix. It does not exhibit a significant fluorescence when in aqueous solution. In contrast, when intercalated into DNA double helix, EB emission increases considerably [29]. If Hylin peptides intercalate into DNA, they would compete with EB for the same hydrophobic site in the DNA double helix, eventually

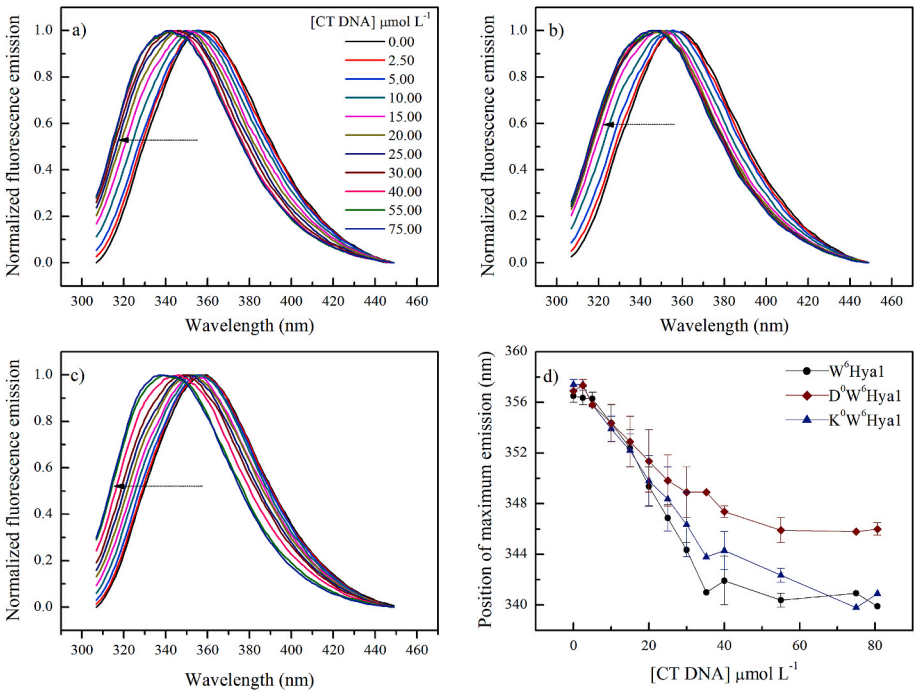


Fig. 7. Typical normalized fluorescence spectra of (a) W^6Hya1, (b) D^0W^6Hya1, (c) K^0W^6Hya1 with increasing CT DNA concentrations. The arrows are guides for eyes, indicating the Trp blue shift. (d) Position of the maximum emission of W^6Hya1 (black circle), D^0W^6Hya1 (wine diamond), and K^0W^6Hya1 (navy triangle) as a function of CT DNA concentration. λ_{exc} = 295 nm. Error bar in (d) indicates standard deviation of at least three experiments with different samples. If not shown, it was found to be smaller than the symbol. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
AMPs Trp blue-shifts due to the presence of CT DNA at the highest lipid concentration used here. (data in Fig. 7). The AMPs net charges are shown in parentheses.

AMP	$\Delta\lambda_{max}$ (nm)
W^6Hya1 (+3)	(16.8 ± 0.2)
D^0W^6Hya1 (+2)	(10.9 ± 0.5)
K^0W^6Hya1 (+4)	(16.5 ± 0.4)

displacing EB to the aqueous environment, hence leading to a decrease of EB intensity emission.

Fig. 8 shows the EB-DNA fluorescence spectra with increasing amounts of Hylin peptides: W^6Hya1 (Fig. 8a), D^0W^6Hya1 (Fig. 8b) and K^0W^6Hya1 (Fig. 8c). It evinces that the Hylin peptides decrease the emission intensity of EB-DNA complex (Fig. 8), indicating that Hylin peptides are binding the DNA by the intercalative mode, thus dislocating EB molecules. Hence, the Stern-Volmer plot (Fig. 8d) and Eq. (4), seems to be appropriated to quantify and compare Hylin peptides abilities to

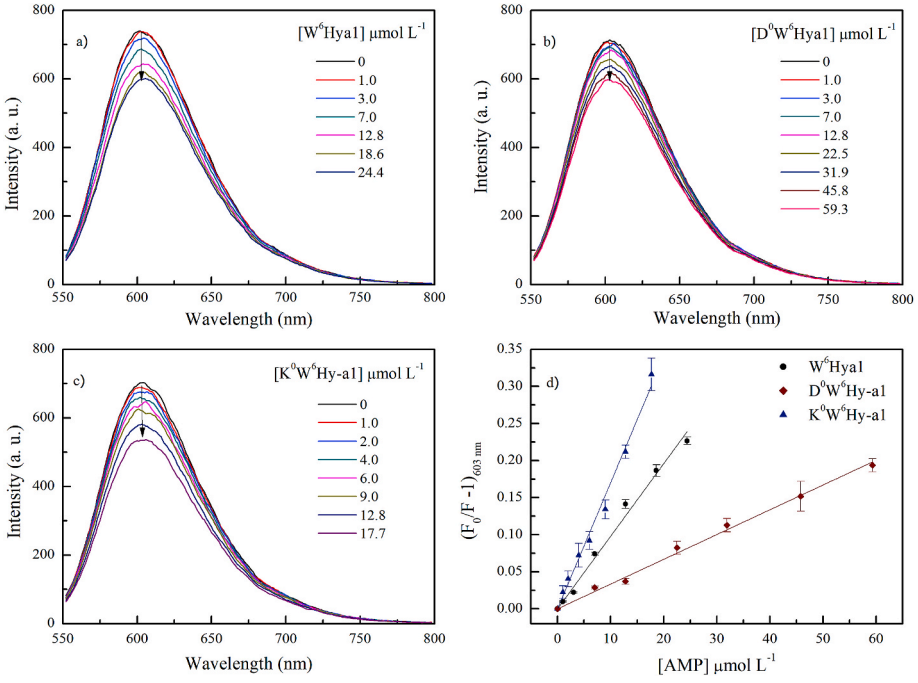


Fig. 8. – Typical Fluorescence spectra of the complex EB-DNA with increasing amounts of: (a) W^6Hya1, (b) D^0W^6Hya1, (c) K^0W^6Hya1. (d) Stern-Volmer plots (Eq. (4)) versus AMPs concentration. λ_{exc} = 545 nm, λ_{ems} = 603 nm. The arrows are guide for the eyes only, indicating increase amounts of AMPs. Error bar in (d) indicates standard deviation of at least three experiments with different samples. If not shown, it was found to be smaller than the symbol.

intercalate into DNA double helix. It is important to note that the Stern-Volmer equation will be used here not to monitor fluorescence quenching, as it is usually applied, but to monitor the decrease in the EB fluorescence due to the decrease in the concentration of EB-DNA complex.

$$F_0/F - 1 = K_{SV}[AMP] \quad (4)$$

Where F_0 and F, are the fluorescence intensities at the maximum emission of the complex EB-DNA (at 603 nm) in the absence of Hylin peptides and with increasing amounts of them, respectively, and K_{SV} is the Stern-Volmer constant.

As mentioned above, we used the Stern-Volmer constant, Table 4, only to compare the ability of the Hylin peptides to dislocate EB from the DNA to the aqueous environment, hence associated with the peptide ability to intercalate into DNA: a higher K_{sv} value indicates a higher intercalative binding mode.

3.2.3. Agarose gel electrophoresis: binding experiments with plasmid DNA (pDNA)

We further investigated Hylin – DNA binding abilities through electrophoretic experiments, which can provide information about interactions of DNA with exogenous molecules such as AMPs. The technique consists of observing DNA pattern migration in agarose gel while it is submitted to a constant electric field. The nucleic acid binding efficiency can be estimated by determining the degree of delayed mobility of a DNA band reflecting in an upshift of the DNA to higher molecular weight.

Fig. 9 displays the electrophoresis mobility of pDNA in agarose gel in the absence and presence of increasing amounts of each of the three AMPs. The results show that AMPs interact with pDNA retarding its migration in a concentration-dependent manner. The greater the relative concentration of peptide the greater the DNA upshift. Corroborating our previous results from assays of Trp fluorescence spectroscopy with CT DNA (Table 3) and binding competition between ethidium bromide and Hylin peptides (Table 4), W^6Hya1 (+3) and K^0W^6Hya1 (+4) show similar retarding in the DNA migration at lower concentrations (see AMP/pDNA weight ratio 2.1 in Fig. 9), and the less cationic peptide D^0W^6Hya1 only induces a marginal effect on DNA migration at 2.1 (peptide:DNA).

At AMP/pDNA weight ratio of 0.6, we did not observe any significant changes in the DNA migration pattern of the three samples when compared to the control, and at the highest peptide concentration (weight ratio of 16.7) there was no DNA migration: pDNAs at the highest AMP/pDNA concentration remains within the well.

4. Discussion

Given the presence of many hydrophobic amino acid residues (Fig. 1), it was expected that Hylin peptides would interact with zwitterionic amphiphilic aggregates, as attested by the present experiments, and previous data [19,22]. As reported before for K^0W^6Hya1 [23], the three peptides studied here seem to laterally diffuse in DPPC membranes, inducing a progressive broadening of DPPC thermograms as the peptide concentration increases. Considering that there is no indication of the coexistence of peptide-bound and peptide-free regions in the bilayer, as observed with K^0W^6Hya1 in the presence of anionic DPPG membranes [23], it is highly likely that the peptides are laterally diffusing in the membrane, causing a time-average or space-average effect in the bilayer DSC profile (Fig. 5). The peptide lateral diffusion seems to happen in the gel and fluid phases of the membrane, as DSC profiles are fairly reversible (Fig. SM3).

For fluid membranes, we determined that the apparent affinity for DPPC membrane follows the decreasing order: D^0W^6Hya1 (+2) > W^6Hya1 (+3) > K^0W^6Hya1 (+4) (Table 2). This finding follows the same decrescent order of hydrophobicity determined by Ref. [19] for these peptides. It indicates that the net charge of the peptide modulates its hydrophobicity, hence, its affinity for zwitterionic membranes: the membrane apparent affinity decreases as the peptide net charge increases.

However, the depth of the peptide penetration into fluid PC bilayer, at least that of the Trp residue in the peptide, follows a somewhat different order: W^6Hya1 (+3) > D^0W^6Hya1 (+2) » K^0W^6Hya1 (+4). That is given by the values of $\Delta\lambda_{max}$ in Table 1, and those of ΔF_∞ in Table 2. It shows that the extra charge at the N-terminal (D^0) somehow prevents a deeper penetration of D^0W^6Hya1 (+2) into the bilayer, though it gives it a higher apparent affinity for PC membranes, as compared with W^6Hya1 (+3).

Therefore, our data show that the presence of an extra positive charge at the N-terminal of the peptide (K^0W^6Hya1), as compared with an extra negative charge (D^0W^6Hya1), drastically decreases both the binding constant and the penetration depth of the peptide into zwitterionic membranes. That is interesting, as it does not agree with previous results with molecular dynamics, which indicated that both D^0W^6Hya1 and K^0W^6Hya1 interacted similarly with zwitterionic lipid structures (micelles of dodecylphosphocholine [22]). It is important to point out that although micelles present similarities with membranes, since they are formed by amphipathic molecules, micelles are much more disordered structures.

The large blue shifts found for the Trp in W^6Hya1 and D^0W^6Hya1 in fluid PC bilayers, around 25 nm, are similar to those found for Trp deep in the bilayer in transmembrane peptides [45–48]. Hence, these two peptides, though diffusing in the membrane, are deeply immersed into the bilayer, either as transmembrane peptides or at the membrane surface.

The most cationic peptide K^0W^6Hya1 (+4) is the one that displays the lowest affinity for fluid PC membranes, and the one that binds at the shallowest bilayer position (Trp location). Accordingly, at 8 mol%, it is the one that less disturb the DPPC gel-fluid transition (Fig. 5c). Therefore, it was noteworthy that this peptide was the most efficient in causing CF leakage through fluid PC vesicles after around 30 min (Fig. 6). Though further investigation is certainly necessary, it is interesting to compare this finding with previous results that showed that although K^0W^6Hya1 binds deeper and stronger in anionic than in zwitterionic membranes, its efficiency in causing CF leakage was found to be much higher in fluid PC than PG vesicles [23]. The authors suggested that the peptide would be located at the surface of zwitterionic vesicles, laterally diffusing on it, triggering the opening of transient membrane polar pores due to the overlapping of irregular lipid packing zones, a mechanism proposed by Ref. [49]. In contrast, in anionic bilayers, K^0W^6Hya1 would be deeply embedded and strongly attached to the bilayer.

Accordingly, we could speculate that due to K^0W^6Hya1 lower affinity for PC vesicles and its shallower position on the membrane, as compared with W^6Hya1 and D^0W^6Hya1, K^0W^6Hya1 would be available to interact with more vesicles in a certain interval of time than the other two peptides, causing transient disruptions on the membrane. Hence, this could possibly explain the higher % of CF release observed after the 2000^{th} second for this peptide as compared with the other peptides. In the same trend, the AMPs Mac1 and aurein present greater affinity to anionic bilayers but they are more efficient to induce CF leakage in zwitterionic vesicles than in anionic [50].

Concerning the interaction of Hylin peptides with CT DNA, the blue

Table 4
Stern-Volmer constants of EB-DNA with Hylin peptides, from Fig. 8d and Eq. (4).

AMP	K_{SV} [$\times 10^4$ L mol^{-1}]
W^6Hya1 (+3)	(0.97 ± 0.03)
D^0W^6Hya1 (+2)	(0.33 ± 0.06)
K^0W^6Hya1 (+4)	(1.7 ± 0.4)

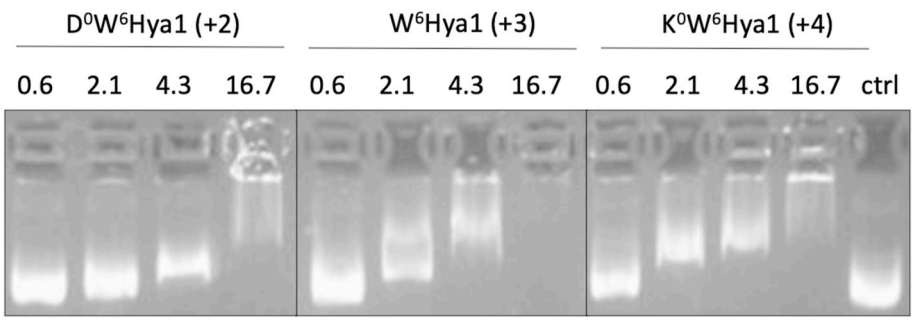


Fig. 9. Inhibition of the plasmid DNA (pOP3BP) mobility in agarose (1% w/v) gel electrophoresis upon addition of increased amount of Hylin peptides. Different amounts of peptides were incubated with 600 ng of pOP3BP plasmid DNA at room temperature for 60 min, and the reaction mixtures were applied into the gel. The relative AMP/pDNA weight ratio are indicated in the lane references and the control consists of plasmid DNA only. The results shown are representative of three experiments.

shifts observed are considerably large and follow a trend similar to those observed for the experiments with EB-peptide competition binding to DNA (see Tables 3 and 4), suggesting that the hydrophobic moiety of the peptides is imbedded into the DNA double helix. Indeed, Hylin peptides have considerable DNA intercalative binding abilities, as attested by the Stern-Volmer (K_{SV}) values in Table 4. As a comparison, synthetic antibiotics from the fluoroquinolone family are also known to intercalate into DNA. Stern-Volmer constants determined by the decrease of the EB-CT DNA fluorescence due to the presence of several fluoroquinolones are in the same order of magnitude, 10^4 L mol^{-1} [51]. The AMP AN5-1 (YSKSLPLSVLNP) also shows ability to decrease EB-CT DNA fluorescence with a K_{SV} of the same order mentioned above [52]. However, other drugs present a K_{SV} of two orders of magnitude higher than those observed by us [53]. Intercalant molecules can exhibit a cytotoxicity, as they can promote breaks and/or prevent the proper functions of the enzymatic cellular machinery, thus interfering in the processes of transcription and/or DNA replication, which can result in cell apoptosis [44,54].

Given the electrostatic interaction between the DNA anionic phosphate groups and the cationic peptides, a groove binding mode between Hylin peptides and DNA might also exist. It is quite common that a charged molecule interacts with DNA through both intercalative and groove binding modes [43,55]. Nonetheless, some molecules only interact via one binding mode, either intercalative or groove binding. For example, the synthetic lipophilic peptides, OA-C1b and LA-C1b, based on the sequence of the highly cationic chensinin-1b (+9) interact with CT DNA presenting a Trp blue shift of 9 nm, but the data indicate that these peptides interact mostly with the anionic phosphate group, since no evidence was found of intercalative binding mode [56]. Similarly, the synthetic aromatic tripeptide Phe-Phe-Phe (FFF), with a null net charge, interacts with CT-DNA but also only exhibits DNA groove binding activity, since it does not displace EB from the DNA double helix [57].

Electrophoresis experiments with the peptides and pDNA follow the same trend as that observed for the Trp blue shifts and K_{SV} values obtained for the peptides interaction with CT-DNA. Namely, W^6Hya1 and K^0W^6Hya1 display nearly the same effect on the pDNA migration pattern, whereas D^0W^6Hya1 is less effective in inducing changes in the pDNA migration (Fig. 9). Similarly, the interaction of D^0W^6Hya1 with CT-DNA yields the smallest values of Trp blue shift and K_{sv} (Tables 3 and 4).

Several factors drive the interaction of AMP with plasmid DNA. For example, the cationic AMP Magainin 2 (+3), which consists in 23 amino acid residues, only prevents total retardation in DNA electrophoresis migration with peptide/pDNA weight higher than 100 [58], whereas Indolicidin an AMP with 13 amino acid residues and a net charge of (+4) is able to promote complete retardation in pDNA electrophoresis experiments with a peptide/pDNA ratio of only 0.6 [59]. Taken this examples as lower and upper limit, we can conclude that the Hylin peptides studied here have a considerable DNA-binding activity since they can induce complete DNA retardation at the relative AMP/pDNA weight ratio of 16.7.

5. Conclusion

This work shows that the net positive charge of AMPs can determine its effective dissociation constant to zwitterionic membranes. From the intrinsic Trp fluorescence experiments, we showed that all three peptides bind zwitterionic bilayers, presenting a higher affinity to fluid than gel membranes. The affinity for fluid DPPC bilayers follows the decrescent order D^0W^6Hya1 (+2) > W^6Hya1 (+3) » K^0W^6Hya1 (+4). Fluorescence data also indicate that the Trp residue in the more positively charged peptide, K^0W^6Hya1, is less deep in the bilayer than the residue in the other two peptides. This finding is supported by DSC data, which shows that both D^0W^6Hya1 and W^6Hya1 disturbs DPPC gel-fluid transition slightly more effective than K^0W^6Hya1. It is important to note that the extra negative charge at the peptide N-terminal keeps the Trp residue in D^0W^6Hya1 in a shallower position in the zwitterionic membrane as compared with the residue in W^6Hya1.

The peptide, K^0W^6Hya1, displays the lowest affinity for PC fluid membranes and is located at the most superficial position in the bilayer. This peptide also happens to be the most efficient in causing pore formation in the membrane, as attested by CF leakage assays. As DSC data indicate that the three peptides laterally diffuse in PC membranes, it would be very interesting to find out why K^0W^6Hya1 can induce a higher leakage in zwitterionic vesicles. It should be important to find out if this is a particular result for these three peptides or if it is part of a more general trend.

The three Hylin a1 analogues studied here present DNA-binding activity. They exhibit a large Trp blue shift in the presence of CT DNA, and intercalate into DNA double helix, since they displace EB from the DNA pocket. W^6Hya1 (+3) and K^0W^6Hya1 (+4) are more efficient in binding DNA than the less cationic peptide D^0W^6Hya1 (+2), as attested by Trp fluorescence blue shifts, electrophoresis experiments, and EB competitive studies. These peptides display a considerable DNA binding activity since they could stop pDNA electrophoretic migration at AMP/pDNA weight ratio of 16.7. Hence, our findings suggest that the antibiotic action of Hylin peptides may also involve DNA-binding in addition to membrane interaction, albeit further investigations are necessary to better comprehend the Hylin peptides mechanism of action in cells.

The results shown here demonstrates that the peptide net charge is relevant and can modulate its interaction with eukaryotic structures such as zwitterionic bilayers and DNA. These findings could help the design of new therapeutic agents, balancing its antimicrobial and toxic effects.

Author contribution

Gabriel S. Vignoli Muniz: Investigation, data analysis, and writing-original draft preparation. **Lilia I. De la Torre:** Investigation, DNA electrophoresis experiments. **Evandro L. Duarte:** Investigation, data analysis and conceptualization **Esteban N. Lorenzón** and **Eduardo M. Cilli**: Peptide synthesis. **Andrea Balan**: DNA electrophoresis experiments. **M. Teresa Lamy**: Supervision, Writing-Reviewing and Editing.

Declaration of competing interest

There is no conflict of interest among authors.

Acknowledgment

This work was supported by the Brazilian agencies CNPq, FAPESP (2017/25930–1). E.M.C., A.B. and M.T.L. are recipient of CNPq research fellowships. G.S.V.M., E.L.D. and M.T.L. are part of the National Institute of Science and Technology Complex Fluids (INCT-FCx), financed by CNPq (465259/2014–6 and 405637/2017–1) and FAPESP (2014/50983–3 and 2018/20162–9). L. I. D. l. T. has scholarship from the Colombian COLCIENCIAS agency.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2020.100827.

Transparency document

Transparency document related to this article can be found online at https://doi.org/10.1016/j.bbrep.2020.100827.

References

[1] G.L. French, The continuing crisis in antibiotic resistance, Int. J. Antimicrob. Agents 36 (2010), https://doi.org/10.1016/S0924-8579(10)70003-0. S3–S7.
[2] G.M. Rossolini, F. Arena, P. Pecile, S. Pollini, Update on the antibiotic resistance crisis, Curr. Opin. Pharmacol. 18 (2014) 56–60, https://doi.org/10.1016/j.coph.2014.09.006.
[3] R. Wu, L. Wang, H.-C.D. Kuo, A. Shannar, R. Peter, P.J. Chou, S. Li, R. Hudlikar, X. Liu, Z. Liu, G.J. Poiani, L. Amorosa, L. Brunetti, A.-N. Kong, An update on current therapeutic drugs treating COVID-19, Curr. Pharmacol. Rep. (2020) 56–70, https://doi.org/10.1007/s40495-020-00216-7.
[4] S. Elnagdy, M. AlKhazindar, The potential of antimicrobial peptides as an antiviral therapy against COVID-19, ACS Pharmacol. Transl. Sci. (2020) 780–782, https://doi.org/10.1021/acsptsci.0c00059.
[5] L.R. Pizzolato-Cezar, N.M. Okuda-Shinagawa, M.T. Machini, Combinatory therapy antimicrobial peptide-antibiotic to minimize the ongoing rise of resistance, Front. Microbiol. 10 (2019), https://doi.org/10.3389/fmicb.2019.01703.
[6] Q. Wu, J. Patočka, K. Kuča, Insect antimicrobial peptides, a mini review, Toxins 10 (2018) 461, https://doi.org/10.3390/toxins10110461.
[7] D. Phoenix, S.R. Dennison, F. Harris, Antimicrobial Peptides, Wiley-VCH, Weinheim, 2013.
[8] D. Drider, S. Rebuffat (Eds.), Prokaryotic Antimicrobial Peptides, Springer New York, New York, NY, 2011, https://doi.org/10.1007/978-1-4419-7692-5.
[9] G. Wang, X. Li, Z. Wang, APD3: the antimicrobial peptide database as a tool for research and education, Nucleic Acids Res. 44 (2016) D1087–D1093, https://doi.org/10.1093/nar/gkv1278.
[10] A.J. Mason, A. Marquette, B. Bechinger, Zwitterionic phospholipids and sterols modulate antimicrobial peptide-induced membrane destabilization, Biophys. J. 93 (2007) 4289–4299, https://doi.org/10.1529/biophysj.107.116681.
[11] K.A. Brogden, Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. 3 (2005) 238–250, https://doi.org/10.1038/nrmicro1098.
[12] N. Raheem, S.K. Straus, Mechanisms of Action for antimicrobial peptides with antibacterial and antibiofilm functions, Front. Microbiol. 10 (2019), https://doi.org/10.3389/fmicb.2019.02866.
[13] M.-A. Sani, F. Separovic, How membrane-active peptides get into lipid membranes, Acc. Chem. Res. 49 (2016) 1130–1138, https://doi.org/10.1021/acs.accounts.6b00074.
[14] J. Lee, D.G. Lee, Antimicrobial peptides (amps) with dual mechanisms: membrane disruption and apoptosis, J. Microbiol. Biotechnol. 25 (2015) 759–764, https://doi.org/10.4014/jmb.1411.11058.
[15] P.R.S. Sanches, B.M. Carneiro, M.N. Batista, A.C.S. Braga, E.N. Lorenzón, P. Rahal, E.M. Cilli, A conjugate of the lytic peptide Hecate and gallic acid: structure, activity against cervical cancer, and toxicity, Amino Acids 47 (2015) 1433–1443, https://doi.org/10.1007/s00726-015-1980-7.
[16] H. Choi, J.-S. Hwang, D.G. Lee, Coprisin exerts antibacterial effects by inducing apoptosis-like death in *Escherichia coli*, IUBMB Life 68 (2016) 72–78, https://doi.org/10.1002/iub.1463.
[17] F.H. Waghu, S. Joseph, S. Ghawali, E.A. Martis, T. Madan, K.V. Venkatesh, S. Idicula-Thomas, Designing antibacterial peptides with enhanced killing kinetics, Front. Microbiol. 9 (2018), https://doi.org/10.3389/fmicb.2018.00325.
[18] M.S. Castro, T.C.G. Ferreira, E.M. Cilli, E. Crusca, M.J.S. Mendes-Giannini, A. Sebben, C.A.O. Ricart, M.V. Sousa, W. Fontes, Hylin a1, the first cytolytic peptide isolated from the arboreal South American frog *Hypsiboas albopunctatus* ("spotted treefrog"), Peptides 30 (2009) 291–296, https://doi.org/10.1016/j.peptides.2008.11.003.
[19] E. Crusca, A.A. Rezende, R. Marchetto, M.J.S. Mendes-Giannini, W. Fontes, M. S. Castro, E.M. Cilli, Influence of N-terminus modifications on the biological activity, membrane interaction, and secondary structure of the antimicrobial peptide hylin-a1, Biopolymers 96 (2011) 41–48, https://doi.org/10.1002/bip.21454.
[20] B.R. da Silva, V.A.A. de Freitas, V.A. Carneiro, F.V.S. Arruda, E.N. Lorenzón, A.S. W. de Aguiar, E.M. Cilli, B.S. Cavada, E.H. Teixeira, Antimicrobial activity of the synthetic peptide Lys-a1 against oral *streptococci*, Peptides 42 (2013) 78–83, https://doi.org/10.1016/j.peptides.2012.12.001.
[21] R.N. Inui Kishi, D. Stach-Machado, J. de L. Singulani, C.T. dos Santos, A.M. Fusco-Almeida, E.M. Cilli, J. Freitas-Astúa, S.C. Picchi, M.A. Machado, Evaluation of cytotoxicity features of antimicrobial peptides with potential to control bacterial diseases of citrus, PloS One 13 (2018), e0203451, https://doi.org/10.1371/journal.pone.0203451.
[22] E. Crusca, A.S. Câmara, C.O. Matos, R. Marchetto, E.M. Cilli, L.M. Lião, A. Lima de Oliveira, NMR structures and molecular dynamics simulation of hylin-a1 peptide analogs interacting with micelles, J. Pept. Sci. 23 (2017) 421–430, https://doi.org/10.1002/psc.3002.
[23] T.A. Enoki, I. Moreira-Silva, E.N. Lorenzon, E.M. Cilli, K.R. Perez, K.A. Riske, M. T. Lamy, Antimicrobial peptide K^0-W^6-Hya1 induces stable structurally modified lipid domains in anionic membranes, Langmuir 34 (2018) 2014–2025, https://doi.org/10.1021/acs.langmuir.7b03408.
[24] D. Marsh, Handbook of Lipid Bilayers, second ed., CRC Press, Taylor & Francis Group, Boca Raton, FL, 2013.
[25] S. Ohki, Cell and Model Membrane Interactions, Springer, New York, 1991. http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3081431. (Accessed 13 July 2020).
[26] G. Rouser, A.N. Siakotos, S. Fleischer, Quantitative analysis of phospholipids by thin-layer chromatography and phosphorus analysis of spots, Lipids 1 (1966) 85–86, https://doi.org/10.1007/BF02668129.
[27] L.R. Gouvea, D.A. Martins, D. da Gama Jean Batista, M. de Nazaré, C. Soeiro, S.R. W. Louro, P.J.S. Barbeira, L.R. Teixeira, Norfloxacin Zn(II)-based complexes: acid base ionization constant determination, DNA and albumin binding properties and the biological effect against *Trypanosoma cruzi*, Biometals 26 (2013) 813–825, https://doi.org/10.1007/s10534-013-9661-z.
[28] J. Marmur, A procedure for the isolation of deoxyribonucleic acid from micro-organisms, J. Mol. Biol. 3 (1961), https://doi.org/10.1016/S0022-2836(61)80047-8, 208-IN1.
[29] J.R. Lakowicz (Ed.), Principles of Fluorescence Spectroscopy, Springer US, Boston, MA, 2006, https://doi.org/10.1007/978-0-387-46312-4.
[30] B. Valeur, Molecular Fluorescence: Principles and Applications, Wiley-VCH, Weinheim ; New York, 2002.
[31] C.C. Vequi-Suplicy, K. Coutinho, M.T. Lamy, New insights on the fluorescent emission spectra of prodan and laurdan, J. Fluoresc. 25 (2015) 621–629, https://doi.org/10.1007/s10895-015-1545-x.
[32] B. Alberts (Ed.), Molecular Biology of the Cell, fifth ed., Garland Science, New York, 2008.
[33] T. Heimburg, Thermal Biophysics of Membranes, Wiley-VCH Verlag, Weinheim, 2007.
[34] G. Denisov, S. Wanaski, P. Luan, M. Glaser, S. McLaughlin, Binding of basic peptides to membranes produces lateral domains enriched in the acidic lipids phosphatidylserine and phosphatidylinositol 4,5-bisphosphate: an electrostatic model and experimental results, Biophys. J. 74 (1998) 731–744, https://doi.org/10.1016/S0006-3495(98)73998-0.
[35] E. Prenner, M. Chiu, Differential scanning calorimetry: an invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions, J. Pharm. BioAllied Sci. 3 (2011) 39, https://doi.org/10.4103/0975-7406.76463.
[36] K.A. Riske, R.P. Barroso, C.C. Vequi-Suplicy, R. Germano, V.B. Henriques, M. T. Lamy, Lipid bilayer pre-transition as the beginning of the melting process, Biochim. Biophys. Acta BBA - Biomembr. 1788 (2009) 954–963, https://doi.org/10.1016/j.bbamem.2009.01.007.
[37] J. Sun, Y. Xia, D. Li, Q. Du, D. Liang, Relationship between peptide structure and antimicrobial activity as studied by de novo designed peptides, Biochim. Biophys. Acta BBA - Biomembr. 1838 (2014) 2985–2993, https://doi.org/10.1016/j.bbamem.2014.08.018.
[38] P.F. Almeida, A. Pokorny, Antimicrobial peptides, chapter 11, binding and permeabilization of model membranes by amphipathic peptides, in: A. Giuliani, A. C. Rinaldi (Eds.), Antimicrobial Peptides, Humana Press, Totowa, NJ, 2010, https://doi.org/10.1007/978-1-60761-594-1.
[39] S.G. Hovakeemian, R. Liu, S.H. Gellman, H. Heerklotz, Correlating antimicrobial activity and model membrane leakage induced by nylon-3 polymers and detergents, Soft Matter 11 (2015) 6840–6851, https://doi.org/10.1039/C5SM01521A.
[40] A.S. Vasilchenko, E.A. Rogozhin, Sub-inhibitory effects of antimicrobial peptides, Front. Microbiol. 10 (2019), https://doi.org/10.3389/fmicb.2019.01160.
[41] C.-F. Le, C.-M. Fang, S.D. Sekaran, Intracellular targeting mechanisms by antimicrobial peptides, Antimicrob. Agents Chemother. 61 (2017), https://doi.org/10.1128/AAC.02340-16.
[42] J. Yan, K. Wang, W. Dang, R. Chen, J. Xie, B. Zhang, J. Song, R. Wang, Two hits are better than one: membrane-active and DNA binding-related double-action mechanism of NK-18, a novel antimicrobial peptide derived from mammalian NK-Lysin, Antimicrob. Agents Chemother. 57 (2013) 220–228, https://doi.org/10.1128/AAC.01619-12.

[43] R. Eckel, R. Ros, A. Ros, S.D. Wilking, N. Sewald, D. Anselmetti, Identification of binding mechanisms in single molecule–DNA complexes, Biophys. J. 85 (2003) 1968–1973, https://doi.org/10.1016/S0006-3495(03)74624-4.
[44] L. Strekowski, B. Wilson, Noncovalent interactions with DNA: an overview, Mutat. Res. Mol. Mech. Mutagen. 623 (2007) 3–13, https://doi.org/10.1016/j.mrfmmm.2007.03.008.
[45] X. Bi, C. Wang, W. Dong, W. Zhu, D. Shang, Antimicrobial properties and interaction of two Trp-substituted cationic antimicrobial peptides with a lipid bilayer, J. Antibiot. (Tokyo) 67 (2014) 361–368, https://doi.org/10.1038/ja.2014.4.
[46] D.M. Pahlke, U. Diederichsen, Synthesis and characterization of β -peptide helices as transmembrane domains in lipid model membranes, J. Pept. Sci. 22 (2016) 636–641, https://doi.org/10.1002/psc.2912.
[47] D. Zweytick, B. Japelj, E. Mileykovskaya, M. Zorko, W. Dowhan, S.E. Blondelle, S. Riedl, R. Jerala, K. Lohner, N-acylated peptides derived from human lactoferricin perturb organization of cardiolipin and phosphatidylethanolamine in cell membranes and induce defects in *Escherichia coli* cell division, PloS One 9 (2014), e90228, https://doi.org/10.1371/journal.pone.0090228.
[48] N.G. Park, U. Silphaduang, H.S. Moon, J.-K. Seo, J. Corrales, E.J. Noga, Structure−activity relationships of Piscidin 4, a piscine antimicrobial peptide, Biochemistry 50 (2011) 3288–3299, https://doi.org/10.1021/bi101395j.
[49] B. Bechinger, The structure, dynamics and orientation of antimicrobial peptides in membranes by multidimensional solid-state NMR spectroscopy, Biochim. Biophys. Acta BBA - Biomembr. 1462 (1999) 157–183, https://doi.org/10.1016/S0005-2736(99)00205-9.
[50] M.-A. Sani, E. Gagne, J.D. Gehman, T.C. Whitwell, F. Separovic, Dye-release assay for investigation of antimicrobial peptide activity in a competitive lipid environment, Eur. Biophys. J. 43 (2014) 445–450, https://doi.org/10.1007/s00249-014-0970-0.
[51] P. Bhattacharya, S. Mukherjee, S.M. Mandal, Fluoroquinolone antibiotics show genotoxic effect through DNA-binding and oxidative damage, Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 227 (2020) 117634, https://doi.org/10.1016/j.saa.2019.117634.
[52] T. Yi, Y. Huang, Y. Chen, Production of an antimicrobial peptide AN5-1 in *Escherichia coli* and its dual mechanisms against bacteria, Chem. Biol. Drug Des. 85 (2015) 598–607, https://doi.org/10.1111/cbdd.12449.
[53] M. Zampakou, M. Akrivou, E.G. Andreadou, C.P. Raptopoulou, V. Psycharis, A. A. Pantazaki, G. Psomas, Structure, antimicrobial activity, DNA- and albumin-binding of manganese(II) complexes with the quinolone antimicrobial agents oxolinic acid and enrofloxacin, J. Inorg. Biochem. 121 (2013) 88–99, https://doi.org/10.1016/j.jinorgbio.2012.12.013.
[54] J.C. Wang, DNA topoisomerases, Annu. Rev. Biochem. 65 (1996) 635–692, https://doi.org/10.1146/annurev.bi.65.070196.003223.
[55] L. Ma, X. Xie, H. Liu, Y. Huang, H. Wu, M. Jiang, P. Xu, X. Ye, C. Zhou, Potent antibacterial activity of MSI-1 derived from the magainin 2 peptide against drug-resistant bacteria, Theranostics 10 (2020) 1373–1390, https://doi.org/10.7150/thno.39157.
[56] W. Dong, X. Luo, Y. Sun, Y. Li, C. Wang, Y. Guan, D. Shang, Binding properties of DNA and antimicrobial peptide Chensinin-1b containing lipophilic alkyl tails, J. Fluoresc. 30 (2020) 131–142, https://doi.org/10.1007/s10895-019-02478-x.
[57] S. Biswas, S. Samui, A. Chakraborty, S. Biswas, D. De, U. Ghosh, A.K. Das, J. Naskar, Insight into the binding of a non-toxic, self-assembling aromatic tripeptide with ct-DNA: spectroscopic and viscositic studies, Biochem. Biophys. Rep. 11 (2017) 112–118, https://doi.org/10.1016/j.bbrep.2017.07.001.
[58] C.B. Park, H.S. Kim, S.C. Kim, Mechanism of action of the antimicrobial peptide Buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions, Biochem. Biophys. Res. Commun. 244 (1998) 253–257, https://doi.org/10.1006/bbrc.1998.8159.
[59] C.-H. Hsu, Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: evidence for multiple conformations involved in binding to membranes and DNA, Nucleic Acids Res. 33 (2005) 4053–4064, https://doi.org/10.1093/nar/gki725.

Anexo 3 - Declaração de Bioética e/ou Biossegurança

Profa. Dra. Andrea Balan
Departamento de Microbiologia
Av. Prof. Lineu Prestes, 1374 – Bio II
Cidade Universitária - Butantã 195
05508-000 – São Paulo – SP – Brasil
Tel. +55-11-3091-7745
abalan@usp.br
#BalanGroup

À
Comissão de Ética
Programa de Pós-graduação em Genética e Biologia Molecular
Instituto de Biologia – UNICAMP

Venho por meio dessa carta informar que todos os experimentos envolvendo camundongos e células apresentados no desenvolvimento do projeto FAPESP 2014/20921-6 e na tese da aluna de doutorado **Lilia Iriarte De la Torre**, intitulado: "**Caracterização Estrutural e Funcional dos Transportadores do Tipo ABC de Açúcares em *Mycobacterium tuberculosis*"**, foram realizados pela Dra. Ana Carolina Ramos Moreno, do Laboratório de Desenvolvimento de Vacinas do Departamento de Microbiologia, da Universidade de São Paulo. A Dra. Moreno tem larga experiência com manuseio de animais, tem os certificados e todos os protocolos utilizados estão de acordo com os Princípios Éticos de Experimentação Animal adotado pela Sociedade Brasileira de Ciência de Animais de Laboratório (SBCAL), aprovados pela Comissão de Ética no Uso de Animais (CEUA) da Universidade de São Paulo.

Os documentos são apresentados em anexo.

Atenciosamente,

São Paulo, 07 de setembro de 2019.

Profa. Dr.a. Andrea Balan
Departamento de Microbiologia
Av. Prof. Lineu Prestes, 1374 – Bio II
Cidade Universitária - Butantã 196
05508-000 – São Paulo – SP – Brasil
Tel. +55-11-3091-7745
abalan@usp.br
#BalanGroup

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 – Cep. 05508-900 São Paulo, SP - Brasil
Telefone :(55) (011) 3091.7733 –e-mail: cep@icb.usp.br

COMISSÃO DE ÉTICA NO USO DE ANIMAIS

Decl. CEUA.057/2014.

DECLARAÇÃO

Em adendo ao Certificado 050/14/CEUA, datado de 10.06.14. e por solicitação do Prof. Dr. **Luis Carlos de Souza Ferreira,** responsável pela linha de Pesquisa, autorizo a inclusão de ***Ana Carolina Ramos Moreno*** ao Protocolo de Pesquisa *"Controle de tumores induzidos pelo vírus papiloma humano (HPV-16) por abordagens vacinais terapêuticas"*, uma vez que se trata de utilização da mesma espécie animal e de métodos experimentais similares ao Projeto.

São Paulo, 24 de julho de 2014.

Prof. Dr. Wothan Tavares de Lima
Coordenador da CEUA
ICB/USP

Profa. Dr.a. Andrea Balan
Departamento de Microbiologia
Av. Prof. Lineu Prestes, 1374 – Bio II
Cidade Universitária - Butantã 197
05508-000 – São Paulo – SP – Brasil
Tel. +55-11-3091-7745
abalan@usp.br
#BalanGroup

DEPARTAMENTO DE MICROBIOLOGIA
UNIVERSIDADE DE SÃO PAULO

Applied Structural Biology Lab
Transporters

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 – CEP. 05508-000 São Paulo, SP Brasil
Telefone :(55) (011) 3091.7733 – e-mail: cep@icb.usp.br

CERTIFICADO

Certificamos que o protocolo registrado sob nº **50** nas fls. **18** do livro **03** para uso de animais em experimentação, sob a responsabilidade do Prof(a) Dr(a) **Luís Carlos de Souza Ferreira,** Coordenador (a) da Linha de pesquisa *"Controle de tumores induzidos pelo vírus papiloma humano (HPV-16) por abordagens vacinais terapêuticas"* do qual participam o(s) aluno(s) **Bruna Maldonado, Natiely Silva Sales, Carina Buzzo de Lima, Lais Helena Teixeira Merlin de Andrade,** está de acordo com os Princípios Éticos de Experimentação Animal adotado pela Sociedade Brasileira de Ciência de Animais de Laboratório (SBCAL) e foi aprovado pela *COMISSÃO DE ÉTICA NO USO DE ANIMAIS* (CEUA) em **10.06.2014, com validade de 4 anos.**

São Paulo, 11 de junho de 2014.

Prof. Dr. WOTHAN TAVARES DE LIMA
Coordenador-CEUA- ICB/USP

Profa. Dra. ANA PAULA LEPIQUE
Secretária- CEUA - ICB/USP

UNICAMP

CEUA/UNICAMP

INFORMAÇÃO

A Comissão de Ética no Uso de Animais da UNICAMP - CEUA/UNICAMP - esclarece que não há necessidade de submeter o projeto de pesquisa "**Caracterização Estrutural e Funcional dos Transportadores do Tipo ABC de Açúcares em Mycobacterium tuberculosis**", de responsabilidade da Profa. Dra. Andrea Balan e da pós-graduanda **Lilia Iriarte De la Torre**, para análise desta comissão.

Justifica-se por se tratar de projeto que utilizará resultados previamente obtidos, e cuja realização de todos os procedimentos experimentais envolvendo animais/células foi previamente realizada pela Dra. Ana Carolina Ramos Moreno, a qual possui devida aprovação pela Comissão de Ética no Uso de Animais – CEUA – ICB/USP (emissão do certificado: 11/06/2014). Não haverá, assim, manipulação *in vivo* na UNICAMP para a execução do projeto.

Campinas, 29 de novembro de 2019.

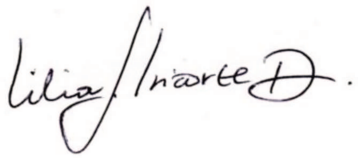
Prof. Dr. WAGNER JOSÉ FÁVARO
Presidente da CEUA/UNICAMP

Anexo 4 - Declaração referente a direitos autorais

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Tese de Doutorado, intitulada: **Caracterização Estrutural e Funcional dos Transportadores do Tipo ABC de Açúcares em *Mycobacterium tuberculosis***", não infringem os dispositivos da Lei nº 9.610/98, nem o direito autoral de qualquer editora.

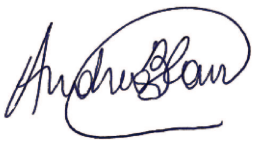
Campinas, 26 de janeiro de 2021.



Assinatura:

Nome da autora: Lilia Iriarte De la Torre

RNE nº: G194253-L

Assinatura: 

Nome da orientadora: Andrea Balan Fernandes

RG nº: 18.361.330-2