The University Of Sheffield.

The Tripartite Tricarboxylate Transporters (TTT) : The neglected family of high-affinity uptake systems

A thesis submitted in part fulfilment for the degree of

Doctor in Philosophy

Department of Molecular Biology and Biotechnology

University of Sheffield

Sponsored by Science without Borders program

Leonardo Talachia Rosa

July 2018

ABSTRACT

Bacteria inhabiting complex environments must possess efficient ways of acquiring nutrients, often present in low concentrations. In this context, Solute Binding Protein (SBP) dependent transporters make use of a periplasmic binding protein to bind substrates with high affinity, and deliver them to their transmembrane counterparts. There are three known families of SBP dependent transporters: The ABC (ATP-Binding Cassette), the TRAP (TRipartite ATP-independent Periplasmic transporters) and the TTT (Tripartite Tricarboxylate Transporter). The latter are very poorly characterized and only a few types of substrates for TTT transporters are currently known. This thesis first presents a review of the TTT transporters. Our database searches reveal a massive overrepresentation of TTT SBPs among α and β-proteobacteria, and highlight the presence of 434 TTT SBPs in the bacterium *Rhodoplanes sp.* Z2-YC6860, the biggest gene family representation described in a bacterial genome to date. We subsequently focus on the characterization of the TTT family in *Rhodopseudomonas palustris,* a model soil non-sulfur purple bacterium. The TTT family in *R. palustris* is formed by two complete tripartite systems, plus five orphan SBPs with no obvious membrane counterparts. From the seven SBPs, we present high-resolution crystal structures for six of them, and two of them were characterized biochemically and physiologically. One of the orphan SBPs (AdpC) is described to bind to dicarboxylic acids ranging from six to nine carbons in length with low µM affinity, and is more expressed under low concentration of adipate. The MatC protein, belonging to the tripartite system MatBAC, is shown to bind to malate with low nM affinity. This study also attempted to provide the first biochemical characterization of a TTT tripartite system, with the reconstitution of the transmembrane components MatBA into proteoliposomes. In summary, this study presents a characterization of the

Tripartite Tricarboxylate Transporter family in bacteria, through a synergistic multidisciplinary approach.

ACKNOWLEDGEMENTS

I would like to thank Professor David Kelly for the opportunity and for the excellent guidance over the past four years. His immense scientific experience, kind personality, concern with the progress and wellbeing of his team and encouragement for us to pursue our own initiatives shall be a model for me in any position the future takes me.

Professor Jeff Green, Professor Gavin Thomas, Dr. Christopher Mulligan and Dr. John Rafferty for sharing their knowledge, expertise, and friendship.

Former and present members in F1, who made the working environment a pleasant place to be at, sharing friendship, knowledge, jokes and also helping to deal with frustrations inherent to the learning process.

My beautiful wife Ana Claudia, whose kindness, patience and positivity amaze me every day, and who remained strong every time I was not. I can't imagine how this would be possible without you.

My unique parents Leila and Antônio, for their unconditional support, encouragement, and for being a model for so many virtues that through them I could develop in myself. I am filled with gratitude and can't thank you enough.

Everyone in the Sheffield Buddhism Centre, who taught me to see myself and the world through the lens of kindness, compassion, stillness, gratitude, truthful communication and awareness.

The present work was accomplished with funding from Brazilian funding agency CNPQ (National Council for Scientific and Technological Development), though PhD studentship in the remit of "Science Without Borders" program (248597/2013-2). May this study reinforce the importance of prioritizing investment in science and research by the Brazilian government.

PUBLICATIONS AND CONFERENCE PRESENTATIONS

PUBLICATIONS

- **Rosa LT**, Dix SR, Rafferty JB, Kelly DJ. Structural basis for high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic-binding protein from the tripartite tricarboxylate transporter (TTT) family in *Rhodopseudomonas palustris*. **The FEBS Journal** 2017. 284(24):4262-4277 DOI: 10.1111/febs.14304
- **Rosa LT**, Bianconi ME, Thomas GH, Kelly DJ. Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): From uptake to pathogenicity. **Frontiers in Cellular and InfectionMicrobiology** 2018. In Press. doi: 10.3389/fcimb.2018.00033
- **Rosa LT**, Springthorpe V, Bianconi M, Thomas G, Kelly D. Massive overrepresentation of solute-binding proteins(SBPs) from the tripartite tricarboxylate transporter (TTT) family in the genome of the α-proteobacterium *Rhodoplanes sp*. Z2-YC6860. **Microbial Genomics** 2018; doi: 10.1099/mgen.0.000176.

CONFERENCE PRESENTATIONS

- **Leonardo Talachia Rosa**, Sam Dixon, John Rafferty, David Kelly (Apr 2018). high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic binding protein from the Tripartite Tricarboxylate Transporter (TTT) family in *Rhodopseudomonas palustris*. Faculty of Science Poster session (Poster)
- **Leonardo Talachia Rosa**, Sam Dixon, John Rafferty, David Kelly (Dec 2017). high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic binding protein from the Tripartite Tricarboxylate Transporter (TTT) family in *Rhodopseudomonas palustris*. CBMNet Factories for advanced biomanufactoring (Poster)
- **Leonardo Talachia Rosa**, Sam Dixon, John Rafferty, David Kelly (Sep 2017). high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic binding protein from the Tripartite Tricarboxylate Transporter (TTT) family in *Rhodopseudomonas palustris*. CBMNet and BIOCATNET: Import and Export of Small Molecules for Biocatalysis (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly (July 2017). Biotechnological potential of novel TTT-family transporters from Rhodopseudomonas palustris. Mechanisms & Methods in Membrane Transporters UK-Taiwan Partnering Kick-off Meeting (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly (june 2017) high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic binding protein from the Tripartite Tricarboxylate Transporter (TTT) family in *Rhodopseudomonas palustris*. CBMNet Membrane Engineering Of Lipids And Proteins For Industrial Biotechnology and Bioenergy (poster)

- Leonardo Talachia Rosa. (January 2017) Metabolite transporters and aromatic compound utilisation by the photosynthetic bacterium *Rhodopseudomonas palustris*. 1st µSheffield symposium. Sheffield, UK (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (January 2017) Biotechnological potential of novel TTT-family transporters from *Rhodopseudomonas palustris*. 3rd year MBB PostGraduate symposium Sheffield, UK (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (July 2016) Biotechnological potential of novel TTT-family transporters from *Rhodopseudomonas palustris*. CBMNet Early Career Researchers Meeting (Networking meeting) Sheffield, UK (Oral) – 2nd place prize
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (July 2016) Biotechnological potential of novel TTT-family transporters from *Rhodopseudomonas palustris*. 17th European Congress on Biotechnology (International Conference) Krakow, Poland (Oral) Doi 10.1016/j.nbt.2016.06.855
- **Leonardo Talachia Rosa** (June 2016) Degradation of aromatic compounds in *Rhodopseudomonas palustris*. 3rd MBB PhD society Retreat (local meeting) Sheffield, UK (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (March 2016) TTT transporters in *Rhodopseudomonas palustris*. Society for General Microbiology (SGM) Annual Meeting 2016 (National conference) Liverpool, UK (Poster)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (September 2015) Characterization of Secondary Soluble Binding Protein Dependant transporters in *Rhodopseudomonas palustris*. 2nd Midlands Microbiology Conference (Regional conference) – Nottingham, UK (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (April 2016) TTT transporters in *Rhodopseudomonas palustris*. Second year MBB PhD student`s symposium (Internal meeting) Sheffield, UK (Oral)
- **Leonardo Talachia Rosa**, John Rafferty, David Kelly. (June 2015) TTT transporters in *Rhodopseudomonas palustris*. 2nd MBB PhD society retreat (local meeting) Sheffield, UK (Poster presentation)

LIST OF NON-STANDARD ABREVIATIONS

ABC	ATP binding Cassette
ATP	adenosine triphosphate
CFE	cell free extract
CMC	critical micelle concentration
CoA	coenzyme A
DDM	dodecyl β-D-maltopyranoside
DM	decyl β-D-maltopyranoside
IM	inner membrane
HGT	horizontal gene transfer
LPS	lipopolysaccharide
LSE	lineage-specific expansion
MFS	major facilitator superfamily
NBD	nucleotide binding domain
OG	N-Octyl-β-D-glucoside
OM	outer membrane
PDB	protein database
PMF	proton motif force
RC	reaction center
RCV	*Rhodobacter capsulatus* vitamins
SBP	solute binding protein
SEC	size exclusion chromatography
Tat	twin-arginine translocase
TCA	tricarboxylic acid cycle
TMD	transmembrane domain
TRAP	triapartite ATP-independent periplasmic transporter
TTT	triapartite tricarboxylate transporter

SUMMARY

CHAPTER 1. TRIPARTITE ATP-INDEPENDENT PERIPLASMIC (TRAP) TRANSPORTERS AND TRIPARTITE TRICARBOXYLATE TRANSPORTERS (TTT): FROM UPTAKE TO PATHOGENICITY

Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. Frontiers in Cellular and Infection Microbiology. 2018;8:33. doi:10.3389/fcimb.2018.00033.

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb. 2018.00033/full#supplementary-material

PREFACE

This literature review provides a comprehensive description of the high-affinity, secondary, Soluble Binding Protein (SBP) dependent transporters from the TRipartite ATP-independent transporters (TRAP) and Tripartite Tricarboxylate Transporter (TTT) families. For each family are described: the discovery and early studies of their respective prototype systems; their genomic organisation; the ubiquity of them among bacteria; and their main physiological functions. The focus is on: pathogenicity and colonisation; the topological, biochemical and energetic properties of the transmembrane components; and the biochemical and detailed structural characterisation of their SBP components and ligand coordination.

AUTHOR CONTRIBUTIONS

The idea for the manuscript was conceived by L Rosa, D Kelly and G Thomas. Literature search, interpretation, discussion and figures were done by L Rosa. M

Bianconi was responsible for the bioinformatics analysis, in close collaboration with L Rosa. L Rosa, D Kelly and G Thomas co-wrote the manuscript.

Manuscript Insert

1.2. THE ATP-BINDING CASSETE (ABC) TRANSPORTERS FAMILY

The ATP Binding Cassete (ABC) transporters were the first class of high affinity transporters depedent on an extracytoplasmic binding-protein to be discovered, and are by far the best characterised and the most numerous compared to the TRAP and TTT families. The characterisation of the first ABC transporters begun in the 1970`s with the description of the galactose (Kalckar 1971) and ribose (Willis and Furlong 1974) transport systems in *E. coli*, when they were known as 'osmotic shock' sensitive transporters, due to the disruption of transport activity upon osmotic shock, which expels the SBPs from the periplasm. Structures of the binding proteins followed shortly after, revealing the now well-known Venus fly-trap like structure for binding proteins (Scheepers *et al.* 2016).

Today, ABC transporters are known to be involved in the import of a vast range of substrates, such as aminoacids and polypeptides, saccharides, metals, vitamins, opines, polyamines and ions, among others (Locher 2016), and spread among all kingdoms of life. Moreover, several additional functions are described for proteins of the ABC family, such as chemotaxis, direct and indirect gene regulation, efflux pumps and multidrug resistance (Wilkens 2015). However, in the present study only the context of solute import in prokaryotes will be looked at in detail.

In bacteria, the ABC importers are formed by three to five distinct proteins: The SBP, which bind the ligand with high affinity, can be free in the periplasm (in Gram-negative bacteria) or in the form of a lipoprotein/fused with the trans-membrane domains (TMDs) (in Gram-positive bacteria and archaea); two TMDs forming a translocation pore; and two Nucleotide Binding Domains (NBDs). The NBD is the most conserved subunit among the ABC transporters, with aminoacid identity ranging from 30% to

50%. It contains: a RecA like domain; a WalkerA (or P-loop) motif specific for NB proteins which binds the nucleotide; an α-helix containing the signature for ABC transporters, LSGGQ (called the C-loop), which contacts the nucleotide in ATP-bound state; a WalkerB domain, which coordinates the water attack to the ATP using a Glu residue; a Q-loop which senses the γ-phosphate; the A-loop, containing aromatic residues that coordinate the ATP adenine; the switch motif, containing an histidine which helps in the ATP catalysis; and a coupling helix which makes the transmission interface with the TMD (Hollenstein *et al.* 2007). The NBDs are organized in dimers, with the two ATP molecules localised in the dimerization interface. Each ATP interacts with a P-loop from one domain and the C-loop from the other domain (Wilkens 2015). The Transmembrane Domains (TMD) are much less conserved in the primary sequence, but show very similar topologies. They have between 6 and 10 transmembrane helixes, and assume a dimeric form, with a pore in the dimerization interface. In the cytoplasmic side of these proteins there is an elongation of α-helix, responsible for interacting with the SBP and the substrate, and in the cytoplasmic side, helix which interact with the RecA-like portion of the NBD (Wilkens 2015).

Although very similar, two topologies and modes of solute translocation were described for the ABC transporters, named type I and type II, reviewed by Locher (2016). In type I ABC importers (Figure 1-1a), the transport cycle starts with the TMD open in the cytoplasmic side. Upon interaction with the SBP, the NBD interacts with ATP, which causes the close approach of the two NBD. This event will cause closing of the TMD in the cytoplasm, and opening on the periplasmic side. This exposes a low affinity binding pocket in the translocation pathway inside the TMD. Migration from the SBP to this domain is thought to happen due to distortion of the high-affinity binding domain upon interaction and also sterical clashes with helices from the TMD.

Cleavage of ATP and release of Pi by the NBD causes it to pull one apart from each other, thus opening the cytoplasmic side of the TMD and releasing the substrate in the cytoplasm. The release of the ADP and interaction with a new ATP molecule initiates the cycle again. Type I ABC importers are related to medium affinity uptake of a diverse range of substrates, such as ions, sugars, oligopeptides, oligosaccharides and aminoacids, having the *E.coli* maltose transporter as the prototype for this class.

Type II ABC importers (Figure 1-1b) contain 3 gates in the TMD translocation pathway, 2 in the cytoplasmic side and one in the periplasmic side. With ATP bound, NBD is closed, making the cyto-gate 2 to be closed as well, while the peri- gate is open. Interaction of the SBP causes migration of the substrate using the same mechanisms described for type I, but the ligand is transferred to a hydrophobic pocket with no experimentally characterised affinity. ATP cleavage pulls the NBD apart, opening cyto-gate 2. This would close cyto-gate I, but the substrate is interfering spatially with this process, and what happens is that the hydrophobic cavity pushes the substrate towards cyo-gate 2, much like in a peristaltic movement. Once substrate is released in the cytoplasm, TMD assumes a relaxed form, avoiding transport in the opposite direction. Type II importers are related to high affinity uptake of substrates present in lower concentration and with higher hydrophobicity, such as iron and heme complexes and vitamins, having the vitamin B12 transporter BtuCD-F from *E.coli* as their prototype (Locher 2016).

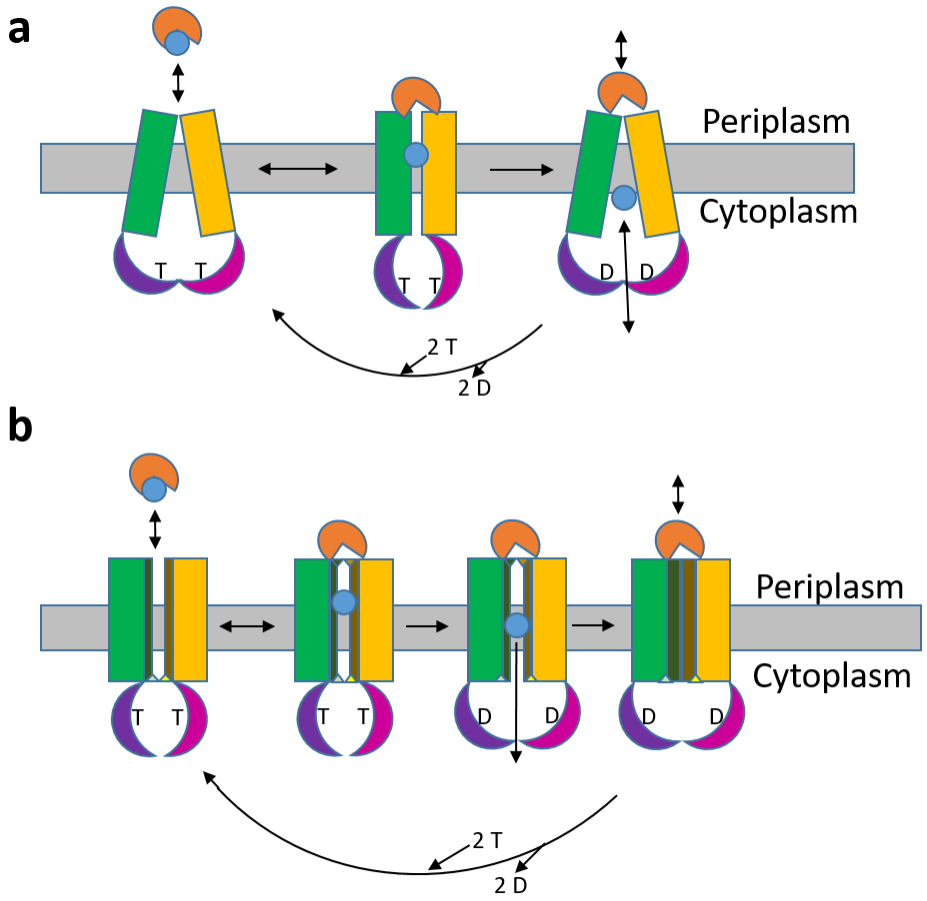


Figure 1-1: Comparison between type I (a) and type II (b) ABC transporters. Legend and colour scheme: TM domains - Green and Yellow; NB domains - magenta and purple; SBP - orange; ligand - blue; cytoplasmic membrane – gray; ATP – T; ADP – D; bi-directional arrows – reversible processes; unidirectional arrows – irreversible processes.

The SBPs associated with ABC transporters show very diverse structures, varying depending on the type of substrate they are adapted to bind. According to the classification performed by Berntsson *et al.* (2010) and updated by Scheepers *et al.* (2016), the SBPs from this family fall into several clusters. Below are presented the particularities of clusters A-G for binding proteins:

- Cluster A: is comprised by Class III ABC transporters, which have the two domains connected by an α-helix, with a low degree of closure upon ligand binding. They usually bind to metals (A-I) or chelated metals (A-II)

- Cluster B: Type I SBPs, thus containing three hinges connecting the domains, and so the N and C terminals are in the same domain. These proteins are involved with uptake of sugars, aminoacids, peptides, aromatic compounds and autoinducers, and might interact with two-component systems' G-proteins and are involved in transcription factor functions, including the well-studied Lac repressor.
- Cluster C: Significantly larger than the other SBPs, this cluster contains proteins ranging from 55 to 70 kDa, which present an additional domain, responsible for enlarging the binding pocket. These proteins are related to uptake of peptides and oligosaccharides and cellobiose. The proteins AppA from *Bacillus subtilis* and OppA from *Lactococcus lactis* are part of this cluster.
- Cluster D: Comprising type II proteins, this cluster presents two small hinges, of 4-5 residues connecting the two domains, and might have a strand located laterally in the protein. These proteins are involved with uptake of a variety of substrates, such as sugars, vitamins, polyamines, tetrahedral anions and metals.
- Cluster E: Do not comprise SBPs from the ABC family, but rather all proteins from TRAP and TTT (subcluster E-II) families. It might contain a long strand which connects the two domains and are part of the two β-strands. In addition, the α-helix positions are well conserved and have a C-terminal α-helix which spans both domains.
- Cluster F: Comprises Type II proteins, where the two strands are comprised of 8-10 residues each, conferring more flexibility to the binding pocket. The proteins in this cluster are involved in the uptake of a variety of substrates, such as compatible solutes, trigonal anions and aminoacids.

With such a structural diversity among the SBPs from the ABC systems, common features in ligand coordination mechanisms are expected to be rare. Indeed, as affirmed by Locher (2016), ligand position and coordination is very variable amongst this family, being described for individual proteins and making ligand prediction from proteins without additional information very difficult. Hydrogen bonds play a very important role, but in many cases the interactions are bridged by water molecules, making it difficult to predict which interactions are likely to happen. Based on the cluster separation above, it is possible to extract from the overall tertiary structure an indication of what class of compound each protein is likely to bind, but no further information than that would be reliable. In some cases, especially inside clusters B and C, some interactions are recurrent and can indicate the general nature of the ligand. In Cluster B, proteins which bind to monosaccharides feature an asparagine, an aspartate and an arginine as conserved residues in the bottom of the binding pocket, while an aromatic residue sits behind these residues interacting with the apolar regions of the sugar (Maqbool *et al.* 2015). In cluster C, composed of large proteins with extended binding pockets, many were shown to bind to peptides and saccharides with high affinity, but low specificity, being adapted to bind to what is available in the environment. As shown with OppA from *Salmonella enterica* (Maqbool *et al.* 2015), the protein chain interacts with the backbone of the oligopeptide ligand, while the variable side chains interact with water molecules, which dissipate their charge and eliminate the problems which would potentially be caused by different ligand sizes. While this protein binds to peptides between 3-5 residues, other Opp homologs in *Lactococcus lactis* were described to bind to peptides of up to 18 residues; Cluster C SBPs from other organisms, such as *Thermatoga maritima*, are adapted to bind to

oligosaccharides; and MppA protein from *E. coli* was described to bind to fragments of the peptideglycan layer, reinforcing that even inside the same cluster, substrate prediction is very limited and coordination is very variable (Maqbool *et al.* 2015).

As suggested by Vetting *et al.* (2015), the characterization of SBPs can provide rich insights regarding microbial metabolism. The genomic localization of the cognate genes is often just upstream of the metabolic pathway they relate to. Furthermore, they are good subjects for high-throughput screening, allowing testing against vast substrate libraries at once using differential scanning fluorescence assays, for instance. In case no ligand is found, it is possible to retrieve adventitious ligands bound to the proteins in the expression host, through protein crystallization or mass spectrometry. Moreover, it is possible to incubate the protein, just before purification, with the cell lysate of the original organism, thus screening it against the whole metabolome, which can include molecules not commercially available or from unconventional classes.

1.3. TRANSPORTERS INVOLVED IN C4-DICARBOXYLATES

C4-dicarboxylates are essential for microbial metabolism. Under aerobic conditions, succinate, fumarate and malate are metabolised by the TCA cycle, culminating in the generation of reducing power which will be further converted to a proton gradient and ATP through respiration (Unden *et al.* 2016). Under anaerobic conditions, C4-dicarboxylates are used in a variety of fermentative or anaerobic respiratory pathways in different bacteria. In facultative anaerobic bacteria such as *Escherichia coli*, C4-dicarboxylates are converted to fumarate, which acts as an electron acceptor to generate succinate (fumarate respiration), which is then secreted (Unden *et al.* 2016). In many anaerobic bacteria, the proton gradient in the cell comes from the energy conservation upon decarboxylation of dicarboxylic acids. For example, in *Lactococcus*

lactis, L-malate is decarboxylated to lactate (Gänzle 2015), while in *Pseudomonas sp.* fumarate is converted to pyruvate (Ogawa *et al.* 2001). Many environmental bacteria, such as the ones belonging to the genus *Pseudomonas, Rhodopseudomonas and Rhizobium*, are unable to metabolize sugars, and use organic acids as their main carbon source (Poole and Allaway 2000, Larimer *et al.* 2004, Rojo 2010). In the latter, the uptake of C4-dicarboxylates, specially L-malate, is linked to their ability for nitrogen fixation in symbiosis with different legume crops (Mitsch *et al.* 2018).

It is not surprising, then, that bacteria have evolved several different transport mechanisms for C4-dicarboxylates, adapted to the particularities of each metabolism. Rather than using ATP, these transporters evolved to use secondary energy gradients, such as the proton-motive force, Na^+ gradient, or the gradient of the organic acids themselves, in the case of many antiporter systems (Saier *et al.* 2016). The literature reports at least 10 different transport families involved in C4-dicarboxylate transport. By far, the best studied systems are the DctA homologues, belonging to the Dicarboxylate/Amino Acid: Cation Symporter (DAACS) family (TC 2.A.23). These proteins are responsible for most of the C4-dicarboxylate uptake in bacteria under aerobic conditions and neutral pH, where molecules are in di-anionic form and are transported in an electrogenic fashion, and have a broad substrate range and Km values ranging from 2 to 15 µM (Janausch *et al.* 2002, Yurgel and Kahn 2004, Unden *et al.* 2016). While in most bacteria transport in the DAACS family is dependent on an H^+ gradient, in bacteria with a genome of high-GC content such as *Corynobacterium glutamicum* it was shown to be dependent on a Na^+ gradient (Holger *et al.* 1991). DctA homologs are present in most of the bacterial kingdom, but are not found in strictly anaerobic bacteria, and have been extensively characterized in *E. coli* and also in *Rhizobium* species, where they are involved in the signaling cascade which culminate

in nitrogen fixation (Yurgel and Kahn 2004, Mitsch *et al.* 2018). In alder root cells, the C4-dicarboxylates were found to be provided to the bacteria through export by the AgDCAT1 protein, a member of the Proton-dependent Oligopeptide Transporter (POT/PTR) Family (TC 2.A.17) (Jeong *et al.* 2004).

Under aerobic, but acidic environments (pH 5-6), transporters from the Sulfate Permease (SulP) family (TC 2.A.53) are found to be responsible for most of the C4-dicarboxylate uptake. Conserved from bacteria to humans, these transporters uptake substrates with a mono-anionic charge, also in an electrogenic, H^+ dependent manner. The DauA protein, from *E. coli*, is the best characterized protein from this family (Unden *et al.* 2016). Under anaerobic conditions, on the other hand, proteins from the DcuAB family (TC 2.A.13), such as DcuA and DcuB from *E. coli*, catalyze the electroneutral antiport of fumarate and succinate to allow for fumarate respiration, but are also capable of electrogenic uptake and efflux and allow for some substrate flexibility (Janausch *et al.* 2002). The average Km in this family was estimated to be 100 µM (Unden *et al.* 2016). Proteins from the DcuC family (TC 2.A.61) were found to have very similar function to the DcuAB family and similar affinities, but are mostly characterized as succinate exporters under conditions of hexose fermentation (Janausch *et al.* 2002, Unden *et al.* 2016).

Also conserved from bacteria to humans, proteins belonging to the Divalent Anion:Sodium Symporter (DASS) family (TC 2.A.47) are involved in the uptake of citrate and di-carboxylic acids (Saier *et al.* 2016). Previously characterized as citrate transporters in chloroplasts and mitochondria, members of this family were shown to be the major uptake system for C4-dicarboxylates in environmental bacteria, such as the DccT protein from *C. glutamicum* (Teramoto *et al.* 2008, Youn *et al.* 2008). The best studied systems are the C4-dicarboxylate Na^+ symporter VcINDY (*Vibrio cholera*)

(Mulligan *et al.* 2014, Mulligan *et al.* 2016) and the TtdT Tartrate-succinate anaerobic antiporter from *E. coli* (Unden *et al.* 2016). Members of this family can have very diverse substrate affinities, where K_m values range from 2 µM to 300 µM for C4-dicarboxylates (Mulligan *et al.* 2016).

Other protein families also associated with the transport of C4-dicarboxylates have either very low substrate affinity and reduced physiological relevance, or are associated with very specific processes and not found represented in many bacteria. SatP, from *E. coli*, is a member of the Acetate uptake transporter (AceTr) family (TC 2.A.96), and catalyzes the symport of either succinate or acetate with H^+, under acidic conditions, with a characterized K_m of 1.8 mM (Unden *et al.* 2016). In some bacteria, symporter L-malate/ H^+ uptake is performed by members of the 2HCT family (TC 2.A.24), such as the MaeN (*Bacillus subtillis*) and MaeP (*Streptococcus bovis*) symporters (Krom *et al.* 2003). In addition, the MleP malate/lactate electroneutral antiporter is crucial for malolactic fermentation in *L. lactis*, although this antiporter activity is also observed in *B. subtillis* as the product of MleN activity, which belongs to the NhaC family (TC 2.A.35) (Krom *et al.* 2003).

While the digestive tract inhabited by *E. coli* and the nitrogen fixation nodules inhabited by *Rhizobium sp.* are enriched in C4-dicarboxylic acids (Unden *et al.* 2016, Mitsch *et al.* 2018), other environments such as soil and wastewaters may offer, instead, a broader variety of substrates present in very low concentrations. In order to successfully uptake these substrates, bacteria evolved to have high-affinity, Soluble Binding Protein (SBP) dependent transporters, involving periplasmic binding proteins which scavenge nutrients and deliver them to their respective membrane counterparts (Rosa *et al.* 2018a). They can be of three different families: the primary ATP-Binding Cassete superfamily (ABC, TC 3.A.1), not found to date to transport C4-dicarboxylic

acids; and the secondary Tripartite ATP-Independent Transporters (TRAP, TC 2.A.56) and Tripartite Tricarboxylate Transporters (TTT, TC 2.A.80). The TRAP prototype system, the DctPQM system of *Rhodobacter capsulatus*, was characterized as a C4-dicarboxylate uptake system, binding to its substrates with K_m values ranging from 50 nM for L-malate to 250 nM for fumarate, and 6.3 µM for D-malate (Kelly and Thomas 2001). DctPQM homologues were further described in several organisms, participating in the uptake of C4-dicarboxylates in many of them. The list of substrates taken up by TRAP systems have expanded considerably beyond C4-dicarboxylates since their first discovery (Vetting *et al.* 2015), as has knowledge regarding the biochemical properties of this family (Mulligan *et al.* 2009, Rosa *et al.* 2018a).

CHAPTER 2. *RHODOPSEUDOMONAS PALUSTRIS*

Rhodopseudomonas palustris is a soil non-sulfur purple bacterium belonging to the α-proteobacteria class, bearer of a robust and versatile metabolism, which allows it to grow under a wide range of environmental conditions. In absence of oxygen, but in the presence of light, it can grow either as a photoautotroph, reducing available CO_2 from the environment, or as a photoheterotroph, using light for ATP generation, but organic compounds as carbon sources. In the presence of oxygen, it can either behave as a chemoautotroph generating energy through inorganic oxidation, or have a chemoheterotroph metabolism, oxidizing organic compounds (Larimer *et al.* 2004, VerBerkmoes *et al.* 2006, Pechter *et al.* 2016).

Due to its versatile metabolism, *R. palustris* has been used as a model organism for the study of several biotechnological potential applications, among which two are noteworthy: The production of hydrogen gas as a by-product of nitrogen fixation, and the catabolism and high-resistance to aromatic compounds. Hydrogen gas has the second highest energy content per mass unit, only behind nuclear power, and can be generated from renewable sources in a sustainable manner by bacteria (McKinlay and Harwood 2010). *R. palustris* contains three independently expressed nitrogenases (Oda *et al.* 2005), which can be regulated to generate hydrogen gas nearly at the maximum theoretical rate (McKinlay and Harwood 2010).

Aromatic compounds represent an industrial challenge but also an opportunity in biotechnology. p-coumaroyl ethers, the major constituents of lignin, represent a great potential for the production of second generation biofuels from residual agriculture material (Austin *et al.* 2015), while polychlorinated biphenyls, polyaromatic

hydrocarbons, benzene and toluene generated as industrial waste accumulate and contaminate soils (Crosby *et al.* 2010, Venkidusamy and Megharaj 2016). While most bacteria cannot cope with more than residual concentrations of aromatic compounds, *R. palustris* has a high capacity of metabolizing them (Harwood and Gibson 1988), and was isolated from several contaminated soil and wastewater environments (Mutharasaiah *et al.* 2012, Venkidusamy and Megharaj 2016, Ottoni *et al.* 2017). This metabolic property has been extensively characterised, where the anaerobic benzoyl-CoA pathway accounts for most of aromatic metabolism in this organism (Pan *et al.* 2008, Porter and Young 2014), although the aerobic protocatechuate/ 3-oxo-adipate pathway is also involved (Harwood and Parales 1996, Kamimura and Masai 2014) (See also section 7.5.1).

In agreement with its highly versatile metabolic capabilities, *R. palustris*' genome contains a wide repertoire of membrane transporters, listed and categorised by Larimer *et al.* (2004). While most Enterobacteria contains around 6% of the genome encoding to transport-related protein, *R. palustris* contains over 15% (Larimer *et al.* 2004, VerBerkmoes *et al.* 2006), and although metabolism is dependent on transport, the biochemical properties and mechanisms of solute uptake capabilities in *R. palustris* have not been extensively characterised.

AIMS OF THIS STUDY

Unlike ABC systems, and to some extent TRAP transporters, the true diversity and substrate range for the poorly characterised TTT class of transport systems is mostly undiscovered. A comprehensive study of TTT transporters in bacteria of biotechnological relevance and versatile metabolism might expand the repertoire of transporters available for industrial processes. This study aims to study the distribution of the TTT transport family among bacteria and then provide a bioinformatics, biochemical, physiological and structural characterisation of the TTT family in the model organism *Rhodopseudomonas palustris*.

CHAPTER 3. METHODOLOGY

3.1. GENETIC ANALYSIS

The genome of *Rhodopseudomonas palustris* strain CGA009 is available online at the GeneBank under the assembly number GCA_000195775.1 . Neighbourhood analysis was done using Artemis software (Rutherford *et al.* 2000). Sequence comparisons were done using BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and operon similarity analysis was performed using the software MultiBLASTGene (Medema *et al.* 2013) .

3.2. STRAINS AND MEDIA

Rhodopseudomonas palustris strains used in this study are described in Table 3-1, while *Escherichia coli* strains are described in Table 3.2

Table 3-1: *Rhodopseudomonas* strains used in this study

Rhodopseudomonas palustris	**Description**	**Source**
CGA009	Environmental isolate of genome sequenced strain	Prof. D.K Newman, California Institute of Technology
Δ*tttBA1*	Strain of CGA009 with an unmarked deletion of the genes rpa2320/21	This study
Δ*tttBA1/matBA*	Strain of CGA009 with unmarked deletions of the genes rpa2320/21 and rpa3495/96	This study

Table 3-2: *Escherichia coli* strains used in this study

Escherichia coli	Description	Source
DH5α™	Cloning strain. F– *Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*	Promega
BL21 (DE3)	Protein expression strain. F– *ompT hsdSB(rB–, mB–) gal dcm* (DE3)	Promega
BL21 Star™ (DE3)	Protein expression strain. As BL21 (DE3) *rne131*	Promega
TOP10	Protein expression strain. *hsdR mcrA lacZΔM15*	Invitrogen
S-17	Conjugative strain. RP4-2 (Tc::Mu, Mn:Tn7) integrated into the chromosome *thi pro hsdR hsdM+ recA* Tp^R (Sm^R)	Simon *et al.* (1983)

R. palustris was routinely grown in liquid Peptone Yeast Extract (PYE) media supplemented with succinate: 5 g/L Peptone (Becton/Dickinson), 5 g/L Yeast Extract (Oxoid), 5 g/L sodium succinate (Oxoid). For growth on solid culture Agar, 1.3% Agar was added to the media prior to autoclaving. RCV, as described by Beatty et al. (1981) was used as minimal media, using succinate as the sole carbon source unless stated otherwise. 1.3% agar was used for solid media. Liquid growth was performed anaerobically under light and growth on plates was performed aerobically in the dark, both at 30 °C. When used, antibiotic concentrations for *R. palustris* cultures were 100 µg/ml Kanamycin, 20 µg/ml Chloramphenicol, 100 µg/ml Gentamycin.

Escherichia coli were cultured in Luria Bertani (LB) broth. When used, antibiotic concentrations for *E. coli* were 50 µg/ml carbenicilin; 50 µg/ml kanamycin; 20 µg/ml gentamycin

3.3. GROWTH CURVES

15 ml of *R. palustris* culture was grown anaerobically in PYE until an OD_{660} of 0.6. The culture was centrifuged at 4500 rpm (16000 x g) for 10 minutes and washed once with Phosphate Buffer 20mM pH 7.4. Growth experiments proceeded in sterile 8ml test-tubes, filled with RCV supplemented with 3 mM final concentration (unless different

concentration is attested) of the subject carbon source and a small bead to homogenize the sample. The initial OD_{660} was adjusted to 0.1. Each carbon source was tested in triplicate; media without carbon source were used as a negative control and media plus succinate were used as a positive control. The cultures were incubated anaerobically at 30 °C, with two 9 W hot white LED lamps pointed at them at a 10 cm distance. OD_{660} measurements were taken twice a day, first in the morning and late in the afternoon, right after homogenizing the samples by turning them gently upside-down for 5 times.

3.4. PHENOTYPE MICROARRAYS FOR MICROBIAL CELLS

R. palustris ability to grow in a range of different carbon sources was assessed using Biolog™ Phenotype Microarray for microbial cells, with carbon source plates PM1 and PM2, prepared accordingly to manufacturer's instructions. Plates were incubated at 30 °C and absorbance was measured at 600 nm using a 96 well plate reader after 24, 48 and 72 hours of incubation. A_{660} higher than 0.110 was considerate a positive growth. Experiments were performed in duplicate. To compare growth of different strains, the absorbance of the mutant strains was subtracted from the wild type absorbance. Subtractions above the value of 0.1 were considered relevant.

3.5. OVEREXPRESSION OF RECOMBINANT PROTEINS OVERVIEW

Overexpression of recombinant proteins was initially designed using pET-21a(+) plasmid (Novagen), shown in Figure 3-1. This plasmid allows expression under the strong inducible promoter of phage T7, active when in presence of lactose or the analogous Isopropyl β-D-1-thiogalactopyranoside (IPTG). Downstream of this promoter is the poly-cloning site, connected to a sequence encoding a 6 Histidine-Tag, before reaching a stop codon. This poly-histidine tag in the C-terminal of the

protein allows further purification through affinity chromatography, using a nickel column, as described in the following sections. The pET21a plasmid contains also an origin of replication in *E. coli* (ori) and a selective marker that confers resistance to β-lactams, such as ampicillin/carbenicillin.

To achieve overexpression of TTT SBPs, the periplasmic localization signal sequence was excluded in primer design using SignalP prediction, and genes were amplified, digested, ligated with the vector, transformed in competent DH5α *E.coli* cells. The resulting plasmid was transformed into competent *E. coli* BL21(DE3) (Novagen), which contains the lysogenic DE3 phage encoding the IPTG inducible T7 polymerase gene, allowing for controlled high level target gene expression from the vector T7 promoter. The expression conditions were optimized and then expression and purification followed. All these steps are described in the next sections in detail. When expression was not good enough in BL21, the alternative *E. coli* strain Rosetta PlysS (Novagen) was used. This strain is derived from the DE3 *E. coli* strain, containing a plasmid encoding genes for alternative tRNAs, and also a deletion in the lactose permease gene. The alternative tRNA can be useful when cloning genes from organisms that use a different set of preferential codons. The lactose permease deletion makes the transcription activation slower, which sometimes improves protein solubility. When the proteins could not be cloned/expressed in pET21, recombinant expression was tried using plasmid pBAD/HisB (Invitrogen) (Figure 3-1), which uses a tunable promoter responsive to L-arabinose, using Top10 *E.coli* strain as a host, which cannot metabolize L-arabinose. Moreover, in this vector the His_{6x} tag is on the N-terminal of the protein.

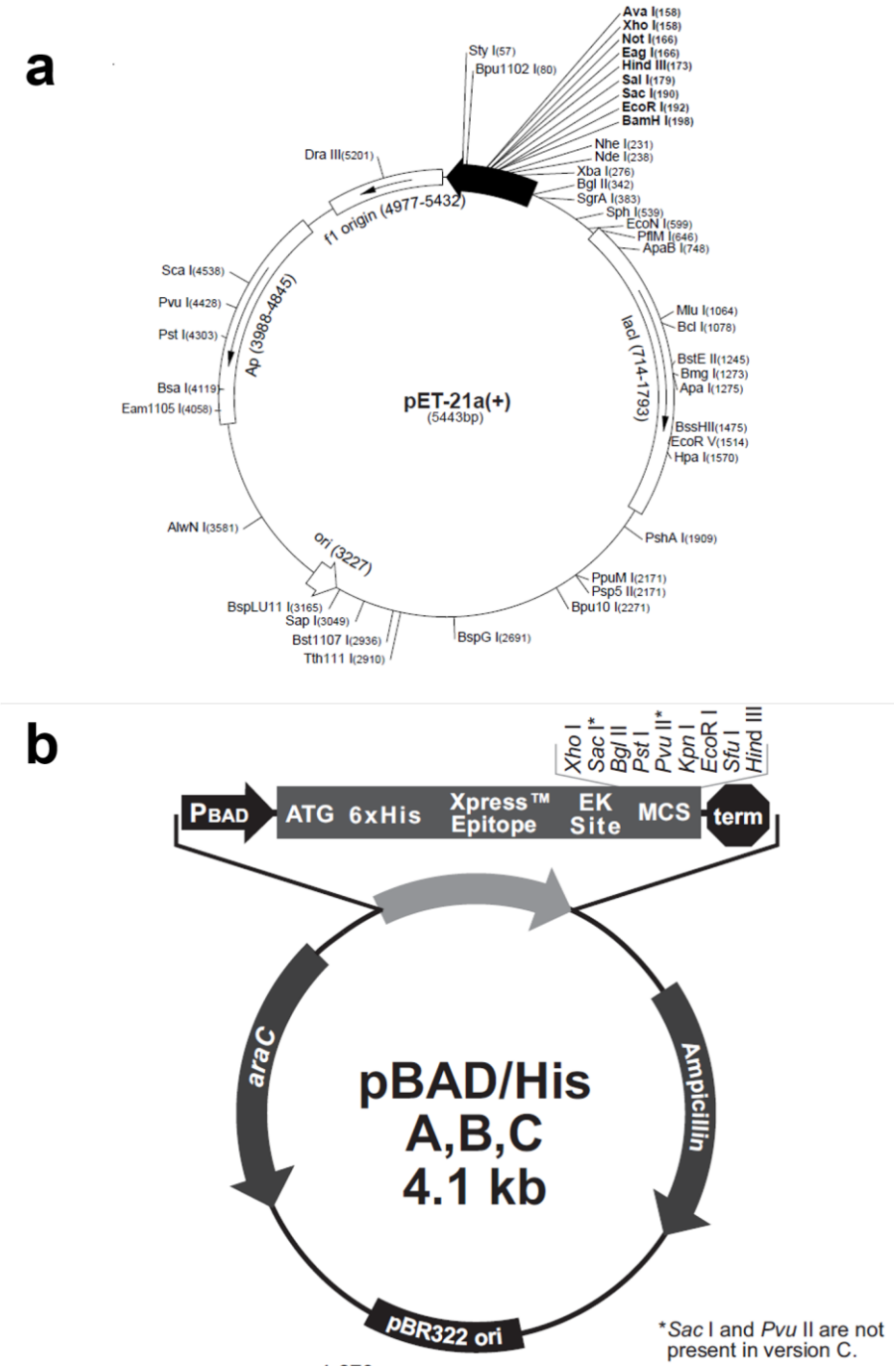


Figure 3-1: Vectors used for overexpression of proteins. (a) pET21a(+) vector (Novagen), contains an β-lactamase gene (Ap), confering resistance to carbenicillin and ampicillin; a polycloning site under the regulation of the T7 promoter, and upstream a 6-histidine coding sequence, which will be added to the C-terminal of the protein to allow affinity chromatography in Ni colloum; a LacI repressor gene, which will block T7 expression in absence of lactose or IPTG. Figure extracted from pET21a-d(+) vectors manual (Novagen – TB036 document) **(b) pBAD/HisB vector (Invitrogen)**, used to overexpress recombinant proteins when pET21a showed limitations. It contains a β-lactamase gene, conferring resistance to Ampicillin and Carbenecillin; a polycloning site under the regulation of the araBAD promoter from *E.coli* and downstream of a 6xHistidine-tag and an antiexpress epitope for antibody recognition; araC repressor, to avoid expression in the absence of L-arabinose. Figure extracted from pBAD/HisA,B and C manual (Invitrogen, Catalog nos. V430-01, V440-01, version J)

For overexpression of the membrane transporter components MatBA, adapted versions of pET and pBAD vectors were provided by Dr. Chris Mulligan (Kent University), and named here pET-TM and pBAD-TM vectors. These vectors contain a N-terminal His_{10x} tag with a kanamycin resistance gene (pET-TM) and a C-terminal His_{8x} tag alongside with an ampicillin resistance gene (pBAD-TM).

3.6. DIRECTED MUTAGENESIS OVERVIEW

Deletion of target genes in *R. palustris* was performed using the suicide vector pK18mobsacB (Schäfer *et al.* 1994), whose map is shown in Figure 3-2a. It contains a levansucrase gene from *Bacillus subtilis (sacB)*. The levansucrase enzyme will convert sucrose into levans, toxic for the bacterial cell. Consequently, cells which have this plasmid incorporated in the genome will be unable to grow in the presence of sucrose (Pelicic *et al.* 1996, Wang *et al.* 2015). Cassettes containing flanking regions of 500bp upstream and downstream of the target gene to be deleted, were constructed. These constructions involved primer design, PCR of flanking fragments, overlapping PCR to join the upstream and downstream regions, digestion with restriction enzymes, ligation inside the vector, transformation into competent DH5α and miniprep plasmid preparation. The vectors containing the deletion cassettes were then introduced into *R. palustris* for homologous recombination, by electroporation or conjugation with *E.coli* SK-17 strain, generating marker-free deletion mutants. When this vector did not show the expected results, another method of producing deletion mutants was performed using the vector pJQ200ks (Quandt and Hynes 1993), which although working in exactly the same way, was more often used for this approach in the literature (Rey *et al.* 2007, Kulkarni *et al.* 2013). All the cited methods are described in the following sections.

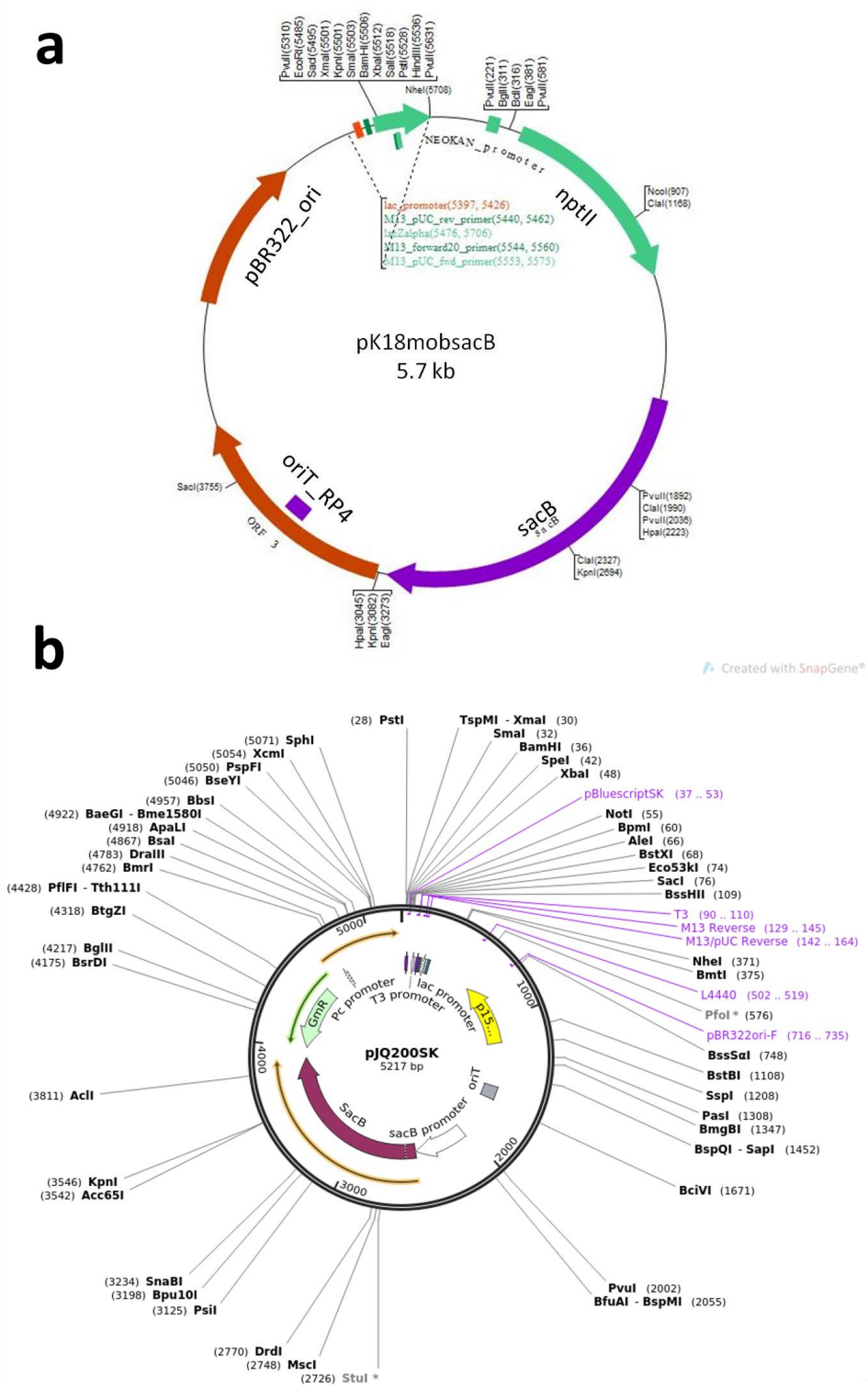


Figure 3-2: Suicide vectors used in this study. (a) pK18mobsacB, contains an orign of replication (pBR322_ori); origin of conjugation (oriT_RP4); neomycin phosphotranferase II gene, conferring resistance to a range of aminoglycoside antibiotics, including kanamycin (nptII); levansucrase gene from *Bacillus subtilis*, which polymerizes sucrose molecules in a toxic polymer, inviabilizing growth in sucrose containing media; polyclonning region. The plasmid was described by Schäfer *et al.* (1994) and the figure was obtained from BVTech Plasmid website (http://www.biovisualtech.com/bvplasmid/aboutus.htm). **(b) pJQ200ks+**, Contains an orign of replication for *E. coli* (oriV); origin of conjugation (oriT); Gentamycin resistance gene; levansucrase gene from *Bacillus subtilis (sacB)*. The plasmid was described by Quandt and Hynes (1993) and the figure was obtained from Addgene website (https://www.addgene.org/).

3.7. GLYCEROL STOCKS

E. coli or *R. palustris* cultures were grown overnight in 5 ml of LB or PYE, respectively, containing antibiotic when appropriate, centrifuged 5000 rpm (3400 x g) for 10 minutes, ressuspended in 1 ml of LB or PYE media containing 50% glycerol, and stored at -80 °C

3.8. DNA AND PLASMID EXTRACTION

R. palustris genomic DNA was obtained using Wizard Genomic DNA extraction kit (promega). Plasmid DNA was obtained using QIAprep® Spin Miniprep Kit (Qiagen)

3.9. PCR

All primers were obtained from Sigma, using Tm °C calculated by Sigma Oligo DNA order system software. The primer sequences, with respective restriction site designed and the annealing temperatures used are shown in Table 3-3 (protein overexpression) and Table 3-4 (Deletion cassette generation).

Amplification reactions were performed in a 20 µl volume with Phusion High-Fidelity DNA polymerase Master Mix (Thermo Scientific) according to manufacturer instructions, using a final concentration of 1 ng/µl template DNA. Construction of directed mutagenesis cassettes were done through overlapping PCR, as described by (Higuchi *et al.* 1988). Screening reactions were performed in a 10 µl volume using MyTaq Mix (Bioline) according to manufacturer instructions.

All reactions were made in a Techne TechGene Thermal cycler (Techne), using a final concentration of 0.2 µM for each primer and further verification was done on Agarose gels. Amplified fragments were purified and concentrated using Gen-Elute^{tm} PCR clean-up kit (sigma).

Table 3-3: Primers used in this study for overexpression of recombinant proteins, with the plasmids used for cloning and resulting plasmids in each case

Gene Number/Name		Sequence	Restriction site	Tm °C	Resulting plasmid / Vector Used
RPA 2319	Forward	5`-ACT CATATG CCGGCTGATTGGCCG-3`	*Nde*I	64	pET21a-rpa2319
tttC1	Reverse	5`-ATTA CTCGAG ATCGAGCTTGGCGCC-3`	*Xho*I		pEt21a
RPA3494	Forward	5`-ATA GCTAGC GAGCCGGACCGGTT-3`	*Nhe*I	70	pET21a-malC
matC	Reverse	5`-ATAT CTCGAG GGGCGTTGTGCCCT-3`	*Xho*I		pEt21a
RPA0686	Forward	5`- ATA GAGCTC A CCCGACTGGCCGGCT -3`	*Sac*I	67	pBAD-rpa0686
tttC3	Reverse	5`- TTA AAGCTT TCAATTCTCTCTGAGGCCGC -3`	*Hind*III		pBAD/HisB
RPA3100	Forward	5`- ATA GAGCTC AATCAGCAGCGCGGC -3`	*Sac*I	63	pBAD-rpa3100
tttC4	Reverse	5`- TTA AAGCTT CTATTGCACCTTGATGCCC -3`	*Hind*III		pBAD/HisB
RPA4515	Forward	5`- ATA GAGCTC A GACTGGCCGACCCGG -3`	*Sac*I	67	pBAD-rpa4515
adpC	Reverse	5`- TTA AAGCTT CTACAGCGGCGCAATTCC -3`	*Hind*III		pBAD/HisB
RPA4580	Forward	5`- ATA GAGCTC AGACGCGGCGGCGTA -3`	*Sac*I	66	pBAD-rpa4580
tttC6	Reverse	5`- TTA AAGCTT TCAGCCCTCCTTGTGAACG-3`	*Hind*III		pBAD/HisB
RPA4694	Forward	5`-AAA CATATG GGCCTGTATCCGACCCG-3`	*Nde*I	66	pEt21a-rpa4694
tttC7	Reverse	5`ATAA CTCGAG TTGCCGCTCGATGCC-3`	*Xho*I		pEt21a
RPA2318/17	Forward	5`- ATT GAGCTC A ATGGCGCAGACAAGGC -3`	*Sac*I	62	pBAD-gtcAB
gtcAB	Reverse	5`- ATA AAGCTT CTAGTGCTTGCGTAGCGC - 3`	*Hind*III		pBAD/HisB
RPA3495/96	Forward	5`- ATACCATGGAAATGGACGAGGCGCCG -3`	*Nco*I	67	pBAD-TM-rpa3495/96
matBA	Reverse	5`- TTAGGTACCCCCCTCGGACTCGACGA - 3`	*Kpn*I		pBAD-TM
RPA3495/96	Forward	5`- AAT GGCGCGCC ATGGACGAGGCGCCG -3`	*Asc*I	67	pET-TM-rpa3495/96
matBA	Reverse	5`- TAA CTCGAG TTACCCCTCGGACTCGACG - 3`	*Xho*I		pET-TM

Table 3-4: Primers used for the construction of deletion cassettes in *R. palustris*

Gene		Primer sequence	Restriction Enzyme	Tm °C
RPA2321/20	Upstream Forward	5`- GCGTCTAGAACCAGACCACCG-3`	*Xba*I	66
tttBA1	Upstream Reverse	5`- AAGGCGAC CCTGCGCCGAACTGACAT-3`		
	Downstream Forward	5`- CGGCGCAGG GTCGCCTTCCACGAGGAC-3`		
	Downstream Reverse	5`- ATA GGATCC TAAGGCACTTCCGCCACC -3`	*Bam*HI	
	Deletion check Upstream	5`- TTCTTCTCTCTTGTTGCACGC-3`		64
	Deletion check downstream	5`- CGCTGAGATGGCACAGC-3`		
RPA3496/95	Upstream Forward	5`- CCATA GGATTC TCAATCACTGGCTGACCGAG -3`	*Bam*HI	66
matBA	Upstream Reverse	5`- CTCGGACTC CGCCTCGTCCATCGGT -3`		
	Downstream Forward	5`- GACGAGGCG GAGTCCGAGGGGTAATTCCAGC -3`		
	Downstream Reverse	5`-CCCATA TCTAGA TGAAGTCCTTGTCAGGCCGAT -3`	*Xba*I	
	Deletion check Upstream	5`- ACCGGGGCACCTTCG-3`		65
	Deletion check downstream	5`- CGTTGAGGTTGTCGGTGTG-3`		

3.10. AGAROSE GEL ELECTROPHORESIS

DNA analyses were routinely done in agarose gels, 0.7% (w/v) in TAE Buffer (40 mM Tris-Acetate, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0), dissolved by heating. Once the gel cooled down, ethidium bromide solution was mixed at 200 ng/ml final concentration. The gel was poured into a gel cast with a gel comb. After the gel had set, 4 µl DNA samples were mixed to 1 µl of 5x loading dye (Bioline) and added to a well. 5 µl of HyperLadderI (Bioline) was added in the first well. Electrophoresis was done in TAE Buffer at 120 V for 35 minutes or until enough band separation. Visualizations of resolved gels were done under UV light, in a Gene Flash Gel Doc System (Syngene).

3.11. EXTRACTION OF DNA FROM AGAROSE GEL

When required, DNA was extracted from agarose gels using the QIAquick® Gel Extraction Kit (Qiagen).

3.12. RESTRICTION DIGESTION

DNA was digested using NEB restriction enzymes (New England Biolabs), following manufacturer instructions for optimal digestion. Reactions were carried out in a 20 µl volume for 1,5 h, and then inactivated at 65 °C for 20 min. DNA was purified using Gen-Elute^{tm} PCR clean-up kit (sigma).

3.13. DNA LIGATION

Ligations of digested fragments were performed using T4 DNA Ligase (New England Biolabs) according to manufacturer instructions. Reactions were carried in a 20 µl final volume for 1 hour at room temperature. 50 ng of vector was used, and the amount of insert to obtain the required 3:1 insert:vector molar ratio was calculated using the formula below. A control with no insert was currently performed to calculate the religation rate.

(ng of vector x kb size of insert x molar ratio of insert:vector)/ (kb size of vector) = ng of insert

3.14. PREPARATION OF *E. COLI* HEAT-SHOCK COMPETENT CELLS

E. coli competent cells were prepared according to (Hanahan 1983). *E. coli* was grown in 100 ml of LB media, up to an OD_{600} of 0.6, corresponding to ~ 4-7 x 10^7 viable cells ml^{-1}. Cells were incubated on ice for 15 min, then centrifuged 10000 rpm (13000 x g) for 10 min at 4 °C. Supernatant was discarded and cells were resuspended in 50 ml of cold RF1 solution (100 mM KCl, 50 mM $MnCl_2*4H_2O$, 30 mM CH_3COOK, 10 mM $CaCl_{2*}2H_2O$, 15 % (w/v) glycerol; adjusted to pH 5.8 with 0.2 M acetic acid). After 15 min on ice, cells were centrifuged as previously. Supernatant was discarded and cells were ressuspended in 8 ml of cold RF2 solution (10 mM MOPS, 10 mM KCl, 75 mM $CaCl_2*2H_2O$, 15% [w/v] glycerol; adjusted to pH 6.8 with 5 M NaOH). After 30 min more on ice, cells were aliquoted in 400 µl samples and stored at -80 °C.

3.15. HEAT-SHOCK TRANSFORMATION IN *E. COLI*

Cells were thawed on ice and aliquoted to 200 µl samples and 2 µl of ligation reaction was added. Cells were incubated 30 min on ice. Heat-shock was performed transferring to 42 °C bath for one minute, and transferring back to ice for 2 min. 1 ml of LB media was added and the culture was incubated at 37 °C for 1 h to cell recovery. Cells were harvested at 13000 rpm (17000 x g) for 5 min and supernatant was discarded. The pellet was resuspended in 100 µl of LB media, spread in LB media containing the appropriated antibiotic and incubated at 37 °C for 16 h.

3.16. DNA SEQUENCING

Samples were sequenced using the "LIGHTRUN™ sequencing service", provided by "GATC Biotech", according to instructions in the company website (https://www.gatc-

biotech.com/en/index.html). Each sample was sequenced using the forward and reverse primers provided in Table 3-3.

3.17. SDS-PAGE

Protein analysis was performed by SDS-PAGE gel. 10 µl of sample was added to 10 µl of the loading buffer (25 µl β-mercaptoethanol added to 475 µl of 2x protein loading buffer: 125 mM Tris-HCl pH 6.8, 4 % [w/v] SDS, 0.05 % [w/v] bromophenol blue, 20 % [w/v] glycerol). Samples were then boiled for 5 minutes at 100^{o}C and loaded to the gel (membrane proteins samples were not boiled). The SDS-PAGE gel consisted of a 12% resolving gel and a 6% stacking gel. The resolving gel was composed as follows: 2.5 ml of 30% acrylamide, 2.35 ml 1 M Tris HCl (pH8.8), 1.25 ml Milli-Q water, 62.25 ml 10% SDS, 6.25 µl N,N,N',N'-Tetramethylethylenediamine (TEMED) and 62.5 µl ammonium persulphate (APS). Once gently mixed by pipetting, the mixture was added to a gel cast and a thin layer of water saturated butanol was poured on top. Once the resolving gel has set the butanol was blotted off using blot paper and the stacking gel was added which was composed of the following: 750 µl of 30% acrylamide, 470 µl 1 M Tris HCL (pH8.8), 2.46 ml MQ water, 37.5 ml 10% SDS, 3.75 µl TEMED and 37.5 µl APS. Once poured a gel comb was added and left to set. Once set the gel comb was removed and the gel placed in the protein 3 electrophoresis apparatus (Biorad). The tank was filled to the appropriate level as marked with 1 x SDS-PAGE running buffer (0.1 M Tris base, 0.38 M glycine, 0.1 % [w/v] SDS). Samples were then added in appropriate volumes along with 10µl EZ-Run™ Prestained *Rec* Protein Ladder (Fisher) molecular weight standard. The gel was run at 180 V for 70 minutes.

Gels were stained by coomassie blue staining solution (0.02% brilliant coomassie blue, 50% methanol, 40% H_2O, 10% acetic acid) and left overnight shaking gently. Stain was then poured off, residual stain being removed by dH_2O and gels were then

left for an extended period in destain solution consisting of 50% methanol, 40% H_2O 10% Acetic acid

3.18. OVEREXPRESSION OPTIMIZATION OF SOLUBLE PROTEINS

Using pET21a

Overexpression optimization was done growing cells in 4x100 ml of LB+carbenicilin 50 µg/ml, in 250 ml flasks, at 250 rpm, up to OD 0.6. Two flasks were induced with IPTG 1 mM, and two remained as a control. The two pairs control/experiment were incubated in 25 °C and 37 °C. 1 mL and 9 mL samples were taken at times 0; 1, 3, 5 and 24 hours. 1 ml samples were centrifuged 13000 rpm (16000 x g) 5 min, and resuspended in 50 µl of TRIS Buffer pH 6.8. 10 µl of these samples were mixed with 10 µl of 2x loading dye and boiled for 5 min. After that, they were centrifuged 5 min 13000 rpm (16000 x g), and 10 µl of the supernatant was injected in the SDS-PAGE gel. To check the protein solubility, the 9 ml samples were centrifuged 4500 rpm (3400 x g) 10 min and lysed by sonication (Soniprep 150: MSE ; 16 µM amplitude sound waves ; 4x 20 s, to not overheat the sample) and centrifuged (13000 rpm/ 40000 x g 30 min) 10 µl of the supernatant were mixed with 10 µl of loading dye and boiled for 5 min. Pellet was ressuspended in 50 µl of water and had the same treatment.

Using pBAD/HisB

Overexpression optimization was done growing cells in 4x100 ml of LB+carbenicilin 50 µg/ml, in 250 ml flasks, at 250 rpm up to OD_{600} 0.6. Two flasks were induced with different final concentrations of L-arabinose (0%, 0.2%, 0.02% and 0.002%) and incubated at 37 °C. Following steps were the same as for pET21a.

Table 3-5: Inducing conditions for overexpression of recombinant proteins used in this study

Plasmid	Expression strain	Inducer concentration	Induction conditions
pET21a-*tttC1*	BL21 (DE3)	1 mM IPTG	1h 37^{o}C
pET21a-*matC*	BL21 (DE3)	1 mM IPTG	3h 37^{o}C
pBAD-*tttC3*	Top10	0.02% L-arabinose	5h 37^{o}C
pBAD-*tttC4*	Top10	0.002% L-arabinose	5h 25^{o}C
pBAD-*AdpC*	Top10	0.2% L-Arabinose	22h 37^{o}C
pBAD-*tttC6*	Top10	0.002% L-arabinose	22h 37^{o}C
pEt21a-*tttC7*	BL21 (DE3)	1 mM IPTG	3h 30^{o}C
pBAD-*gtcAB*	Top10	0.02% L-arabinose	22h 30^{o}C
pBAD-TM-*matBA*	Top10	0.02% L-arabinose	22h 37^{o}C

3.19. SOLUBLE PROTEIN PURIFICATION

For purification using the pET21a system, *E. coli* strains were grown in 2 x 1 L of LB media + carbenicilin 50 µg/ml to OD_{600} 0.6 in 2L flasks at 250 rpm and then induced with IPTG 1 mM for the optimized time (different for each construct). For the pBAD/HisB system, instead of IPTG the optimized concentration of L-arabinose was added. The culture was centrifuged at 10000 rpm (16000 x g) for ten minutes, and the supernatant was discarded. Cells were resuspended in 10 ml of binding buffer (20 mM sodium phosphate buffer, 0.5 M NaCl and 20 mM Imidazole made up to pH7.4) and lysed by sonication (Soniprep 150: MSE ; 16 µM amplitude sound waves ; 5 x 20 s, to not overheat the sample) and centrifuged (18000 rpm/ 40000 x g 30 min). The cell free extract (supernatant) was filtered in 0.45 µm pores.

Purification was performed by affinity chromatography. The cell free extract was passed through a HisTrap HP Nickel affinity column (GE Healthcare) via an AKTA design chromatography system (GE Healthcare) according to the manufacturer's instructions using column binding and elution buffer (20 mM sodium phosphate buffer, 0.5 M NaCl and 500 mM Imidazole made up to pH7.4.). UV absorbance determined the fraction which contained the highest amount of purified protein and purity was

checked by SDS-PAGE gel. Elution was performed in a total volume of 50 ml, with 3 ml fractions, in an elution buffer gradient from 0% to 100%.

3.20. PROTEIN DIALYSIS

Desalting of proteins was performed by transferring the protein solution to an 8 kDa molecular-weight cut-off dialysis bag, and dialysing it against 2 L of cold dialysing buffer (50 mM TRIS-HCl Buffer, pH 7.2; 100 mM NaCl) for 2 hours, and then dialysing again against 2 L of cold dialysing buffer for 16 hours.

3.21. UNFOLDING AND REFOLDING OF SBP`S WITH UREA

Protein solution was transferred to 8 kDa molecular-weight cut-off dialysis bag and dialyzed against 1 L of cold unfolding buffer (50 mM TRIS-HCl Buffer, pH 7.2; 100 mM NaCl; Urea 6 M) for 16 hours. Refolding was performed by dialyzing the protein solution against 2 L of cold dialysing buffer (50 mM TRIS-HCl Buffer, pH 7.2; 100 mM NaCl) for 8 hours, changing the buffer every 2 hours, with a final overnight dialysis. Protein was transferred to a Falcon tube and centrifuged for 15500 x g for 10 minutes at 4 °C to remove precipitated misfolded protein.

3.22. PROTEIN EXPOSURE TO *R. PALUSTRIS* CELL-FREE EXTRACT

In order to increase the chances that the solute binding protein would bind to its cognate ligand, protein was exposed to a complex mixture of compounds during purification. Overproduction was performed as described previously until immobilization of protein in the Ni^{2+} column. In parallel, 1 L of *R. palustris* was grown in PYE media to OD_{660} 1.0 and a cell free extract prepared as described above. 0.1 g of a mixture of humic acids (Sigma) was added to the *R. palustris* cell free extract, the solution was filtered through a 0.45 µm pore filter, and this mixture was passed through

the nickel column containing the bound protein. 40 ml of binding buffer was then passed to wash the column and protein purification followed as described above.

3.23. PROTEIN CONCENTRATION DETERMINATION

Protein concentration was determined by measuring the UV absorbance of the purified protein solution using a UV-Vis spectrophotometer UV2401 (SHIMADZU) and the following formula, where A is the light absorbance at specified wavelength and ε is the molar extinction coefficient of the protein:

$$\text{Protein concentration (M)} = \frac{A_{280} - A_{320}}{\varepsilon\ (M^{-1}\ cm^{-1})}$$

3.24. SIZE EXCLUSION CHROMATOGRAPHY

Size exclusion chromatography of Rpa4515 was performed using a 24 ml Superdex200 increase column at 0.5 ml/min flow rate in an AKTA instrument, using TF Buffer (Tris-HCl 50 mM; NaCl 0.1 M pH7.4). A calibration curve was generated using thyroglobulin (669 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and aprotinin (6.5 kDa) and k_{av} was calculated using the formula K_{av}=(Ve-Vo)/(Vt-Vo), where Ve is the elution volume for a given protein, Vo is the void volume and Vt is the total volume of the column.

Size exclusion chromatography of Rpa3494 was performed using a 24 ml 1.6 x 60 cm HiLoad Superdex200 column equilibrated with 0.5 M NaCl in 50 mM Tris-HCl pH 8.0 plus 1 mM L-malate at 1.5 ml/min flow rate in an AKTA instrument. A calibration curve was generated using Ferritin (440 kDa), aldolase (150 kDa), ovalbumin (44 kDa) and ribonuclease (13.5 kDa) .

Size exclusion chromatography of Rpa3495/96 (MatBA) was permormed using a 11 ml Superdex200GL column equilibrated with Tris 20 mM pH 7.4, 0.2 M NaCl, 2 mM DDM and run at a 0.5 ml/min flow rate in an AKTA instrument. A calibration curve was generated using thyroglobulin (669 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and aprotinin (6.5 kDa), both in presence and absence of DDM in the buffer.

3.25. TRYPTOPHAN FLUORESCENCE ASSAY

Intrinsic tryptophan fluorescence was measured in a Cary Eclipse fluorimeter (Varian Ltd, UK). 3 ml of 0.2 µM protein in TRIS-HCl Buffer, pH 7.2, was transferred to a 3 ml Quartz curvette and set to 30 °C at maximum stirring with the magnetic flea. Scan assays were done by exciting the sample at 280 nm (5 nm slit width) and measuring fluorescence between 300 nm and 500 nm (20 nm slit width). In order to observe a quench in fluorescence due to interaction, substrate was added at an excess of 1 µM final concentration.

Ligand titrations were typically performed with 0.2 µM protein with λ_{ex} 280 nm λ_{em} 340 nm using 5 nm excitation and 20 nm emission slit widths respectively. Ligand was titrated in and the subsequent change in fluorescence emission monitored. Three separate titrations were used to calculate K_d values by fitting to the quadratic equation for tight binding as described in (Smart *et al.* 2009). Emission was measured for 30 seconds for each substrate concentration, and given fluorescence values were obtained from the mean in this period, excluding the initial and final 5 seconds. Substrate concentration ranged from 0.01 µM to 20 µM, or until saturation was reached. A control was performed with addition of buffer instead of substrate, and control values were subtracted from the samples.

3.26. Differential Scanning Fluorescence assay

Experiments were performed in triplicate in a final volume of 50 µl, with final concentration of protein 5 µM, ligand 500 µM, SYPRO orange 1x (dye; Invitrogen); 1 mM phosphate buffer pH 7.4 (Giuliani *et al.* 2011). Experiments were performed in a Q-PCR machine, starting with 25 °C and increasing one degree/ minute, until a final temperature of 95 °C, taking a read at endpoint each 1 °C. Data were processed according to Structural genomics consortium, Oxford (ftp://ftp.sgc.ox.ac.uk/pub/biophysics).

The ligand library produced comprised 84 compounds at 3 mM stock concentration in DMSO, dispensed in triplicates into 96-well plates. 1 µl of substrate was added to each 50 µl assay, giving a final concentration of 60 µM. The library comprised mostly compounds containing carboxyl groups, from different chemical classes, as described in Table 3-6.

Table 3-6: Compounds present in the ligand library used for Differential Scanning Fluorescence assay.

Organic acids	**Nitrogenated compounds**	**Aromatic acids**	**Sugars**	**Aminoacids**
2-hydroxy-butyrate	2,4,dinitrophenol	4-aminobenzoate	L-arabinose	L-alanine
2-isopropyl-malate	3,3,diaminobenzidine	4-chlorobenzoate	D-arabinose	L-arginine
acetate	4-acetamidophenol	4-chlorocinnamate	D-erythrose	L-aspartate
aconitate	4-aminobutyrate	4-hydroxybenzoate	D-fructose	L-cysteine
adipate	4-nitrocinnamate	4-hydroxybutyrate	L-fucose	
methyl-malonate	*p*-phenylene-diamine	4-hydroxyphenylacetate	D-galactose	L-glutamate
butyrate	4-nitrophenylacetonitrile	4-hydroxyphenyl-lactate	D-glucose	L-glutamine
cis-*cis*-muconate	4-nitrotoluene	4-hydroxyphenylpropionate		L-histidine
citrate	4-phenylbutyronitrile	4-hydroxyphenylpyruvate	D-lactose	L-isoleucine
crotonate	6-amino-1-hexanol	4-methoxycinnamate	D-lyxose	L-leucine
fumarate	adenine	4-phenylbutyrate	D-maltose	L-methionine
gluconate	adipamide	5-phenylvalerate	D-manitol	L-phenylalanine
glutaconate	adiponitrile	benzoate	L-rhamnose	L-proline
glutarate	aminolevulinate	caffeate	D-ribose	L-threonine
itaconate	benzamide	cinnamate	D-ribulose	L-tryptophan
lactate	fumaronitrile	coumarate	D-sucrose	L-valine
levulinate	glutaronitrile	ferrulate	D-tylose	
malate	hexamethylenediamine	gallate	D-xylose	
malonate	indole-3-acetamide	phenylacetate		
mesaconate	mandelamide	phenylpyruvate		
methyl-2-oxo-butanoate	metronidazole	protocatechuate		
methyl-heptanoate	orotate	syringate		
methyl-hexanoate	4-nitrophenylsulphate	vanillate	**Others**	**Vitamins**
propanoate	4-nitrophenol	vanillin	Adonitol	ascorbate
methyl-octanoate	sulphanilamide		Barbitone	biotin
n-valerate	*trans*-nitrostyrene		EDTA	nicotinate
oxalate	triethanolamine		Fucitol	pyroxidine
oxaloacetate			TAPS	thiamine
succinate			TES	D-panthotheate
tartrate			glyceraldehyde	nicotinamide
α-ketoglutarate			Taurine	
β-hydroxypyruvate			Ectoine	

The Spectrum Collection library (Microsource Discovery Systems), comprising 2000 compounds, (50% approved drugs, 30% natural compounds and 20% known

bioactives) was in aliquots of 5 mM in DMSO displayed in 96 well plates. 1 µl of each compound was used for each 50 µl assay.

3.27. Isothermal Titration Calorimetry

ITC analyses were carried out in a Nano ITC calorimeter (TA instruments). Protein was concentrated to 100 µM and dialyzed against TF Buffer (Tris-HCl 50 mM pH7.4; NaCl 0.1M). The dialysis buffer was used to prepare substrate solutions at 1mM. The cell reaction mixture contained 300 µl of protein. Titrations were carried out at 30 °C by 25 x 2 µl injections, with a 4 minutes' interval between injections. After optimizing the baseline values and discounting the values of ligand dilution, integrations, fitting into independent-binding model, K_d determination and statistics were performed using NanoAnalyze software (TA instruments).

3.28. Protein concentration

Protein was concentrated using VIVASPIN 2 ml 30,000 MWCO PES (Sartorius stedim), according to manufacturer instructions.

3.29. Sitting drop crystallization trials

Purified and dialysed protein was concentrated to at least 10 mg/ml, and tested against the commercial screens PACT, AS, PEG, JCSG, Proplex and PhClear (Qiagen). 200 nl of protein were added to the same volume of mother liquor in sitting drop vapour diffusion experiments, using a Matrix_Hydra II Plus One crystallisation robot.

3.30. Hanging drop experiments

Once identified, a favourable condition from the crystallization trial was optimized in a 24-well plate, varying slightly the components within it. 1 µl of protein was added to 1 µl of mother liquor in the inner side of the lid, hanging against the reservoir.

3.30.1. SEEDING CRYSTALLIZATION TRIALS

When crystals with diffraction patterns unsuitable for data collection were obtained, they were used as microseeds as catalysts of the nucleation process (Bergfors 2007). 6 µl of reservoir were added to the crystals, and these were smashed repeatedly using a glass piston. The total volume was transferred to an Eppendorf, where an additional 44 µl of reservoir were added. In the presence of a small glass bead, the solution was vortexed for two minutes. The resulting liquid was serial diluted and stored at -80 °C. In the trials, both for sitting drop (final volume 0.6 µl) and hanging drop (final volume 6 µl), the proportions were added as 1 seed :3 protein: 2 reservoir. If too many crystals were obtained with the stock solution, the experiment was repeated with a 10 x dilution from the serial dilution.

3.30.2. IODINATION OF CRYSTALS

In hanging-drop plates, a crystal was immersed in 1 µl of reservoir, and 100 µl of reservoir was added to the bottom of the well. A fragment of I_2 was stuck in the dropping glass with the use of paraffin. The experiment was left for 4 hours, to allow vaporization of the iodine and diffusion into the crystal.

3.30.3. IMPREGNATION OF CRYSTALS WITH HEAVY-METALS

When a diffraction pattern did not retrieve phase solving information, it was attempted to impregnate it with heavy metals, which could give a strong enough signal once diffracted with X-rays. To do so, solutions similar to the reservoir were prepared, with a 10% increasing in precipitant concentration and heavy metal concentration between

0.05 and 0.1 mM. Crystals were immersed in these solutions for one hour to allow interaction with the heavy metals. In this study the following metals were used; Osmium ($OsCl_3$), Mercury ($HgCl_2$), Gold ($KAu(CN)_2$), Platinum (K_2PtCl_4) and Silver (AgCl).

3.31. X-RAY DATA COLLECTION AND STRUCTURE DETERMINATION

Data collection and computational modelling were performed by Dr. John Rafferty, from the Molecular Biology and Biotechnology Department, University of Sheffield.

Crystals were flash-cooled in liquid nitrogen in the presence of glycerol 20% (v/v) as cryoprotectant added to the mother-liquor and stored in liquid nitrogen.

Data were collected at the Diamond Light Source (Harwell, UK) on Beamline station I04-1 and I03. Data processing was done with XDS and merging with XSCALE (MPI for Medical Research, Heidelberg). Molecular replacement was performed with PHASER software, part of the Collaborative Computational Project, Number 4 (CCP4) software suite (Winn *et al.* 2011). The resultant electron density maps were analysed using COOT (Emsley and Cowtan 2004) and the protein models refined using REFMAC5 (Murshudov *et al.* 1997) software. Validation was performed within COOT and confirmed with MOLPROBITY server (Chen *et al.* 2010). Figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and Ligplot+(Laskowski and Swindells 2011).

3.32. MEMBRANE PREPARATION

For membrane preparation, cells were grown to OD_{600} 0.8 and harvested at 16000 x g for 10 minutes. The pellet was ressuspended in a minimum of 20 mL, or 5ml/L of culture, of Lysis Buffer (20 mM Tris HCl pH 7.4; 0.1 M NaCl; 1 mM EDTA ; 0.1 mg/ml of Lysozyme; a tablet of PierceTM Protease inhibitor EDTA free mini tablets (Thermo)) and disrupted by passing it twice through a French pressure cell (SLM Aminco

instruments), at 1000 psig cell press (482 psig gauge pressure). The resulting suspension was centrifuged at 20 000 rpm (42000 x g) for 20 minutes in JA-20 rotor to get rid of cell debris. The supernatant was then centrifuged at 200 000 x g for two hours. The resulting pellet will contain the total membrane preparation.

3.33. MEMBRANE PROTEIN EXPRESSION OPTIMIZATION

For each condition tested, 600 ml of cells were grown to OD_{600} 0.8 and, after induction, the total membrane preparation of the culture was performed. The pellet was ressuspended in 2mL of Lysis buffer + 1% sarkosyl (Sodium N-lauryl sarcosinate) and centrifuged for 200 000 x g for 30 minutes. The supernatant, containing the inner membrane preparation, was assayed through western blot to assess expression of membrane proteins.

A combination of inducer concentration (0.4 mM and 1 mM IPTG for pET-TM; 0.2%, 0.02%, 0.002%, 0.0002% L-arabinose, for pBAD-TM), temperature (37ºC and 25ºC) and time (1, 3, 5 and 24h) were tested.

3.34. WESTERN BLOT

For western blot assays, two identical SDS-PAGE gels were run, as described in session 3.18, where only one of them was coomassie stained. Transference of proteins from the second gel to the 45 µm pore nitrocellulose membrane (BioRad) was performed in transfer buffer (Tris 3.03 g/L; Glycine 14.4 g/L; methanol 10% (v/v)) in the mini trans-blot cell assembly (BioRad) for 1 h at 100 V, in the presence of a cooling unit and constant stirring.

After transference, the membrane was washed with TBST Buffer (PBS Buffer 20 mM pH 7.4, Tween-20 0.05% (v/v)). Membrane was blocked by incubation with 10 ml TBST + 4% (w/v) BSA, either 1h at Room Temperature or overnight at 4ºC. Membrane was incubated 3 times with 10 ml TBST Buffer for 5 min for washing, and then with

primary antibody (Anti-His-Tag IgG2 monoclonal antibody from mouse, GE Healthcare) diluted 2000 times in 8 ml of TBST Buffer, and left for either 1 h at room temperature or overnight at 4ºC. Membrane was washed as before, and incubated for 1h at RT with Secondary antibody (ECL™ Peroxidase labelled (HPR) Anti-mouse antibody from rabbit). After washing, developing solution (1.4 ml of 1:1 solution A: solution B – ECL™ Prime Western Blotting Detection Reagent – GE HealthCare) was added to the membrane, and results were visualised with ChemiDoc instrument (BioRad), under automatic exposure in the chemi blot option.

3.35. MEMBRANE PROTEIN PURIFICATION

Each 1 g of membrane was resuspended using a homogenizer (GPE) in 5 ml of Ni-NTA load Buffer (Tris 50 mM pH 7.4; NaCl 0.1 M). Two ml of DDM 200 mM was added to the solution, homogenized, and then another 3 ml of Ni-NTA load buffer was added. The mixture was incubated on ice for 1 h, eventually inverting the tube. The solution was centrifuged at 200 000 x g for 30 min. During this time, 0.5 ml of either 50% His-Pur Ni-NTA resin (Thermo) (Standard) or Talon metal affinity resin (Clontech) was poured into a 10 ml polipreparative column (BioRad), washed with 5 Column volumes (CV) of MQ water, then 5 CV of Ni-NTA load Buffer. The supernatant from the centrifugation step was poured into the column and incubated on ice for 1 h, with periodic inversion of the tubes. The column was left for resin sedimentation, and the flowthrough was released. Then, it was washed with 10 CV of DDM wash Buffer (Tris 50 mM pH 7.4; NaCl 0.1 M; 2mM DDM (0.5% w/v); glycerol 20%; 40 mM Imidazole), and subsequent 10 CV of either DM or OG wash buffer (Tris 50 mM pH 7.4; NaCl 0.1 M; glycerol 20%; 40 mM Imidazole; DM 0.1% (w/v) (2.07 mM) or OG 1% (w/v) (3.42 mM)). Elution was performed with 4.5 CV of either DM or OG elution Buffer (Tris 50 mM pH 7.4; NaCl 0.1 M; glycerol 20%; 500 mM Imidazole; DM 0.1% (w/v) (2.07 mM)

or OG 1% (w/v) (3.42 mM)), and fraction were collected as follow: void fraction: 1/5 CV; fraction 1: 2 CV; fraction 2: 2 CV. Protein concentration was assayed using the A_{280} method, with 72 kDA MW and extinction coefficient of 63285 M^{-1} cm^{-1} for MatBA from pBAD-TM construct.

3.36. MASS-SPECTROMETRY PROTEIN FINGERPRINT

The bands from the gel to be analysed were cut, chopped into small pieces in order to increase surface availability and initially de-stained in 0.5 ml acetonitrile 1:1 $(NH_4)_2CO_3$ buffer (Ammonium bicarbonate 100 mM pH 8.0). Supernatant was discarded and gel pieces were washed twice with 0.5 ml of AmCO3 buffer for 10 min each time. Then, gels were washed twice with 0.5 ml of acetonitrile 1:1 $(NH_4)_2$CO3 buffer, for 10 minutes each time, and final dehydration was done by incubating the gel pieces with 0.5 ml acetonitrile for 10 min. Reduction of the cysteines was then performed by adding 0.2 ml of 50mM TCEP in AmCO3 buffer and incubating for 20 min at 70°C. The supernatant was discarded, and after cooling down the gel pices, reduced cysteines were acetylated by incubation with 0.2 ml of 50 mM Iodoacetamide in AmCO3 buffer, shaking in the dark. Gel pieces were then washed and dehydrated as before. Trypsin digestion was then performed by incubation of gel pieces in 0.2 ml of 1 ng/µL Trypsin in AmCO3 buffer overnight at 37°C. In the next day, 0.1 ml of acetonitrile was added and incubated for 15 min at 37°C. The supernatant containing digested peptides was collected in a separated tube. The gels were incubated with 50 µl of 0.5% formic acid, shaking for 15 min, and then 0.1 ml acetonitrile was added and incubated for additional 15 min. The supernatant was transferred to the collection tube, and the extraction with formic acid and acetonitrile was repeated. A final incubation with 0.1 ml acetonitrile shaking for 15 min was performed, resulting in a final volume of 0.7 ml of digested peptides in the collection tube. Peptides were dried in a SpeedVac concentrator. Then,

the peptide pellet was resuspended in 40 µL 0.5% formic acid and 20 µL of the peptide solution were introduced into a vial for HPLC-MS analysis.

HPLC-Mass-spectrometry identification of peptides was performed by the “Biological Mass Spectrometry Facility”, from the University of Sheffield. HPLC was performed on a Dionex Ultimate 3000 system (ThermoFisher Scientific) using an Acclaim™ PepMap™ 100 C18, 20 mm X 75 µm trap column (3 µm, 100 Å,) and an EASY-Spray™ C18, 150 mm X 50 um (2 µm, 100 Å) analytical nano-column (Thermo Scientific) heated at 45°C. Peptides were separated for 45 min using a gradient of 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in 80 % acetonitrile (mobile phase B). The gradient was run at a flow rate of 250 nL/min as follows: 0–5 min at 3% B, then linear increase to 40% B at 40 min, followed by a linear increase to 90% B at 45 min.

MS analysis was performed on a LTQ Orbitrap Elite (Thermo Scientific) hybrid ion trap-orbitrap mass spectrometer equipped with an EASY-Spray source (Thermo Scientific), with an ion transfer capillary at 250°C and a voltage of 1.8 kV. MS survey scans in positive ion mode were acquired in the FT-orbitrap analyzer using a *m/z* window from 375 to 1600, a resolution of 60´000, and an automatic gain control target setting of $1x10^6$. The 20 most intense precursor ions were selected for the acquisition of tandem mass spectra in the dual cell linear ion trap (normal scan rate) using collision induced dissociation (CID). Charge states 1+ were excluded for precursor selection. Normalized collision energy was set to 35%, activation time to 10 ms, isolation width to 2 *m/z*, and automatic gain control target value was set to $1x10^4$.

MaxQuant search for the identification of peptides/proteins was performed using default parameters, the *Escherichia coli* BL21-DE3 database (downloaded from

Uniprot Proteomes on the 2016-09-01) and a database including the theoretical sequences of the expressed proteins. Identification results were imported into excel files for further interpretation.

3.37. LIPOSOME RECONSTITUTION OF MEMBRANE PROTEINS

3.37.1. LIPID PREPARATION FOR RAPID DILUTION METHOD

A 100 ml round-bottomed flask was washed with dH_2O, followed by ethanol and dried under a N_2 stream. It was then washed with chloroform and dried the same way. A 4 ml vial of 25mg/ml *E. coli* polar lipids (Avanti Polar lipids), giving 100 mg of total lipids, was transferred to the flask, and dried under the N_2 stream for 30 min after appearing to be fully dried. Lipids were solubilized in pentane, and dried for 30 min under the N_2 stream after appearing to be fully dried. Lipids were solubilized in Inside Buffer (Tris 50 mM pH 7.4; NaCl 1 mM; KCl 99 mM), at a final concentration of 20 mg/ml (5 ml). The liposomes were snap frozen in ethanol immersed in dry ice, and thawed for three times, separated in 0.4 ml aliquots, and stored at -80ºC.

3.37.2. RAPID DILUTION METHOD FOR PROTEOLIPOSOME PREPARATION

2 ml of the protein elution fraction in DM or OG were mixed with 400 µl of prepared lipids at room temperature. The mixture was completely transferred to 65 ml of ice-cold Inside Buffer, and centrifuged for 1.5 h at 45 000 rpm (235000 x g) in a Ti type 45 rotor. The resulting pellet was ressuspended in 1 ml of Inside Buffer (Tris 50 mM pH 7.4; NaCl 1 mM; KCl 99 mM), and frozen-thawed 3 times in dry ice immersed in ethanol.

3.37.3. BIOBEADS REMOVAL METHOD FOR PROTEOLIPOSOME PREPARATION

Twelve mg of lipids prepared as in section 3.39.1 were used for each reconstitution. After being extruded 11 times with a lipid extruder (See section 3.40 for further details), the lipids were diluted to a 4 mg/ml final concentration, and titrated with aliquots of 2 µl/ml of a 10% solution of Triton X100 detergent, while monitoring the absorbance at 540 nm wavelength. After saturation of liposomes was reached, indicated by a decrease in absorbance, 5 more titrations of Triton X100 were added.

To the lipids detergent mixture, 0.5 mg of protein purified as in section 3.37 was added, and incubated in the rocker for 15 min. After that, 100 mg of Biobeads (Non polar polystyrene beads – BioRad) were added, and incubated for 1 h at room temperature. Further 100 mg of Biobeads were added, and incubation was for 1 h at 4 ºC. After adding more 100 mg of Biobeads, the mixture was incubated on rocker at 4 ºC overnight. Next morning, another 100 mg of biobeads was added, and incubated for 2 h at 4 ºC. The solution was poured into a 10 ml polipreparative column (BioRad) for separation of the biobeads and proteoliposomes. The flow-through was centrifuged at 70 000 rpm (200000 x g) in a TLA100.3 rotor for 30 min at 15 ºC. The pellet, containing the proteoliposomes, was resuspended for a final concentration of 10 mg/ml in Inside Buffer (1.2 ml), snap frozen in liquid nitrogen and stored at -80 ºC.

3.38. UPTAKE OF RADIOLABELLED SUCCINATE INTO PROTEOLIPOSOMES

Before usage, proteoliposomes in 1 ml of Inside Buffer were extruded 11 times in a lipid extruder (Stratech), using 400 nm Nucleopore Track-Etch membranes (Whatman). The solution was centrifuged at 70 000 rpm (200 000 x g) in TLA 100.3 rotor at 15ºC and the pellet resuspended in 100 µl of Inside Buffer.

For each uptake reaction, 0.22 µm GSWP nitrocellulose membranes (Millipore) were placed in six holders of a 1225 sampling manifold (Millipore). 600 µl of outside buffer (Tris 50 mM pH 6.4; NaCl 99mM; KCl 1mM; valinomycin 1 mM; MatC 5 µM; succinate 900 nM; ^{3}H-succinate 100 nM) were prepared. 10 µl of proteoliposomes were transferred into the 600 µl reaction and quickly vortexed for homogenization, marking time zero. At 30 s, and 1, 2, 3 and 5 minutes, 100 µl of the reaction were transferred to an Eppendorf containing 100 µl of stop buffer (50 mM Tris pH 7.4; Succinate 1 mM; Choline Chloride 0.1 M), aiming to stop uptake. The resulting 200 µl were then quickly transferred to one of the filters in the vacuum manifold and washed with 3 ml of cold stop buffer. Control reactions were performed in three different ways with protein-free liposomes, no proteoliposomes and reaction buffer without SBP (MatC).

At the end of the reaction, filters were transferred to scintillation counter inserts (Thermo) and 10 ml of filter counter solution (Perkin-Elmer) was added. To the 6th filter, 10 µl of the original reaction were added, representing 10% of the total counts possible in the analysis. After 20 minutes incubation, scintillation was measured in a LS6000TA Scintilation counter (Beckman) for 1 min.

3.39. PREPARATION OF ELECTROPORATION COMPETENT *R.PALUSTRIS*

R. palustris was grown anaerobically in 50 ml of RCV, at 30 °C, up to OD_{600} 0.6. The culture was centrifuged at 10000 rpm (16000 x g) for 10 minutes, at 4 °C and resuspended in 50 ml of chilled sterile distilled water. Centrifugation and resuspension were repeated as before once again. After the third centrifugation, cells were resuspended in 10 ml of chilled sterile 10% glycerol. Cells were centrifuged again as before and resuspended in 1 ml of chilled sterile 10% glycerol. 40 µl aliquots were transferred to chilled sterile microfuge tubes and stored at -80 °C.

3.40. ELECTROPORATION OF *R. PALUSTRIS*

1 µg of plasmid and 80 µl of chilled sterile deionized water were added to 40 µl of competent *R. palustris*. Culture was transferred to 2 mm sterile electroporation curvette, and electroporated as follows: 2.5 kV voltage, 25 uFD capacitor, 200 ohms resistor. The culture was then transferred to a microfuge tube and 1 ml of RCV media was added. The culture was incubated anaerobically under light at 30 °C for 20 hours and plated in RCV containing the appropriated selective pressure. Plates were incubated at 30 °C for 6 days.

3.41. CONJUGATION WITH *E. COLI* S17-1 STRAIN

Plasmids containing the deletion cassette were transformed into *E.coli* S17-1 selected on LB plates supplemented with kanamycin. After selecting positive colonies, a conjugation experiment between S17-1 strain and *R. palustris* was performed. 80 ml of two days cultured *R. palustris* in PYE media were centrifuged 4500 rpm (3400 x g) for 10 min and resuspended in 200 µl of PYE media. A streak of S17+pK18mobsacB was resuspended in 30 µl of LB and mixed with *R. palustris*. The mix was dropped in 50 µl drops onto PYE media plates and incubated 6 h 30 °C. The drops were then resuspended and streaked on PYE + kanamycin (70 µg/ml) plates and incubated at 30 °C for 6 days.

3.42. SELECTION OF KNOCK-OUT MUTANTS IN *R. PALUSTRIS*

After transferring the plasmid with the deletion cassettes into the cells by electroporation or conjugation, *R. palustris* transconjugants were plated in RCV plates supplemented with 70 µg/ml of kanamycin at 30 °C under light for 6 days. Obtained colonies were stabbed in RCV plates containing kanamycin and RCV plates with 10% sucrose. Colonies which were kanamycin resistant and sucrose sensitive should have gone through the first recombination event, having incorporated the vector in the

genome. These colonies were used to inoculate 5 ml of PYE + kanamycin. After two days, serial dilutions were plated onto RCV plates with 10% of sucrose and left for at least 14 days at 30 °C. The grown colonies were stabbed in RCV + sucrose plates and RCV + kanamycin plates. Those which grew on sucrose media, but not in kanamycin media should have undergone the second recombination event and have a marker-free deletion of the target gene. Positive colonies were tested by colony PCR.

3.43. RNA EXTRACTION

RNA extractions and quantitative real time PCR were performed using modified versions of the methods explained in Guccione *et al.* (2008).

R. palustris broth cultures were grown in triplicate under normal conditions to an OD_{660} nm of 0.6. 10 ml of culture was mixed with a chilled 5% v/v phenol in ethanol solution to stabilize the RNA. Samples were then centrifuged at 12,000 x g for 10 minutes at 4°C, the cell pellets were stored at -80°C.

Total RNA was extracted by re-suspending the pellet in 1 ml Tri-Reagent (Sigma) with 3.2 µl GlycoBlue (Ambion) to visualize the RNA, 200 µl chloroform was added to the samples to separate the RNA from the rest of the cellular materials and the samples were centrifuged at 15,000 xg for 15 minutes at 4°C. The upper aqueous layer containing the RNA was removed and added to 500 µl chilled isopropanol with 90 µl of 5 M ammonium acetate and incubated at -20°C for 1 hour. Samples were centrifuged at 15,000 x g for 15 minutes at 4°C and the supernatant removed from the now blue stained RNA. The RNA was washed with chilled 75% v/v ethanol and re-suspended in 100 µl nuclease free water and stored at -20°C. Purified RNA samples were DNAse treated using the 'Turbo-DNA Free' kit (Ambion) to remove any contaminating DNA, according to the manufacturer's instructions. The RNA

concentration and purity were determined by using a 'Genova nano micro-volume spectrophotometer' (Jenway), and diluted to 20 ng/µl aliquots.

3.44. RT-PCR ASSAY

Gene specific primers were designed to amplify a 100-300 bp fragment within the gene of interest, the *rpoD* gene was used as an internal control. All primers were diluted to 25 µM in nuclease free water.

Each reaction was carried out in a 25 µl volume in a MicroAmp® 96-well optical reaction plate (ABI prism). Reactions were performed using the Brilliant III Ultra-fast SYBR Green RT-PCR kit (Agilent). Each reaction contained 10 µl Agillent SYBR Green Mix, 1 µl of each primer, 0.2 µl of DTT, 1.8 µl nuclease free water, 1µl of Block RT/RNAse enzyme and 10 µl of 20 ng/µl RNA or the proper dilution of genomic DNA. Each reaction using RNA was repeated in triplicate; reactions using genomic DNA for the standard curve were replicated in duplicate.

PCR amplification was carried out in a Stratagene MX3005p thermal cycler (Agilent) at 50°C for 10 min and 95°C for 3 min followed by 40 cycles of 95°C for 15 s; 60°C for 20s. A last cycle of 95°C for 60s, 55°C for 30 s and 95°C for 30 s was performed in the end. Data was collected with the associated MxPRO QPCR software (Agilent). A standard curve for each gene was generated using a series of *R.palustris* genomic DNA dilutions. Gene expression between cultures was calculated as relative to *rpoD* expression. The data was analysed as described previously (Guccione *et al.,* 2008)

Primers for this assay were designed using Primer3 software (Untergasser *et al.* 2012), aiming for 20 bp length, with a melting temperature of 58 °C to amplify a region between 200 and 400 bp inside the gene, at least 100 bp after the mRNA start position. Primers are shown in Table 3-7.

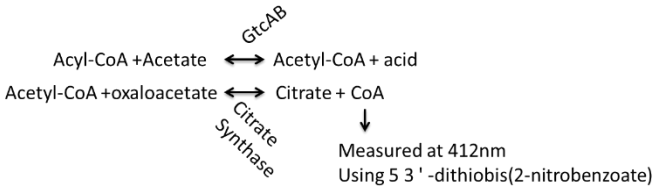
Table 3-7: Primers designed for RT-PCR Analysis

Primers used in RT-PCR		
Gene Number		**Sequence**
RPA3494	Forward	5`- GGCCAGCCGATCGAAATC -3`
matC	Reverse	5`- GGCGTGAAGTCCTTGTCA -3`
RPA2047	Forward	5`- GCGTCAAGGAATTCGAAGTC -3`
dctP	Reverse	5`- CATGATCTTGAAGCCGTTGT -3`
RPA4515	Forward	5` - CGGCCAGTCCTTTGTCAT - 3`
adpC	Reverse	5`- GATCGTAGCCCATGGTCC- 3`
rpoD	Forward	5` - CGACTTCCTGCGCAACTATC - 3`
rpa1288	Reverse	5`- GGTTGGTGTACTTCTTGGCG - 3`

3.45. CoA TRANSFERASE ACTIVITY ASSAY

The enzymatic activity of GtcAB was performed in accordance to Buckel *et al.* (1981), in an indirect measurement of free Coenzyme A produced according to equation 1. Each assay was composed of 100 mM potassium phosphate buffer pH 7.0, 200 mM sodium acetate, 1 mM oxaloacetate, 1 mM dithio-*bis*-(2-nitrobenzoate), 20 µg citrate synthase; and 0.1 mM glutaryl-CoA or succinyl-CoA. The reaction was started upon addition of the enzyme. Absorbance was followed over time at 412 nm. The preparation of the CoA derivatives was performed according to Simon and Shemin (1953).

Equation 1: Indirect measurement of CoA transferase activity by production of free Coenzyme A



For the measurement of glutaconyl-CoA production, each assay contained 100 mM potassium phosphate buffer pH 7.0; 1 mM glutaconate; and 0.1 mM acetyl-CoA. Reaction was started with addition of enzyme, and absorbance was measured at 263 nm.

CHAPTER 4. MASSIVE OVER-REPRESENTATION OF SOLUTE-BINDING PROTEINS (SBPS) FROM THE TRIPARTITE TRICARBOXYLATE TRANSPORTER (TTT) FAMILY IN THE GENOME OF THE A-PROTEOBACTERIUM *RHODOPLANES SP.* Z2-YC6860.

Rosa LT, Springthorpe V, Bianconi ME, Thomas GH, Kelly DJ. Microbial Genomics. 2018;4(5):e000176. doi:10.1099/mgen.0.000176.

Suplementary information for this manuscript can be found online at http://mgen.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000176#tab5

PREFACE

In order to identify the ubiquity of the Tripartite Tricarboxylate Transporters (TTT) family among bacteria, we performed a high-throughput BLAST search using characterised TTT systems as query. The results showed an overrepresentation of this family among α and β-proteobacteria, and that the bacterium *Rhodoplanes sp*. Z2-YC6860, belonging to the α-proteobacteria class, contains 434 TTT SBP genes in its 8.2 Mbp genome, representing 6% of its total encoded proteins. Using various bioinformatics tools, we provide an insightful genome plot of this bacterium, highlighting aspects involving the TTT SBP, alongside with high throughput sequence analysis of these genes. Taken together, our data suggest the TTT SBP might be involved in a wide range of signalling pathways, and that a Lineage-Specific Expansion (LSE) event was majorly responsible for this expansion.

AUTHOR CONTRIBUTIONS

The idea for the manuscript was conceived by L Rosa, G Thomas and D Kelly. High-throughput BLAST Analysis were performed by M Bianconi, under close collaboration with L Rosa. Genome plot and gene density of CG islands were provided by V Springthorpe. Genomic neighbourhood analysis was performed by L Rosa. Protein sequence analyses were performed by L Rosa, M Bianconi and V Springthorpe. L Rosa, M Bianconi, G Thomas and D Kelly co-wrote the manuscript.

MANUSCRIPT INSERT

CHAPTER 5. STRUCTURAL BASIS FOR HIGH-AFFINITY ADIPATE BINDING TO ADPC (RPA4515), AN ORPHAN PERIPLASMIC BINDING PROTEIN FROM THE TRIPARTITE TRICARBOXYLATE TRANSPORTER (TTT) FAMILY IN *RHODOPSEUDOMONAS PALUSTRIS*

Rosa, L. T., Dix, S. R., Rafferty, J. B. and Kelly, D. J. FEBS J, (2017) 284: 4262-4277. doi:10.1111/febs.14304

PREFACE

In this manuscript, we characterise the protein AdpC, one of the five orphan Solute Binding Protein (SBP) from the TTT family in *R. palustris*. Ligand library screens showed that AdpC binds to di-carboxylates ranging from six to nine carbons in length, and isothermal titration calorimetry showed adipate was the highest affinity ligand for this protein. We show that AdpC expression is increased under low concentrations of substrate, and repressed as this concentration increases, in conformity with a high-affinity uptake system, only required under certain conditions. We obtained two high-resolution crystal structures of AdpC, both with adipate and 2-oxo-adipate coordinated in the binding pocket. The structure revealed the presence of two water molecules bridging hydrogen bonds between the proximal carboxylic group of the substrate and the α-chain of AdpC, and the absence of positive residues in the binding pocket to counter-act the negative charge of the ligand.

The seven homologs of *tctC* predicted by Antoine *et al.* (2003) in *Rhodopseudomonas palustris* were identified using BLAST analysis, using TctC of *Salmonella typhimurium* as a query (Sweet *et al.* 1979). The *tctC* homologs were named *tttC* genes and numbered from 1 to 7, where those present in a complete TTT system were numbered *tttC1* and 2, and the remaining numbered according to their appearance in the genome. Analogously, using the *tctBA* genes as a query (Widenhorn *et al.* 1988), it was possible to identify the two TTT transmembrane systems as RPA2321-2320 (*tttBA1*) and RPA3496-3495 (*tttBA2*). Given their function characterised in this study, *tttC5* was renamed AdpC, and the *tttBAC2* system was renamed *matBAC*. The original gene number and correlation to our proposed naming system is shown in Figure 5-1.

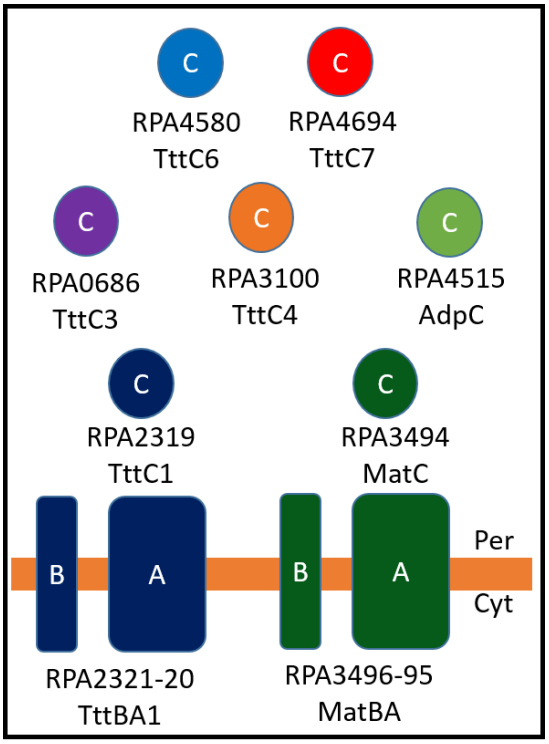


Figure 5-1: Proteins comprising the TTT family in *Rhodopseudomonas palustris*. Two complete sets of genes are found, in addition to five periplasmic binding proteins not obviously related to any transmembrane protein in genetic analysis, so-called "orphan" proteins. Proteins of similar colour are located in the same operon. The orange line in the bottom represents the cytoplasmic membrane.

AUTHOR CONTRIBUTIONS

The idea for this manuscript was conceived by L Rosa and D Kelly. All genetic, biochemical and microbiology work, as well as all figures, were done by L Rosa. Crystallisation, data collection and structural analysis were performed by L Rosa and J Rafferty. Structural model was built by S Dix and J Rafferty. L Rosa and D Kelly co-wrote the manuscript, and J Rafferty commented on it.

CHAPTER 6. HIGH-AFFINITY UPTAKE OF C4-DICARBOXYLIC ACIDS BY A TRIPARTITE TRICARBOXYLATE TRANSPORTER SYSTEM FROM *RHODOPSEUDOMONAS PALUSTRIS*

6.1. THE MATBAC TRIPARTITE SYSTEM

As illustrated in Figure 5-1, *R. palustris* contains two complete tripartite systems, composed by RPA2319-21 and RPA3494-96, named in this study as *tttBAC*1 and *matBAC*, respectively. The *matBAC* operon is shown in Figure 6-1. This chapter will focus on the biochemical characterisation and potential physiological roles associated with MatBAC, which our findings indicate to be a C4-dicarboxylate transport system, with sub-nanomolar affinity for malate, and for this reason this system was named as Malic Acid Transport system (Mat). Initially, the characterisation of the SBP component of the system, MatC, will be shown, followed by our attempts to reconstitute the membrane components, MatBA, into liposomes, and provide the first energy-coupling data for the TTT family.

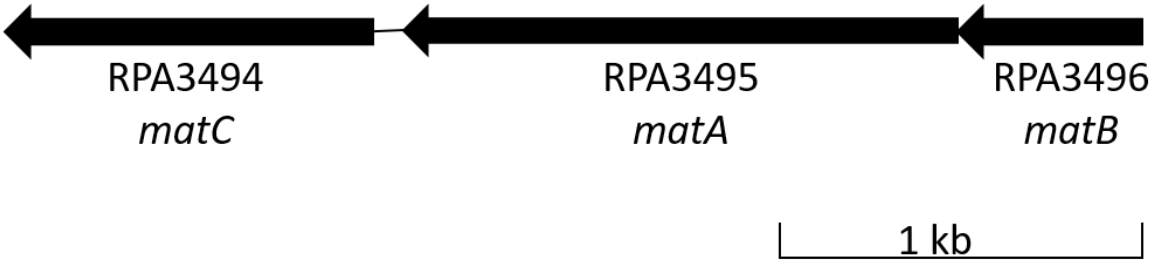


Figure 6-1: Genetic organisation of the *matBAC* system (RPA3494-3496).

6.2. MATC DOES NOT OLIGOMERISE UPON LIGAND BINDING

MatC, the SBP component of the MatBAC system, was cloned, overexpressed and subjected to binding assays. MatC is a protein composed of 335 amino acid residues (aa.), of which 307 are present in the 32.6 kDa mature protein, after signal peptide cleavage. After cloning in pET21a, the mature 307 aa. protein fused with a hexa-

histidine (His_{6x}) tag resulted in a 34 kDa protein. Figure 6-2b shows successful purification of MatC in the SDS-PAGE gel.

In order to test if dimerization was necessary for optimal protein binding (see next section for ligand characterization), as suggested for some components of TRAP transporters (Gonin *et al.* 2007), MatC was subjected to size exclusion chromatography analysis, in the presence of 1 mM D/L-malate racemic mixture. If the protein binds to the substrate as a monomer, a peak around 34 kDa would be expected, and, if dimerization happened upon ligand binding, the peak would be present at around 68 kDa. As shown in Figure 6-2c, MatC was eluted after 16.22 ml of buffer, which in the calibration curve provided in Figure 6-2d, would be equivalent to 39 kDa, inside the margin of error acceptable for the technique, suggesting that the protein binds to the substrate in a monomeric form.

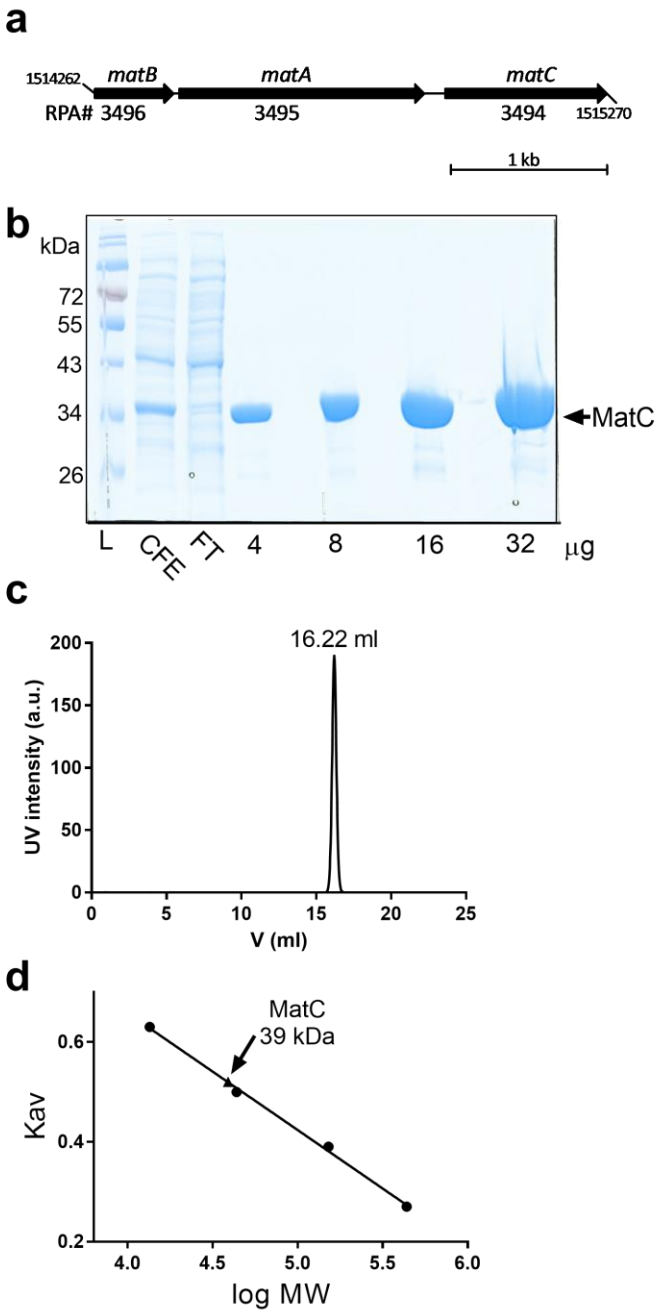


Figure 6-2: Purification of the MatC recombinant protein. a) genomic localization and identity of the *matBAC* operon b) 12% polyacrylamide SDS-PAGE gel of MatC purification after Nickel affinity chromatography and size exclusion chromatography. Samples were loaded in the following order: Molecular weight marker (L); Cell-free-exract (CFE); Flow-through before initiate the imidazole elution (FT); increasing amounts of MatC after Nickel affinity and siz exclusion chromatographies. **c) Size exclusion chromatography of MatC**. The protein was submitted to separation in the presence of Tris 50mM pH 8.0; NaCl 0.5 M; malate 1 mM. Elution of protein was observed after 16.22 ml of Buffer, which in the **corresponding calibration curve (d)** would be equivalent to 39 kDa, near to the estimated size for a MatC monomer, of 34 kDa. The calibration curve was generated using ferritin (440 kDa), aldolase (150 kDa), ovalbumin (44 kDa) and ribonuclease (13.5 kDa) and Kav was calculated using the formula Kav=(Ve-Vo)/(Vt-Vo).

6.3. Differential Scanning Fluorescence assays reveal MatC binds to C4-dicarboxylic acids

MatC was screened against a library of 84 compounds comprising organic acids, amino acids, vitamins, aromatic compounds, sugars and nitrogenated compounds, as described in Rosa *et al.* (2017), using a differential scanning fluorescence assay. A significant shift of 9 °C, 7 °C and 6 °C in the melting temperature of MatC was observed when in the presence of the C4-dicarboxylic acids D/L-malate, succinate and fumarate, respectively, as shown in Figure 6-3. However, no stabilization was observed with molecules of similar structure, such as the C4-carboxylates tartrate, oxaloacetate, mesaconate, acetoacetate, aspartate, crotonate or butyrate (Figure 6-3b/c). The absence of a shift in the presence of butyric acid suggests both carboxylic groups are essential for substrate coordination. Furthermore, although a hydroxyl group seems to improve protein stabilization, the substitution of this group for a carbonyl or methyl group in oxaloacetate and mesaconate, respectively, does not result in stabilization, nor does the addition of another hydroxyl group on carbon 3, as shown in the presence of tartrate. The absence of thermal shift in the presence of malonate and glutarate suggests also that MatC binds strictly substrates with four carbons in length. MatC did not show thermal stabilization upon interaction with citrate, found in several organisms to be a common ligand for the TTT family (Sweet *et al.* 1984, Herrou *et al.* 2007, Brocker *et al.* 2009, Batista-Garcia *et al.* 2014).

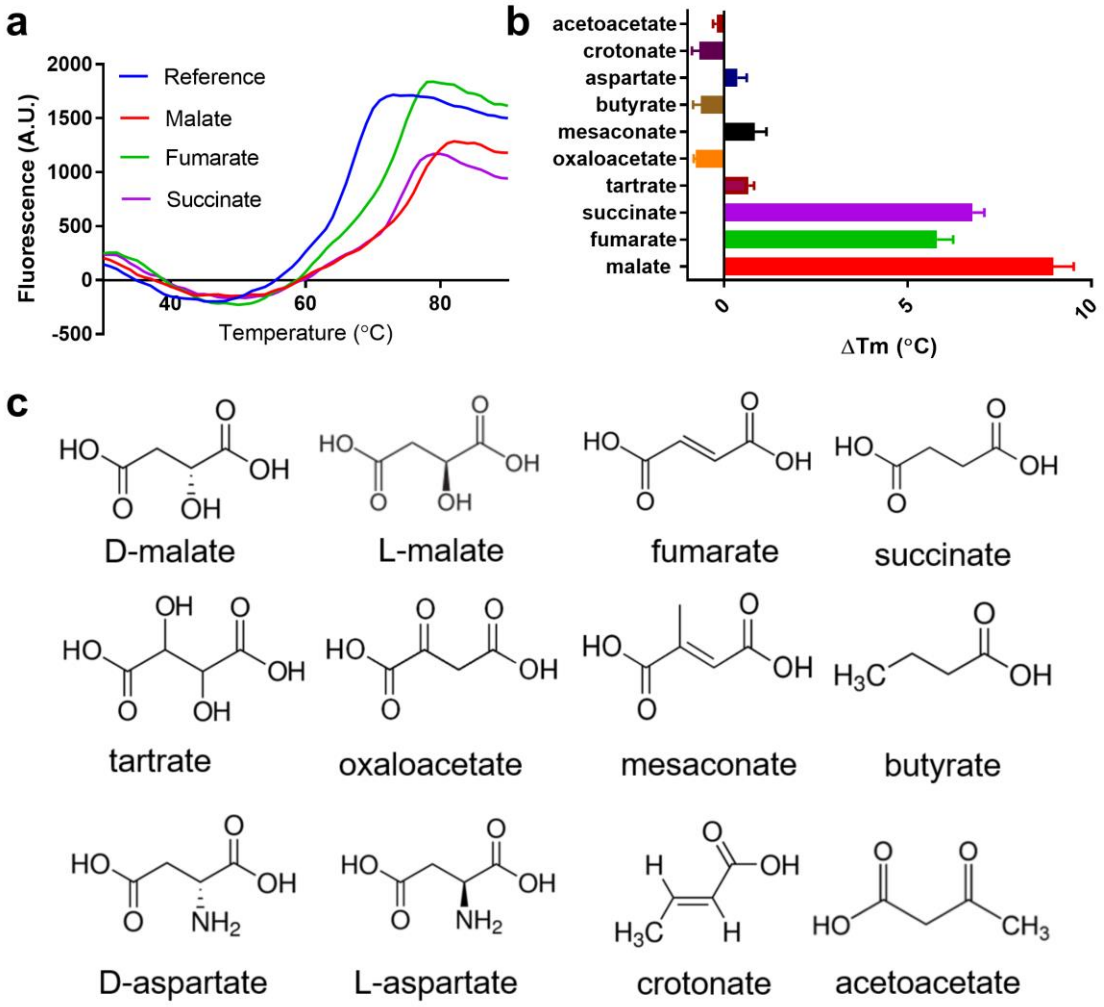


Figure 6-3: Differential Scanning Fluorescence assay of MatC. a) Thermal shift profile of MatC when in contact with malate, succinate and fumarate. Assays were performed with 5 µM protein in TF Buffer, 1x SYPRO orange dye concentration and 60 µM ligands. **b) Graphical representation of MatC thermical stabilization in the presence of different C4-carboxylates.** Each representation is the result of three independent analysis **c) Structure of the C4-carboxylates used for DSF assay.**

6.4. MatC binds to malate with sub-nanomolar affinity, but with much weaker affinity to succinate and fumarate

Following the differential scanning fluorescence assay, further characterization was performed using tryptophan fluorescence in order to determine the affinity constants

for each of the C4 dicarboxylic acid compounds. A clear quench could be observed in intrinsic tryptophan fluorescence when in the presence of 1 mM malate, fumarate or succinate (Figure 6-4a), with a reduction of 21.2%, 13.0% and 8.7% in the fluorescence, respectively. Confirming the results from the differential scanning fluorescence assay, no quench was observed in the presence of tartrate, oxaloacetate, mesaconate, butyrate, acetoacetate, aspartate, crotonate or citrate (data not shown). In order to determine binding affinity, titrations were performed with the first three substrates, as described in the methods section. MatC was shown to bind very tightly to both isomeric forms of malate, with a K_d value of 21 ± 2 nM for L-malate and 27 ± 2 nM for D-malate, the highest affinity described for any C4-dicarboxylate transporter, and apparently not distinguishing between the two enantiomeric forms as previously described for the other transport families (Figure 6-4b) (Janausch *et al.* 2002, Yurgel and Kahn 2004, Unden *et al.* 2016). Lower affinities were observed for succinate and fumarate, with a K_d of 110 ± 40 nM and 400 ± 40 nM, respectively, suggesting that malate could be the most physiologically relevant substrate for this system. These results suggest that the hydroxyl group on carbon 2 must play an important role in substrate coordination in the protein binding pocket (Figure 6-3c).

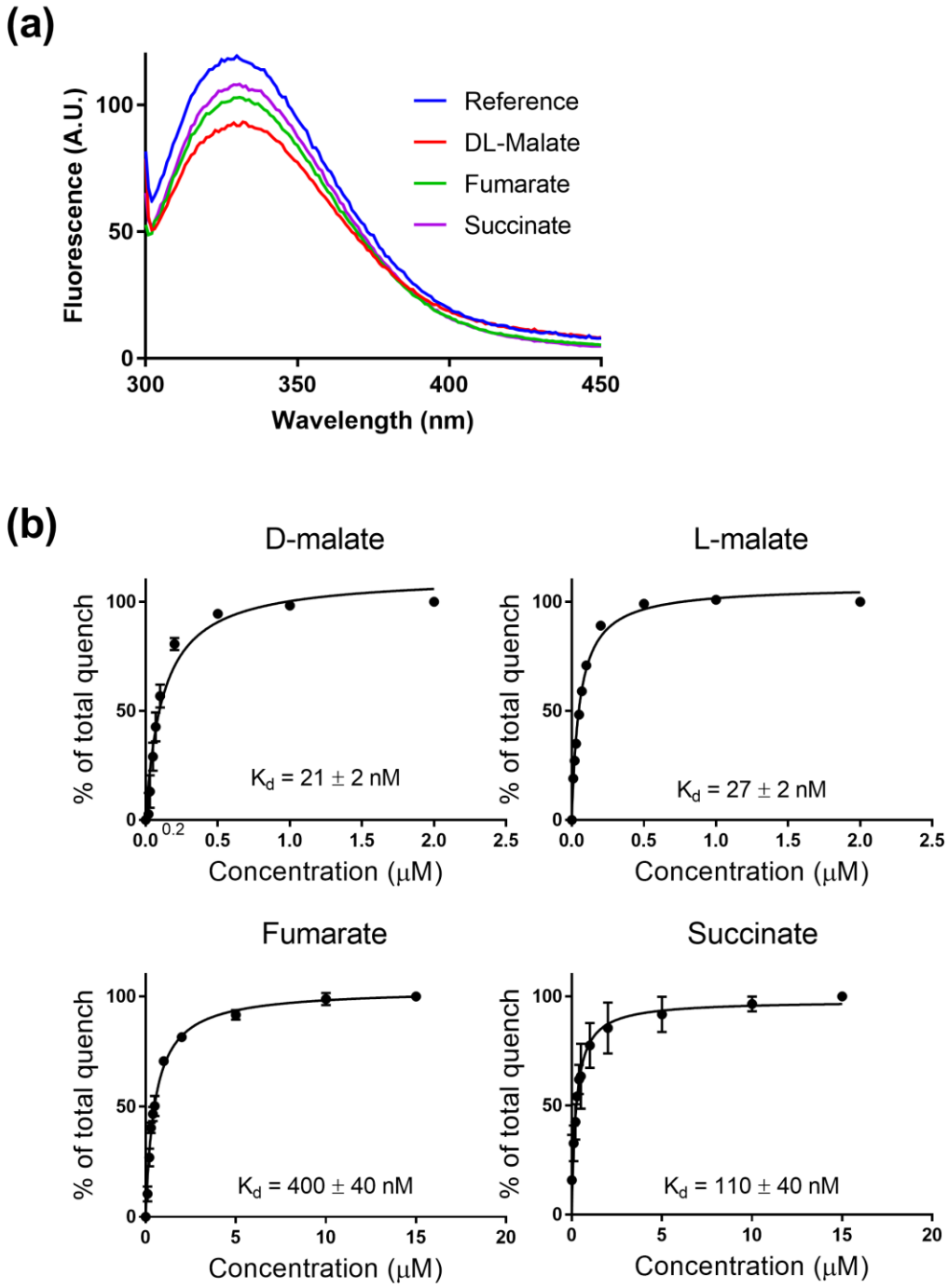


Figure 6-4: Evidence of binding of MatC to C4-dicarboxylates using the Intrinsic Tryptophan fluorescence assay. a) Screening of emission spectra of MatC in the presence of C4-dicarboxylates. b) Titration of MatC with C4-dicarboxylates. Protein was assayed in 0.2 μM concentration in 3 ml of TF buffer (Tris 50 mM pH 7.4, NaCl 0.1 M), at 30 °C and with maximum stirring. Titrations were made in triplicate, with increasing substrate concentration added every 20 s. Ex/Em wavelengths were 280/300-450 for the initial screen and 280/335 nm for the titrations. Km calculations were performed using the quadratic equation for tight binding as previously described (Smart *et al.* 2009).

6.5. RT-PCR SHOWS A SMALL INCREASE IN EXPRESSION OF *MATC* IN THE PRESENCE OF C4-DICARBOXYLATES

In order to investigate how *matC* would respond in the presence of C4-dicarboxylic acids, an RT-PCR experiment was designed to look for differences in gene expression under growth conditions with addition of 10 mM of DL-malate, fumarate or succinate. The results, summarized in Figure 6-5, suggest a small increase in the transcription levels relative to the control with pyruvate. An increase in expression of 2.67 ± 0.88, 2.47 ± 0.04 and 1.62 ± 0.06 –fold for *matC* was observed when malate, succinate and fumarate were used as sole carbon source, respectively. These results support the data obtained in binding assays and suggest a physiological relevance of the *matBAC* system under these conditions.

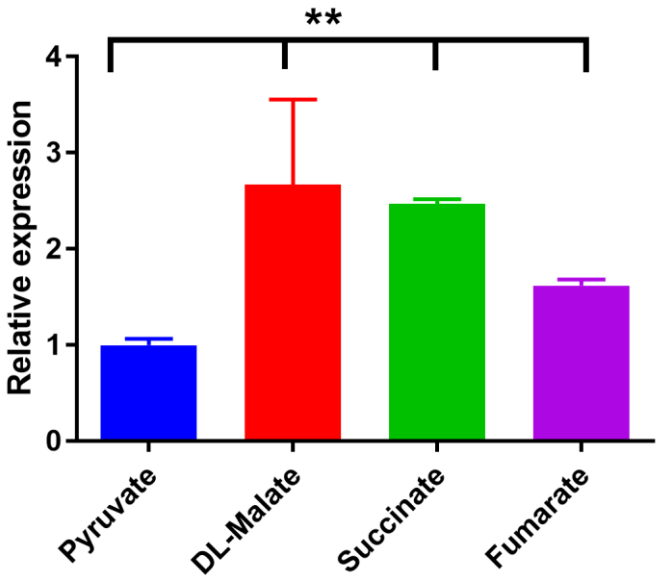


Figure 6-5: RT-PCR analysis of *matC* expression upon induction with C4-dicarboxylates. Experiments are a result of 3 biological replicates each with three technical replicates, after growing cells with 10 mM of substrate, except for *matC* measurements under malate growth, where six technical replicates were performed from three biological replicates. Statistics were performed with Student`s t-test, where ** $p<0.01$.

6.6. THE 2.1 Å STRUCTURE OF MATC

MatC was subjected to protein crystallization in the presence of 1mM of a racemic mixture of D and L-malate. A high-resolution protein crystal structure was obtained with D-malate coordinated in the binding pocket, as shown in Figure 6-6. Statistics are provided in Table 7-1 (See next chapter) . Three identical molecules were present in the asymmetric unit. MatC is a monomer comprising of 307 aa. in the mature form (without the signal sequence) and 318 in our His-tagged recombinant construct, whereas residues 13 to 306 are defined in the 2.1 Å crystal structure map. MatC protein is composed of nine α-helixes, nine β-strands and three 3_{10} helixes. This secondary structure is organized in two globular domains separated by a hinge, formed by β4 and β9, in a “venus fly-trap” like domain characteristic for binding proteins, revealing MatC as a Type II binding protein according to the Fukami-Kobayashi *et al.* (1999) classification and cluster E-II according to the classification proposed by Scheepers *et al.* (2016). Each domain is composed of a β-sheet surrounded by α-helixes. Domain 1 is comprised of residues 1-110 and 237-312, while domain two is comprised of residues 111-236. In domain 1, the architecture of the sheet is β2-β1-β3-β9-β4, and in domain 2 it is β6-β5-β7-β4-β8. Although the structure shows two cysteines, Cys190 and Cys155, adjacent in space and with the sulphydryl groups directed to each other, no electron density suggesting a disulphide bond was observed, in contrast to those described for the TTT SBPs BugD, BugE and Bug27 (Huvent *et al.* 2006a, Huvent *et al.* 2006b, Herrou *et al.* 2007). However, *in vivo* MatC is predicted to be located in the periplasm, which might suggest that under the periplasmic oxidative conditions, this disulphide bond would be formed.

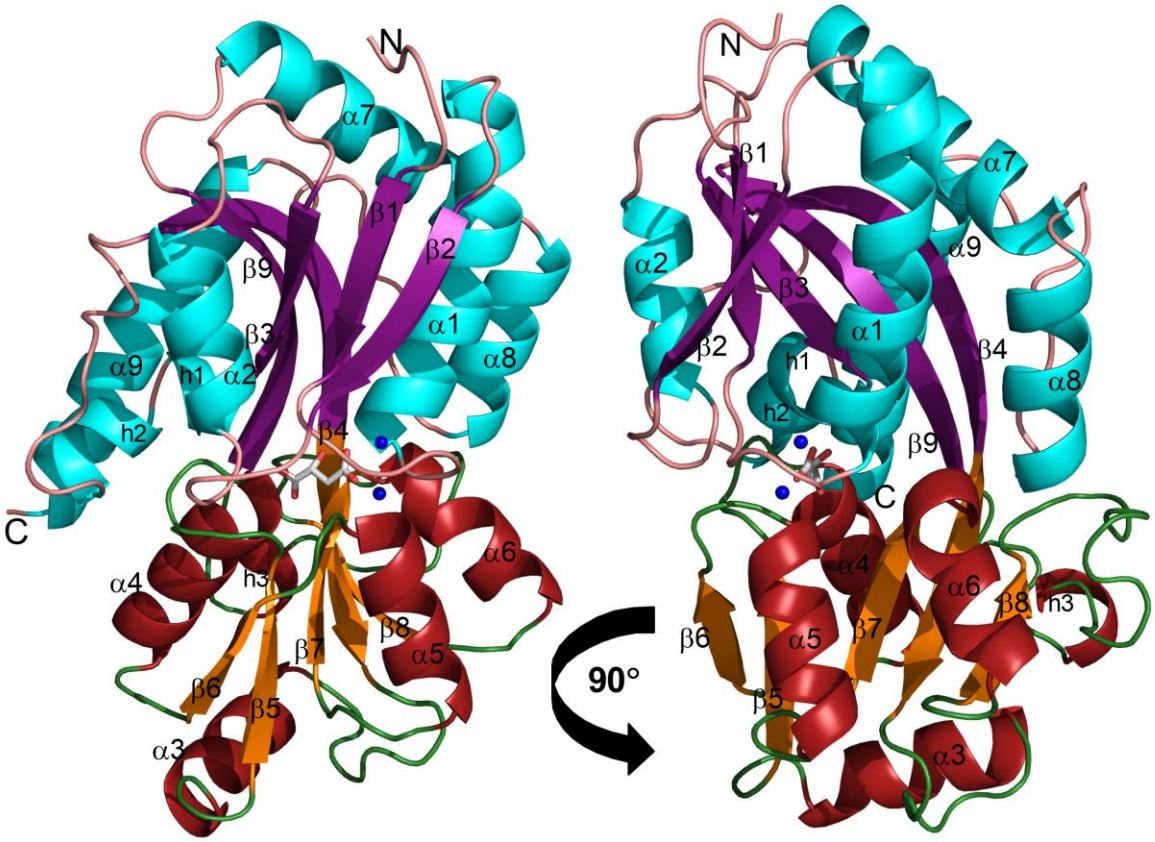


Figure 6-6: Overall structure of D-malate bound MatC. The protein is folded in a typical Venus-flytrap like domain for binding proteins, with two globular domains composed of 5 strand β-sheets surrounded by α-helixes and separated by a cleft where the substrate is coordinated. The two domains are connected by a hinge formed by β4, which is part of both β-sheets, and β9, characterizing this protein in Cluster E-II for binding proteins classification (Scheepers *et al.* 2016). Two well conserved β-loops, between β1 and α1 and β6 and α5 form a "pincer-like" structure, which coordinate the ligand proximal carboxylic group with the bridging of two conserved water molecules.

6.7. D-MALATE COORDINATION IN MATC CRYSTAL STRUCTURE

MatC was crystalized in the presence of 1 mM of a racemic mixture of D and L-malate, and is shown to contain a D-malate molecule coordinated in the protein binding pocket (Figure 7a). In agreement with previously described structures for members of the TTT family (Herrou *et al.* 2007), the ligand proximal carboxylic group is coordinated by two spatially conserved water molecules, which form hydrogen bonded bridges between the ligand and a "pincer-like" structure composed by two β-loops, between β1 and α1 and β6 and α5 (in MatC), respectively. These hydrogen bonds, as described for AdpC

(Rosa *et al.* 2017), are mostly made with the main chain of aliphatic residues, rather than side chains. In the D-malate molecule, both oxygens 4 and 5, in the proximal carboxylic group, are bridged by a water molecule, which interacts with the main chain carbonyl oxygen of Ala25 and nitrogen of Asp31. Oxygen 4 is further coordinated by another water molecule, which interacts with the main chain nitrogen of Gly175 and carbonyl oxygen of Ala25 (Figure 6-7a). In addition to water coordination, oxygen 4 on the malate has hydrogen bonds with the main chain's nitrogen of Thr174, and oxygen 5 with both the main chain's nitrogen and the side chain oxygen of Thr30. Most of these residues are conserved among AdpC, BugD and BugE, maintaining the described conservation and relevance of this "pincer-like" domain among the TTT family (Huvent *et al.* 2006a). A third water molecule is involved in coordination of oxygen 2 in the distal carboxylic group, as seen in all previous structures. This water bridges hydrogen bonds between the oxygen and the sidechain carboxamide of Gln192, and also with the sidechain carboxylic group of Asp191. Oxygen 2 makes further direct hydrogen bonds with the main chain's nitrogen and side chain oxygen of Ser151, and Oxygen 1 with the main chain's nitrogen in Val150. The Carbon 2 hydroxyl group in malate is coordinated also by a hydrogen bond with the side chain oxygen in Thr30 and with the side chain of Gln192. In addition, a few residues are involved also in hydrophobic interactions. These are: Phe24, described as a hydrophobic platform which interacts with the linear carbon chain of the ligand in all TTT transporter (Rosa *et al.* 2017), Gly28, Pro29, His80, Ser149, and Gly173.

6.8. COMPARISON BETWEEN MATC AND BUGD CRYSTAL STRUCTURES

There is a striking 60% similarity and 39% identity between MatC and BugD (Huvent *et al.* 2006a), which is also reflected by the similarity of their ligands, as BugD was crystalized with a bound L-aspartate molecule. Aspartate and malate have similar

structures and binding interactions, with an amine in aspartate replacing the hydroxyl group on carbon 2 in malate. The substrate coordination overlaps almost entirely in both proteins, as shown in Figure 6-7 (blue and magenta dots for MatC and BugD, respectively), and only two residues involved in hydrogen bond coordination aren't conserved: Val150 and Gly175 (in MatC), both substituted in BugD for alanine, which seem to be silent mutations, as the coordination involves only main chain atoms and the hydrogen bond coordination is maintained. Curiously, there is no evidence for binding of Aspartate by MatC, as tested through differential scanning fluorescence and tryptophan fluorescence assays. Surprisingly subtle variations obtained in specificity are also seen in AdpC (Rosa *et al.* 2017).

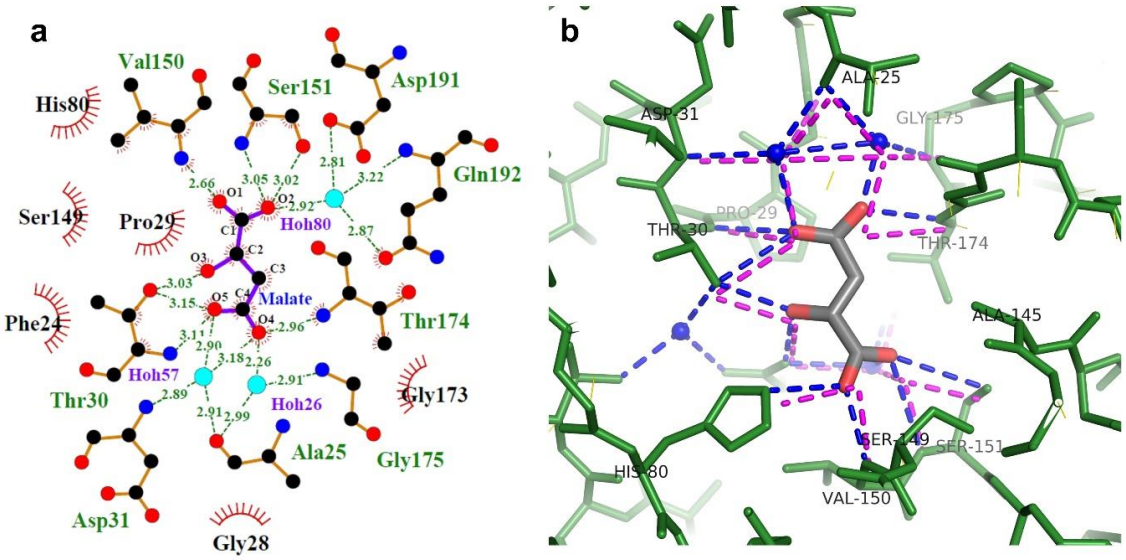


Figure 6-7: D-Malate coordination in MatC binding pocket. a) Coordination of malate molecule using the software LigPlot+ (Laskowski and Swindells 2011). Traces represent hydrogen bond interactions with the respective distances, red dashes represent hydrophobic interactions**.** Coordination of the α-carboxylic group by two water molecules is a conserved feature among the TTT family, bridging hydrogen bonds with the main chain of residues **b) Comparison between D-malate coordination in MatC (structure and blue dashes) and Aspartate coordination in BugD (Magenta dashes). D**ashes represent hydrogen bonds, and blue spheres represent water molecules in the MatC structure.

6.9. KNOCK-OUT MUTANTS OF THE *TTTBA1* AND *MATBA* SYSTEMS SHOW NO DISTINCTION IN GROWTH PHENOTYPE

In order to characterize the importance of the TTT family in C4-dicarboxylic acids uptake, a marker-free double mutant in the transmembrane components of the *tttBA1* and *matBA* was generated (Figure 6-8a). The option to delete both TTT systems was made in order to rule out the possibility that one TTT SBP would be able to interact with more than one transmembrane component, as suggested previously (Rosa *et al.* 2018a). As shown in Figure 6-8b, no phenotypical difference was observed between the wild type and the mutant strains, when using 10 mM of D/L-malate, succinate or fumarate as the sole carbon source, suggesting the TTT family is not essential for the uptake of C4-dicarboxylates under these conditions. Only a minor decrease in the growth rates of the *ΔtttBA1* is observed, which is not maintained in the double mutant and is likely to be a result of small methodological variation, given the final OD_{660} of the wild type strain is no more than 0.25 under growth with fumarate.

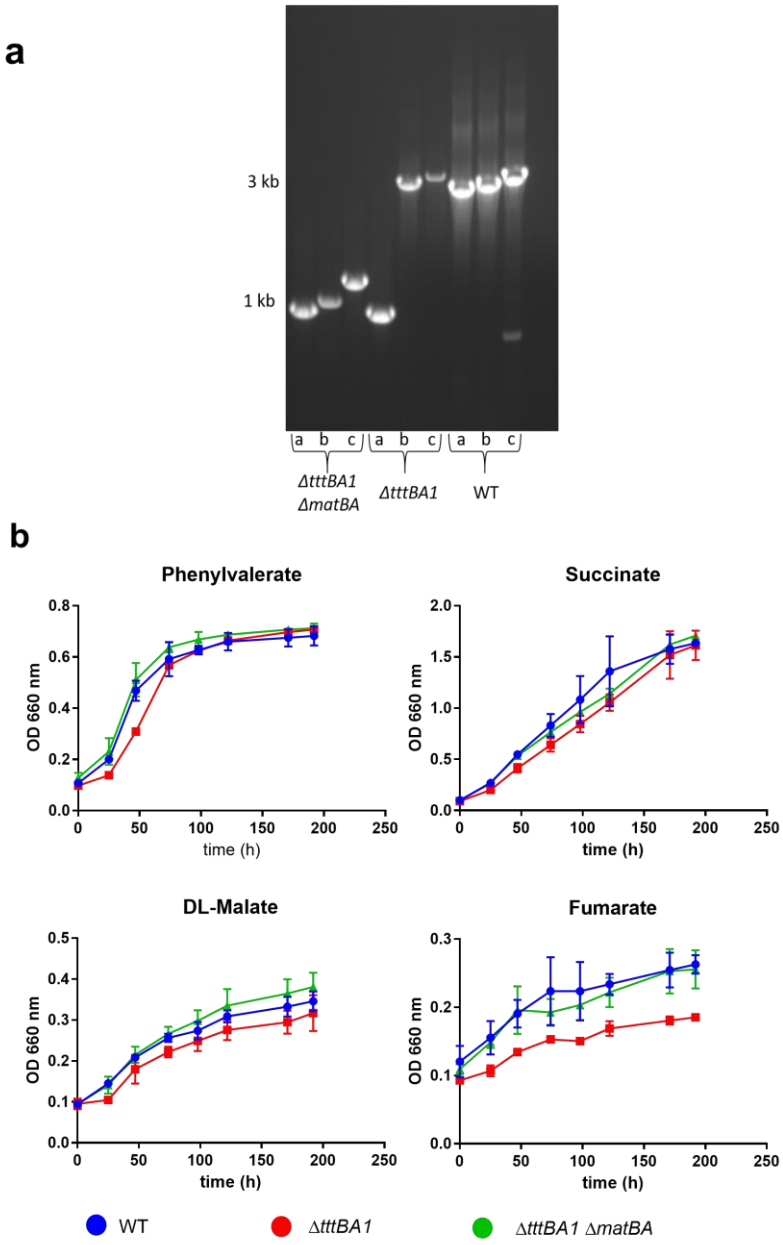


Figure 6-8: Genotypic and phenotypic characterisation of the *R. palustris ΔtttBA1 ΔmatBA* double mutant. a) amplification of deletion regions in *R.palustris* genomic DNA. Primer pair "a" refers to the flaking region of *tttBA1*, expected to be 2.9kb in the wild type (WT) and 0.9 kb in case the *tttBA1* genes are deleted. Primer pair "b" refers to the flanking regions of *matBA*,expected to be 3 kb in the WT and 1 kb in case the *matBA* genes are deleted. Primer pair "c" refers amplifies a region including 100 bp upstream of each side in primer "b", in order to ensure the amplification is not an artefact due to plasmid maintenance in the genome, and is expected to be 3.2 kb in the WT and 1.2 kb in case *matBA* is deleted in the genome **b)Growth experiments comparing *Rhodopseudomonas palustris* wild type and label-free knockout mutants lacking the tttBA1/2 transmembrane components.** Growth was performed photoanaerobically in the presence of 10 mM substrates, in biological triplicates, and absorbance measured at 660 nm**.** No significant distinction is observed between each strain in the presence of succinate, fumarate, malate or phenylvalerate (used as a control), apart from a small decrease in the growth of *ΔtttBA1* with fumarate.

As described in the discussion session, this lack of phenotypic difference was predicted, and is a result of the redundancy of C4-dicarboxylic acids uptake by different transport systems.

6.10. BIOCHEMICAL CHARACTERISATION OF THE MATBA SYSTEM

As detailed in Rosa *et al.* (2018a), knowledge about the physiology of the transmembrane components of the TTT family is poor, relying on the domain prediction by Winnen *et al.* (2003), the computational models provided by Batista-Garcia *et al.* (2014) and the physiological studies provided by Hosaka *et al.* (2013). This study describes, then, the first attempt to biochemically characterise such transmembrane components, using the MatBAC system as a model, investigating the uptake of ^{3}H-succinate in proteoliposomes. This approach was undertaken under co-supervision of Dr. Christopher Mulligan, from the University of Kent. The reconstitution and transport experiments were performed in Dr. Mulligan's laboratory in May 2018.

6.10.1. OPTIMISATION OF MATBA RECOMBINANT EXPRESSION

The MatBA complex is predicted to have a molecular weight of 70.1 kDa, formed by the larger subunit TttA, of 52.6 kDa, and the smaller subunit TttB, of predicted 17.5 kDa. The study of the SiaPQM TRAP transporter by Mulligan *et al.* (2012) showed that the co-expression of the two transmembrane components in the native genetic construction retrieved higher overexpression and improved stability of the complex compared to the separated expression of each component. Thus, the subunits *matBA* were expressed in a single construct, preserving the intergenic regions, and inserted into two alternative cloning vectors: pBAD-TM vector, a modified version of pBAD,

contain a C-terminal His_{8x}-Tag, and, consequently, the MatA subunit only would contain this tag. The pET-TM vector, on the other hand, contain a N-terminal His_{10x}-Tag, and consequently only the MatB subunit would be tagged. As a consequence, the purification of the MatBA complex would depend on the close interaction between the two subunits. Although both systems were constructed simultaneously, the optimisation in each vector is shown separately for didactic purposes.

6.10.1.1. MatBA optimisation in pET-TM vector

As detailed in the methods section, overexpression of MatBA was optimised in terms of temperature, time of induction and inducer concentration. After overexpression, a total membrane fraction was prepared, in order to avoid masking of the membrane components in the SDS-PAGE gels from the much more abundant soluble proteins in the cell. Even then, given the high protein levels of major membrane protein components, such as porins and channels, visualisation would be hard in Coomassie stained gels, and so MatBA expression was assessed through western blot, with an anti his-tag primary antibody.

In the pET-TM-MatBA construct, the MatB subunit contains a His_{10x}-tag in its N-terminal portion, resulting in a 21 kDa protein. Using 0.4 mM IPTG, initial optimisation of time and temperature showed induction for 1 h at 37 °C as the optimal expressing condition (Figure 6-9, top). This condition was then used for optimisation of IPTG concentration, which confirmed 0.4 mM IPTG concentration as the optimal for expression (Figure 6-9, bottom).

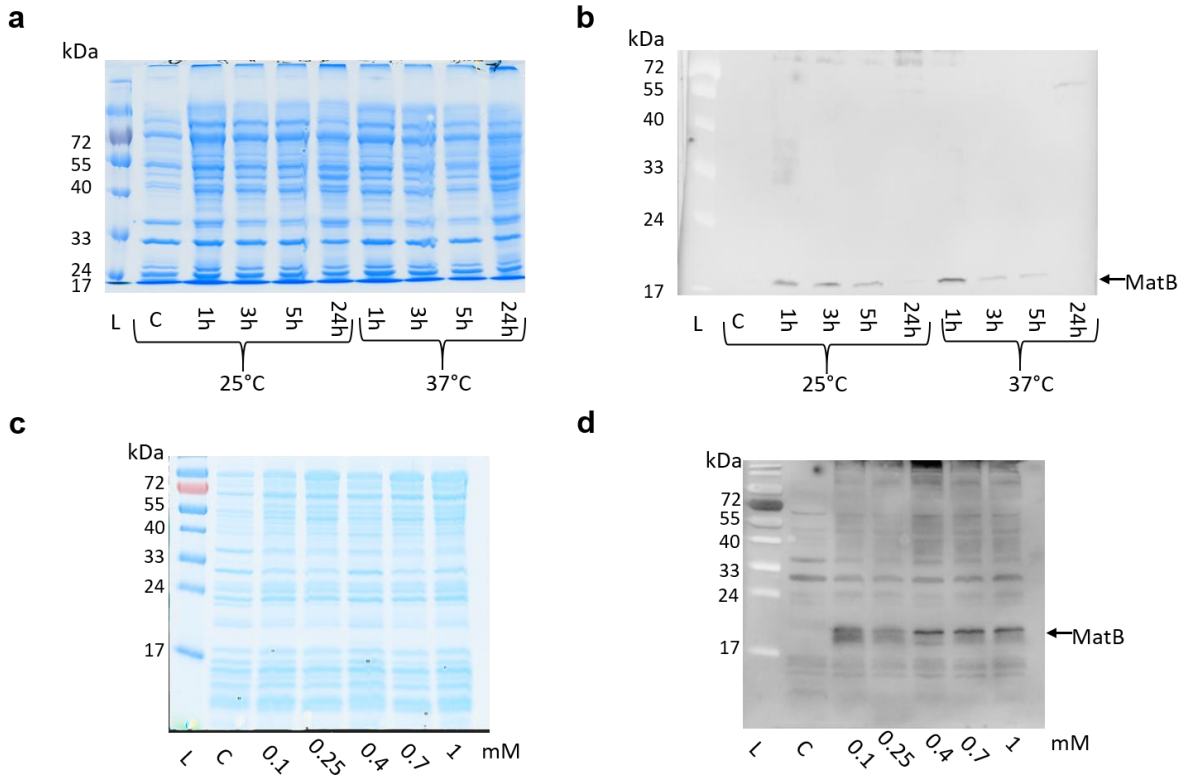


Figure 6-9: Optimisation of MatBA overexpression in pET-TM vector. The His-tagged MatB subunit is expected to have 21 kDa, and is indicated by an arrow in the western blots (**b** and **d**). Optimisation of time and temperature, performed with 0.4 mM IPTG (**a** and **b**), and optimisation of IPTG concentration, performed at 37 °C for 1 h (**c** and **d**). Visualisation was performed using comassie staining of the 12% polyacrylamide gel (**a** and **c**) and western blot with Anti-His primary antibody (**b** and **d**). Leader (L); Control (C)

Purification of MatBA was attempted with membranes from a 6 L culture batch with Ni^{2+} affinity chromatography (Figure 6-10a), which retrieved protein in a purity high enough for us to attempt liposome reconstitution. While TttB ran with the expected size of 21kDa, the TttA subunit ran with an apparent molecular weight of 40 kDa rather than the 52.6 kDa expected. Many membrane proteins are described to run in the SDS-PAGE gel with an apparent lower molecular weight, suggested due to the incomplete unfolding state of the preoteins given their high hydrophobicity, and also the higher number of SDS molecules interacting with the hydrophobic regions (Heller 1978, Walkenhorst *et al.* 2009), which would result in a more negative charge when compared to a soluble protein of the same size (Rath *et al.* 2009). The bands

corresponding to the MatB subunit were seen to be more intense than the ones representing MatA, suggesting the interaction between the two subunits in the complex was not strong enough to ensure co-purification. Alternatively, it could be the case that MatBA assembled in a different stoichiometry than the suggested 1:1, with higher copy number of MatB, diverging from the results found for the TRAP system SiaPQM in *Vibrio cholera* (Mulligan et al. 2012*)*.

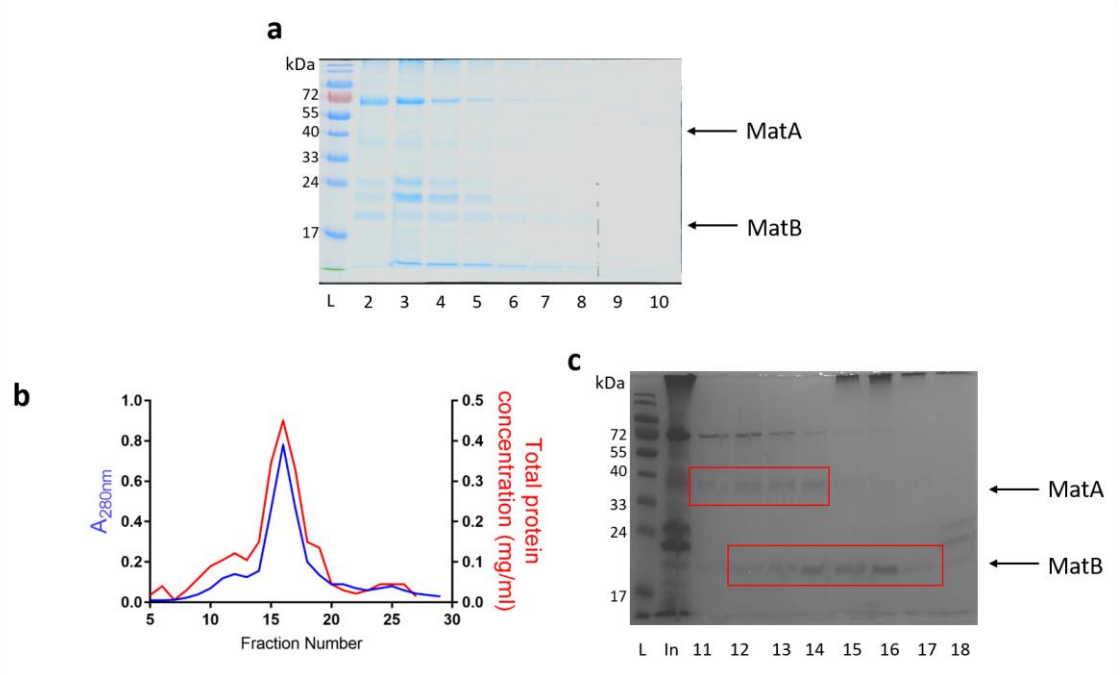


Figure 6-10: Purification of MatBA after overexpression from the pET-TM vector. (a) 12% polyacrylamide SDS-PAGE gel of MatBA purification using Nickel affinity chromatography, eluted in 500 µl fractions. His-tagged MatB subunit is expected to have 21 kDa, and MatA subunit is expected to have 52.6 kDa, but our observations is that its apparent molecular weight is ~40 kDa. **(b)** Size exclusion chromatography of MatBA, performed using a 11ml Superdex200GL column equilibrated with Tris 20 mM pH 7.4, 0.2 M NaCl, 2 mM DDM and run at a 0.5 ml/min flow rate in an AKTA instrument. A calibration curve was generated using thyroglobulin (669 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and aprotinin (6.5 kDa). A_{280} absorbance is shown in blue, and calculated total protein concentration is shown in red. **(c)** 12% polyacrylamide SDS-PAGE gel of fraction 11-18 from the SEC experiment. Elution of MatA and MatB are indicated by a red square in the gel, and an arrow in the right side. A 72 kDa band is also observed. leader, L; Injected sample, In.

In order to test whether the proteins were forming a stable complex in solution, Size exclusion chromatography was performed in the purified protein (Figure 6-10b). The

elution profile showed a single peak, assembled in a normal-shaped curve, suggesting the complex was forming in a stable way. Assuming this peak is only composed by MatBA and adding together the protein amounts retrieved in fractions 11-17, a total of 1.9 mg of protein per 6 L batch was purified.

However, when these fractions were assessed in SDS-PAGE gel (Figure 6-10c), it was shown that a third band, of 72 kDa, was observed in the elution profile. Although this is the approximate size of the MatBA complex, it is highly unlikely that the two proteins would remain as a complex under the highly denaturing conditions of the SDS-PAGE gel, and so that band is more likely to be a sample contaminant (See Mass spec fingerprint in next section). Moreover, the elution profile of the MatA and MatB subunits are slightly shifted, with most of the larger subunit eluted in fractions 11-14 and the smaller subunit in fractions 13-17. This is likely a result of weak interactions inside the complex, leading the proteins to separate in the presence of DDM detergent (Mayer *et al.* 2009, Hu *et al.* 2012). We speculated that a possible reason for this event to occur was the presence of the N-terminal His_{10x}-tag in MatB, which could be impairing the contact between the two subunits. In order to assess if this was the case, expression was attempted with a second vector, pBAD-TM, which contains a C-terminal His_{8x}-tag, present in the C-terminal of the MatA subunit.

6.10.1.2. MatBA Optimisation in PBAD-TM vector

In parallel to the previous construct, the pBAD-TM-MatBA was assayed in order to find the optimal expression conditions. An initial trial was performed at 37 °C with 0.002% L-arabinose, for different times (Figure 6-11, top), where 3-24 hours inductions seemed to produce similar relative amount of protein. As the cell concentration after 24 h growth is substantially higher than after 3 h, and slightly higher than after 5 h (data not shown), we assumed a 24 h induction time would retrieve a higher protein

amount. We then assessed different temperature and L-arabinose concentrations (Figure 6-11, bottom) and concluded that induction for 37 °C for 24 h with 0.002% L-arabinose would be the optimal condition for MatBA expression. Consistently with the pET-TM construction, the His_{8x}-tagged TttA protein, of predicted 54.6 kDa, had an apparent molecular weight of ~40 kDa when ran in the SDS-PAGE gels.

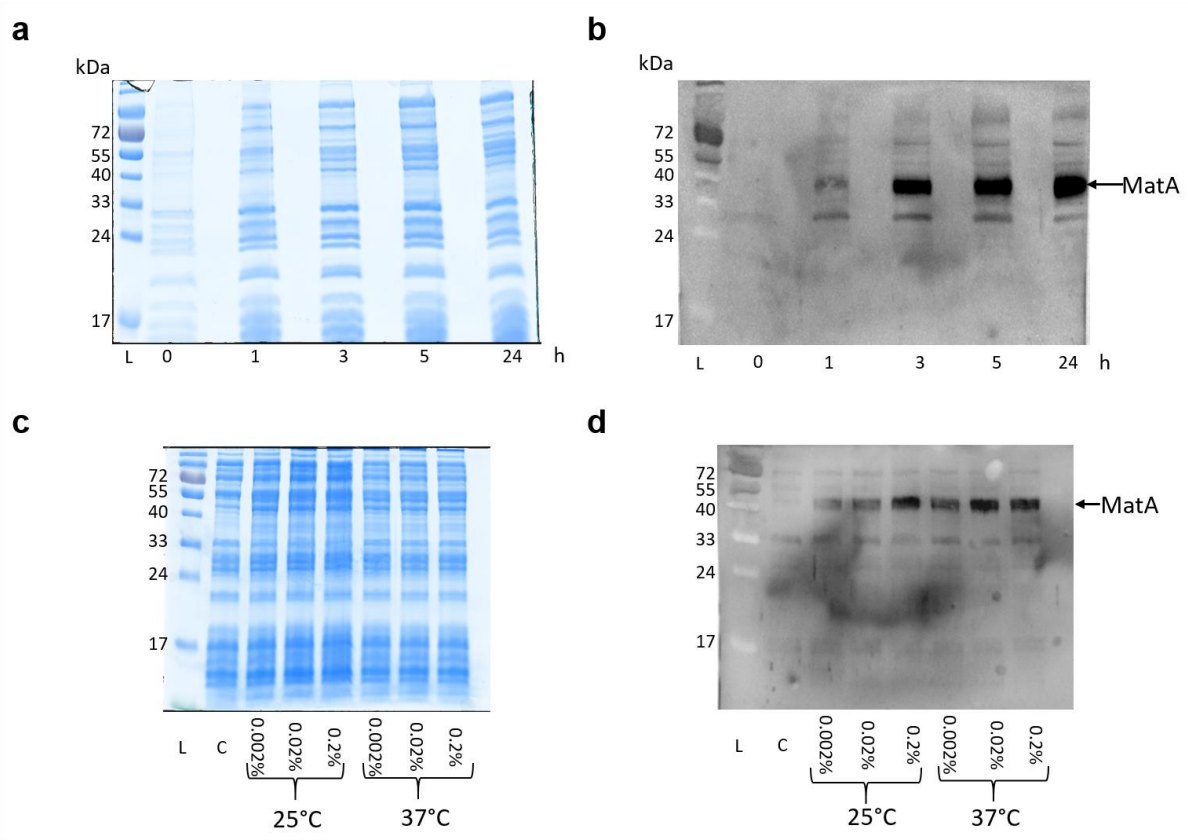


Figure 6-11: Optimisation of MatBA overexpression in pBAD-TM vector. The His-tagged MatA subunit is expected to have 54.6 kDa, but was observed to have an apparent molecular weight of ~40 kDa in the gel, indicated by an arrow in the Western blots (b and d). Optimisation of time of induction, performed at 37 °C and 0.002% L-arabinose (a and b), and optimisation of temperature and inducer concentration, performed with a 24 h induction time (c and d). Visualisation was performed using coomassie staining of the 12% polyacrylamide SDS-PAGE gel (a and c) and western blot with Anti-His primary antibody (b and d). Leader (L); Control (C)

Similarly to the previous construct, purification of MatBA was attempted from the membranes of a 6L batch overexpression, which showed protein pure enough to allow

the reconstitution into liposomes intended, although some impurities could still be found (Figure 6-12a). Using the extinction coefficient of the MatBA 1:1 complex (63285 M^{-1} cm^{-1}) and assuming all the present protein would be MatBA, a total amount of 1.61 mg of protein was calculated for the 6 L batch.

In order to see how stable was the complex in DDM solution, a size exclusion chromatography of fraction 2-6 was performed. A reasonable normal curve was observed (Figure 6-12b), although a second peak was seen in fractions 14-16 and a third peak in fractions 22-24. Assessing the elution profile in SDS-PAGE gel showed co-elution of the two subunits between fractions 11-15, and the brightness of the coomassie staining in the two bands is more similar to the assumed 1:1 stoichiometry in the MatBA complex (Figure 6-12c) (Mulligan *et al.* 2012). The gel still revealed two additional bands, of 55 kDa and 72 kDa, as also seen in the pET-TM purification, which could represent a tight interaction between the two subunits even in denaturing conditions, but mass-spectrometry fingerprint of these samples revealed otherwise (see below).

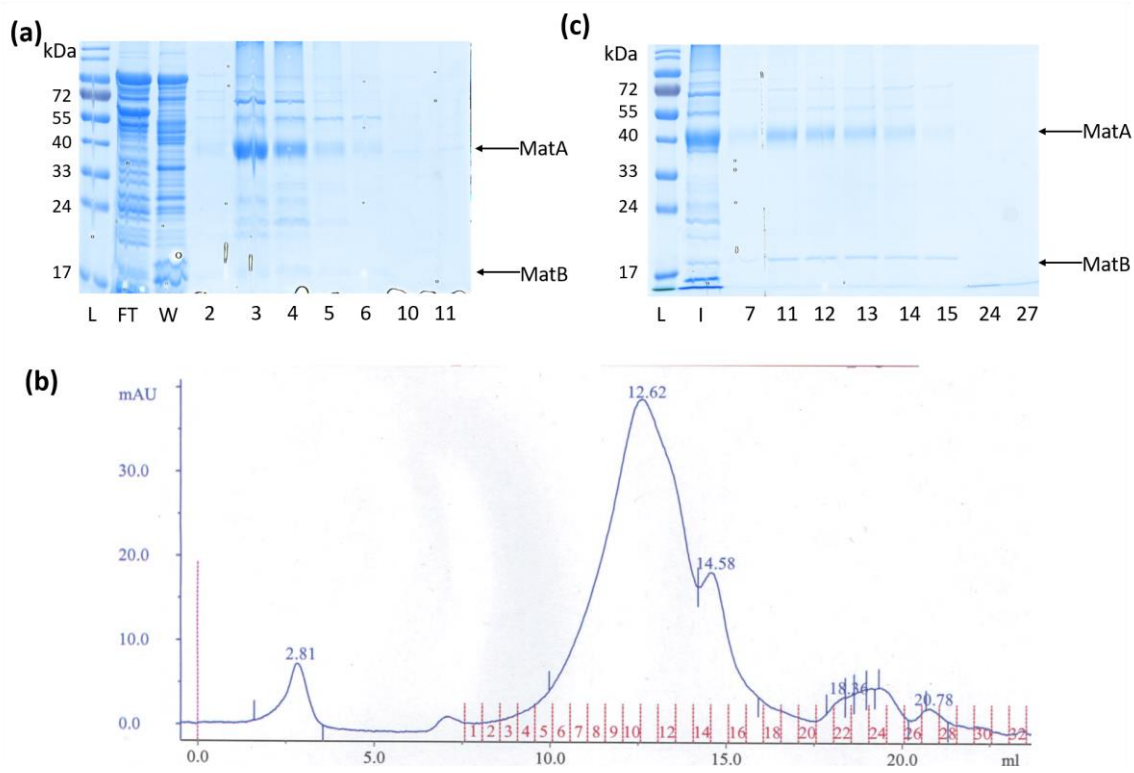


Figure 6-12: Purification of MatBA after overexpression from the pBAD-TM vector. (a) 12% polyacrylamide SDS-PAGE gel of MatBA purification using Nickel affinity chromatography, eluted in 500 µl fractions. His-tagged MatA subunit is expected to have 54.6 kDa, but our observations is that its apparent molecular weight is ~40 kDa. MatB subunit is expected to have 21 kDa **(b)** Size exclusion chromatography of MatBA, performed using a 11ml Superdex200GL column equilibrated with Tris 20 mM pH 7.4, 0.2 M NaCl, 2 mM DDM and run at a 0.5 ml/min flow rate in an AKTA instrument. A calibration curve was generated using thyroglobulin (669 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and aprotinin (6.5 kDa). A_{280} absorbance is shown in blue, and fraction numbers in red spikes in the bottom. **(c)** 12% polyacrylamide SDS-PAGE gel of fractions 7-15, 24 and 27 from the SEC experiment. Elution of MatA and MatB are indicated by an arrow in the right side. A 52 kDa and a 72 kDa band are also observed. leader, L; Injected sample, I.

Two calibration curves were performed for the 11 ml Superdex200GL column used for MatBA purification, either in presence or absence of DDM detergent. In presence of DDM, the elution of MatBA would be equivalent to a complex of 400 kDa, and in the calibration curve without DDM the elution profile would be equivalent to a complex of 560 kDa. These data suggest that MatBA complex forms an oligomer of five to seven monomers. However, this assumption requires further validation, either with cross-linking assays or even cryo-Electron Microscopy, and there is a high probability of this being an artefact, resulting from the poor condition of the column, interference of the

detergent or interaction between the hydrophobic regions of MatBA in a non-native environment, i.e., absence of the lipids naturally surrounding the protein (see discussion session)

The bands of apparent molecular weight 17 kDa, 40 kDa, 55 kDa and 72 kDa were extracted from the SDS-PAGE gel and submitted to trypsin digestion and subsequent protein mass spectrometry fingerprint. The results showed that the 17 kDa and the 40 kDa represented, as expected, the MatB and MatA proteins, respectively. 91.22% of all identified peptides from the 17 kDa band belonged to MatB based on the iBAC values, altogether with traces of MatA accounting for 0.09%. In the 40 kDa band, 93% of the identified peptides belonged to MatA, while MatB accounted for 0.03%. The 55 kDa and 72 kDa bands, however, are significantly related to contaminants from the expression host. In the 52 kDa band, 74.66% accounted for an uncharacterised protein from *E. coli* (A0A140N6F9) and 17.84% accounted for an *E. coli* pyruvate kinase, while MatA and MatB accounted for 4.99 % and 0.54%, respectively. In the 72 kDa band, 93.96% of the identified peptides were of *E. coli* glutamine-fructose-6-phosphate aminotransferase, while MatA and MatB accounted for 2.34% and 0.24%, respectively.

Assuming that all protein found in the SEC fractions 11-15 were MatBA complex and using the 1:1 stoichiometry complex extinction coefficient (63285 M^{-1} cm^{-1}), the total amount of protein purified in the peak was of only 96 µg of protein per 6 L batch expression. This represents only 6% of the total protein calculated from the Ni-NTA purification, representing a 94% loss. However, our calculations assumed all protein present is composed of the MatBA complex, which we demonstrate is a false premise, used only for estimations, but still it is a considerable loss. Given that it is shown the complex is stable and co-purified, our choice was to use the elution profile from the

Ni-NTA purification for the reconstitution in proteoliposome attempts, without further purification in SEC.

6.10.2. RECONSTITUTION OF MATBA INTO PROTELIPOSOMES AND SUCCINATE UPTAKE ASSAYS

Once the right conditions for the expression of the MatBA complex were stablished and the protocols for the purification of a stable and pure enough complex were set, several batches of MatBA containing membranes were obtained, in order to provide enough material for the proteoliposome reconstitutions. Eight batches of six litres each were grown, adding up to a total of 48 L of culture. Each of these batches were stored separately, and Figure 6-13 shows that the MatA subunit could be visualised through immunoblot in all of them.

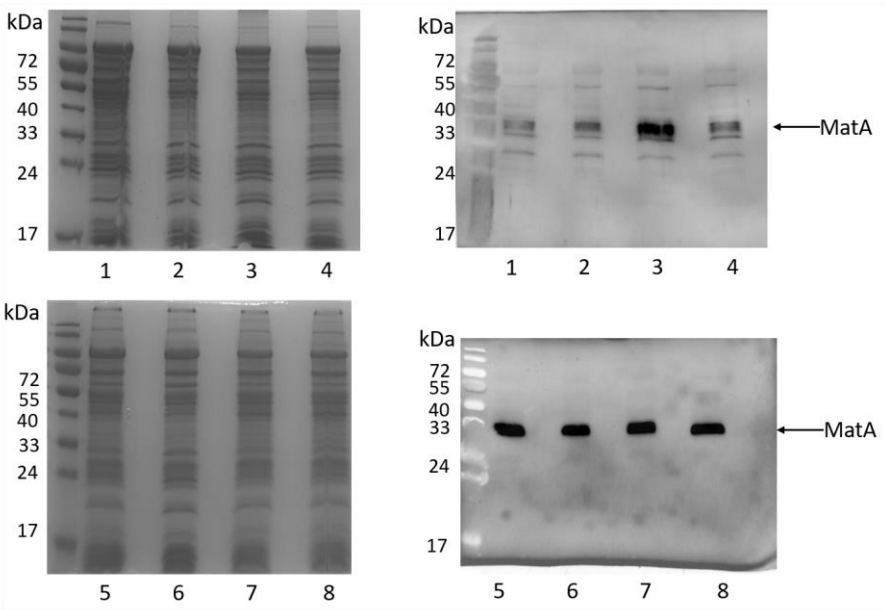


Figure 6-13: 12% Polyacrylamide SDS-PAGE gels (left) and western blot visualisation (right) of each of the eight batches of MatBA containing membranes used for proteoliposome reconstitution, each from a 6 L batch culture. The His-tagged MatA subunit is expected to have 54.6 kDa, but has an apparent molecular weight of ~ 40 kDa in the gel, and is indicated by an arrow in the right side of the western blots.

In order to assess the energy requirements of the TTT family, the MatBA complex was reconstituted into proteoliposomes and uptake assays were performed with ^{3}H-succinate. Several different reconstitution methods are described in the literature, each one with their limitations and particularities, and a brief overview of them is presented by Seddon *et al.* (2004). Each membrane protein seems to respond better to a different reconstitution method and detergent composition, and thus it is important that different methods are attempted. In this study, we attempted the rapid dilution method and the bio beads removal method for proteoliposome formation, benefiting from Dr. Mulligans experience with the characterisation of TRAP transporters membrane components (Mulligan *et al.* 2009, Mulligan *et al.* 2012).

Although the highest ligand affinity for MatC was found in the presence of malate, ^{3}H-succinate was found to be considerably less costly, and more easily available. Considering that MatC showed an affinity of 0.1 µM in the presence of succinate, it was speculated that this would also be a natural substrate for the MatBAC system, and thus it was the selected radiolabelled substrate for this study.

6.10.2.1. Rapid dilution reconstitution

The rapid dilution method consists in solubilizing the membrane protein with an non-ionic detergent at 1.5 times the Critical Micelle Concentration (CMC) (Rigaud and Levy 2003, Seddon *et al.* 2004). Once purified, and mixed with the prepared lipids, the 3.5 ml of the mixture protein: detergent: lipid is quickly diluted into 65 ml of buffer. This dilution will bring the detergent concentration much below the CMC, allowing the lipids to form liposomes, which will hopefully incorporate the membrane proteins among them, forming proteoliposomes. The buffer used to dilute the sample will consequently be the buffer found in the interior of the proteoliposome.

The first attempt of reconstituting the MatBA complex was performed in the presence of 2.7 mM Decyl β-D-maltopyranoside (DM), which has a CMC of 1.8 mM. Once the proteoliposomes were formed, they were freeze-thawed to induce formation of multilamellar, large proteoliposomes, which were then extruded into 400 nm pore filters to produce standardised proteoliposomes, used in the uptake assays.

As detailed in Rosa *et al.* (2018a), the transmembrane proteins from the TTT family do not contain any ATPase predicted domains, and are described as secondary transporters, and for this reason our assays did not test the ATP requirements of MatBA. Varying the composition of the inside buffer, used in the process of the proteoliposome formation, and the outside buffer, added at the start of the uptake assay, it was possible to mimic the three most important secondary energy gradients present in the bacterial cell, as shown schematically in Figure 6-14. The pH gradient (ΔpH), was formed by a 1-unit pH difference between inner and outer buffers, making the outside proton concentration 10 times higher than the inside. The sodium gradient (ΔNa^+) was created by an excess of sodium outside, where charge and osmotic balance was maintained through presence of excess of potassium (K^+) on the inside. The membrane potential or electrical charge gradient (ΔΨ) was created through addition of 1 mM valinomycin, which inserts into the membrane, allowing the free passage of K^+ ions. Alteration of the outer buffer composition allowed for testing each of these gradients separately, or in combination.

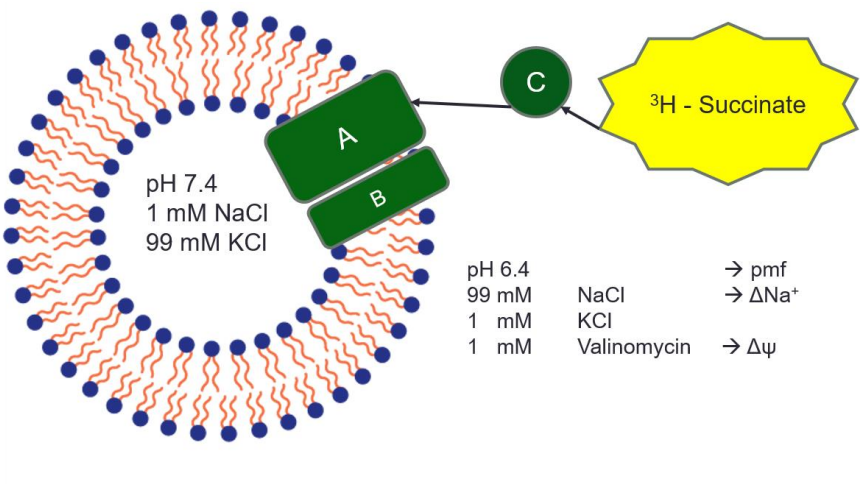


Figure 6-14: Schematic representation of the ^{3}H-succinate uptake assay using MatBA in reconstituted proteoliposomes. The *E. coli* polar lipids 400 nm liposomes are represented as blue polar heads with orange apolar tails, MatBA is represented as green retangles embedded in the membrane, the MatC SBP is represented as a green circle and the radiolabelled succinate represented in yellow. The diagram is not represented to scale. The buffer constitution inside the proteoliposomes is Tris-HCl 50 mM pH 7.4, 1 mM NaCl and 99 mM KCl. The three most relevant secondary energy gradients from the bacterial membrane are mimicked by having an external buffer with Tris-HCl 50 mM pH 6.4 (pmf), 99 mM NaCl (ΔNa$^+$), 1 mM KCl and 1 mM Valinomycin, which will allow the free flow of K$^+$ ions through the membrane (ΔΨ)

The initial attempt was to use all three gradients at once, to check if the MatBA complex was successfully reconstituted and active. However, as shown in Figure 6-15a, no uptake activity was observed when compared with protein-free liposomes, suggesting the proteins were not active. As discussed above, each membrane protein responds differently to multiple detergents and reconstitution methods. Thus, we attempted to repeat the rapid dilution method purifying the MatBA with a different detergent, Octyl β-D-glucopyranoside, at 35 mM, 1.5 times its CMC (20 mM). performing the uptake assay with the three gradients, as before, we could see an incremental accumulation of ^{3}H-succinate into the proteoliposomes, in a linear pattern, compatible with what would be expected if uptake was active (Figure 6-15b-blue). However, when we prepared protein-free liposomes in the same manner (Mixing OG containing Buffer with the lipids, but no protein), our observations showed a similar pattern, with very close final scintillation counts when compared to the assay (Figure

6-15b-Red). Our suggestion is that the OG detergent, having a much higher CMC than DM, would not have been completely removed from the mixture during the dilution procedure (Seddon *et al.* 2004), and would be consequently destabilising the formed proteoliposomes, leading to non-specific incorporation of the radiolabelled substrate.

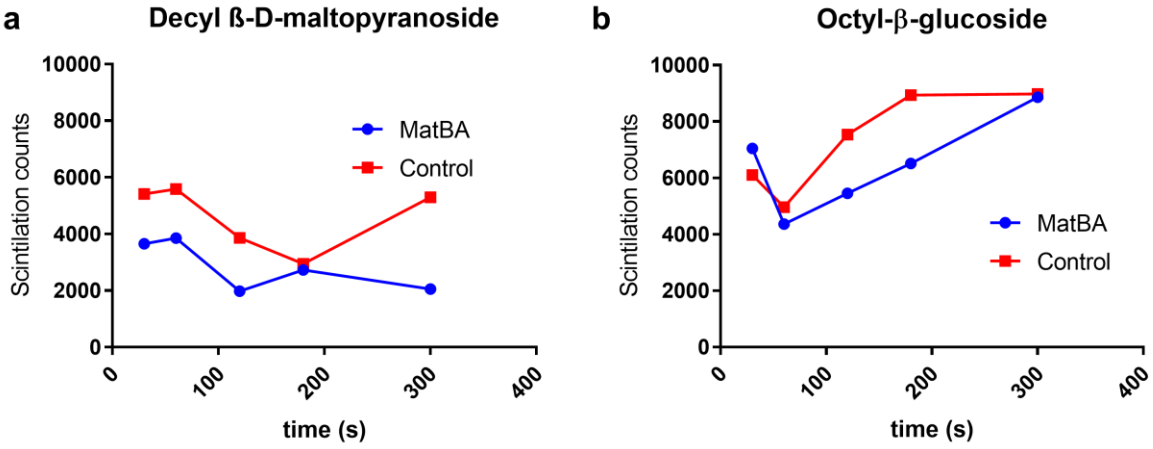


Figure 6-15: Radiollabelled succinate uptake assay with MatBA complex reconstituted into proteoliposomes through the rapid dilution method, using either Decyl-β-D-maltopyranoside (DM) (a) or Octyl-β-glucoside (OG) (b). While no uptake rate is observed in (a), a linear uptake rate is shown in (b-blue). However, the protein-free control (b-red) showed a similar final count, suggesting the proteoliposomes in (b) are leaky and leading to misincorporation of radiollabelled substrate.

Given that our attempts of reconstitution with two different detergents failed to show uptake activity, we moved on to a different reconstitution method, based on the Bio beads removal of detergent from solution.

6.10.2.2. Biobeads detergent removal

Instead of making use of a rapid dilution procedure to quickly drop the detergent concentration in the sample, which can be a too drastic approach for some membrane proteins, thhe Biobeads detergent removal method makes use of polystyrene Biobeads, which will interact and chelate the detergent molecules slowly and gently, allowing incremental stabilisation of the proteoliposomes (Rigaud and Levy 2003, Seddon *et al.* 2004).

Once prepared as described in methods section, the *E. coli* polar lipids are extruded in 400 nm pore filters in order to form unilamellar and uniform liposomes. Before protein addition, however, the liposomes are titrated with Triton X-100 detergent, in order to destabilise the liposome, but not solubilise them completely, creating an environment unstable enough that the proteins, when added, can intercalate between the lipids. The titration of liposomes is assessed at 540 nm absorbance, as shown in Figure 6-16a. With an initial OD_{540} of ~ 0.8, the incremental addition of Triton X-100 cause the liposomes to swell, increasing the absorbance, until it reaches a saturation point, at the OD_{540} of ~ 1.0. Beyond this point, the addition of further detergent will cause de solubilisation of the lipids and liposome disruption, causing the absorbance to decrease. Although protein incorporation into liposomes was shown to happen regardless of Triton X-100 detergent, experiments assessing different points in the saturation curve showed the highest incorporation rate when protein was added slightly after the saturation point, and thus we added the purified MatBA in DM solution to the liposomes 5 aliquots of detergent after the saturation point (marked with an arrow in Figure 6-16a).

After protein addition, several incubations with the polystyrene BioBeads followed, in order to remove both Triton X100 and DM residual concentrations. The resulting proteoliposomes were freeze-thawed three times and extruded again before the experiments.

Performing the transport assays similarly to what was done before, no uptake could be observed in the proteoliposomes formed, when compared to the liposome control without protein (Figure 6-16b)

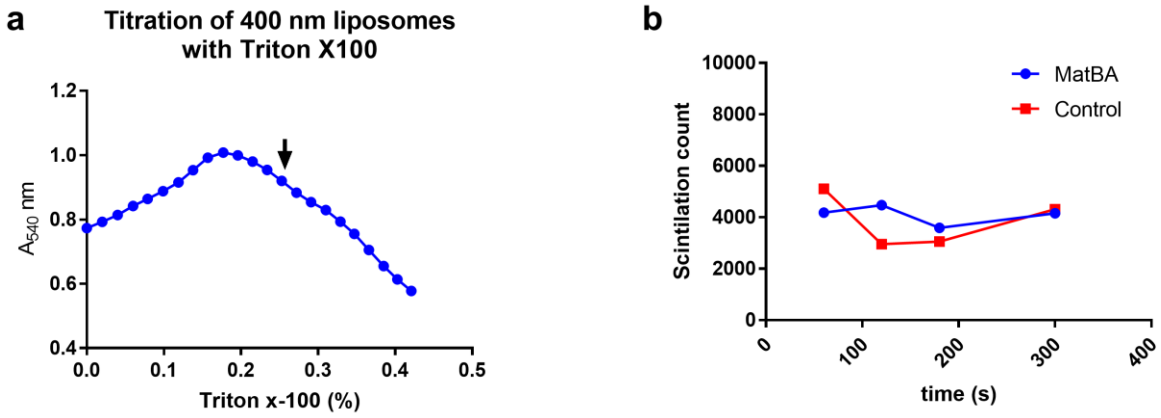


Figure 6-16: Reconstitution of MatBA into proteoliposomes using the Biobeads removal method. (a) Titration of liposomes with Trtom X-100 detergent, monitored by the A_{540}. As the detergent is added, absorbance increases as a result of liposomes swelling, until it reaches a saturation point, and solubilisation of lipids and liposomes disruption starts, culminating in a decrease in absorbance. Triton X-100 was added until 5 aliquots after the saturation point, indicated by a black arrow, point in which the purified MatBA complex was added. (b) Radiollabelled succinate uptake assay with MatBA complex reconstituted into proteoliposomes through the biobeads detergent removal method. No uptake of radiollabelled substrate was observed, when in comparison with the protein-free liposomes.

We showed that we could overexpress and successfully purify a stable MatBA complex using DDM detergent and attempted to reconstitute it into proteoliposoomes using different detergents and methods. In order to assess the successful incorporation of MatBA, the remaining proteoliposome samples used for the uptake assays were run in a SDS-PAGE gel with 1% SDS, and Figure 6-17 shows a clear presence of the complex in the reconstitution using the rapid dilution method, and a weaker, but still evident presence in the Biobeads reconstitution method when compared to the protein-free liposome controls. We show that the lack of activity observed in the uptake assays is not due, thus, to the absence of incorporated protein.

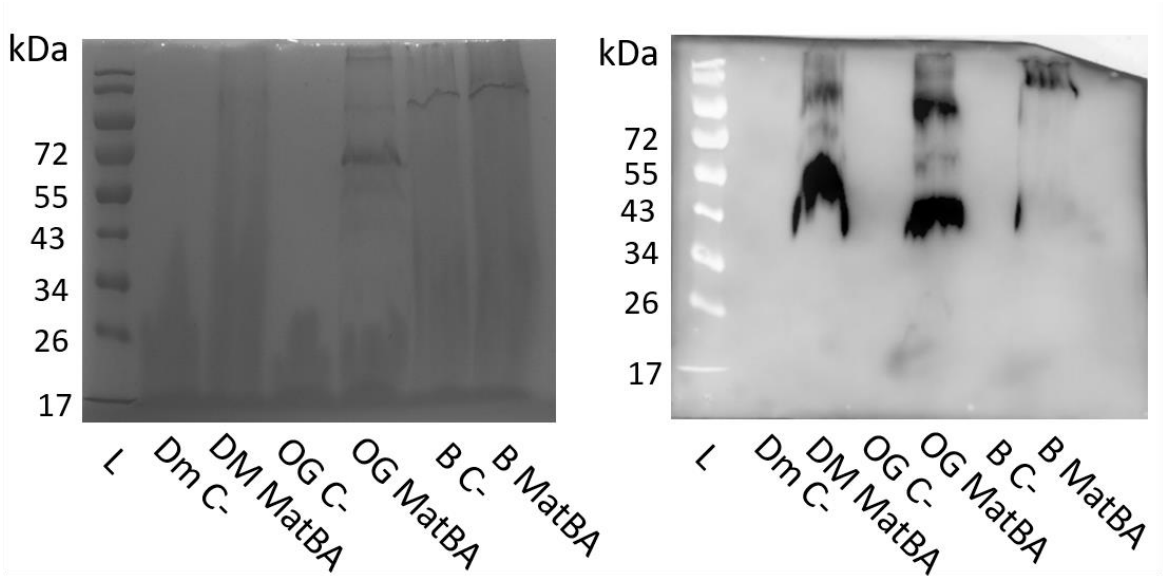


Figure 6-17: Monitoring of the MatBA reconstitution into proteoliposomes, through comassie stained 12% polyacrylamide SDS-PAGE 1% SDS gel (left) and western blot (right). His-tagged MatA is expected to have 52.6 kDa, but have an apparent molecular weight of ~40 kDa when running on the gel. The high lipid content of the sample decreases overall resolution, however immuno visualisation can be observed clearly in the MatBA containing proteoliposomes reconstituted through the rapid dilution method either with DM and OG detergent ("DM MatBA" and "OG MatBA", respectively) in comparison with the protein-free liposomes ("DM C-" and "OG C-", respectively). A weaker, but also present imunnovisualisation is observed in the proteoliposome reconstituted using the Biobead removal method ("B tttBA2") in comparison to its respective control ("B C-"). leader (L)

Several factors can influence the stability, incorporation and activity of membrane proteins (see discussion below) and certainly there are several approaches which could be taken in order to improve the MatBA activity, however time constrictions were prohibitive and so may this study serve as a foundation for the further study of the MatBA uptake system.

6.11. DISCUSSION

In this chapter we aimed to study the transport properties and physiological roles of the MatBAC uptake system, one of the two complete tripartite systems from this family in *Rhodopseudomonas palustris*, which contains also 5 orphan SBPs. This

characterisation was done through through genetic analysis, binding assays, protein crystallography, overexpression and reconstitution of the membrane components into proteoliposomes and uptake of radiolabelled substrate.

Upon screening of MatC against a library of compounds (provided in Rosa *et al.* (2017)) using Differential Scanning Fluorescence, we identified protein thermal stabilization in the presence of the C4-dicarboxylates malate, succinate and fumarate. Furthermore, the screening showed MatC has high ligand specificity, only allowing C4 dicarboxylate binding with restricted substitution at C2. This degree of specificity is present in some of the TTT SBP characterized so far, such as the therephtalate binding protein ThpC (Hosaka *et al.* 2013) and the sulpholactate binding protein SlcH (Denger and Cook 2010), but is not seen in others such as AdpC, which binds molecules from six to nine carbons in length (Rosa *et al.* 2017) or in Bug27, which is found to bind to nicotinate, nicotinamide, benzoate and citrate (Herrou *et al.* 2007). MatC showed no indication of binding to citrate, the prototype ligand which gave name to the family and binding of which is characterized in many organisms such as *Salmonella typhimurium*, *Bordetella pertussis, Corynobacterium glutamicum* and *Lactococcos lactis* (Sweet *et al.* 1979, Herrou *et al.* 2007, Brocker *et al.* 2009, Pudlik and Lolkema 2012).

Titration of MatC with substrates in tryptophan intrinsic fluorescence analysis revealed a binding affinity of ~20 nM for malate, not distinguishing between D and L-isomers. This affinity is two orders of magnitude higher than the highest affinity described previously for DctA homologs and for members of the DAAS family (Yurgel and Kahn 2004, Mulligan *et al.* 2014), demonstrating the effectiveness of solute binding protein mediated transport. In addition, K_d values for MatC are at least half of the 50 nM affinity described for the *R. capsulatus* DctP homolog, belonging to the

TRAP transporters family (Kelly and Thomas 2001), making the MatBAC from *R. palustris* the uptake system with highest reported affinity for C4-dicarboxylates, to the best of our knowledge. MatC also showed higher affinity than the sub micro molar affinity found for other components in the TTT family (Sweet *et al.* 1979, Sweet *et al.* 1984, Herrou *et al.* 2007, Rosa *et al.* 2017). Although Sweet *et al.* (1984) demonstrated that succinate and malate were capable of displacing citrate binding in *S. typimirium* TctC when present in millimolar concentrations, this is the first high-affinity, dedicated TTT system described to bind C4-dicarboxylates.

The RT-PCR analysis of *matC* suggests an increase in expression upon presence of C4-dicarboxylates. However, deletion of both TTT transmembrane systems did not affect the growth phenotype of *R. palustris* in the presence of 10 mM of substrates. C4-dicarboxylic acids are the most used and physiologically relevant carbon source for soil phototrophic bacteria such as *Rhodopseudomonas palustris*, which are poorly capable of metabolizing sugars (van Niel 1944, Larimer *et al.* 2004). Thus, it is expected that this bacterium would have evolved to optimize the uptake of C4-dicarboxylates. *R. palustris* contains 16% of genome encoding for transport-related proteins (Larimer *et al.* 2004), and often it contains redundant transport systems for many substrates (Salmon *et al.* 2013). Indeed, the genomic analysis provided by Larimer *et al.* (2004) describe one DauA homolog (RPA4326), two DctA homologs (RPA0552 and RPA2448), and eight different TRAP transporters, having a high potential for redundant uptake of C4-dicarboxylates under different environmental conditions. Although the DctA and DauA are conventionally associated with uptake under aerobic conditions, their activity does not require oxygen, and it might be that these homologs are regulated to respond to different conditions in *R. palustris*. Furthermore, we show that MatBAC is a high-affinity uptake system, and thus it is

likely that it will have more physiological relevance under low substrate concentrations, where a higher affinity and higher specificity would be needed, a condition challenging to access using growth experiments.

MatC was successfully crystalised in the presence of 1 mM of a racemic mixture of D/L-malate, and a high-resolution protein structure was obtained with D-malate coordinated in the binding pocket. This protein is folded in a “Venus fly-trap” like conformation, characteristic of binding proteins, with two globular domains, each composed of a 5 strand β-sheet surrounded by α-helixes, separated by a cleft where D-malate is coordinated. The two domains are connected by a hinge formed by β4 and β9. These features include MatC in cluster E-II of binding protein classification, in accordance with the proposed of Scheepers *et al.* (2016), together with the remaining binding proteins of the TTT family.

Coordination of D-malate is mediated by a “pincer-like” structure formed by two β-loops between β1 and α1 and β6 and α5, which coordinate the proximal carboxylate group in D-malate through hydrogen bonds bridged by two water molecules, equally conserved. This domain is a signature among the TTT family, suggesting that carboxylates must comprise most of the ligands for this family, which has been proved right so far (Rosa *et al.* 2018a). Similarly to what was previously described in AdpC (Rosa *et al.* 2017), no positively charged residues are present in the binding pocket in order to counter-act the negative charges of the substrate carboxylic groups. In addition to the dissipation provided by the water molecules bridging hydrogen bonds with these groups, the charges might be neutralized by the N-terminal of α1 and α5, in the proximal carboxylic group, and α4 in the distal carboxylic group.

The additional hydrogen bonds provided by the hydroxyl group in the carbon 2 of malate are likely responsible for the higher affinity of MatC for this substrate, when compared to succinate and fumarate, both with no substitution in C2. Moreover, it is possible to conjecture why tartrate, oxaloacetate, mesaconate and butyrate, being carboxylic acids of similar length, did not show any indication of binding. Although oxaloacetate contains an oxygen also in carbon 2, the keto group would not allow hydrogen bonds with the Thr30 oxygen, which would cause instead an electric repulsion. The same is valid for mesaconate, which shows a methyl group in place of the hydroxyl. For butyrate, the absence of a second carboxyl group would disrupt the coordination, reinforcing the hypothesis suggested in Rosa *et al.* (2017) that not only one, but two carboxylic groups would be required for binding in this family, as the only exemption to this is the binding of nicotinate by Bug27 (Herrou *et al.* 2007). Tartrate would simply have an additional hydroxyl group which wouldn`t fit sterically in the binding pocket, likely to clash with Thr174.

Our binding assays show that MatC does not distinguish between the D and L isomers of malate. Given both the occupancy of D-malate in the binding pocket and the L-aspartate coordination in the BugD protein, we speculate that L-malate coordination is very similar to its D isomer, and all the hydrogen bonds and hydrophobic interactions found in D-malate coordination would be maintained. It might be the case, however, that in order to preserve the spatial positioning of the C2 hydroxyl and thus maintain the hydrogen bond with Thr30, L-malate is likely to fit the pocket with its C1 buried, and C4 being coordinated by the “pincer-like” structure and two water molecules, similar to what is found in BugD L-aspartate coordination (Huvent *et al.* 2006a).

A surprising conservation in substrate coordination was observed between MatC and *B. pertussis* L-aspartate binding protein BugD (Huvent *et al.* 2006a), where most of

the residues and hydrogen bonds involved in coordination were conserved. Aspartate diverges from malate only in the substitution of the C2 hydroxyl group for an amine group. However, MatC showed no indication of binding to aspartate either by DFS or tryptophan fluorescence assay. While the neutral C2 hydroxyl group of malate provides further potential for hydrogen bonding with the main structure, the binding of an aspartate molecule in the binding pocket would place a positively charged amino group in a position where there is no clear potential for neutralization of this charge by surrounding residues in the structure. Given their structural similarity, malate and aspartate would be undistinguishable in the electron density map, regardless of their resolution. The characterization of aspartate binding by BugD comes uniquely from the fortuitous co-purification of this ligand with the protein and is not confirmed by biochemical analysis (Huvent *et al.* 2006a). We suggest the possibility that L-malate, rather than L-aspartate, would be the ligand present in the BugD structure.

This study also attempted to perform the first biochemical characterisation of the transmembrane components of a TTT tripartite system. We successfully expressed the MatBA complex from *R. palustris* in *E. coli* using a single construct, preserving the intergenic regions and adding a C-terminal His_{8x}-tag in the MatA subunit. Relying on the interactions between the two subunits, we show the complex can be purified using Ni-NTA affinity chromatography, and that the two subunits co-elute in size exclusion chromatography. Using the MatBA-containing membranes from 48 L of bacterial cultures, we reconstituted the TTT complex in proteoliposomes, using either the rapid dilution method or the Biobeads detergent removal method. However, we could not retrieve uptake activity of ^{3}H-succinate in any case. Several factors could be impairing stability and activity of the MatBA complex (Rigaud and Levy 2003). The solubilisation of the proteins with detergents may disrupt the protein native fold, which would not be

refolded correctly once into liposomes. The elution of the MatBA complex with a molecular weight compatible with an oligomer reinforce that this might be the case, although the complex was still successfully purified and did not form aggregates. An alternative to overcome this problem would be to attempt the reconstitution of MatBA using the total membrane of the expression host, without prior purification with detergents, as successfully shown by Groeneveld *et al.* (2010). Another alternative for purifying membrane proteins while maintaining the lipidic environment would be to make use of styrene-maleic-acid nanodiscs (SMA) (Stroud *et al.* 2018). However the subsequent proteoliposome reconstitution is challenging and so far not mentioned in the literature. In addition, the *E. coli* polar lipids may not offer the lipid composition needed for stabilisation and activity of the MatBA complex, naturally found in the much more complex membrane of *R. palustris*. In a parallel study (not presented here), we attempted to destabilise the outer and inner membrane of *R. palustris* with different detergents, and found that the concentration needed for membrane destabilisation was much higher than that required for *E. coli* membranes (our unpublished observations), indicating a different and robust lipid composition of the membrane. In the case the proteins were active in the proteoliposomes, an independence on SBP would offer an additional limitation: it could be argued that the protein incorporation into two distinct orientations could create an artefact during the uptake assay, where the amount of substrate transported in each direction would be the same, and thus no total uptake would be observed. However, the requirement of the SBP for substrate transport in TRAP and TTT systems (Sweet *et al.* 1979, Mulligan *et al.* 2009, Hosaka *et al.* 2013) exclude this limitation.

As detailed in Rosa *et al.* (2018a), the predicted topology of the TTT transmembrane proteins is very similar to those of TRAP systems, despite the absence of sequence

similarity. However, while Mulligan *et al.* (2009) shows clear biochemical evidence that the TRAP SiaPQM system from *Haemophilus influenza* is energised by a Na^+ gradient and uses the ΔΨ to increase the free energy available, the physiological studies performed by Hosaka *et al.* (2013) suggests that the TTT family is rather energized by the proton-motif force in the membrane, and that Na^+ absence in the medium does not disrupt substrate uptake. In addition, experiments performed by Brocker *et al.* (2009) suggests the involvement of divalent cations in uptake, corroborated by the early TTT studies from Sweet *et al.* (1979) which indicate an improved affinity of TctC in the presence of these ions. Thus the importance of a complete biochemical characterisation of the transmembrane components of the TTT family in clarifying the energy requirements and physiological relevance of the TTT uptake systems, and this study may hopefully lay the foundations for it.

In this study, we show the characterization of the *Rhodopseudomonas palustris* MatBAC system, the first system from the Tripartite Tricarboxylate Transporter family shown to bind to C4-dicarboxylates, and the transport system with the highest affinity described to date to these compounds. Together, these data increase our understanding of the complex C4-dicarboxylate metabolism in bacteria, and provide further knowledge on the still poorly investigated TTT transport family.

CHAPTER 7. CHARACTERISATION OF THE REMAINING TTT SBPS FROM *RHODOPSEUDOMONAS PALUSTRIS*

PREFACE

As illustrated previously in Figure 5-1, *R. palustris* contains seven SBP from the TTT family. Chapter 5 shows the characterisation of AdpC, while Chapter 6 focus on the characterisation of the MatBAC system. This chapter will focus on the characterisation of the remaining components of the TTT family in *R. palustris* through a variety of approaches, including genetic, biochemical, physiological and structural characterisation. It will end by comparing all of them based on sequence, structure and function, and suggesting future works.

Noteworthy in this study is the description of the protein structure of six out of the seven TTT SBP in this organism. In order to facilitate comparison, the data collection and refinement statistics are shown for all of them in Table 7-1, provided by Dr. John Rafferty and Sam Dix, from the University of Sheffield.

Table 7-1: Data collection and refinement statistics for the six *R. palustris* TTT SBP crystallised in this study. Model building and table provided by collaborators Sam Dix and Dr. John Rafferty (University of Sheffield)

Data collection	**TttC1***	**MatC***	**TttC4***	**AdpC**	**TttC6***	**TttC7***
Wavelength (Å)	1.69997	0.97949	0.97625	0.92819	0.92819	0.97626
Resolution range (Å)	59.74-1.69 (1.74-1.69)	63.54-2.11 (2.17-2.11)	63.47-2.96 (3.14-2.96)	59.0-1.78 (1.81-1.78)	33.5-1.05 (1.07 – 1.05)	52.9-1.49 (1.52-1.49)
Space group	P 21 21 21	P 21 21 21	P 1 21 1	C 2 2 21	P 1 21 1	C 2 2 21
Cell dimensions (a,b,c) (Å) (α,β,γ) (°)	37.5, 59.7, 117.3, 90, 90, 90	47.6, 80.8, 205.8, 90, 90, 90	34.1, 81.5, 101.2, 90, 90, 90	47.3, 168.5, 82.5, 90, 90, 90	50.5, 38.5, 67.9, 90, 99.7, 90	50.2, 116.6, 105.88, 90, 90, 90
Total reflections	133847 (479)	76223 (4478)	22723 (3695)	61065 (3514)	176315 (647)	97453 (4914)
Unique reflections	23394 (320)	41492 (2471)	11675 (1886)	32168 (1825)	90810 (357)	51096 (2530)
Multiplicity	5.7 (1.5)	1.8 (1.8)	1.9 (2.0)	1.9 (1.9)	1.9 (1.8)	1.9 (1.9)
Completeness (%)	77.6 (15.0)	88.9 (65.3)	99.7 (98.3)	100 (100)	75.7 (6.1)	99.9 (100)
Mean I/σI	18.4 (1.7)	12.4 (1.6)	6.3 (1.5)	9.9 (2.0)	29.9 (4.7)	11.6 (1.4)
R_{merge}	0.087 (0.500)	0.046 (0.591)	0.090 (0.742)	0.057 (0.4)	0.01 (0.18)	0.034 (0.515)
Anomalous completeness	71.6 (4.0)	-	-	-	-	-
Anomalous multiplicity	2.9 (1.2)	-	-	-	-	-
Anomalous correlation	0.560 (0.000)	-	-	-	-	-
Anomalous slope	1.82	-	-	-	-	-
Refinement						
resolution (Å)	59.74-1.69 (1.74-1.69)	63.54-2.11 (2.17-2.11)	63.47-2.96 (3.14-2.96)	58.9-1.78 (1.81-1.78)	33.5-1.05 (1.07 – 1.05)	52.9-1.49 (1.52-1.49)
R_{work}/R_{free}	0.288/ 0.355	0.193/ 0.245	0.343/ 0.396	0.185/ 0.216	0.169/ 0.182	0.201/ 0.229
B-factor protein ($Å^2$)				18.8		

Outer shell data in parenthesis
*Refinement in progress

7.1. TTTC1 - RPA2319

7.1.1. GENOME ANALYSIS

The *rpa2319* SBP gene is located immediately downstream of *rpa2321-2320*, which encodes for one of the complete TTT transmembrane systems, *tttBA1*, strongly suggesting that they are co-transcribed in an operon. Downstream of the SBP are two genes annotated as *gtcA* and *gtcB*, two subunits of a possible glutaconate CoA-transferase, and BLAST results showed that the amino-acid sequence also has 31% identity to a 3-oxo-adipate CoA-transferase from *Desulfomonile tiedjei.* The genic proximity, shown in Figure 7-1, suggests that these genes are involved in the same pathway, where the TTT system transports the substrate, glutaconate or 3-oxo-adipate, and the CoA transferase promptly converts it to its CoA derivative. Furthermore, *rpa2316* encodes an enoyl-CoA hydratase, which may further act on this CoA-derivative. In our search in the literature, we could not find any description of any type of glutaconate nor 3-oxo-adipate transporter, which means if confirmed, this would be a novel transport system. Upstream in the reverse strand, *rpa2322* codes for a transcriptional regulator from the broad LysR family, however the spacing from the *rpa2321-16* operon and proximity from *rpa2323* suggests this regulator is involved in a different pathway.

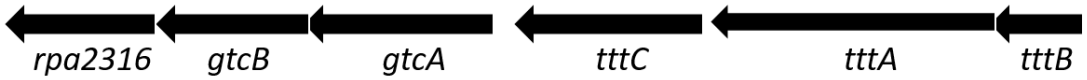


Figure 7-1: Genetic neighbourhood of *rpa2319* (*tttC*). Upstream of it, the two transmembrane domains of the tripartite transporter (tttBA). Downstream, genes coding for a CoenzymeA transferase (gtcAB), and an enoyl CoA-hydratase (*rpa2316*)

Analysis using the software MultiBLASTGene revealed several operons co-transcribing *gtcA*, *gtcB* and a hydratase gene, but none of them contained a TTT transport system in the neighbourhood (data not shown).

7.1.2. PRODUCTION OF KNOCKOUT MUTANT OF THE TRANSMEMBRANE PROTEINS RPA2320/21

In order to characterise the function of the TttBA1/TttC1 transporter, the generation of a marker-free deletion mutant in rpa2320/21 (*tttBA1*) was aimed. Attempts to generate the mutant using the pK18mobSacB plasmid failed, as all the colonies screened after the sucrose selection were seen to be wild type colonies, not the 1:1 wild-type: mutant ratio expected from the protocol. A successful attempt was then performed using the vector pJQ200ks+, generating the CGA009 Δ*tttBA1* strain. Figure 7-2 shows the gel confirmation of two screened colonies, which were further confirmed through sequencing.

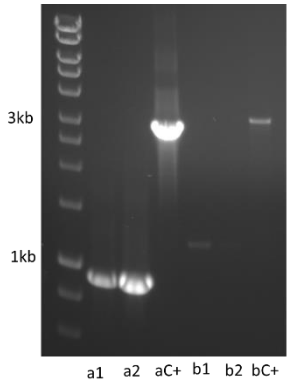


Figure 7-2: 1% agarose gel showing screening of deletion mutants CGA009 Δ*tttBA1*, corresponding to one of the transmembrane systems for the TTT family. Amplifications "a" were done with the primers used to amplify the flanking regions of the gene, and primers "b" correspond to a region flanking the deleted region, but outside the deletion cassete region. It is shown that colonies 1 and 2 show amplifications around 1kB, referring to the upstream+downstream of the deleted region, while the Wild type (C+) show an amplification of 3kB.

7.1.3. GROWTH EXPERIMENTS

In order to test whether *R.palustris* has a transporter for glutaconate and glutarate, a growth experiment was set-up in minimal media, using these compounds as the sole carbon source. The results are shown in Figure 7-3. *R. palustris* was able to grow when the sole carbon source was glutarate or glutaconate, albeit with a slower rate than when using succinate. These results support our hyphothesis that *R. palustris* has a transporter for these molecules, otherwise they would not enter the cell and could not be metabolized. The experiment was repeated with the CGA009 Δ*tttBA1* strain, and if TttBA1 was responsible for C5-dicarboxylic acid uptake, the deletion mutant would not be able to use these compounds as sole carbon sources. However, as shown in Figure 7-3, the mutant showed no growth defect when grown with glutarate, glutaconate or adipate as sole carbon sources.

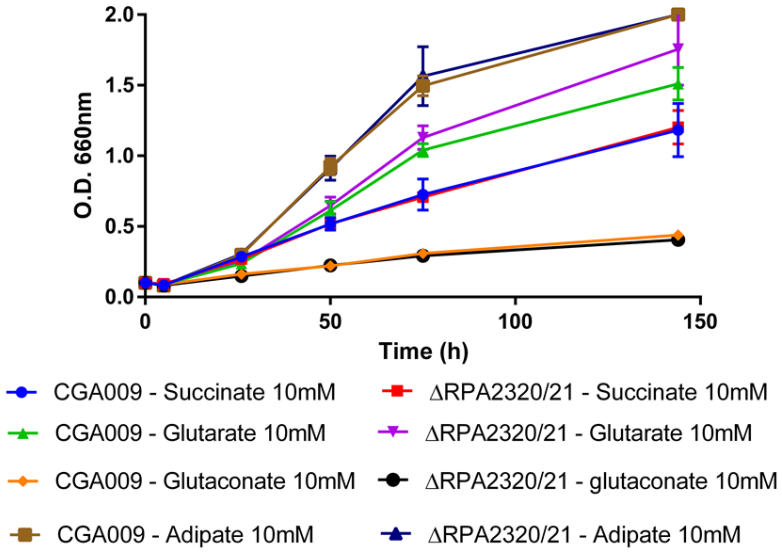


Figure 7-3: Growth experiment comparing the use of glutarate, glutaconate, succinate and adipate between the wild type and the CGA009 Δ*tttBA1* mutant.

7.1.4. PHENOTYPE MICROARRAY

In order to try to identify whether TttBA1 proteins are essential for carbon source uptake, a comparative Phenotype Microarray was performed between the CGA009 Δ*tttBA1* mutant and the wild type, as described in methodology section. Results after 24 and 48 hours were considered for the analysis; however, the results obtained after 72 hours supplied the most conclusive data and resolution. After three days of incubation, the wild type strain plates showed absorbance above 0.110 for 45 out of the 192 carbon sources tested, listed in Table 7-2. Most of them, as expected, are comprised of organic acids, however there were also a few sugars and aminoacids.

Table 7-2: List of carbon sources present in the Phenotype Microarray plates in which an A_{660} higher than 0.110 was observed, meaning *R.palustris* is able to metabolize them.

3-hydroxy-2-butanone	glycerol	Propionate
3-O-β-D-Galactopyranosyl-D-arabinose	L-Glutamine	Putrescine
Acetate	L-isoleucine	Pyruvate
Acetoacetate	L-lactate	Sebaceate
bromo succinate	L-leucine	Succinamate
Caproate	L-malate	Succinate
Citramalate	L-pyroglutamate	Turanose
Dextrin	Malate	Tween 20
D-fructose	Maltose	Tween 40
D-fructose-6-phosphate	Mannan	Tween 80
D-lactate methyl ester	mono methyl-succinate	α-hydroxy-glutarate-Ɣ-lactone
D-Malate	Octopamine	α-keto-glutarate
D-xylose	Oxalomalate	β-cyclodextrin
Formate	Palatinose	β-hydroxy-butyrate
Fumarate	p-hydroxy-phenyl-acetate	Ɣ-hydroxy-butyrate

Subtracting the CGA009 Δ*tttBA1* absorbances from the wild type, no differences in A_{660} greater than 0.1 were observed, suggesting that the TttBA1 proteins are not involved, or at least are not essential, for the metabolism of the tested carbon sources. These results, however, provide rich information regarding the catabolic capabilities of *R. palustris* CGA009.

7.1.5. PROTEIN OVERPRODUCTION IN *E. COLI*

In order to obtain large amounts of protein for biochemical work, the SBP *tttC1* (*rpa2319*) gene was cloned into pET21a(+) and inserted into the BL21 *E. coli* strain for overexpression. TttC1 showed good overexpression, the optimal conditions being 1h induction with IPTG at 37 ºC. The protein was purified by nickel affinity chromatography, and eluted from 190 mM to 250 mM of imidazole. As shown in Figure 7-4, the protein was obtained in high yields, with an average of 7 mg/L of protein for a batch of BL21 culture.

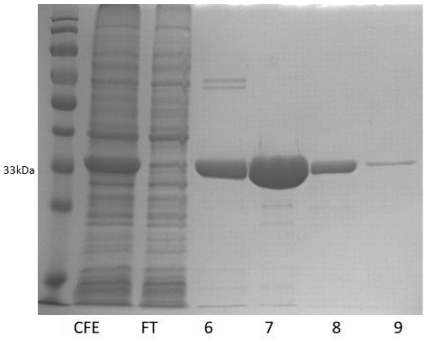


Figure 7-4: 12% polyacrylamide SDS-PAGE gel showing purified fractions of TttC1 after affinity chromatography. Cell-Free extract, CFE; Flow-through, FT; Fractions 6-9. The protein was obtained pure and in high yields.

7.1.6. DIFFERENTIAL SCANNING FLUORESCENCE

Differential Scanning Fluorescence (DSF) assays were performed to search for binding, both with native protein and renatured protein after urea denaturation (see section 3.21). As the protein denatures with increasing temperature, the Sypro Orange dye (Invitrogen) binds to the hydrophobic regions of the protein and emits fluorescence. Binding of the ligand stabilizes the protein, which starts denaturing at a

higher temperature. Six different substrates comprising different chain lengths and carboxyl content were initially assayed: Glutarate, glutaconate, citrate, adipate, 2-oxoglutarate and acetoacetate. However, no statistically significant (>2 °C) differences in melting temperature were observed, suggesting that TttC1 might not bind any of these substrates.

Once TttC1 was shown not to bind to the expected substrates, a screening was done using the thermofluor library built by our group (section 3.26). Screening the denatured/renatured protein against this library, however, retrieved no significant results. Finding a substrate for a binding protein might prove very difficult once the expected substrates and the standard screening procedure do not retrieve any positive hint. One possible next step is to expand the ligand library available. In order to do so, several ligand libraries were analysed in detail, and the Spectrum Collection library (Microsource Discovery Systems) seemed to be the most promising one, containing 2000 compounds, (50% approved drugs, 30% natural compounds and 20% known bioactives). TttC1 was screened against these 2000 compounds, and no change in melting temperature higher than 2 °C was found. Although being a viable option, commercial ligand libraries usually are optimized for drug discovery purposes, and so may not contain simple molecules and metabolites, believed to bind to the TTT SBP`s.

7.1.7. TRYPTOPHAN FLUORESCENCE ASSAY

Screening using Differential Scanning Fluorescence assay is an easy way of testing several compounds at once. On the other hand, although a thermal stabilization of the protein is a strong indicator of binding, the lack of stabilization does not necessarily mean that binding did not occur. Once the genetic neighbourhood of TttC1 strongly

suggested binding of C5-dicarboxylic acids, these substrates were once again screened using a more sensitive technique, the tryptophan fluorescence assay. TttC1 was initially assayed with glutaric and glutaconic acids. No significant quenching was observed, however, reinforcing that the protein didn`t bind to the predicted substrates. The assay was expanded to compounds of similar structures; adipate, malate, malonate, oxaloacetate and succinate. Again, no quenching was observed with any of them (data not shown).

One possible reason for the lack of fluorescence change is that the protein, being expressed in the *E. coli* cytoplasm, would bind to a ligand from the host metabolome. Having a ligand already bound would block the actual ligand from binding and thus prevent the quenching. To test this hypothesis, the protein was denatured with Urea 6 M and renatured, as described in methods section 3.21. This should release any pre-bound ligand. The assay was repeated using the same substrates, now with the renatured TttC1, but the same negative results were observed.

7.1.8. CRYSTALLIZATION AND STRUCTURE DETERMINATION

As described in section 3.29, crystallization trials were set against 5 different commercial screens. TttC1 was the first protein presented in this study to be expressed and crystalized. After two weeks, small crystals could be seen in two different conditions: JESG screen, well H8 (0.2 M NaCl; 0.1 M bis-TRIS pH 5.5; 25% w/v PEG3350) and PEG screen, well E12 (0.2 M Ammonium iodide; 20% w/v PEG3350). The X-ray shot and the crystal quality were enough to retrieve a 1.8 Å resolution diffraction pattern from the PEG E12 condition crystal. The iodine ions present in the buffer showed stable interactions with positively charged residues,

allowing for Single wavelength Anomalous Dispersion (SAD) de novo phase determination (Abendroth *et al.* 2011). Statistics are provided in Table 7-1.

Using the RefMac program (Murshudov *et al.* 1997) and the primary sequence of the protein, nearly 70% of the residues were fit into the electron density. Structural similarity searches using the Dali server retrieved the highest matches with BugE (PDB accession code 2DVZ), Bug27 (PDB accession code 2QPQ-A) and BugD (PDB accession code 2F5X-B), three Bug proteins from *Bordetella pertussis*, referred to in Huvent *et al.* (2006b), Huvent *et al.* (2006a) and Herrou *et al.* (2007), respectively. BugE was crystallized with a glutamate molecule interacting with the binding pocket; Bug27 binds to nicotinate, but was crystalized with a citrate bound to the open conformation due to crystallization conditions; and BugD was crystalized with an aspartate in the binding pocket. Table 7-3 shows the highest scores in Dali match results. After these three Bug proteins, it is possible to see a big drop in %id, confirming TttC1 is indeed closely related to the Bug proteins.

Table 7-3: Results for Dali server comparison for the TttC1 structure. The highest matches are with three Bug proteins from *Bortedella pertussis* BugE, Bug27 and BugD. After them, the identity of results to TttC1 drops considerably.

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	2dvz-A	35.1	2.1	271	300	42	PDB	MOLECULE: PUTATIVE EXPORTED PROTEIN;
2:	2qpq-A	35.1	1.9	271	296	37	PDB	MOLECULE: PROTEIN BUG27;
3:	2qpq-C	34.8	2.0	272	296	37	PDB	MOLECULE: PROTEIN BUG27;
4:	2f5x-B	33.1	2.3	270	300	36	PDB	MOLECULE: BUGD;
5:	2f5x-C	33.0	2.5	270	299	36	PDB	MOLECULE: BUGD;
6:	2f5x-A	32.9	2.5	271	299	36	PDB	MOLECULE: BUGD;
7:	2qpq-B	32.5	2.2	266	291	36	PDB	MOLECULE: PROTEIN BUG27;
8:	4x9t-A	28.8	3.6	267	288	26	PDB	MOLECULE: UNCHARACTERIZED PROTEIN UPF0065;
9:	3un6-A	13.8	3.6	213	291	12	PDB	MOLECULE: HYPOTHETICAL PROTEIN SAOUHSC_00137;
10:	2x26-B	13.8	3.2	210	291	10	PDB	MOLECULE: PERIPLASMIC ALIPHATIC SULPHONATES-BINDING PROTEIN
11:	2x26-A	13.5	3.2	210	308	10	PDB	MOLECULE: PERIPLASMIC ALIPHATIC SULPHONATES-BINDING PROTEIN
12:	4ddd-A	13.4	3.6	219	300	10	PDB	MOLECULE: IMMUNOGENIC PROTEIN;
13:	3e4r-A	13.1	3.6	212	291	12	PDB	MOLECULE: NITRATE TRANSPORT PROTEIN;
14:	3hn0-A	13.0	3.1	202	280	9	PDB	MOLECULE: NITRATE TRANSPORT PROTEIN;
15:	2g29-A	13.0	3.7	223	385	8	PDB	MOLECULE: NITRATE TRANSPORT PROTEIN NRTA;
16:	3ksi-A	13.0	3.7	214	289	12	PDB	MOLECULE: NITRATE TRANSPORT PROTEIN;
17:	3ksx-A	12.9	3.7	212	289	12	PDB	MOLECULE: NITRATE TRANSPORT PROTEIN;
18:	2i48-A	12.9	3.7	228	399	10	PDB	MOLECULE: BICARBONATE TRANSPORTER;
19:	2i4c-A	12.9	3.6	226	401	10	PDB	MOLECULE: BICARBONATE TRANSPORTER;
20:	2i49-A	12.8	3.6	226	398	10	PDB	MOLECULE: BICARBONATE TRANSPORTER;

Superposing their structures, it was possible to complete the TttC1 structural model, shown in Figure 7-5. The asymmetric unit comprises of a single molecule, a monomer of 301 amino-acids, excluding Histidine tag and signal peptide, organized in a Venus flytrap-like conformation, with two globular domains connected by a hinge, surrounding a cleft where the substrate binds. The overall structure is formed by 10 α-helix (H1-10) and 9 β-strands (β1-9) and one short helix (h1) in the order, from N-terminal to C-terminal: β1-α1-β2-α2-β3-α3-β4a-β4b-α4-β5-α5-β6-α6-β7-α7-β8-h1-β9-α8-α9-α10, as numbered in Figure 7-5. Domain 1 is comprised of amino-acids 1-91 and 235-301, having six α-helixes organized around a β-sheet composed of 5 strands. Domain two is comprised of amino-acids 108-226, with a disulphide bond between residues C136 and C143, similar to the one found between residues C142 and C178 in BugD, C136 and C165 in Bug27 and C81 and C140 in BugE. This domain is organized with 4 α-helixes, also surrounding a β-sheet composed of five strands. The hinge connecting the two domains is comprised of a loop and β4, usually seen in TTT structures as a single strand crossing both domains (Herrou *et al.* 2007), but in TttC1, as in AdpC (Rosa *et al.* 2017), the interdomain portion of the structure was not assigned as part of the strand, and thus β4 was separated in β4a and β4b. These features together classify the TttC1 protein as a Cluster E-II SBP (Scheepers *et al.* 2016).

The protein was crystallised in the open conformation, without substrate bound to it. Bug27, obtained in the same opened form, was estimated to have a rotation angle between the opened and closed form of 24.7° (Herrou *et al.* 2007), and the high similarity between the two structures suggests the same might apply to TttC1. Comparing the binding pocket structure, some interesting features could be found. TttC1 shows a conserved Phenylalanine (Phe16) which is described to interact with the carbon chain of the substrates in BugD and BugE (Huvent *et al.* 2006a, Huvent *et*

al. 2006b). However, it lacks an important Threonine (Thr33 in BugE and Thr22 in Bug27) which is described to coordinate one of the carboxylic groups in the substrate. Instead, it shows an Alanine (Ala22), which lacks the hydroxyl group responsible for this coordination. This difference is illustrated in Figure 7-6.

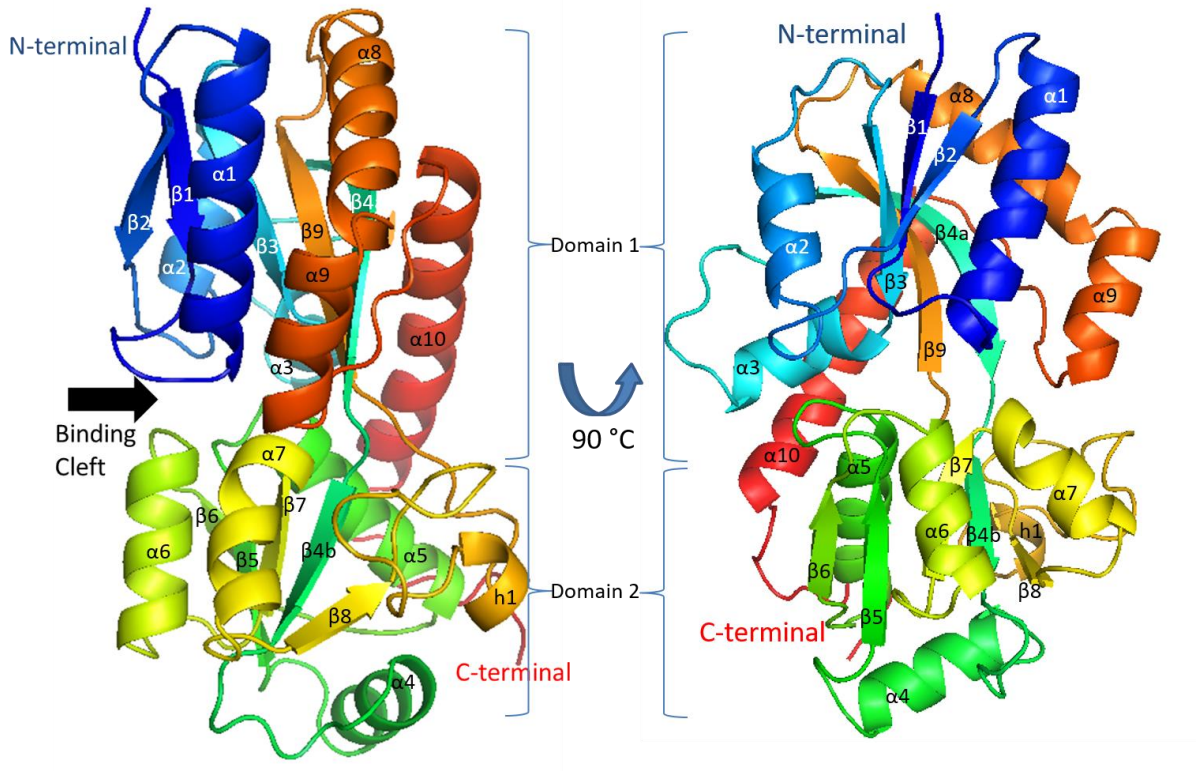


Figure 7-5: Overall structure of TttC1 in the openned state with no inbound ligand. The protein is comprised of two-globular domains of α-helixes folded around a 5-stranded β-sheet, separated by a cleft for substrate binding, connected by a hinge of three β-sheets. α-helixes and β-strands are numbered in the figure.

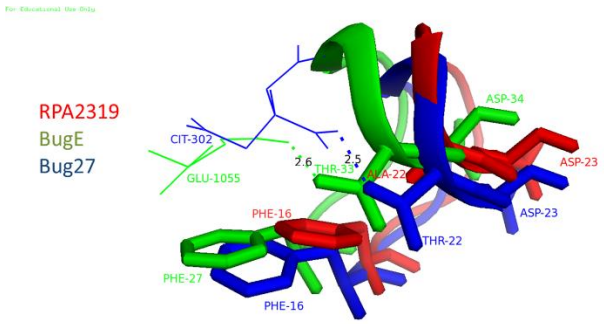


Figure 7-6: Comparison between the binding pocket of Rpa2319 (TttC1), BugE and Bug27. All of them show a conserved phenylalanine, described to interact with the carbon chain in the substrate. Rpa2319, however, has an Alanine (Ala22) replacing an important Threonine, believed to coordinate carboxyl groups in the substrates.

Multi-alignment assays were performed with the seven *R. palustris* Ttt SBP`s, BugE, BugD, Bug27, TctC from *Polaromonas* (PDB accession Code 4x9T), therepthalate transporter (TphC) from *Comamonas* sp. (Hosaka *et al.* 2013) and a low identity LacI transporter (see Figure 7-20 in section 7.6). The results showed that TphC and TttC7 also have the substitution of threonine for an Alanine, suggesting they might be related in the kind of substrate they transport. TttC1 lacks also an Alanine (Ala14 in BugD) described in BugD as responsible for two water-mediated interactions with aspartate oxygen. However, Bug27 also does not share this residue. Other key amino-acids in substrate coordination in BugE also are substituted in TttC1. Tyr231, which mediates hydrogen bonding with the substrate carboxyl group, was replaced by a phenylalanine, lacking the hydroxyl, a replacement also observed in TphC. Ser73, which hydrogen bonds another oxygen in the substrate, is changed to a glycine in most of the other TTT SBP`s. Ser137, also an oxygen coordinator, was replaced by a Cysteine, which is involved in disulphide bonding in TttC1. Together, this information helps in the characterization of TttC1, allowing some initial prediction about the potential substrate it binds to, suggesting a ligand different from the linear carboxylic acids characterized so far.

Given that none of the previous methods could elucidate what the substrate for TttC1 is, an alternative approach of exposing the protein to a complex mixture of compounds was attempted (section 3.22), in the hope that the protein would bind to a ligand present in the mixture, which could then be further characterised using mass spectrometry or crystalizing the protein again, now in presence of substrate, as suggested by Vetting *et al.* (2015). This strategy is shown schematically in Figure 7-7.

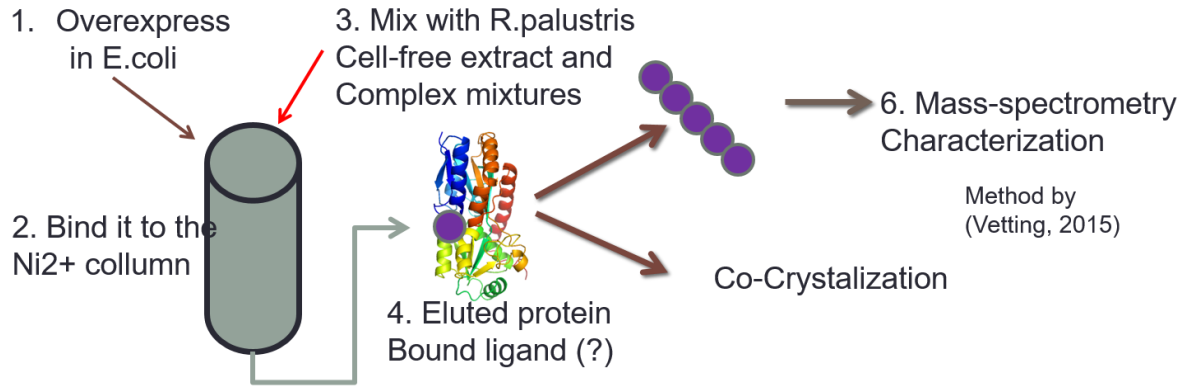


Figure 7-7: Alternative approaches to determinate protein ligands, as suggested by Vetting *et al.* (2015). The protein is overproduced in *E coli* and imobilised in a Nickel affinity column. A complex mixture composed of *R.palustris* cell free extract and humic acid is passed through the protein in the hope that it will find its specific ligand and bind to it. Protein is then eluted and ligand is characterised by mass spectrometry or co-crystalization.

Several unsuccessful attempts were performed to characterise the protein using mass spectrometry after exposure to a cell-free extract of *R. palustris*. Also, after being exposed to this mixture of compounds, a new hanging drop experiment with TttC1 was performed in the conditions which crystalized the protein previously. After eight months of incubation, new protein crystals were observed. A high-resolution dataset was obtained (1.9 Å) and the structure elucidated using the previous data as a model. Once again, no substrate was found in the binding pocket, suggesting TttC1 did not bind to any ligands among the complex mixture to which it was exposed.

7.1.9. OVERPRODUCTION OF GTCAB PROTEIN IN *E. COLI*

The suspicion that TttC1 was involved in the transport of C5-dicarboxylic acids was raised due to the presence of a downstream CoA transferase, GtcAB (Rpa2318/17), which BLAST searches indicated as being involved with glutaconate and glutarate metabolism. However, no further data was available to reinforce these BLAST results. In order to verify the GtcAB activity, this protein was overproduced in *E. coli*. Overexpression was attempted in pET21a, and although the protein expressed very well, solubility tests showed that inclusion bodies were formed and all protein was in the insoluble fraction of the cell free extract (not shown). In order to overcome this problem, overexpression was attempted in the pBAD-HisB vector. Again, the expression was very good. The solubility trials showed that although the majority of the protein was insoluble, a reasonable amount was still in the soluble fraction. The protein was expressed for 24 h with 0.02% of L-arabinose at 30 ºC. The protein was purified by Nickel affinity chromatography, eluted from 164 mM to 335 mM of imidazole. As shown in Figure 7-8, the protein was obtained in a yield of 36 mg/L of Top10 culture.

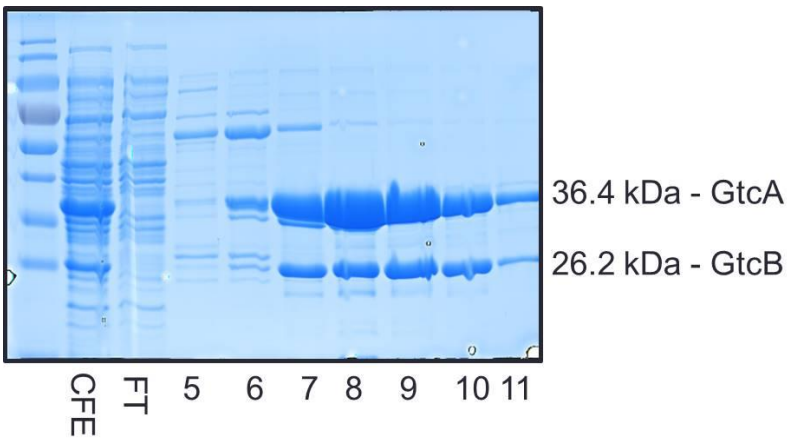


Figure 7-8: 12% polyacrylamide SDS-PAGE gel showing purified fractions of GtcAB after affinity chromatography. Cell-Free extract, CFE; Flow-through FT; Fractions 5-11. The protein was obtained pure and in high yields.

7.1.10. ENZYMATIC ACTIVITY OF GTCAB

In order to characterise the activity of GtcAB, a transferase activity assay was performed, through indirect measurement of acetyl-CoA production, as described in methodology section 3.45. No increase in absorbance was observed in the presence of succinyl-CoA or glutaryl-CoA, suggesting succinate and glutarate, respectively, are not the substrates for gtcAB. Another assay was performed to measure the production of glutaconyl-CoA from acetyl-CoA and glutaconate, but no activity was observed as well. This enzyme is potentially connected metabolically to the transport activity of TttC1, but with a substrate other than those suggested by BLAST searches.

7.2. TTTC3 - RPA0686

7.2.1. GENOME ANALYSIS

Upstream of the *tttC3* gene (*rpa0686*), *rpa0685* encodes for a transcriptional regulator from the IclR family. Taking into account the proximity of the two genes, it is possible that they are co-transcribed in an operon. However, as reviewed by Molina-Henares *et al.* (2006), this family can be involved in a wide range of metabolic processes, making it hard to determine specific functions only with the genomic sequence. Further upstream in the same strand, *rpa0682* was shown in BLAST results to encode a protein related to ABC transporters. So it is possible that Rpa0686 acts as a sensor protein, which would then trigger Rpa0685 and activate expression of an ABC uptake system for a given substrate.

7.2.2. PROTEIN OVERPRODUCTION IN *E. COLI*

In order to obtain large amounts of protein for biochemical work, *tttC3* gene was cloned into pET21a(+) and inserted into the BL21 *E.coli* strain for overexpression, but could not be expressed in this vector. It was then cloned into pBAD/His B vector and transformed into Top10 *E.coli* strain for overexpression.

TttC3 showed poor overexpression, the optimal conditions being 5h induction with 0.02% of L-arabinose at 37 ºC. The protein was purified by Nickel affinity chromatography, eluted from 160 mM to 220 mM of imidazole. As seen in Figure 7-9, the protein was obtained in a yield of 1 mg/L of Top10 culture. In the purification gel, two protein bands could be seen. Because they are co-eluted and proportional in quantity, it is unlikely that it is a product of contamination, but rather be a product of degradation of protein. The eluted fractions were sent for analysis by mass spectrometry. The analysis, in Figure 7-10, showed a single peak of 35.4 kDa. As the expected size was 35.6 kDa, the difference is too small to be significant, and we can conclude the protein was pure and the two bands in the gel were due to some artefact.

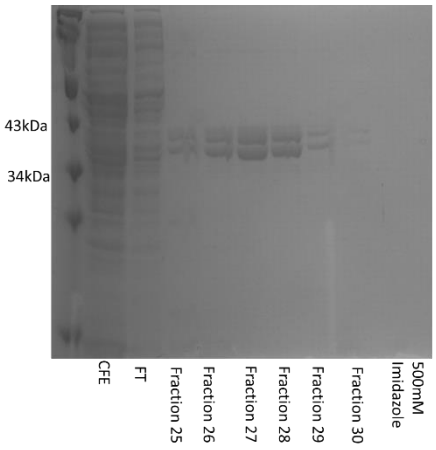


Figure 7-9: 12% polyacrylamide SDS-PAGE gel showing purified fractions of TttC3 after affinity chromatography. Cell-Free extract, CFE; Flow-through; Fractions 25-30, Flow-through after wash with Elution Buffer.

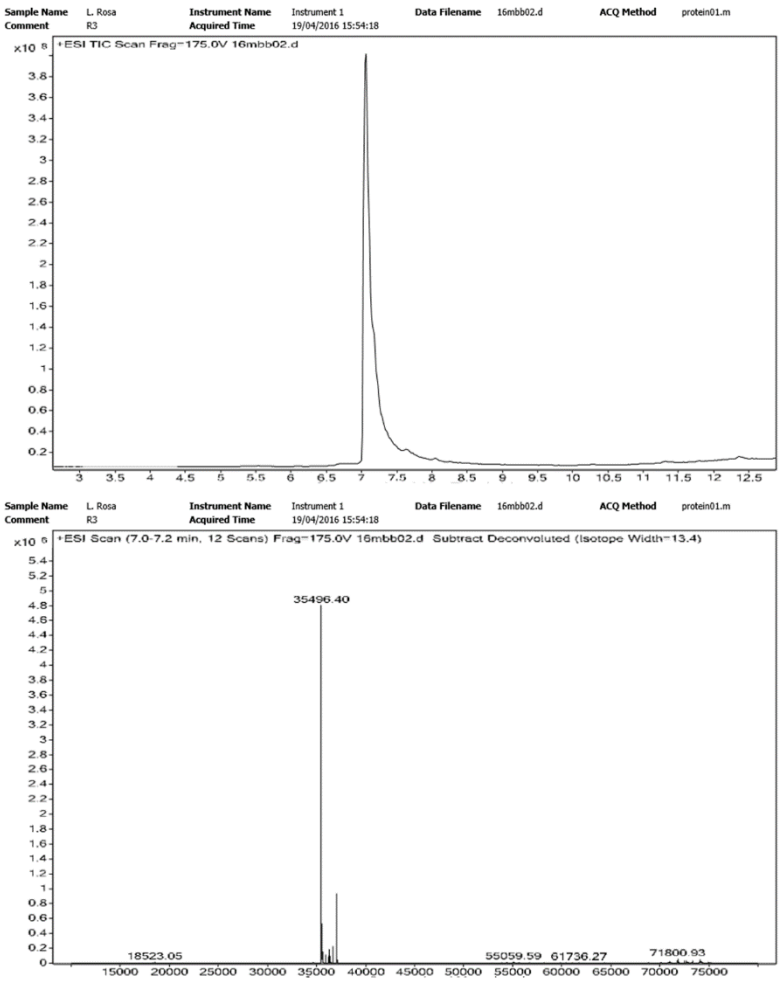


Figure 7-10: Mass spectrometry analysis of TttC3 purity. The chromatogram showed a single elution peak after 7 minutes (top), and the MS showed a single peak of 35.4 kDa, very similar to the expected 35.6kDa (bottom).

7.2.3. DIFFERENTIAL SCANNING FLUORESCENCE

As the genome analysis retrieved little information about the function of TttC3, the protein was then subjected to differential scanning fluorescence assay with the thermofluor database, as described in methods section 3.26. No shift higher than 2 °C was observed for any tested compound.

7.2.4. CRYSTALLIZATION

Crystal trials were set using PACT and JCSG screens, however no crystals were observed after three year's incubation.

7.3. TTTC4 - RPA3100

7.3.1. GENOME ANALYSIS

The neighbourhood of TttC4, shown in Figure 7-11, lacks any obvious association that could give a hint about its function. However, it is worth noting that *rpa3101* codes for a chemotaxis receptor which was shown by transcriptomics and proteomics to be overexpressed under growth with aromatic compounds (Pan *et al.* 2008). However, from the way the genes are disposed, it is unlikely that they are co-transcribed and BLAST searches were inconclusive regarding the genes function. Analysis using the software MultiBLASTGene showed that the genes RPA3096-99 are found together in several organisms, but their function did not seem characterized in any of them (Data not shown).

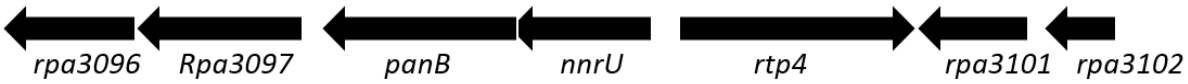


Figure 7-11:Genetic neighbourhood of RPA3100.

7.3.2. PROTEIN OVERPRODUCTION IN *E. COLI*

In order to obtain large amounts of protein for biochemical work, the *tttC4* gene was cloned into pET21a(+) and inserted into the BL21 *E.coli* strain for overexpression but

could not be expressed in this vector. It was then cloned into pBAD/His B vector and transformed into Top10 *E.coli* strain for overexpression.

TttC4 showed reasonable overexpression, the optimal conditions being 5 h induction with 0.002% of L-arabinose at 25 ºC. The protein was purified by Nickel affinity chromatography, eluted from 160 mM to 220 mM of imidazole. As shown in Figure 7-12, the protein was obtained in a yield of 4 mg/L of Top10 culture

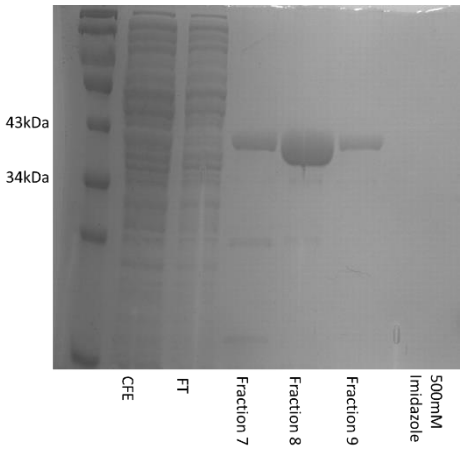


Figure 7-12: 12% polyacrylamide SDS-PAGE gel showing purified fractions of TttC4 after affinity chromatography. Cell-Free extract, CFE; Flow-through; Fractions 25-30, Flow-through after wash with Elution Buffer. The protein was obtained pure and in high yields.

7.3.3. DIFFERENTIAL SCANNING FLUORESCENCE

As the genome analysis retrieved little information about the function of TttC4, the protein was then subjected to differential scanning fluorescence assay with the thermofluor database, as described in section 3.26. No shift higher than 2 °C was observed for any tested compound.

7.3.4. CRYSTALLIZATION AND STRUCTURE DETERMINATION

Once expressed, crystallization trials were set with TttC4. After two weeks, some conditions retrieved crystals. One of them, from the PACT screen, position F4 [K thiocyanate 0.2 M; Bis-TRIS propane 0.1 M pH6.5; PEG3350 20% m/v] retrieved needle-like crystals, which were shot and retrieved low resolution diffraction, around 4 Å. Hanging drop experiments were set in this condition, but crystal formations were not observed. After nine months incubation, very fine needle-shaped crystals were observed in Proplex D12 condition [0.1 M Tris pH 8.5; 20% mw PEG6000], which showed a 3 Å resolution. Statistics are provided in Table 7-1. Further optimisations have been set to obtain a higher-resolution structure without success up to date.

7.4. TTTC6 - RPA4580

7.4.1. PROTEIN OVERPRODUCTION IN *E.COLI*

In order to obtain large amounts of protein for biochemical work, *tttC6* gene was cloned into pET21a(+) and inserted into the BL21 *E.coli* strain for overexpression. Rpa4580 could be expressed inducing cells with IPTG for 24 hours at 37 °C, however yields were very low and attempts to purify the protein did not retrieve enough material for subsequent analysis. Because *R. palustris* has a high CG content in the genome and the proteins are being expressed in a host which has lower GC content, the bad expression of *TttC6* could be due to differences in preferential codons. In order to test this hypothesis, the vectors containing the cloned gene was transferred into a variant strain of *E. coli*, Rosetta pLysS, which contains a plasmid encoding for rare codons and has the gene coding for the lactose transporter deleted, allowing more subtle

expression. This approach did not result in any improvement for the two proteins` expression.

The *tttC6* gene was then cloned into pBAD/HisB vector and transformed into Top10 *E. coli* strain for overexpression. TttC6 showed good overexpression, the optimal conditions being 22 h induction with 0.002% of L-arabinose at 37 ºC. The protein was purified in Nickel affinity chromatography, eluted from 160 mM to 220 mM of imidazole. As shown in Figure 7-13, the protein was obtained in high yields, with an average of 8 mg/L of protein for a Top10 culture.

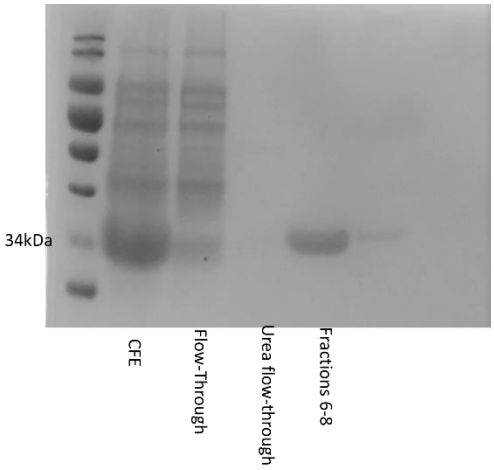


Figure 7-13: 12% polyacrylamide SDS-PAGE gel showing purified fractions of TttC6 after affinity chromatography. Cell-Free extract, CFE; Flow-through; Flow through after washing with Urea, Fractions 6-8. The protein was obtained pure and in high yields.

7.4.2. DIFFERENTIAL SCANNING FLUORESCENCE

TttC6 was screened against the thermofluor library as described in section 3.26. No indication of thermal stabilisation was observed in presence of any of the tested compounds. As the genetic neighbourhood lacks evidence as to what might be the

physiological role of the protein, further characterisation was attempted with crystalisation trials.

7.4.3. CRYSTALLIZATION AND STRUCTURE DETERMINATION

As described in methods, five different commercial screens were set for TttC6. After one year without any sign of crystallisation, crystals were finally formed in PACT screen, condition B11 [0.2 M CaCl2; 0.1 M MES pH 6.0; 20% m/v PEG6000]. The crystals retrieved 1.8 Å resolution diffraction, and molecular replacement was possible using AdpC as a model (Rosa *et al.* 2017). Statistics are provided in Table 7-1. The protein is a 333 residues peptide without the signal sequence, and residues 38 to 333 are present in the crystal structure. Domain 1 is comprised of residues 38-136 and 261-333, with a β-sheet formed by strands β2- β1- β3- β9- β4a. Domain 2, comprised by residues 137-260, contains a β-sheet comprised of strands β6- β5- β7- β4b- β8. The overall structure is shown in Figure 7-14 and the overall secondary structure of TttC6 is [β1- α1- β2- α2- β3- α3- h1- β4a-][β4b- α4- β5- α5- β6- α6- β7- α7- β8- h2][β9- α8- α9- α10]. No disulphide bonds are observed in this protein, which contains only one cysteine.

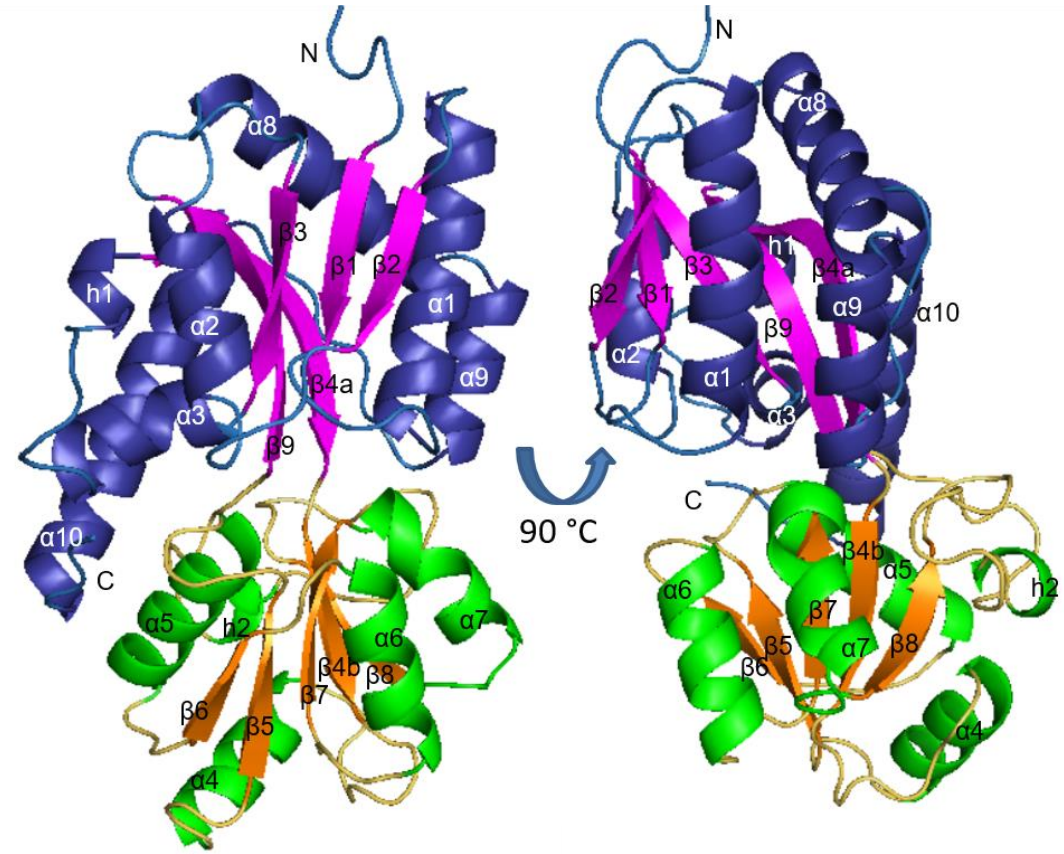


Figure 7-14: Overall crystal structure of TttC6. The protein is a 333 residues peptide without the signal sequence, and residues 38 to 333 are present in the crystal structure. Domain 1 is comprised of residues 38-136 and 261-333, with a β-sheet formed by strands β2- β1- β3- β9- β4a. Domain 2, comprised by residues 137-260, contains a β-sheet comprised of strands β6- β5- β7- β4b- β8.

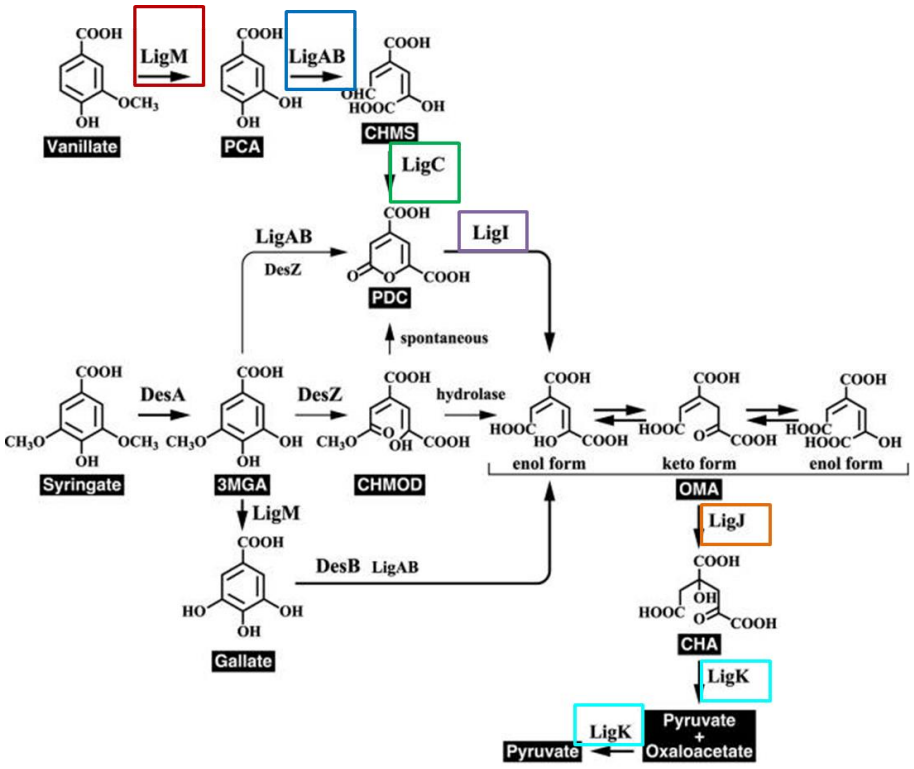
7.5. TTTC7 - RPA4694

7.5.1. GENOME ANALYSIS

TttC7 is one of five orphan TTT SBP`s in the *R.palustris* genome, meaning there is no obvious transmembrane transporter domains in the surrounding genomic regions. This SBP, then, could interact with one of the two nonspecific transmembrane domains, as suggested by (Antoine *et al.* 2003), or act as a part of a two-component system, as described by (Antoine *et al.* 2005). Downstream of this gene, there were several genes on the same strand, apparently cotranscribed, that might act in the

same pathway. BLAST comparisons using these downstream sequences retrieved several *lig* genes as the best match. These genes are reported in the literature to encode enzymes that participate in the protocatechuate degradation pathway, being very well characterized in *Leptospira biflexa* (Figueira *et al.* 2011) and *Sphingomonas paucimobilis* SYK-6 (Hara *et al.* 2003). Figure 7-15 shows the protocatechuate degradation pathway, extracted from Figueira *et al.* (2011), edited to relate to the reconstruction of the pathway operon in *Sphingomonas paucimobilis and Rhodopseudomonas palustris.* In the pathway, the enzymes are signalled by the same color correponding to the respective gene in *S. paucimobilis*, and the corresponding homolog in *R. palustris*, according to the BLAST search. TttC7 (RPA4694) sits just upstream of the whole catabolic pathway genes, suggesting it might act to scavenge substrates for it.

a



b

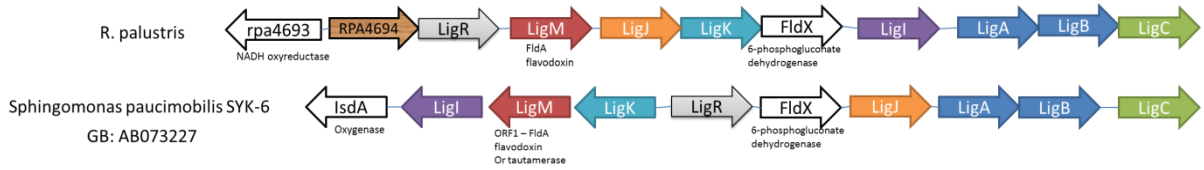


Figure 7-15: Protocatechuate degradation pathway, extracted from (Figueira, Croda et al. 2011), edited to relate to the reconstruction of the pathway operon in *Sphingomonas paucimobilis* and *Rhodopseudomonas palustris* (a). In the pathway, the enzymes are signalled by the same colour corresponding to the respective gene in *S. paucimobilis*, and the corresponding homolog in *R.palustris*, according to the BLAST search (b). *tttC7* (RPA4694) sits just upstream of the whole catabolic pathway, strongly suggesting it acts scavenging substrates for it. PCA, protocatechuate; 3-MGA, 3-O-Methylgallate; CHMS, 4,carboxy-2-hydroxymuconate-6-semialdehyde; PDC, 2-pyrone-4,6-dicarboxylate; OMA 4-oxalo-mesaconate; CHA, 4-carboxy-4-hydroxy-2-oxoadipate.

A search using the software MultiBLASTGene showed that indeed these genes are grouped together in many organisms, reinforcing the idea that they are in the same operon or encode the same pathway enzymes. Once again, most of the genes

retrieved in the cluster are annotated in the database as part of the protocatechuate degradation pathway or related compounds such as gluconate, extradiol, muconelactone, shikimate, oxalomesaconate, phthalate, 4-hydroxybenzoate, succinyl benzoate and vanillate. In very few organisms, however, a homolog of the TTT SBP is found among these genes. Among the top 18 hits for homolog operons, shown in Figure 7-16, only *Acidovorax sp.* and *Burkholderiales bacterium* show a homolog protein of TttC7 (RPA4694).

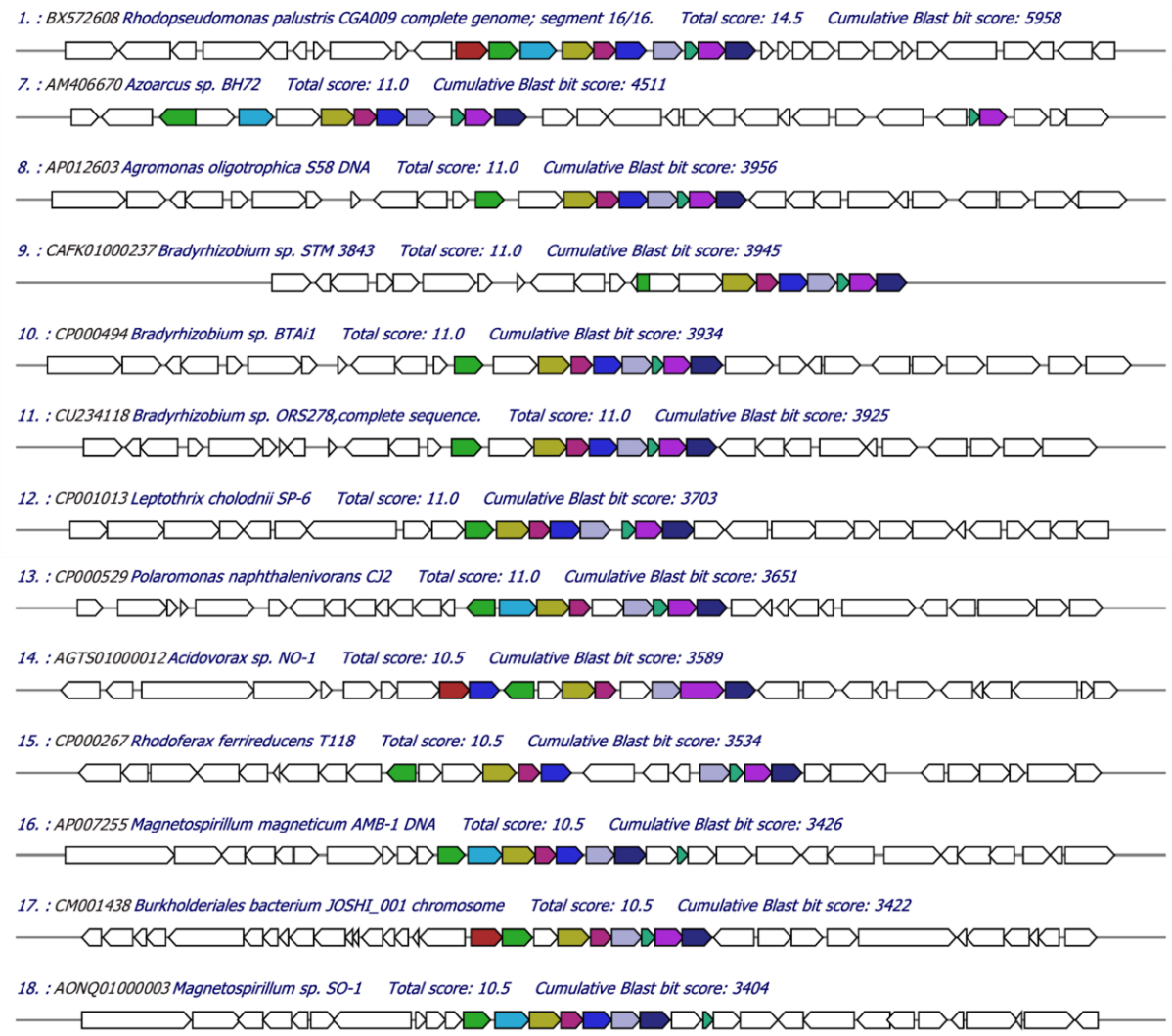


Figure 7-16: MultiBLASTGene (Medema *et al.* 2013) analysis of *tttC7* neighbourhood. Homolog genes are represented using the same colour. *tttC7*, red; *ligR*, green; *fldA*, light blue; *ligJ*, yellow; *fldZ*, dark pink; *fldX*, dark blue; *ligI*, grey; *ligA*, cyan; *ligB*, magenta; *ligC*, dark purple. Results 2-6 were omitted once they were related to other strains of *R. palustris* and showed identical gene organisation.

7.5.2. PROTEIN OVERPRODUCTION IN *E.COLI*

In order to obtain large amounts of protein for biochemical work, *tttC7* gene was cloned into pET21a(+) and inserted into the BL21 *E.coli* strain for overexpression. TttC7 showed good overexpression, the optimal conditions being 3 h induction with IPTG at 30 ºC. The protein was purified by Nickel affinity chromatography, eluted from 220 mM to 270 mM of imidazole. As seen in Figure 7-17, the protein was obtained in good yields, with an average of 3 mg/L of protein for a batch of BL21 culture.

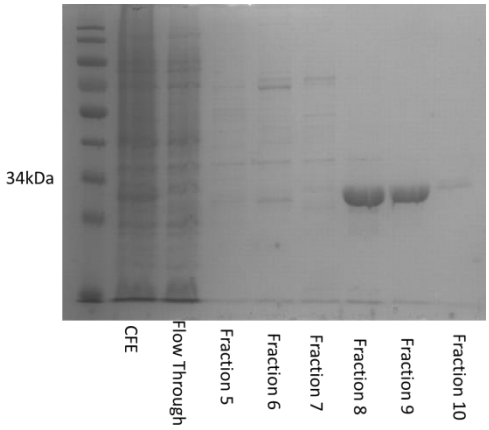


Figure 7-17: 12% polyacrylamide SDS-PAGE gel showing purified fractions of TttC7 after affinity chromatography. Cell-Free extract, CFE; Flow-through; Fractions 5-10. The protein was obtained pure and in high yields.

7.5.3. DIFFERENTIAL SCANNING FLUORESCENCE

Differential Scanning Fluorescence (DSF) assays were performed with TttC7, using both native protein and renatured protein. Ten different substrates were assayed, comprising substrates present in the protocatechuate degradation pathway and substrates with analoguous structures. Protocatechuate, vanillate, gallate, syringate, pthalate, citrate, mesaconate, coumarate 4-hydroxy-benzoate and phenylacetic acid were tested. No significant ΔTm (>2 °C) was observed, suggesting that TttC7 does

not bind to these substrates. This protein was then screened against the thermofluor ligand library, and once again no change in melting temperature higher than 2 °C was observed. As described for TttC1, this protein was also screened against the Spectrum library, comprised of 2000 compounds, and none of them showed a promising result.

7.5.4. TRYPTOPHAN FLUORESCENCE

Using the same concepts applied to TttC1, In order to check whether TttC7 has a role predicted by the genetic analysis, the tryptophan fluorescence was assayed using protocatechuate, vanillate and syringate as potential substrates. However, these molecules have an aromatic structure, which fluoresces in the same region as the protein itself, making it impossible to distinguish between the protein and substrate fluorescence, and consequently masking any possible quenching. In order to further analyse binding properties of TttC7, Isothermal Titration Calorimetry assays were performed.

7.5.5. ISOTHERMAL TITRATION CALORIMETRY

Still aiming to investigate the binding properties of TttC7 to protocatechuate and similar molecules, isothermal titration calorimetry assays were performed, as described in section 3.27. Attempts to titrate the protein were performed with protocatechuate, syringate, gallate, vanillate, 2-oxo-adipate and *cis-cis*-muconate. No significant heat change could be observed in the presence of any of them (data not shown), reinforcing the differential scanning fluorescence assay that the protein does not interact with these substrates, despite the strong evidence suggested by the genomic context.

7.5.6. CRYSTALLIZATION AND STRUCTURE DETERMINATION

As described in the methods section, crystallization trials were set against 5 different commercial screens. After two weeks, crystal formation could be observed in the Ammonium Sulphate trial, in a buffer comprising 0.1 M citric acid pH 4.0 and ammonium sulphate 1.6 M. The crystal was X-ray shot, and the diffraction pattern revealed they were indeed protein crystals. However, the diffraction pattern was poor, and further optimization was needed. Hanging-drop experiments using small variations of the crystallization condition retrieved again poor quality, needle-shaped crystals, which were not enough to retrieve a good diffraction pattern. These crystals were used then in seeding experiments against a JSCG trial in sitting drop experiments. This approach was very successful, as crystals were obtained in five different conditions, with the most promising ones being D12 [0.04 M KPO4; 16% (m/w) PEG8000; 20% (v/v) glycerol], G10 [0.15 M KBr; 30%(m/w) PEG MME 2000] and H9 [0.2 M LiSO4; 0.1 M bis-Tris pH 5.5; 25%(w/v)PEG3350]. X-ray analysis of H9 condition crystals revealed a high-resolution diffraction pattern (<2 Å), shown in Figure 7-18, but no signals that could allow the phase solving of the structure. A search on PDB database showed that the highest sequence similarity within the databank was with Bug27 structure. So it was attempted to solve the structure using molecular replacement, with Bug27 and TttC1 as models. The proteins, however, were not similar enough to allow it. Expression in cells using the heavy-atom analogue of methionine, selenomethionine, would not be suitable with this particular protein, as it only contains one methionine, in the C-terminal region, which can vary in conformation in the crystals as it is not part of a core region in the protein, so it would not generate a signal strong enough. Crystals from position G10 showed a surprisingly large size,

and it was attempted to retrieve the structure phase from the sulphur atoms in the protein, but no success was achieved for the same reasons, added to the fact that the protein contains no cysteines. Iodination of the crystals were also attempted, but the Iodine ions did not interact enough with the protein. Impregnation of the crystals was performed with the heavy metals osmium, mercury, gold, platinum and silver, but again no phase solving could be retrieved from the diffraction pattern. It was also tried to use seeding under the conditions that worked for TttC1, but no crystallization was found. When the structure of AdpC was finally elucidated (Rosa *et al.* 2017), it was used successfully as a model for molecular replacement (Molecular replacement performed by Sam Dixon, University of Sheffield). It was necessary to divide the two globular domains of AdpC and leave only the main chain of the residues, without the side chains, but finally a statistically significant phase for TttC7 was obtained. Statistics are provided in Table 7-1. TttC7 is a peptide of 332 residues without the signal peptide, where residues 36 to 332 are shown in the structure. Domain 1 is comprised of residues 36-134 and 259-332, and domain 2 is comprised of residues 135-258. β-sheet in domain one is formed by β strands β2- β1- β3- β9- β4a and in domain two by β6- β5- β7- β4b- β8. The overall secondary structure of TttC7 can be described as [β1- α1- β2-α2- β3- α3- β4a-][β4b- α4- β5- α5- β6- α6- β7- α7- β8-h1][β9- α8- α9- α10]. As the protein contain no cysteines, no disulphide bonds are observed, contrasting with TttC1 and the Bug proteins (Huvent *et al.* 2006a, Huvent *et al.* 2006b). The overall structure of Rpa4694 is shown in Figure 7-19.

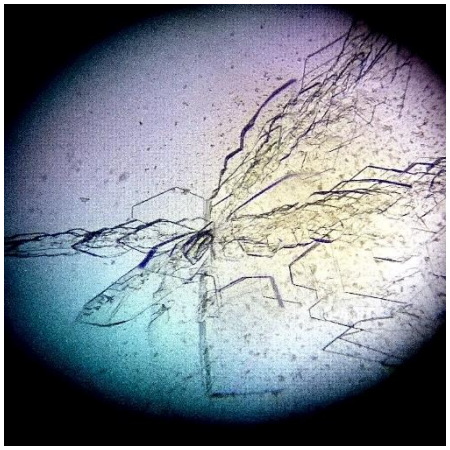


Figure 7-18: Crystals of RPA4694 after micro seeding experiment in condition H9 of JCSG screen.

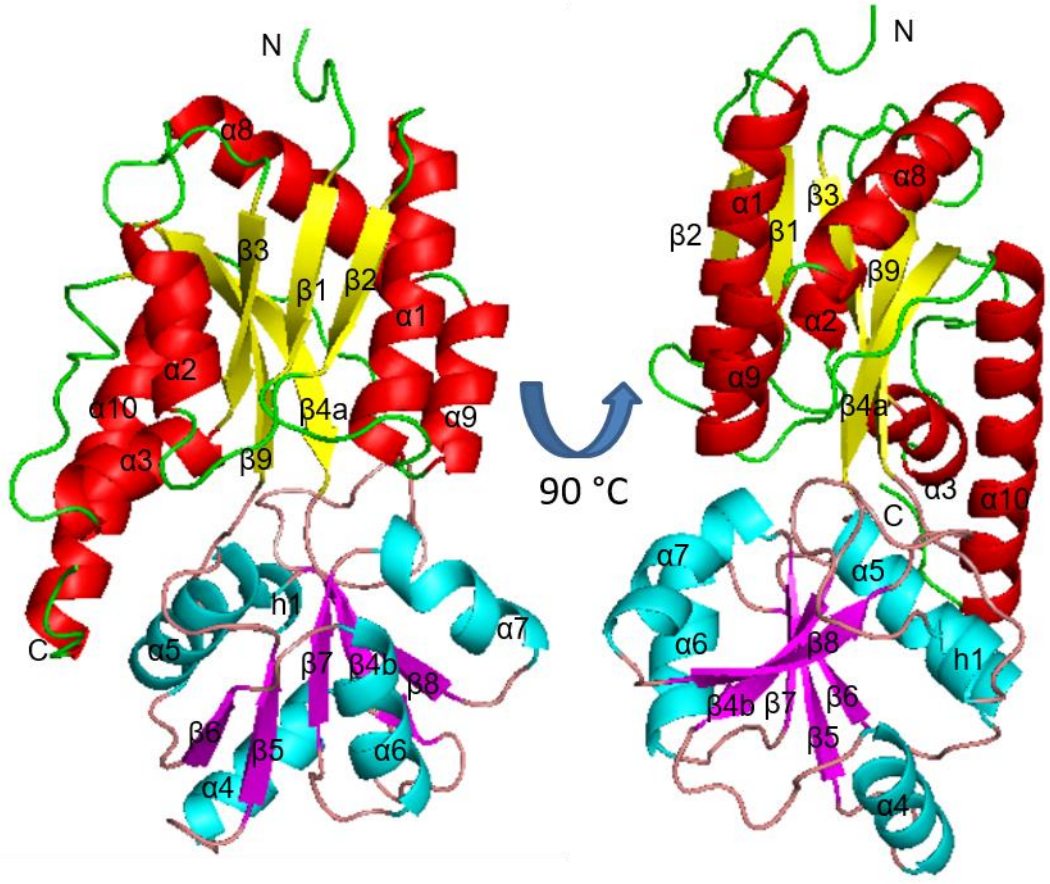


Figure 7-19: Crystal structure of TttC7, a peptide of 332 residues without the signal peptide, where residues 36 to 332 are shown in the structure. Domain 1 is comprised of residues 36-134 and 259-332, and domain 2 is comprised of residues 135-258. B-sheet in domain one is formed by β strands β2- β1- β3- β9- β4a and in domain two by β6- β5- β7- β4b- β8.

The structure and alignment performed for TttC7, shows us some interesting data. TttC7, as TttC1 and ThpC, has a substitution of the key Threonine (Thr33 in BugE) for an Alanine in the binding pocket. However, it has conserved the Alanine (Ala14 in BugD) described to coordinate water-mediated interactions with the carboxyl groups in the substrate. Comparing with BugE, it has an Asparagine replacing Y231, which maintains the potential to the hydrogen bond performed in BugE; like most of the TTT SBP`s, TttC7 has a glycine replacing Ser73; and has an Isoleucine replacing S137, that would lose the potential for hydrogen bonding in this position.

7.6. SEQUENCE AND STRUCTURAL ANALYSIS OF THE TTT FAMILY

Herrou *et al.* (2007) conjectured that the nature of the ligand could potentially be extrapolated from the binding pocket primary sequence of SBP`s. In order to see to what extent this can be applied to this study, a protein sequence alignment, containing only the regions described so far to be involved in ligand coordination in the TTT structures available is shown in Figure 7-20. In addition to the seven TttC proteins, the alignment shows the three Bug proteins of elucidated structure (BugC, BugE and Bug27)(Antoine *et al.* 2003, Huvent *et al.* 2006a, Huvent *et al.* 2006b), the TctC homolog from *Polaromonas* sp. (Unpublished, PDB accession code 4X9T) and the TphC, described to bind to terephthalate, from *Comamonas* sp. (Hosaka *et al.* 2013). All the TttC proteins have conserved the two glycine residues needed for coordination of water molecules (Gly28 in the [AKGG] motif and Gly173 in the [FSGT] motif, for MatC), demonstrating that the presence of at least one carboxylic group will possibly be a common feature among these protein`s ligands. Thr30 (in MatC) is also present in 4 of the TttC proteins, helping to further stabilize the proximal carboxylic group and indicating possible hydrogen bonds with side chains, as was the case for all the 4 TTT SBPs characterized so far. In three other proteins, this residue is substituted for an

Alanine, indicating that perhaps the ligand would be of more apolar nature. In fact, this substitution is present in TphC, described to bind to terephthalate (Hosaka *et al.* 2013); TttC1, potentially involved in the metabolism of 3-oxo-adipate, a classical intermediate in aromatic degradation pathway (Harwood and Parales 1996); and TttC7, suggested to be involved in the uptake of protocatechuate and related aromatic compounds. Five out of the seven TttC proteins show also the Phe24 residue conserved, which acts coordinating the aliphatic portions of ligands characterized so far. TttC3 and TttC6, instead, have it substituted by a tyrosine, which wouldn't interfere much with this role, and might provide an additional hydrogen bond either with another residue in the structure or with the gama-carboxyl group in the ligand.

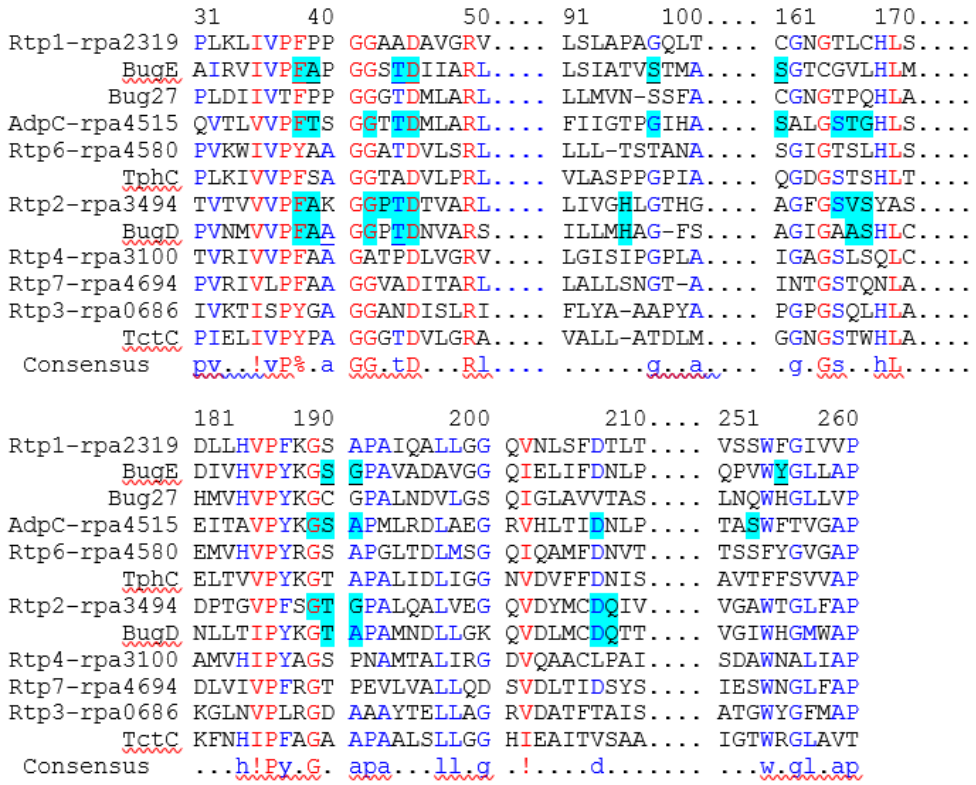


Figure 7-20: Protein sequence alignments for SBP members of the TTT family, generated using MultAlin software. In addition to the 7 SBP from *R. palustris*, are present the three Bug proteins from *B. pertussis* with solved structure (BugD, BugE, Bug27), the terephthalate binding protein from *Comomonas sp.* (TphC), and an uncharacterized TctC from *Polaromonas sp.* (PDB 4X9T). Conserved Residues are shown in red, less conserved residues in blue and non-conserved residues in black. Residues with light-blue background are found in crystal structures to be involved in ligand coordination.

The positioning of α4, altogether with the downstream α-loop-β, seems to be the limiting factor for determination of the binding cleft depth, as shown by comparison between MatC and AdpC (Rosa *et al.* 2017). However, no prediction from the primary or secondary sequence seems to be possible to determine how large the ligand is likely to be. From the six protein structures presented in this study, four of them were obtained in the Apo form, without a substrate bound in the binding cleft. In fact, the two proteins containing a bound substrate had it added in the crystallization buffer. The absence of inbound ligands and metabolites from the host strain in such a high frequency is unexpected for periplasmic binding proteins, since the binding of metabolites from the host strain frequently helps to bring insights into their physiological function (Vetting *et al.* 2015) and are often a disturbance when affinity measurements are required, but the binding site is already occupied (Salmon *et al.* 2013).Table 7-4 provides a comparison of all TTT SBP crystal structures elucidated so far, in terms of primary sequence similarity and orientation in space through the Root-Mean Square Deviation of atomic positions (RMSD). All structures seem to be relatively conserved, except the TctC from *Polaromonas sp.,* which seems to show poor sequence similarity and high RMSD with all the others. For this reason, calculations of the average are given considering either inclusion and exclusion of the values relating to TctC. Without taking into account the TctC data, it is possible to see two different clusters in the structures in terms of RMSD. Comparison between proteins containing a coordinated ligand show an average of 1.53 Å. Similarly, comparing structures in the open, unliganded conformation results in a RMSD of 1.74 Å. When comparing open and closed structures, on the other hand, this value jumps to 2.56 Å, an indication of the spatial change upon ligand binding. The average primary

sequence similarity does not seem to change when we are comparing sequences only inside the same organism or between *R. palustris* and *B. pertussis*.

Table 7-4: Comparison between all available crystal structures of SBP from the TTT family. On the top part of the "Non-applicable" (NA) diagonal is shown the Root-Mean Square Deviation (RMSD), a measurement of how conserved the structure is in the 3D space. In the bottom of the NA diagonal is the comparison of primary sequence similarity. Colour representation: Yellow, comparison between two liganded structures; Green, comparison between two open structures; Orange, comparison between closed and open structures; Blue, proteins belonging to the same organism; Purple, proteins belonging to different organisms. "No TctC" average exclude any data involving the TctC protein from the calculations. "Only TctC" average used only data involving the TctC protein for the calculations.

RMSDs (Å)	**BugD (B)**	**BugE**	**Bug27 (B)**	**TctC**	**AdpC**	**Rpa1**	**Rpa2**	**Rpa3**	**Rpa4**	**Rpa6**	**Rpa7**
SI (%)	**2F5X**	**2DVZ**	**2QPQ**	**4X9T**	**5OEI**						
BugD (B)	NA	1.33	3.00	3.89	1.75	2.19	1.27			2.54	3.11
BugE	34	NA	2.62	3.56	1.54	2.02	1.56			2.24	3.02
Bug27 (B)	25	32	NA	2.2	2.75	1.84	2.45			1.6	1.8
TctC	15	19	28	NA	3.96	3.09	3.52			2.7	2.32
AdpC	30	33	27	18	NA	1.87	1.75			2.17	2.89
Rpa1	32	41	36	24	35	NA	2.28			1.33	1.98
Rpa2	41	34	25	9	33	32	NA			2.69	3.18
Rpa3	27	31	31	27	31	35	31	NA			
Rpa4	32	35	32	30	32	36	34	32	NA		
Rpa6	29	30	35	25	35	41	24	35	32	NA	1.89
Rpa7	25	25	33	29	30	32	19	36	37	34	NA

Average	**Closed**	**Open**	**Mixed**		**Same organism**	**Different Organism**
No TctC	1.53	1.74	2.56		31.23	31.62
Total		2.07	2.8			28.64
Only TctC		2.57	3.73			22.4

7.7. DISCUSSION

From the analysis of the genetic neighbourhood of the TttC proteins, the TttC7 genetic neighbourhood had by far the strongest suggestion on the physiological role of an SBP, located just upstream of a protocatechuate degradation pathway gene cluster, confirmed with high-throughput genome comparison of operons using MultiBLAST

gene software and literature searches (Nichols and Harwood 1997, Contzen and Stolz 2000, Hara *et al.* 2003). However, the collected data suggest that this protein does not bind to protocatechuate, syringate, vanillate or any other related compound. If, on the other hand, the binding to aromatic compounds would be confirmed, it would be interesting to study whether TttC7 acts by regulating the operon or interacts with a promiscuous transmembrane transporter for substrate intake. From the seven TttC genes, five of them contained a transcription factor or a transmembrane signaling protein in the vicinity, reinforcing the growing hypothesis that most of these periplasmic binding proteins are actually involved in signaling processes rather than transport itself, as recently reviewed by Piepenbreier and Fritz (2017), and although there are cases where it applies to the TTT transporters (Widenhorn *et al.* 1988, Winnen *et al.* 2003, Brocker *et al.* 2009), further studies are needed to clarify if it is indeed a pattern in this family.

All seven SBPs were successfully cloned and expressed in *E. coli*, retrieving yields good enough to perform the desired binding experiments. The differential scanning fluorescence assay was capable of revealing the nature of the ligand for TttC5 (AdpC) and TttC2 (MatC), but it was surprising that no thermal stabilisation was found with any ligand in the case of 5 proteins. In addition, four protein structures were obtained in an open, unliganded conformation, showing they did not bind any metabolite from the expression host, or even from the original organism, in the case of TttC1, a fact which discards also Mass-spectrometry approaches for ligand identification.

This study also aimed to produce mutants of *R. palustris* CGA009 strain independently lacking each of the seven *tttC* genes, and each of the two *tttBA* genes. We successfully produced marker-free mutants in both transmembrane TTT systems; however, no phenotype was identified through growth experiments, either using the predicted

substrates or an array of carbon sources. At least for MatBAC, it was shown that this occurrence happens due to ligand redundancy among different transport systems, evidenced also by other studies in *R. palustris* (Salmon *et al.* 2013). Although among the objectives of this project was to generate also knock-out mutants in the SBP of the TTT family, genomic editing in *R*. *palustris* proved to be far more complex than initially thought, with the two homologous recombination events needed rarely occurring.

On the other hand, the comparison of primary sequences of all *R. palustris* TTT SBPs adds to the information provided by the genetic neighbourhood of their genes, helping to improve the indications of what ligand might be expected for them. This was valid for TttC1 and TttC7, which showed substitution only shared with the terephthalate binding protein TphC, indicating they might bind to aliphatic aromatic compounds. However, this information is not much more than just speculative, and is yet to be proven experimentally.

Comparing the 3D organisation of all crystal structures available for TTT SBP, it was possible to see a clear grouping between proteins containing or not-containing a coordinated substrate in the binding pocket, through the RMSD values. Although it might seem obvious that the presence of a ligand will increase the probability that the SBP is found in a closed conformation, it was still not experimentally shown how this process is driven. Upon the release of the first unliganded Bug protein structure, Herrou *et al.* (2007) proposed a model in which the substrate would be recognised by domain 1, using the carboxylic group as an anchoring point when interacting with one of the β-turns of the “pincer-like” structure; domain 2 would then perform a spatial change to provide interactions with the second part of the “pincer-like” structure, being able to distinguish between a correct or an incorrect ligand in the pocket and allow further coordination of the two conserved water molecules. The fact that none of the

unliganded proteins was obtained in a partially closed conformation indicates that the SBP exist either as a closed or open conformation, but do not exist in an intermediary state. This speculation is in agreement with the recent findings for the TRAP family, provided by PELDOR analysis, which failed to identify an intermediary state for the closure of SiaP, even when titrating the protein with sialic acid (Glaenzer *et al.* 2017).

Before this study, only four crystal structures of TTT SBPs were available in the database. Adding six more, accompanied by comparisons between them, under the light of a critical literature revision, we hope this study contributes to enriching the scientific knowledge regarding the Tripartite Tricarboxylate Transporter family, bringing to surface the unexplored and widespread potential of this high-affinity transport system.

CHAPTER 8. FINAL DISCUSSION

In this study, we aimed to explore and further characterise the Tripartite Tricarboxylate Transporters (TTT), a third family of SBP dependant high affinity transporters, in addition to the ABC and TRAP transporters. Information about this family was initially generated nearly forty years ago, with the discovery of the TctC citrate binding protein in *S. typhimurium* (Sweet *et al.* 1979). It was followed by a twenty year gap until the TTT SBP genomic overrepresentation in *B. pertussis* by Antoine *et al.* (2003) was published, which pointed out the importance of this family in the bacterial metabolism. Since then, new insights into the TTT family were again scarce and often restricted to the background of main research topics, despite their abundance in many environmental genomes (Wubbeler *et al.* 2014) and participation in unique bacterial metabolic processes (Hosaka *et al.* 2013).

We suggest that the TTT family is far more physiologically relevant than has been considered so far, especially in environmental bacteria belonging to the β and α-proteobacteria classes, where we show a consistent TTT SBP overrepresentation in the genome (Rosa *et al.* 2018b). In Chapter IV, we provide a high-throughput genomic search among all available bacterial genomics, evidencing this overrepresentation, and describe the genome of the uncultured soil bacterium *Rhodoplanes sp.* Z2-YC6860, which was shown to contain 434 TTT SBP in the genome, as compared to only 9 transmembrane counterparts. Many of these SBP genes were found in the vicinity of two-component systems, chemotaxis receptors and transcription factors, reinforcing the potential role for this family to be involved in several processes other than substrate import. Taken together, the array organisation of the genes, the low copy number in low-GC genomic islands and the high rate of intragenic BLAST hits

for these sequences strongly suggests that this overrepresentation is caused mainly by a Lineage-Specific Expansion, and that similar overrepresentations in α and β-proteobacteria are possibly a case of convergent evolution, triggered by common evolutionary pressure.

In order to provide a biochemical and structural characterisation of the TTT family, we focused on the α-proteobacterium *Rhodopseudomonas palustris*, a model organism relevant for its multiple biotechnological potential (Gall *et al.* 2013, Mukhopadhyay *et al.* 2013, Austin *et al.* 2015, Venkidusamy and Megharaj 2016). The TTT family in this organism is composed of two complete systems, here named TttBAC1 and MatBAC, and five orphan SBP, named TttC3-7. From the genomic analysis, there was strong evidence that at least two SBP, TttC1 and TttC7, were involved in the metabolism of aromatic compounds. However, our biochemical characterisation of these proteins did not confirm this hypothesis. From the 7 TTT SBP, we managed to obtain high resolution crystal structures (<2 Å) of 5 of them, and low resolution structure of a sixth one. They all share a very similar topology, characteristic for the TTT family. They are composed of two globular domains of β-sheets surrounded by α-helixes, connected by a hinge and separated by a binding cleft in the middle. Two of these structures were obtained with substrates present in the binding pocket. Chapter V described the adipate binding protein AdpC, and Chapter VI described the malate binding protein MatC. Both proteins share a conserved substrate coordination mechanism, also observed in the *B. pertussis* proteins BugD (Huvent *et al.* 2006a) and BugE (Huvent *et al.* 2006b), where two water molecules bridge hydrogen bonds between the substrate proximal carboxylate group and the main chain of two conserved loops in the protein structure. The sequence comparison performed by Huvent *et al.* (2006a) showed that the aliphatic aminoacid residues involved in this coordination are very

well conserved, and that the few proteins that do not contain them would have them mutated to a tyrosine or a threonine, where the residue's side chains would then perform a coordination similar to the one provided by the water molecules. An interesting future approach would be to perform several site-directed mutagenesis experiments and see how the binding affinities are disrupted or improved, which could lead to potential rational design of binding sites, used for the uptake of particular substrates in industry, or to use it as a biosensor when coupled to a reporter protein (Medintz and Deschamps 2006). The site directed mutagenesis could also provide insight into the interaction interface between the SBP and the transmembrane components, an area so far unexplored. Neither AdpC nor MatC were shown to oligomerise upon substrate binding, as shown by Size Exclusion Chromatography, and there is no evidence in the literature which would suggest otherwise for any TTT system, which might indicate that no oligomerisation is required for interaction of the SBP with the membrane counterparts.

Chapter V describes the orphan SBP AdpC, shown to bind to di-carboxylic acids ranging from 6 to 9 carbons in length. Isothermal titration calorimetry experiments showed that AdpC had a high affinity for adipate, and that the affinity would decrease as the substrate carbon chain length increased, being 40 times lower for azelate. The double knock-out mutant in the TttBA1 and MatBA membrane systems did not impair growth in adipate at 10 mM concentration. Indeed, our RT-PCR analysis showed that the *adpC* gene is repressed in high concentrations of substrate, and overexpressed under very low concentrations, being compatible with a high affinity transport/ signalling system. It is intriguing how a binding pocket with very little empty spaces and water molecules can be flexible enough to coordinate such a variety of substrates,

and further crystallisation of AdpC with different substrates shall provide us further insight on this aspect.

Chapter VI describes the biochemical and physiological characterisation of the MatBAC system, where MatC was shown by tryptophan fluorescence to bind with nM affinity to the C4-dicarboxylic acids malate, aspartate and fumarate. Given the considerable flexibility of AdpC, it was surprising to observe MatC did not bind any other C4 compounds with different functional groups. The deletion of both TTT transmembrane systems did not impair the growth of *R. palustris* under 10 mM concentration of C4-dicarboxylic acids, which are likely to be a consequence of redundant uptake system, not rarely found in *R. palustris* (Salmon *et al.* 2013). In fact, our genome searches revealed potential C4-dicarboxylate transporters of 5 different protein families. Among them, a DctP homologue was biochemically characterised by other members of our group and confirmed to bind to these same compounds. RT-PCR done with Both TTT and TRAP SBP's, however, did not show substantial difference in expression of neither, but a subtle 2.7 times overexpression at maximum. This correlates with our observations of AdpC expression, and we suggest extending the transcription analysis in MatC to similar substrate concentrations used for AdpC.

While the study of the SBP components of the TTT family provided insightful knowledge regarding the biochemistry, structural and physiological properties of these proteins, information regarding the transmembrane components of the TTT remains elusive. While Batista-Garcia *et al.* (2014) produced two structural models of the TctA subunit, and indicated potential interfaces for Na^+ cation binding, the physiological studies performed by Hosaka *et al.* (2013) showed a ΔpH dependency for TTT substrate uptake, and that absence of sodium in the media did not disrupt transport. Thus, a more detailed and focused biochemical and structural characterisation of the

TTT transmembrane components is required in order to elucidate the energy requirements and mode of action of this transport family. We successfully overexpressed the MatBA in *E. coli* and showed that the complex could be co-purified and reconstituted into liposomes. However, the radiolabelled succinate uptake assays showed no indication of transport activity with any of the methods tested. Our observations will, however, lay the ground for future work with the MatBA system, in attempts of reconstitution using different methods such as the total vesicles reconstitution (Groeneveld *et al.* 2010) and in further expression optimisations.

The ligands for the five remaining SBP from *R. palustris* remain unidentified, despite our diverse ligand library screening and parallel approaches using genomic analysis, mass spectrometry, co-crystallisation with environmental chemical mixtures and phenotype microarray experiments. Nevertheless, four crystal structures were obtained in apo form, which will contribute towards a more robust structural database for the TTT family, which before this study contained only 4 structures in total. With the availability of further structures, it may be possible to perform docking experiments for substrate prediction.

In summary, the present study offers a discussion on the literature available for the TTT family, together with the description of potentially the biggest overrepresentation of a gene family described in prokaryotes. Also, it offers the structural, biochemical and physiological characterisation of two TTT SBP, alongside with four SBP apo structures, and the first attempt to biochemically characterise the membrane components of the TTT family. May this study highlight the importance of the Tripartite Tricarboxylate Transporters family, highlighting its potential and relevance for future studies.

REFERENCES

Abendroth, J., A. S. Gardberg, J. I. Robinson, J. S. Christensen, B. L. Staker, P. J. Myler, L. J. Stewart and T. E. Edwards (2011). "SAD phasing using iodide ions in a high-throughput structural genomics environment." Journal of Structural and Functional Genomics **12**(2): 83-95.

Antoine, R., I. Huvent, K. Chemlal, I. Deray, D. Raze, C. Locht and F. Jacob-Dubuisson (2005). "The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction." J Mol Biol **351**(4): 799-809.

Antoine, R., F. Jacob-Dubuisson, H. Drobecq, E. Willery, S. Lesjean and C. Locht (2003). "Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*." J Bacteriol **185**(4): 1470-1474.

Austin, S., W. S. Kontur, A. Ulbrich, J. Z. Oshlag, W. Zhang, A. Higbee, Y. Zhang, J. J. Coon, D. B. Hodge, T. J. Donohue and D. R. Noguera (2015). "Metabolism of Multiple Aromatic Compounds in Corn Stover Hydrolysate by *Rhodopseudomonas palustris*." Environ Sci Technol **49**(14): 8914-8922.

Batista-Garcia, R. A., A. Sanchez-Reyes, C. Millan-Pacheco, V. M. Gonzalez-Zuniga, S. Juarez, J. L. Folch-Mallol and N. Pastor (2014). "A novel TctA citrate transporter from an activated sludge metagenome: structural and mechanistic predictions for the TTT family." Proteins **82**(9): 1756-1764.

Bergfors, T. (2007). Succeeding with seeding: some practical advice. Evolving Methods for Macromolecular Crystallography: The Structural Path to the Understanding of the Mechanismof Action of CBRN Agents. R. J. Read and J. L. Sussman. Dordrecht, Springer Netherlands**:** 1-10.

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

Brocker, M., S. Schaffer, C. Mack and M. Bott (2009). "Citrate Utilization by *Corynebacterium glutamicum* Is Controlled by the CitAB Two-Component System through Positive Regulation of the Citrate Transport Genes *citH* and *tctCBA*." J Bacteriol **191**(12): 3869-3880.

Buckel, W., U. Dorn and R. Semmler (1981). "Glutaconate CoA-transferase from *Acidaminococcus fermentans*." Eur J Biochem **118**(2): 315-321.

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

Contzen, M. and A. Stolz (2000). "Characterization of the genes for two protocatechuate 3, 4-dioxygenases from the 4-sulfocatechol-degrading bacterium *Agrobacterium radiobacter* strain S2." J Bacteriol **182**(21): 6123-6129.

Crosby, H. A., E. K. Heiniger, C. S. Harwood and J. C. Escalante-Semerena (2010). "Reversible N epsilon-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in *Rhodopseudomonas palustris*." Mol Microbiol **76**(4): 874-888.

Denger, K. and A. M. Cook (2010). "Racemase activity effected by two dehydrogenases in sulfolactate degradation by *Chromohalobacter salexigens:* purification of (S)-sulfolactate dehydrogenase." Microbiology **156**(Pt 3): 967-974.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Figueira, C. P., J. Croda, H. A. Choy, D. A. Haake, M. G. Reis, A. I. Ko and M. Picardeau (2011). "Heterologous expression of pathogen-specific genes *ligA* and *ligB* in the saprophyte *Leptospira biflexa* confers enhanced adhesion to cultured cells and fibronectin." BMC Microbiol **11**: 129.

Fukami-Kobayashi, K., Y. Tateno and K. Nishikawa (1999). "Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history." J Mol Biol **286**(1): 279-290.

Gall, D. L., J. Ralph, T. J. Donohue and D. R. Noguera (2013). "Benzoyl coenzyme a pathway-mediated metabolism of meta-hydroxy-aromatic acids in *Rhodopseudomonas palustris*." J Bacteriol **195**(18): 4112-4120.

Gänzle, M. G. (2015). "Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage." Current Opinion in Food Science **2**: 106-117.
Giuliani, S., A. Frank, D. Corgliano, C. Seifert, L. Hauser and F. Collart (2011). "Environment sensing and response mediated by ABC transporters." BMC Genomics **12**(1): 1-14.
Glaenzer, J., M. F. Peter, G. H. Thomas and G. Hagelueken (2017). "PELDOR Spectroscopy Reveals Two Defined States of a Sialic Acid TRAP Transporter SBP in Solution." Biophys J **112**(1): 109-120.
Gonin, S., P. Arnoux, B. Pierru, J. Lavergne, B. Alonso, M. Sabaty and D. Pignol (2007). "Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding." BMC Struct Biol **7**: 11.
Groeneveld, M., R. G. J. Detert Oude Weme, R. H. Duurkens and D. J. Slotboom (2010). "Biochemical Characterization of the C4-Dicarboxylate Transporter DctA from *Bacillus subtilis*." Journal of Bacteriology **192**(11): 2900-2907.
Guccione, E., M. Del Rocio Leon-Kempis, B. M. Pearson, E. Hitchin, F. Mulholland, P. M. Van Diemen, M. P. Stevens and D. J. Kelly (2008). "Amino acid-dependent growth of *Campylobacter jejuni*: key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate." Molecular Microbiology **69**(1): 77-93.
Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J Mol Biol **166**(4): 557-580.
Hara, H., E. Masai, K. Miyauchi, Y. Katayama and M. Fukuda (2003). "Characterization of the 4-carboxy-4-hydroxy-2-oxoadipate aldolase gene and operon structure of the protocatechuate 4,5-cleavage pathway genes in *Sphingomonas paucimobilis* SYK-6." J Bacteriol **185**(1): 41-50.
Harwood, C. S. and J. Gibson (1988). "Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*." Appl Environ Microbiol **54**(3): 712-717.
Harwood, C. S. and R. E. Parales (1996). "The beta-ketoadipate pathway and the biology of self-identity." Annu Rev Microbiol **50**: 553-590.
Heller, K. B. (1978). "Apparent molecular weights of a heat-modifiable protein from the outer membrane of *Escherichia coli* in gels with different acrylamide concentrations." Journal of Bacteriology **134**(3): 1181-1183.
Herrou, J., C. Bompard, R. Antoine, A. Leroy, P. Rucktooa, D. Hot, I. Huvent, C. Locht, V. Villeret and F. Jacob-Dubuisson (2007). "Structure-based mechanism of ligand binding for periplasmic solute-binding protein of the Bug family." J Mol Biol **373**(4): 954-964.
Higuchi, R., B. Krummel and R. K. Saiki (1988). "A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions." Nucleic Acids Research **16**(15): 7351-7367.
Holger, E., W. Brita and K. Reinhard (1991). "Na+-dependent succinate uptake in *Corynebacterium glutamicum*." FEMS Microbiology Letters **77**(1): 61-66.
Hollenstein, K., R. J. P. Dawson and K. P. Locher (2007). "Structure and mechanism of ABC transporter proteins." Current Opinion in Structural Biology **17**(4): 412-418.
Hosaka, M., N. Kamimura, S. Toribami, K. Mori, D. Kasai, M. Fukuda and E. Masai (2013). "Novel tripartite aromatic acid transporter essential for terephthalate uptake in *Comamonas sp*. strain E6." Appl Environ Microbiol **79**(19): 6148-6155.
Hu, S.-H., A. E. Whitten, G. J. King, A. Jones, A. F. Rowland, D. E. James and J. L. Martin (2012). "The Weak Complex between RhoGAP Protein ARHGAP22 and Signal Regulatory Protein 14-3-3 Has 1:2 Stoichiometry and a Single Peptide Binding Mode." PLOS ONE **7**(8): e41731.
Huvent, I., H. Belrhali, R. Antoine, C. Bompard, C. Locht, F. Jacob-Dubuisson and V. Villeret (2006a). "Crystal structure of *Bordetella pertussis* BugD solute receptor unveils the basis of ligand binding in a new family of periplasmic binding proteins." J Mol Biol **356**(4): 1014-1026.

Huvent, I., H. Belrhali, R. Antoine, C. Bompard, C. Locht, F. Jacob-Dubuisson and V. Villeret (2006b). "Structural analysis of *Bordetella pertussis* BugE solute receptor in a bound conformation." Acta Crystallogr D Biol Crystallogr **62**(Pt 11): 1375-1381.

Janausch, I. G., E. Zientz, Q. H. Tran, A. Kröger and G. Unden (2002). "C4-dicarboxylate carriers and sensors in bacteria." Biochimica et Biophysica Acta (BBA) - Bioenergetics **1553**(1): 39-56.

Jeong, J., S. Suh, C. Guan, Y. F. Tsay, N. Moran, C. J. Oh, C. S. An, K. N. Demchenko, K. Pawlowski and Y. Lee (2004). "A nodule-specific dicarboxylate transporter from alder is a member of the peptide transporter family." Plant Physiology **134**(3): 969-978.

Kalckar, H. M. (1971). "The Periplasmic Galactose Binding Protein of *Escherichia coli*." Science **174**(4009): 557-565.

Kamimura, N. and E. Masai (2014). The Protocatechuate 4,5-Cleavage Pathway: Overview and New Findings. Biodegradative Bacteria: How Bacteria Degrade, Survive, Adapt, and Evolve. H. Nojiri, M. Tsuda, M. Fukuda and Y. Kamagata. Tokyo, Springer Japan**:** 207-226.

Kelly, D. J. and G. H. Thomas (2001). "The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea." FEMS Microbiology Reviews **25**(4): 405-424.

Krom, B. P., J. B. Warner, W. N. Konings and J. S. Lolkema (2003). "Transporters involved in uptake of di- and tricarboxylates in *Bacillus subtilis*." Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology **84**(1): 69-80.

Kulkarni, G., C. H. Wu and D. K. Newman (2013). "The general stress response factor EcfG regulates expression of the C-2 hopanoid methylase HpnP in *Rhodopseudomonas palustris* TIE-1." J Bacteriol **195**(11): 2490-2498.

Larimer, F. W., P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. Torres, C. Peres, F. H. Harrison, J. Gibson and C. S. Harwood (2004). "Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*." Nat Biotechnol **22**(1): 55-61.

Laskowski, R. A. and M. B. Swindells (2011). "LigPlot+: multiple ligand-protein interaction diagrams for drug discovery." J Chem Inf Model **51**(10): 2778-2786.

Locher, K. P. (2016). "Mechanistic diversity in ATP-binding cassette (ABC) transporters." Nat Struct Mol Biol **23**(6): 487-493.

Maqbool, A., R. S. Horler, A. Muller, A. J. Wilkinson, K. S. Wilson and G. H. Thomas (2015). "The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity." Biochem Soc Trans **43**(5): 1011-1017.

Mayer, C. L., W. K. Snyder, M. A. Swietlicka, A. D. VanSchoiack, C. R. Austin and B. J. McFarland (2009). "Size-exclusion chromatography can identify faster-associating protein complexes and evaluate design strategies." BMC Research Notes **2**: 135-135.

McKinlay, J. B. and C. S. Harwood (2010). "Photobiological production of hydrogen gas as a biofuel." Curr Opin Biotechnol **21**(3): 244-251.

Medema, M. H., E. Takano and R. Breitling (2013). "Detecting sequence homology at the gene cluster level with MultiGeneBlast." Mol Biol Evol **30**(5): 1218-1223.

Medintz, I. L. and J. R. Deschamps (2006). "Maltose-binding protein: a versatile platform for prototyping biosensing." Current Opinion in Biotechnology **17**(1): 17-27.

Mitsch, M. J., G. C. diCenzo, A. Cowie and T. M. Finan (2018). "Succinate Transport Is Not Essential for Symbiotic Nitrogen Fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*." Applied and Environmental Microbiology **84**(1).

Molina-Henares, A. J., T. Krell, M. Eugenia Guazzaroni, A. Segura and J. L. Ramos (2006). "Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors." FEMS Microbiol Rev **30**(2): 157-186.

Mukhopadhyay, M., A. Patra and A. K. Paul (2013). "Phototrophic Growth and Accumulation of Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by Purple Nonsulfur Bacterium *Rhodopseudomonas palustris* SP5212." Journal of Polymers.

Mulligan, C., C. Fenollar-Ferrer, G. A. Fitzgerald, A. Vergara-Jaque, D. Kaufmann, Y. Li, L. R. Forrest and J. A. Mindell (2016). "The bacterial dicarboxylate transporter VcINDY uses a two-domain elevator-type mechanism." Nat Struct Mol Biol **23**(3): 256-263.
Mulligan, C., G. A. Fitzgerald, D.-N. Wang and J. A. Mindell (2014). "Functional characterization of a Na^+-dependent dicarboxylate transporter from *Vibrio cholerae*." The Journal of General Physiology **143**(6): 745-759.
Mulligan, C., E. R. Geertsma, E. Severi, D. J. Kelly, B. Poolman and G. H. Thomas (2009). "The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter." Proc Natl Acad Sci U S A **106**(6): 1778-1783.
Mulligan, C., A. P. Leech, D. J. Kelly and G. H. Thomas (2012). "The Membrane Proteins SiaQ and SiaM Form an Essential Stoichiometric Complex in the Sialic Acid Tripartite ATP-independent Periplasmic (TRAP) Transporter SiaPQM (VC1777–1779) from *Vibrio cholerae*." Journal of Biological Chemistry **287**(5): 3598-3608.
Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of Macromolecular Structures by the Maximum-Likelihood Method." Acta Crystallographica Section D **53**(3): 240-255.
Mutharasaiah, K., V. Govindareddy and K. Chandrakant (2012). "Biodegradation of 2-Chlorophenol by *Rhodopseudomonas palustris*." Bioremediation Journal **16**(1): 1-8.
Nichols, N. N. and C. S. Harwood (1997). "PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*." J Bacteriol **179**(16): 5056-5061.
Oda, Y., S. K. Samanta, F. E. Rey, L. Wu, X. Liu, T. Yan, J. Zhou and C. S. Harwood (2005). "Functional Genomic Analysis of Three Nitrogenase Isozymes in the Photosynthetic Bacterium *Rhodopseudomonas palustris*." Journal of Bacteriology **187**(22): 7784-7794.
Ogawa, J., C.-L. Soong, M. Ito and S. Shimizu (2001). "Enzymatic production of pyruvate from fumarate — an application of microbial cyclic-imide-transforming pathway." Journal of Molecular Catalysis B: Enzymatic **11**(4): 355-359.
Ottoni, J. R., L. Cabral, S. T. P. de Sousa, G. V. L. Júnior, D. F. Domingos, F. L. Soares Junior, M. C. P. da Silva, J. Marcon, A. C. F. Dias, I. S. de Melo, A. P. de Souza, F. D. Andreote and V. M. de Oliveira (2017). "Functional metagenomics of oil-impacted mangrove sediments reveals high abundance of hydrolases of biotechnological interest." World Journal of Microbiology and Biotechnology **33**(7): 141.
Pan, C., Y. Oda, P. K. Lankford, B. Zhang, N. F. Samatova, D. A. Pelletier, C. S. Harwood and R. L. Hettich (2008). "Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics." Mol Cell Proteomics **7**(5): 938-948.
Pechter, K. B., L. Gallagher, H. Pyles, C. S. Manoil and C. S. Harwood (2016). "Essential Genome of the Metabolically Versatile Alphaproteobacterium *Rhodopseudomonas palustris*." Journal of Bacteriology **198**(5): 867-876.
Pelicic, V., J. M. Reyrat and B. Gicquel (1996). "Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on *mycobacteria*." J Bacteriol **178**(4): 1197-1199.
Piepenbreier, H. and G. Fritz (2017). "Transporters as information processors in bacterial signalling pathways." Molecular Microbiology **104**(1): 1-15.
Poole, P. and D. Allaway (2000). "Carbon and nitrogen metabolism in *Rhizobium*." Adv Microb Physiol **43**: 117-163.
Porter, A. W. and L. Y. Young (2014). Chapter Five - Benzoyl-CoA, a Universal Biomarker for Anaerobic Degradation of Aromatic Compounds. Advances in Applied Microbiology. S. Sariaslani and G. M. Gadd, Academic Press. **88:** 167-203.
Pudlik, A. M. and J. S. Lolkema (2012). "Substrate Specificity of the Citrate Transporter CitP of *Lactococcus lactis*." Journal of Bacteriology **194**(14): 3627-3635.
Quandt, J. and M. F. Hynes (1993). "Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria." Gene **127**(1): 15-21.
Rath, A., M. Glibowicka, V. G. Nadeau, G. Chen and C. M. Deber (2009). "Detergent binding explains anomalous SDS-PAGE migration of membrane proteins." Proc Natl Acad Sci U S A **106**(6): 1760-1765.

Rey, F. E., E. K. Heiniger and C. S. Harwood (2007). "Redirection of metabolism for biological hydrogen production." Appl Environ Microbiol **73**(5): 1665-1671.
Rigaud, J. L. and D. Levy (2003). "Reconstitution of membrane proteins into liposomes." Methods Enzymol **372**: 65-86.
Rojo, F. (2010). "Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment." FEMS Microbiology Reviews **34**(5): 658-684.
Rosa, L. T., M. E. Bianconi, G. H. Thomas and D. J. Kelly (2018a). "Tripartite ATP-Independent Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT): From Uptake to Pathogenicity." Frontiers in Cellular and Infection Microbiology **8**(33).
Rosa, L. T., S. R. Dix, J. B. Rafferty and D. J. Kelly (2017). "Structural basis for high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic-binding protein from the tripartite tricarboxylate transporter (TTT) family in *Rhodopseudomonas palustris*." Febs j **284**(24): 4262-4277.
Rosa, L. T., V. Springthorpe, M. E. Bianconi, G. H. Thomas and D. J. Kelly (2018b). "Massive over-representation of solute-binding proteins (SBPs) from the tripartite tricarboxylate transporter (TTT) family in the genome of the alpha-proteobacterium *Rhodoplanes sp*. Z2-YC6860." Microb Genom.
Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream and B. Barrell (2000). "Artemis: sequence visualization and annotation." Bioinformatics **16**(10): 944-945.
Saier, M. H., Jr., V. S. Reddy, B. V. Tsu, M. S. Ahmed, C. Li and G. Moreno-Hagelsieb (2016). "The Transporter Classification Database (TCDB): recent advances." Nucleic Acids Res **44**(D1): D372-379.
Salmon, R. C., M. J. Cliff, J. B. Rafferty and D. J. Kelly (2013). "The CouPSTU and TarPQM transporters in *Rhodopseudomonas palustris*: redundant, promiscuous uptake systems for lignin-derived aromatic substrates." PLoS One **8**(3): e59844.
Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach and A. Pühler (1994). "Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*." Gene **145**(1): 69-73.
Scheepers, G. H., J. A. Lycklama a Nijeholt and B. Poolman (2016). "An updated structural classification of substrate-binding proteins." FEBS Letters **590**(23): 4393-4401.
Seddon, A. M., P. Curnow and P. J. Booth (2004). "Membrane proteins, lipids and detergents: not just a soap opera." Biochim Biophys Acta **1666**(1-2): 105-117.
Simon, E. and D. Shemin (1953). "The preparation of S-succinyl coenzyme A." J. Amer. Chem. Soc **75**: 2520.
Simon, R., U. Priefer and A. Puhler (1983). "A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria." Nat Biotech **1**(9): 784-791.
Smart, J. P., M. J. Cliff and D. J. Kelly (2009). "A role for tungsten in the biology of *Campylobacter jejuni*: tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter." Mol Microbiol **74**(3): 742-757.
Stroud, Z., S. C. L. Hall and T. R. Dafforn (2018). "Purification of membrane proteins free from conventional detergents: SMA, new polymers, new opportunities and new insights." Methods.
Sweet, G. D., C. M. Kay and W. W. Kay (1984). "Tricarboxylate-binding proteins of *Salmonella typhimurium*. Purification, crystallization, and physical properties." J Biol Chem **259**(3): 1586-1592.
Sweet, G. D., J. M. Somers and W. W. Kay (1979). "Purification and properties of a citrate-binding transport component, the C protein of *Salmonella typhimurium*." Can J Biochem **57**(6): 710-715.
Teramoto, H., T. Shirai, M. Inui and H. Yukawa (2008). "Identification of a gene encoding a transporter essential for utilization of C4 dicarboxylates in *Corynebacterium glutamicum*." Applied and Environmental Microbiology **74**(17): 5290-5296.
Unden, G., A. Strecker, A. Kleefeld and O. Kim (2016). "C4-Dicarboxylate Utilization in Aerobic and Anaerobic Growth." EcoSal Plus.
Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm and S. G. Rozen (2012). "Primer3—new capabilities and interfaces." Nucleic Acids Research **40**(15): e115-e115.
van Niel, C. B. (1944). "The culture, general phisiology, morphology, and classification of the non-sulfur purple and brown bacteria." Bacteriological Reviews **8**(1): 1-118.

Venkidusamy, K. and M. Megharaj (2016). "A Novel Electrophototrophic Bacterium *Rhodopseudomonas palustris* Strain RP2, Exhibits Hydrocarbonoclastic Potential in Anaerobic Environments." Frontiers in Microbiology **7**(1071).

Venkidusamy, K. and M. Megharaj (2016). "A Novel Electrophototrophic Bacterium *Rhodopseudomonas palustris* Strain RP2, Exhibits Hydrocarbonoclastic Potential in Anaerobic Environments." Front Microbiol **7**: 1071.

VerBerkmoes, N. C., M. B. Shah, P. K. Lankford, D. A. Pelletier, M. B. Strader, D. L. Tabb, W. H. McDonald, J. W. Barton, G. B. Hurst, L. Hauser, B. H. Davison, J. T. Beatty, C. S. Harwood, F. R. Tabita, R. L. Hettich and F. W. Larimer (2006). "Determination and comparison of the baseline proteomes of the versatile microbe Rhodopseudomonas palustris under its major metabolic states." J Proteome Res **5**(2): 287-298.

Vetting, M. W., N. Al-Obaidi, S. Zhao, B. San Francisco, J. Kim, D. J. Wichelecki, J. T. Bouvier, J. O. Solbiati, H. Vu, X. Zhang, D. A. Rodionov, J. D. Love, B. S. Hillerich, R. D. Seidel, R. J. Quinn, A. L. Osterman, J. E. Cronan, M. P. Jacobson, J. A. Gerlt and S. C. Almo (2015). "Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes." Biochemistry **54**(3): 909-931.

Walkenhorst, W. F., M. Merzlyakov, K. Hristova and W. C. Wimley (2009). "Polar residues in transmembrane helices can decrease electrophoretic mobility in polyacrylamide gels without causing helix dimerization." Biochimica et Biophysica Acta (BBA) - Biomembranes **1788**(6): 1321-1331.

Wang, P., Z. Yu, B. Li, X. Cai, Z. Zeng, X. Chen and X. Wang (2015). "Development of an efficient conjugation-based genetic manipulation system for *Pseudoalteromonas*." Microbial Cell Factories **14**: 11.

Widenhorn, K. A., J. M. Somers and W. W. Kay (1988). "Expression of the divergent tricarboxylate transport operon (*tctI*) of *Salmonella typhimurium*." J Bacteriol **170**(7): 3223-3227.

Wilkens, S. (2015). "Structure and mechanism of ABC transporters." F1000Prime Reports **7**: 14.

Willis, R. C. and C. E. Furlong (1974). "Purification and Properties of a Ribose-binding Protein from *Escherichia coli*." Journal of Biological Chemistry **249**(21): 6926-6929.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallographica Section D **67**(4): 235-242.

Winnen, B., R. N. Hvorup and M. H. Saier, Jr. (2003). "The tripartite tricarboxylate transporter (TTT) family." Res Microbiol **154**(7): 457-465.

Wubbeler, J. H., S. Hiessl, J. Schuldes, A. Thurmer, R. Daniel and A. Steinbuchel (2014). "Unravelling the complete genome sequence of *Advenella mimigardefordensis* strain DPN7T and novel insights in the catabolism of the xenobiotic polythioester precursor 3,3'-dithiodipropionate." Microbiology **160**(Pt 7): 1401-1416.

Youn, J.-W., E. Jolkver, R. Krämer, K. Marin and V. F. Wendisch (2008). "Identification and Characterization of the Dicarboxylate Uptake System DccT in *Corynebacterium glutamicum*." Journal of Bacteriology **190**(19): 6458-6466.

Yurgel, S. N. and M. L. Kahn (2004). "Dicarboxylate transport by rhizobia." FEMS Microbiol Rev **28**(4): 489-501.