Microbiology (2016), **162**, 823–836 DOI 10.1099/mic.0.000266

Allantoin transport protein, PucI, from *Bacillus subtilis*: evolutionary relationships, amplified expression, activity and specificity

Pikyee Ma,† Simon G. Patching,† Ekaterina Ivanova, Jocelyn M. Baldwin, David Sharples, Stephen A. Baldwin and Peter J. F. Henderson

School of BioMedical Sciences and the Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Correspondence
Peter J. F. Henderson
p.j.f.henderson@leeds.ac.uk

Received 8 January 2016
Accepted 26 February 2016

This work reports the evolutionary relationships, amplified expression, functional characterization and purification of the putative allantoin transport protein, PucI, from *Bacillus subtilis*. Sequence alignments and phylogenetic analysis confirmed close evolutionary relationships between PucI and membrane proteins of the nucleobase-cation-symport-1 family of secondary active transporters. These include the sodium-coupled hydantoin transport protein, Mhp1, from *Microbacterium liquefaciens*, and related proteins from bacteria, fungi and plants. Membrane topology predictions for PucI were consistent with 12 putative transmembrane-spanning α-helices with both N- and C-terminal ends at the cytoplasmic side of the membrane. The *pucI* gene was cloned into the IPTG-inducible plasmid pTTQ18 upstream from an in-frame hexahistidine tag and conditions determined for optimal amplified expression of the PucI(His_6) protein in *Escherichia coli* to a level of about 5 % in inner membranes. Initial rates of inducible PucI-mediated uptake of ^{14}C-allantoin into energized *E. coli* whole cells conformed to Michaelis–Menten kinetics with an apparent affinity ($K_m$app) of 24 ± 3 μM, therefore confirming that PucI is a medium-affinity transporter of allantoin. Dependence of allantoin transport on sodium was not apparent. Competitive uptake experiments showed that PucI recognizes some additional hydantoin compounds, including hydantoin itself, and to a lesser extent a range of nucleobases and nucleosides. PucI(His_6) was solubilized from inner membranes using n-dodecyl-β-D-maltoside and purified. The isolated protein contained a substantial proportion of α-helix secondary structure, consistent with the predictions, and a 3D model was therefore constructed on a template of the Mhp1 structure, which aided localization of the potential ligand binding site in PucI.

INTRODUCTION

Allantoin (or 5-ureidohydantoin) is a naturally occurring compound and a major metabolic intermediate in most living organisms, including bacteria, fungi, plants and animals. The compound name originates from its presence in fluid of the allantois embryonic secretory organ (Inman & Downs, 2007; Arora & Papaioannou, 2012) and it is produced in the degradation pathway of purine nucleobases by action of a urate oxidase (or uricase) enzyme on uric acid (Pizzichini *et al.*, 1996; Xi *et al.*, 2000; Johnson *et al.*, 2009). The gene for urate oxidase in humans (and in some other primates), however, is non-functional, so uric acid is the end product of purine catabolism and this is excreted in the urine (Johnson *et al.*, 2009). In most other mammals, allantoin is excreted in the urine, whilst in fish it is degraded further to ammonia before excretion. In bacteria, purines are used as secondary sources of nitrogen under nutrient-limiting conditions and their degradation ends in formation of the primary nitrogen source ammonia, which can then be utilized.

In the case of the Gram-positive soil bacterium *Bacillus subtilis*, the pathway for purine degradation and its controlling genes have been well characterized (Fig. 1; Nygaard *et al.*, 1996, 2000; Schultz *et al.*, 2001; Saxild *et al.*, 2001; Beier *et al.*, 2002; Goelzer *et al.*, 2008). The pathway is activated under conditions where the preferred sources of nitrogen (ammonia or glutamate) are not available, therefore requiring utilization of other compounds. The *B. subtilis* genes encoding the enzymes and transporters

†These authors contributed equally to this work.

Abbreviations: DDM, n-dodecyl-β-D-maltoside; IMAC immobilized-metal-ion affinity chromatography; NCS-1, nucleobase-cation-symport-1.

Twelve supplementary figures, one table and methods are available with the online Supplementary Material.

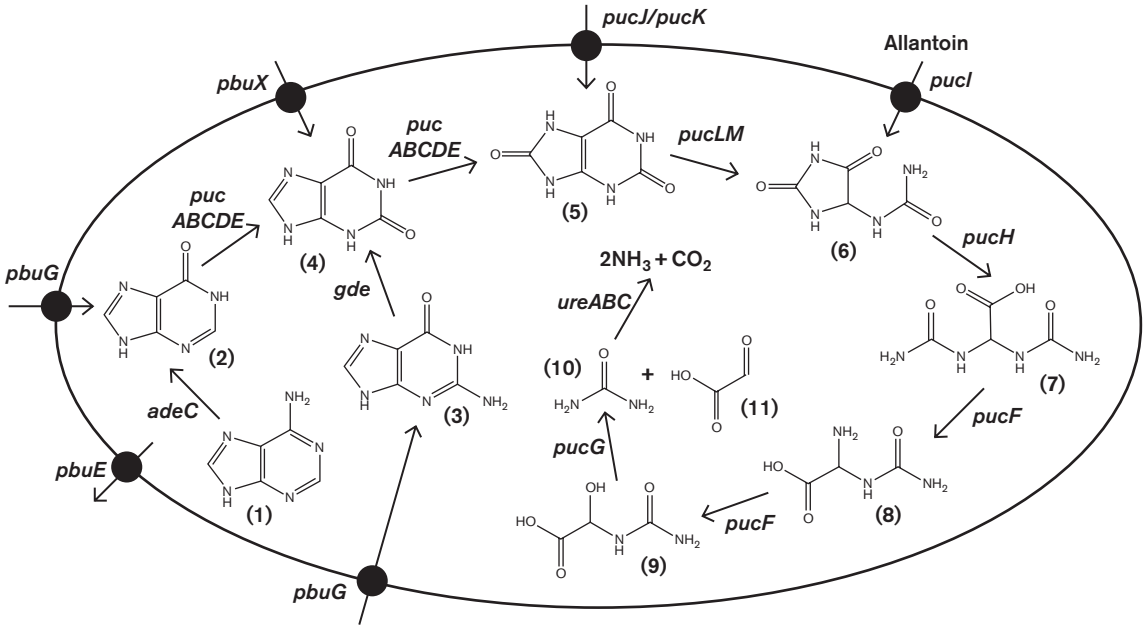

Fig. 1. Metabolic pathway for catabolism of purine nucleobases in *B. subtilis*. The large oval represents the inner cell membrane of *B. subtilis* containing transport proteins for purine nucleobases and their catabolites (black circles) encoded by the given genes: *pbuG* (hypoxanthine and guanine), *pbuX* (xanthine), *pucJ*/*pucK* (uric acid), *pucI* (allantoin) and *pbuE* (adenine and hypoxanthine efflux). Compound structures are numbered as follows: 1, adenine; 2, hypoxanthine; 3, guanine; 4, xanthine; 5, uric acid; 6, allantoin; 7, allantoic acid; 8, ureidoglycine; 9, ureidoglycolic acid; 10, urea; 11, glyoxylic acid. The genes encoding the enzymes for the conversion reactions are labelled on the reaction arrows. For simplicity, the other reactants and reaction products involved in each conversion are not shown on the diagram, but are listed as follows: *adeC* (adenine deaminase, $H_2O \rightarrow NH_3$), *gde* (guanine deaminase, $H_2O \rightarrow NH_3$), *pucABCDE* (xanthine dehydrogenase, $O_2 \rightarrow H_2O_2$), *pucLM* (uricase, $O_2 \rightarrow CO_2$), *pucH* (allantoinase, $O_2 \rightarrow CO_2$), *pucF* (allantoic acid aminohydrolase, $H_2O \rightarrow CO_2 + NH_3$), *pucG* (ureidoglycolase). This diagram was constructed based on information given in: Nygaard *et al.* (1996), (2000); Schultz *et al.* (2001); Saxild *et al.* (2001); Beier *et al.* (2002); Nygaard & Saxild (2005); Goelzer *et al.* (2008).

necessary for a complete degradation of purine nucleobases to ammonia have been identified as follows: *adeC* (adenine deaminase), *gde* (guanine deaminase), *pbuG* (hypoxanthine/guanine transporter), *pbuX* (xanthine transporter), *pucABCDE* (xanthine dehydrogenase), *pucF* (allantoic acid aminohydrolase), *pucG* (ureidoglycolase), *pucH* (allantoinase), *pucI* (allantoin transporter), *pucJK* (uric acid transport), *pucLM* (uricase) and *ureABC* (urease) (Fig. 1). The *gde*, *puc* and *ure* genes constitute a regulon that is controlled by the PucR gene product for which allantoic acid, allantoin and uric acid can all serve as effector molecules (Schultz *et al.*, 2001). The induction of *pucR* gene expression under nitrogen-limiting conditions is also under the control of the global nitrogen state regulatory protein TnrA (Wray *et al.*, 1996; Fisher, 1999; Beier *et al.*, 2002). The *pbuG* and *pbuX* genes are under the control of the PurR regulon and the *xpt–pbuX* operon, respectively (Saxild *et al.*, 2001; Christiansen *et al.*, 1997). A purine base efflux pump encoded by the *pbuE* gene has also been identified, which controls the intracellular pool of purine bases including adenine and hypoxanthine (Nygaard & Saxild, 2005) (Fig. 1). PbuE is a member of the DHA1 family of the major facilitator superfamily.

B. subtilis is able to utilize a number of different compounds, including allantoin, as its sole nitrogen source (Fisher, 1993; Cruz-Ramos *et al.*, 1997; Nygaard *et al.*, 2000; Beier *et al.*, 2002; Goelzer *et al.*, 2008) and since mutants in the *pucI* gene were demonstrated not to grow on allantoin, this was evidence to suggest that it encodes an allantoin transporter (Schultz *et al.*, 2001). Based on protein sequence alignments, PucI from *B. subtilis* has also been classified as a member of the widespread nucleobase-cation-symport-1 (NCS-1) family of secondary active transport proteins (de Koning & Diallinas, 2000; Pantazopoulou & Diallinas, 2007; Weyand *et al.*, 2010; Witz *et al.*, 2014). A structural paradigm for the NCS-1 family is the sodium-coupled hydantoin transport protein, Mhp1, from *Microbacterium liquefaciens* (Suzuki & Henderson, 2006; Jackson *et al.*, 2013) for which crystal structures have been determined with the protein in three different conformations (Weyand *et al.*, 2008; Shimamura *et al.*, 2010). Mhp1 has provided a principal model for the alternating access mechanism of membrane transport and for the mechanism of ion-coupling (Shimamura *et al.*, 2010; Weyand *et al.*, 2011; Adelman *et al.*, 2011; Shi, 2013; Kazmier *et al.*, 2014). Direct experimental evidence for

the transport of allantoin by PucI has not yet been demonstrated, however. Here we use protein sequence alignments, phylogenetic analysis and membrane topology predictions to illustrate how PucI is evolutionarily related to other putative bacterial allantoin permeases, Mhp1 and additional NCS-1 family transport proteins from eukaryotes. We achieved the amplified expression in *Escherichia coli* of PucI with a C-terminal hexahistidine tag, characterized its transport of ^{14}C-labelled allantoin in energized *E. coli* whole cells, established its specificity of ligand recognition and accomplished its solubilization in detergent and purification. The protein contains a high proportion of α-helix, as predicted from its sequence, and so a more precise model of its 3D structure and ligand binding site are derived from comparisons with Mhp1.

METHODS

General. For details of general materials, sources, methods and equipment see Supplementary Methods (available in the online Supplementary Material).

Cloning of the *pucI* gene and amplified expression of PucI protein in *E. coli*. Cloning and amplification of expression of the PucI protein in *E. coli* was achieved using a strategy that we and others have found successful with a wide range of bacterial and archaeal membrane proteins (Ward *et al.*, 1999; Saidijam *et al.*, 2003; Szakonyi *et al.*, 2007; Ma *et al.*, 2008, 2013; Bettaney *et al.*, 2013). This involved design of PCR primers (forward, 5′-CCGGAATTCGCATATGAAATTAAAAGAGAGTCAGCAGCAATCCA-3′, and reverse, 5′-AAAACTGCAGCTTCAGCCTGGCGGACCTGCGCATGTT-3′) to extract and amplify the *pucI* gene from *B. subtilis* genomic DNA, introducing *Eco*RI and *Pst*I restriction sites at the 5′ and 3′ ends, respectively. The restriction-digested PCR product was ligated into plasmid pTTQ18 (Stark, 1987) immediately upstream from a hexahistidine tag coding sequence and the resultant pTTQ18-pucI(His_6) construct was used to transform *E. coli* BL21(DE3) cells. For full details of the cloning procedure see Supplementary Methods. Expression tests for PucI were performed using three different types of media for cell growth (LB, 2TY, M9 minimal) and comparing uninduced versus induced cells using 0.5 mM IPTG for induction (Fig. 2). Cell growth curves showed a significant decrease in growth rate and maximum cell density achieved in induced versus uninduced cells, indicative of toxicity from amplified expression of heterologous genes (Fig. 2a). SDS-PAGE analysis of total membrane preparations from these cells (Fig. 2b) revealed an amplified protein band migrating at a molecular mass position of ~36 kDa in the induced samples. The effect of induction time ranging from 0.1 to 22 h on amplified expression revealed an optimum of 3 h by SDS-PAGE and Western blot analysis using an antibody to the $(His)_6$ epitope (Fig. 2c, d). The Western blot signal (Fig. 2d) also confirmed the presence of a hexahistidine tag on the amplified protein of 36 kDa apparent molecular mass; therefore, the C-terminal end was intact. To provide sufficient quantities of protein for purification and further analysis, these culture conditions were then scaled up to volumes of 10 l in flasks or 30 l in a fermentor, and the resultant cells were used to prepare inner membranes. For full details of cell growth and membrane preparation procedures see Supplementary Methods. SDS-PAGE analysis confirmed amplified expression of the PucI(His_6) protein at around 5 % of total protein in inner membranes as measured by densitometric analysis of the Coomassie-stained gel (Fig. S1). Since the level of expression of the amplified protein appeared higher in membranes coming from cells grown in minimal medium than in the rich media, minimal medium was used for all further cell cultures. The appearance at ~36 kDa in SDS-PAGE gels instead of the expected 54 kDa is a phenomenon routinely observed with membrane proteins (Ward *et al.*, 2000; Ma *et al.*, 2008, 2013; Rath *et al.*, 2009; Findlay *et al.*, 2010; Bettaney *et al.*, 2013), probably a result of only partial unfolding of a membrane protein in SDS (Findlay *et al.*, 2010).

Whole-cell transport and competition assays. Measurements of uptake of ^{14}C-allantoin (synthesized in-house; Patching, 2009) into energized whole cells of *E. coli* were performed using a method based on that of Henderson *et al.* (1977). Cells were grown in M9 minimal medium supplemented with glycerol (20 mM) and carbenicillin (100 μg ml^{-1}) in volumes of 50 or 100 ml at 37 °C in 250 or 500 ml baffled conical flasks with aeration at 200 r.p.m. to an OD_{680} of ~0.4–0.6. The cells were then either left uninduced or induced with IPTG (0.5 mM) and grown for a further 1 h. After harvesting by centrifugation (2500 ***g*** 10 min, in Falcon tubes using a bench-top instrument), the cells were washed three times with 40 ml transport buffer (150 mM KCl, 5 mM MES, pH 6.6) and then resuspended in the same buffer to an OD_{680} of 2.0. The basic method for the assay is described as follows. Cells were energized for building up the proton gradient to drive allantoin transport by incubating aliquots of the suspension (493.4 μl) with 20 mM glycerol (5 μl of 2 M) and with bubbled air in a bijou bottle held in a water jacket at 25 °C. After exactly 3 min, ^{14}C-allantoin at a concentration of 50 μM [1.6 μl of a 15.6 mM solution with specific activity 2.5 μCi ml^{-1} (92.5 kBq ml^{-1})] was added with brief mixing. At times of exactly 15 s and 2 min after adding the radiolabelled substrate, 200 μl aliquots were transferred to cellulose nitrate filters (0.45 μm pore size), pre-soaked in transport buffer, on a vacuum manifold and washed immediately with transport buffer (3 ml) three times. The filters were transferred to scintillation vials with 10 ml Emulsifier-Safe liquid scintillation fluid (Perkin Elmer) and incubated overnight. The level of ^{14}C radioactivity retained by the cells was measured by liquid scintillation counting (Packard Tri-Carb 2100TR instrument). Background counts were measured from washing filters under vacuum in the absence of cells or radiolabelled substrate. Standard counts were measured by transferring 1 μl radiolabelled substrate stock solution directly to a washed filter in the vial. The uptake of radiolabelled substrate into the cells was calculated using the following equation: uptake [nmol (mg cells)$^{-1}$]=(cell counts−background counts) × (total assay volume/sample taken volume) × [(1 (mg cells)$^{-1}$] × (moles of standard/standard counts), where dry weight of cells (mg) = total assay volume (ml) × OD_{680} × 0.68. The times of sampling and/or of added ^{14}C-allantoin concentration were varied for kinetic analyses. To test the effect of potential competing compounds on ^{14}C-allantoin uptake, the unlabelled compound was added from a stock solution in 100 % DMSO to the cells prior to the energization period, the final concentration of the unlabelled compound was 500 μM and the final concentration of DMSO was 2 %. Relative uptake values were measured as a percentage of those obtained from samples in the absence of any added unlabelled compound or DMSO.

Protein solubilization, purification and circular dichroism spectroscopy. Inner membrane preparations were solubilized for up to 4 h at 4 °C in a buffer containing 20 mM Tris (pH 8.0), 1 % n-dodecyl-β-D-maltoside (DDM), 20 % (v/v) glycerol and 300 mM sodium chloride (Table S1). Immobilized-metal-ion affinity chromatography (IMAC) was then performed to purify the PucI(His_6)-tagged protein as described in Supplementary Methods and Fig. S10. Far-UV circular dichroism spectroscopy analysis of purified protein (0.05 mg ml^{-1}) in potassium phosphate buffer (10 mM, pH 7.6) with 0.05 % DDM was performed using a Jasco J-715 spectropolarimeter at a temperature of 18 °C with constant nitrogen flushing as described in Supplementary Methods and Fig. S10.

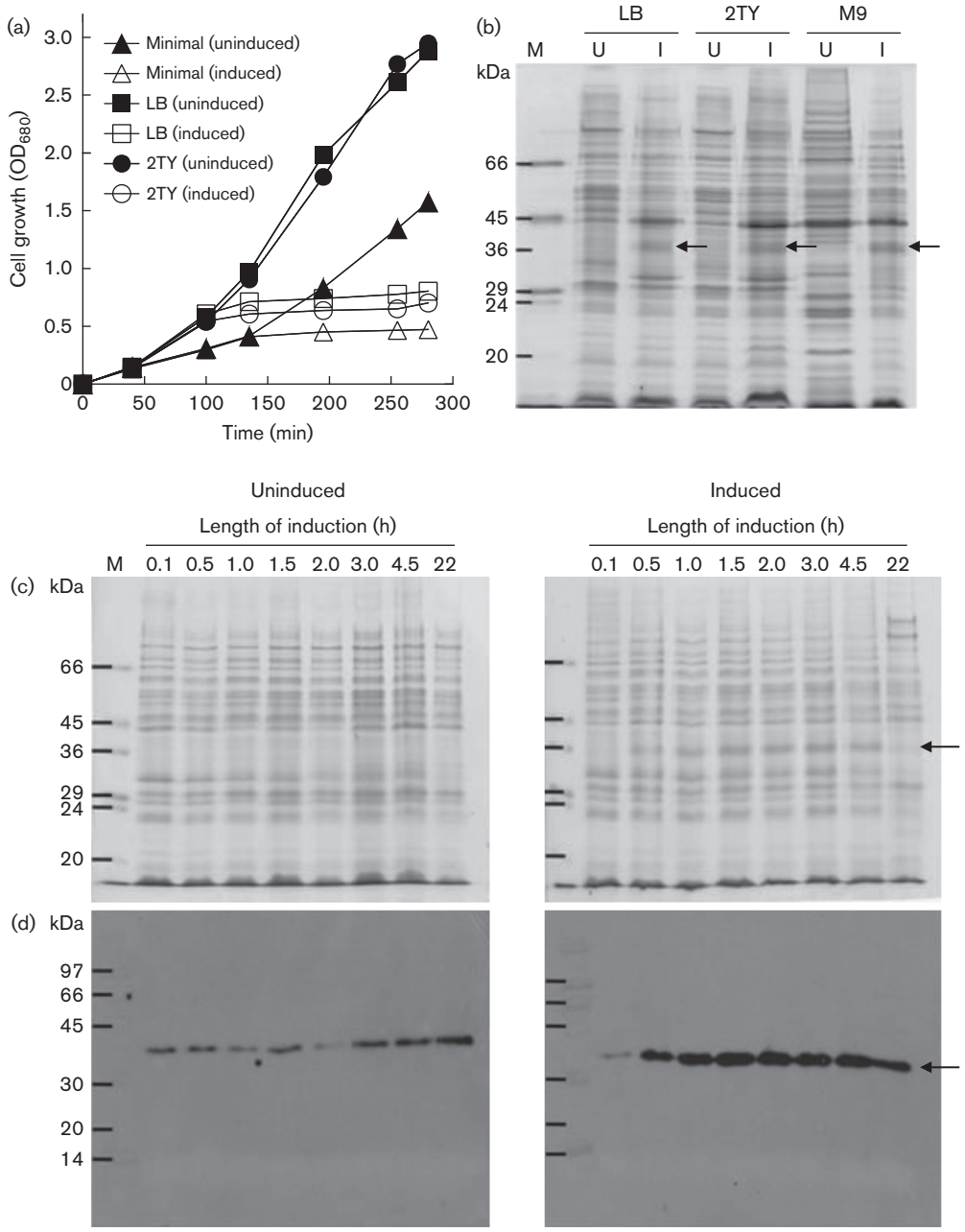

Fig. 2. Induced amplified expression of the PucI(His$_6$) protein in *E. coli*. (a) Growth curves for BL21(DE3) cells containing the pTTQ18-pucI(His$_6$) construct in LB (squares), 2TY (circles) and M9 minimal (triangles) media supplemented with 20 mM glycerol left uninduced (black shapes) or induced with 0.5 mM IPTG at OD_{680} 0.4–0.6 (white shapes). (b) SDS-PAGE analysis of total membranes prepared from the cells described above. The arrows indicate the position of the amplified PucI(His$_6$) protein. M, Molecular mass markers; I, induced; U, uninduced. (c, d) Effect of induction time on amplified expression of the PucI(His$_6$) protein in *E. coli* shown by SDS-PAGE (c) and Western blot (d) analysis of total membrane preparations from cultures of BL21(DE3) cells containing the pTTQ18-pucI(His$_6$) construct in M9 minimal medium supplemented with 20 mM glycerol uninduced (left) or induced with 0.5 mM IPTG at OD_{680} 0.4–0.6 (right) and grown for the given further length of time, from 0.1 to 22 h. The arrows indicate the position of the amplified PucI(His$_6$) protein. M, Molecular mass markers.

Homology modelling and ligand docking. Homology modelling of *B. subtilis* PucI was performed using SWISS-MODEL server (http://swissmodel.expasy.org/; Biasini *et al.*, 2014) and the crystal structure of *M. liquefaciens* hydantoin transport protein Mhp1 with bound benzylhydantoin (PDB 4D1B; Simmons *et al.*, 2014) as the model template, with which it shares an overall 51.2 % sequence homology (25.1 % identical, 26.9 % highly similar). The output model from the server was optimized by manual corrections using the

Crystallographic Object-Oriented Toolkit (Coot) (http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/; Emsley & Cowtan, 2004), followed by energy minimization using Swiss-PDBViewer (http://spdbv.vital-it.ch/; Guex & Peitsch, 1997). The Ramachandran plot assessment tool RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php; Lovell *et al.*, 2003) was used to validate the predicted structure, which had most of the Ramachandran plot within the favourable region and only 0.6 % outliers. The PucI model was manually docked with (*S*)-allantoin based on superimposition with benzylhydantoin bound in the Mhp1 crystal structure.

RESULTS AND DISCUSSION

Evolutionary relationships of PucI

The PucI protein from *B. subtilis* strain 168, also known as ALLP, comprises 490 aa residues and has a close evolutionarily relationship with other putative allantoin transporters from a number of different bacteria. A search of the UniProt KnowledgeBase (http://www.uniprot.org/) for proteins encoded by '*allP*' genes identified 20 bacterial putative transport proteins, including PucI itself. A multiple protein sequence alignment of these 20 proteins revealed a significant number of conserved residues with 26.7 % identical and a further 24.5 % highly similar (combined total 51.2 %). A further search based on sequence similarity with PucI from *B. subtilis* identified homologous putative allantoin permeases in 23 other different species of bacteria. A multiple protein sequence alignment of the 24 proteins revealed 38 identically conserved residues and over 100 further residues that are highly similar in all of the proteins (Fig. S2). These conserved positions, which are predominated by aromatic, charged and polar residues, are spread throughout the entire lengths of the sequences, and may represent important structural and functional motifs for achieving allantoin recognition and transport in a distinct group of related allantoin transport proteins. A sequence alignment between PucI from *B. subtilis* and Mhp1 from *M. liquefaciens* also revealed a significant number of conserved residues with 25.1 % identical and a further 26.9 % highly similar (combined total 52.0 %) (Fig. S3), thus confirming a close evolutionary relationship between PucI and Mhp1, and classification of PucI as a NCS-1 family transport protein.

Further sequence alignments (Figs S4 and S5) and a phylogenetic analysis (Fig. 3) were performed using experimentally characterized NCS-1 family transporters from bacteria, fungi and plants to reveal the wider evolutionary relationships of PucI. Interestingly, PucI shows closest evolutionary relationship with the plant transporters AtNCS1 (PLUTO) from *Arabidopsis thaliana* (adenine/guanine/uracil) (Mourad *et al.*, 2012; Witz *et al.*, 2014), CtNCS1 from *Chlamydomonas reinhardtii* (adenine/guanine/uracil/allantoin) (Schein *et al.*, 2013), ZmNCS1 from *Zea mays* (adenine/guanine/cytosine) and SvNCS1 from *Setaria viridis* (adenine/guanine/hypoxanthine/cytosine/allantoin) (Rapp *et al.*, 2016). Indeed, the values of sequence homology between PucI and the plant proteins

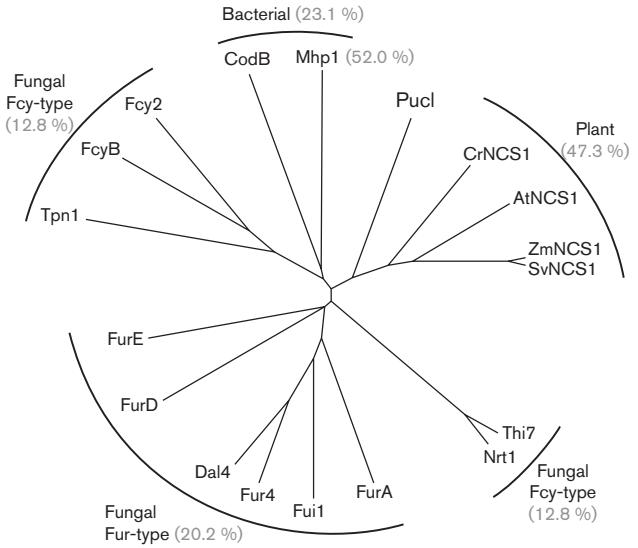

Fig. 3. Evolutionary relationships of PucI with NCS-1 family transporters. Phylogenetic tree showing evolutionary relationships of PucI from *B. subtilis* (P94575) with experimentally characterized bacterial, fungal (Fur-type and Fcy-type) and plant NCS-1 family transport proteins. The NCS-1 proteins are: Mhp1 from *M. liquefaciens* (D6R8X8), CodB from *E. coli* (P0AA82), FurA from *Asp. nidulans* (Q5BFM0), FurD from *Asp. nidulans* (A6N844), FurE from *Asp. nidulans* (Q5ATG4), Fur4 from *Sac. cerevisiae* (P05316), Dal4 from *Sac. cerevisiae* (Q04895), Fui1 from *Sac. cerevisiae* (P38196), FcyB from *Asp. nidulans* (C8V329), Fcy2 from *Sac. cerevisiae* (P17064), Thi7 from *Sac. cerevisiae* (Q05998), Tpn1 from *Sac. cerevisiae* (P53099), Nrt1 from *Sac. cerevisiae* (Q08485), AtNCS1 (PLUTO) from *Ara. thaliana* (Q9LZD0), CrNCS1 from *C. reinhardtii* (A8J166), ZmNCS1 from *Z. mays* (B4FJ20), SvNCS1 from *Set. viridis* (V9SBV7). Protein sequences were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and aligned using the online multiple sequence alignment tool CLUSTAL Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers *et al.*, 2011). The resultant neighbour-joining phylogenetic tree was exported in Newick format and drawn using the online tool Phylodendron (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html). The overall sequence homology (percentage of identical plus highly similar residues) of PucI with Mhp1 and with each group of proteins is given in parentheses. These values were calculated from the sequence alignments shown in Figs S3, S4 and S5.

AtNCS1 and CrNCS1 (overall 51.2 %) are very similar to those between PucI and Mhp1. Note that the two recently characterized grass proteins ZmNCS1 and SmNCS1 share an extremely high sequence homology with each other (90.4 % identical, 4.6 % highly similar) (Rapp *et al.*, 2016). Grouping of Mhp1 with the other characterized bacterial NCS-1 protein, cytosine transporter CodB from *E. coli* (Danielsen *et al.*, 1995), reduces the overall sequence homology with PucI from 52.0 to 23.1 %, and moves it further away from PucI and the plant NCS-1 transporters in the phylogenetic tree (Fig. 3). When CodB is removed

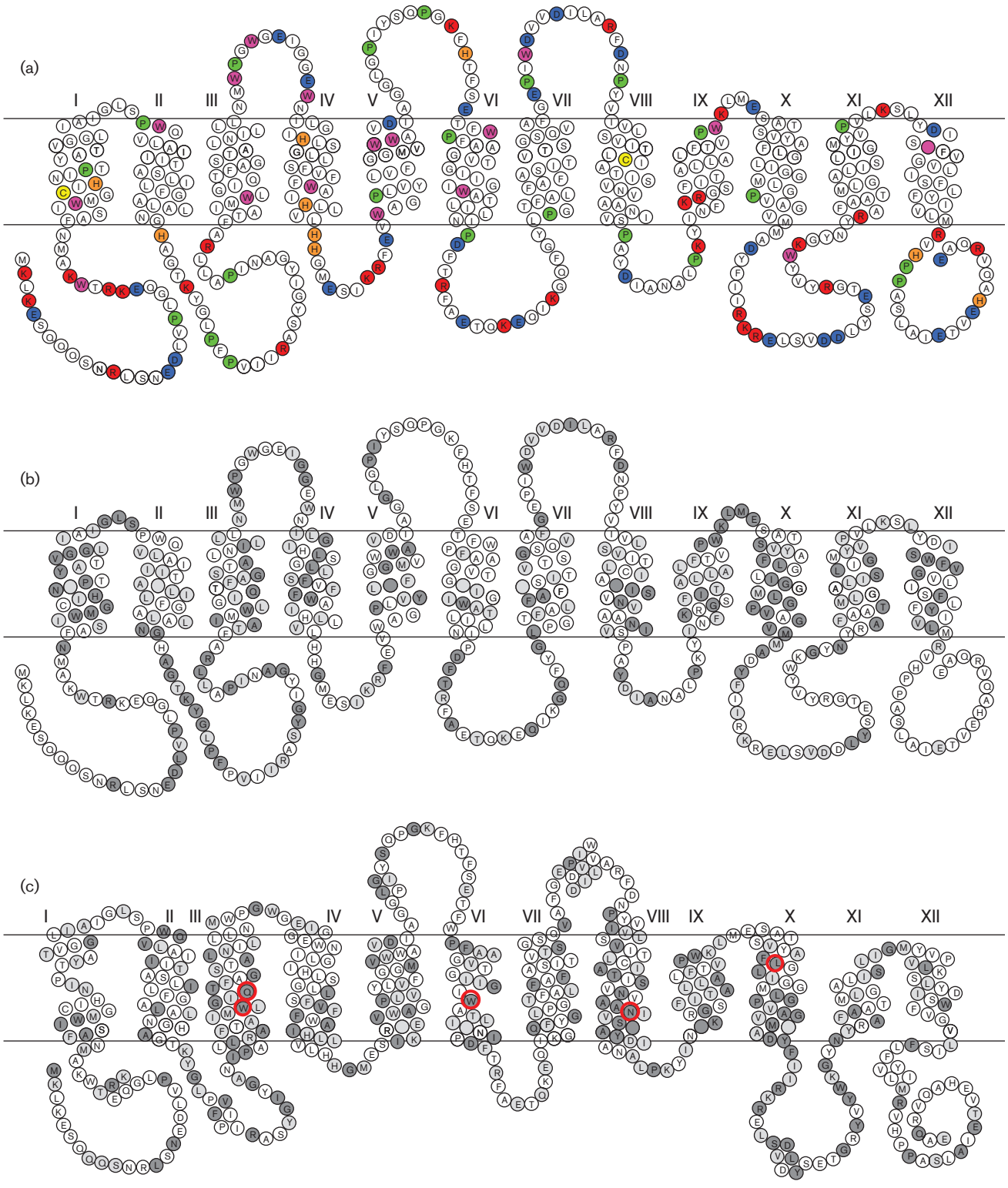

Fig. 4. Putative membrane topology of the PucI protein from *B. subtilis* and conserved residues shared with bacterial putative allantoin permeases and with the hydantoin transport protein Mhp1 from *M. liquefaciens*. (a, b) The putative membrane topology of the PucI protein (Bsu3645, P94575, ALLP_BACSU) from *B. subtilis* (strain 168) based on analyses of its sequence using the membrane topology prediction tools TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/; Krogh *et al.*, 2001) and TOPCONS consensus prediction server (http://topcons.cbr.su.se/; Bernsel *et al.*, 2009) (Fig. S6). The putative 12 transmembrane-spanning α-helices are drawn from the N- to the C-terminal end left to right and labelled using roman numerals I–XII. (a) Colours highlight specific types of amino acid residue as follows: positively charged (arginine and lysine,

red), negatively charged (aspartic acid and glutamic acid, blue), cysteine (yellow), tryptophan (pink), histidine (orange), proline (green). (b) Grey shading shows residues that are conserved between PucI from *B. subtilis* and 19 other proteins encoded by *allP* genes. (c) This diagram is based on an alignment between the sequences for PucI from *B. subtilis* and the hydantoin transport protein Mhp1 from *M. liquefaciens* and on the positions of helices in the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B; Simmons *et al.*, 2014) (Fig. S3). Residues are shaded to show those that are conserved between PucI and Mhp1. In (b) and (c), identical residues are shown as dark grey and highly similar residues are pale grey. In (c), the red circles highlight residues identical with those in Mhp1 that are involved in direct binding interactions with the hydantoin substrate: Trp117, Gln121, Trp220, Asn318, Leu363.

from this analysis, Mhp1 aligns best with PucI and the plant transporters. PucI also shows more distant evolutionary relationships with fungal Fur-type and Fcy-type NCS-1 transporters with overall sequence homologies of 20.2 and 12.8 %, respectively (Fig. 3). The Fur-type transporters FurA (allantoin), FurD (uracil/uric acid), FurE (uracil/uric acid/allantoin), Fur4 (uracil), Dal4 (allantoin), Fui1 (uridine), and the Fcy-type transporters FcyB (purines/cytosine), Fcy2 (purines/cytosine), Thi7 (thiamine), Tpn1 (pyridoxine), Nrt1 (nicotinamide riboside) from *Aspergillus nidulans* and *Saccharomyces cerevisiae*, used in the phylogenetic analysis have been well characterized in terms of function and substrate specificity (Hamari *et al.*, 2009; Krypotou *et al.* 2015). Studies also suggest that these two distinct fungal NCS-1 subfamilies and the plant homologues originated through independent horizontal transfers from prokaryotes and demonstrate that transport activities in NCS-1 family proteins have appeared independently by convergent evolution (Hamari *et al.*, 2009; Krypotou *et al.* 2015). This is discussed further below in the context of substrate recognition and solute binding by PucI.

Predicted topology of PucI in the membrane

Membrane topology analyses of PucI using the amino acid sequence with online tools TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), which uses a hidden Markov model (Krogh *et al.*, 2001), and the TOPCONS consensus prediction server (http://topcons.cbr.su.se/; Bernsel *et al.*, 2009) both predicted 12 transmembrane-spanning α-helices with the N- and C-terminal ends at the cytoplasmic side of the membrane (Fig. S6). The limits for all 12 transmembrane helices predicted by TMHMM and TOPCONS were very similar, differing by only 1 or 2 residues in the large majority of cases. A diagram of the putative membrane topology of PucI based on these predictions is shown in Fig. 4(a), which also has some specific types of amino acids coloured. The combined number of positively charged residues (arginine and lysine) is closely similar to the combined number of negatively charged residues (aspartic acid and glutamic acid) at 29 and 28, respectively, but the positively charged residues dominate on the cytoplasmic side of the membrane consistent with the positive inside rule of von Heijne (1992), whilst negatively charged residues are found mostly in the loops on both the cytoplasmic and periplasmic sides of the membrane. The protein has 2 cysteine residues, 1 in each of putative helices I and VIII, 17 tryptophans, 9 histidines and 22 prolines. Indeed, the amino acid composition of PucI is similar to the average pattern of percentage amino acid composition in all the secondary transport proteins from *E. coli* (Saidijam & Patching, 2015). The topology diagram has also been shaded to show the residues that are conserved between PucI and other bacterial transporters encoded by *allP* genes (Fig. 4b). A topology diagram was also drawn based on the sequence alignment between PucI and Mhp1 (Fig. S3), and on the positions of helices in the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B; Simmons *et al.*, 2014) (Fig. 4c). The diagram is shaded to show residues that are conserved between PucI and Mhp1. The diagrams based on topology prediction and on the crystal structure of Mhp1 have some similarities and also some interesting differences. They have the same general overall fold with 12 transmembrane-spanning α-helices, both the N- and C-terminal ends at the cytoplasmic side of the membrane and a long extracellular region at both ends. The diagrams have different numbers of residues in the transmembrane helices, which have different start and end positions in the majority of cases. The diagram based on Mhp1 shows an overall greater number of residues in transmembrane helices, which consequently provides smaller extramembrane loops. This diagram also shows breaks in the first and sixth transmembrane helices and six internal and external helical regions revealed by the crystal structure. Although the total number of conserved residues in the diagrams is very similar, the distribution of conserved residues is somewhat different. The evolutionary relationship of PucI to the other putative allantoin transporters in bacteria may be distinguished from its relationship to Mhp1, which will be discussed further below.

PucI-mediated transport of ^{14}C-allantoin into energized whole cells

In order to test experimentally the function of PucI as a transporter of allantoin, we used ^{14}C-labelled allantoin in uptake assays with energized *E. coli* whole cells containing the PucI-expressing construct with or without added IPTG inducer, or the empty plasmid pTTQ18. A time-course of ^{14}C-allantoin (50 μM) uptake over 60 min revealed negligible uptake into cells containing pTTQ18, a modest uptake into uninduced cells containing pTTQ18-pucI(His_6) and a much higher uptake into induced cells containing pTTQ18-pucI(His_6) (Fig. 5a). This was a clear demonstration of mediation by PucI of ^{14}C-allantoin

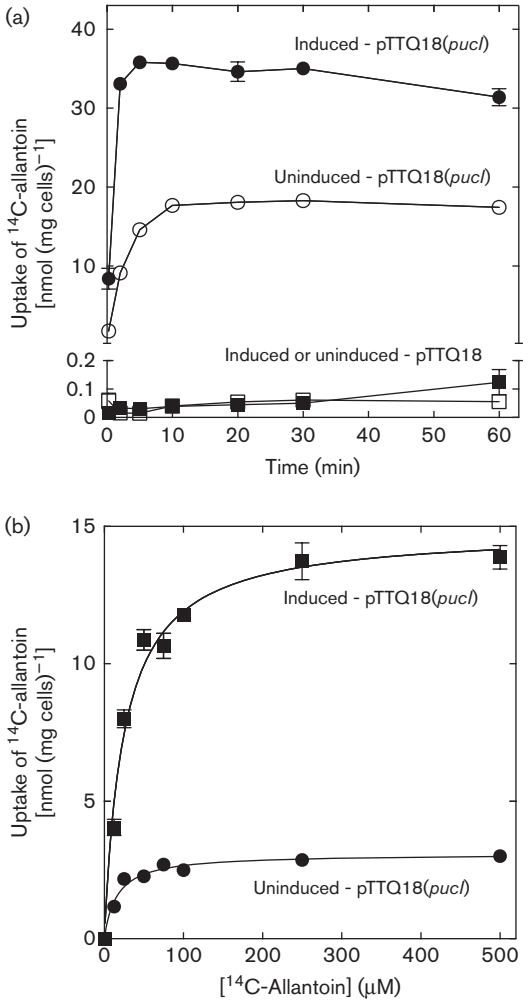

Fig. 5. PucI-mediated ^{14}C-allantoin uptake into energized whole cells. (a) Time-course of ^{14}C-allantoin uptake. Uptake of ^{14}C-allantoin (50 μM) at time points over 60 min into energized *E. coli* BL21(DE3) cells containing the empty plasmid pTTQ18 (squares) or the construct pTTQ18-pucI(His$_6$) (circles) that were uninduced (white shapes) or induced with IPTG (black shapes). (b) Concentration dependence of initial rate PucI-mediated ^{14}C-allantoin uptake. Michaelis–Menten plots for the uptake of ^{14}C-allantoin at concentrations up to 500 μM after 15 s into energized *E. coli* BL21(DE3) cells containing the construct pTTQ18-pucI(His$_6$) that were uninduced (black circles) or induced with IPTG (black squares). The plot for induced cells was analysed by a least-squares fit to obtain values for the apparent affinity of initial-rate transport ($K_m$app) and the maximum velocity (V_{max}) of 24.4 ± 3 μM and 14.8 nmol (mg cells)$^{-1}$, respectively. For all measurements, cells were cultured in minimal medium with 20 mM glycerol and induced at OD_{680} 0.4–0.6 with 0.5 mM IPTG for 1 h. Harvested cells were washed three times with assay buffer (150 mM KCl, 5 mM MES, pH 6.6) and resuspended to an OD_{680} of 2.0. Cells were energized with 20 mM glycerol and bubbled air for 3 min, followed by incubation with ^{14}C-allantoin at the given concentrations and removal of aliquots for analysis at the given times. The data points represent the mean of triplicate measurements and the error bars represent SEMs.

uptake into cells and, therefore, confirmation that PucI is a transport protein for allantoin. The negligible uptake into cells containing the empty plasmid pTTQ18 demonstrates that there is no contribution from any endogenous active transporters of allantoin in *E. coli*. The relatively small but significant uptake into uninduced cells is the result of some leaky expression from the pTTQ18-pucI(His$_6$) construct. Examination of the appearance of allantoin transport after addition of IPTG at times from 0.1 to 22 h revealed that 0.5 to 2 h was sufficient for optimal expression (Fig. S7), so a 1 h length of induction was used for all subsequent transport experiments. A possible effect of sodium ions on PucI-mediated ^{14}C-allantoin transport was investigated by performing the assay in the presence of a range of concentrations of sodium chloride from 0 to 150 mM (Fig. S8). The uptake of ^{14}C-allantoin was essentially the same over the entire range of sodium chloride concentrations for measurements after both 15 s and 2 min. This suggests that the transport of allantoin by PucI is not dependent on sodium, unlike the transport of 5-aryl-substituted hydantoins by the homologous Mhp1 protein. PucI is, therefore, likely to be a proton-coupled symporter. Note that the absolute values for ^{14}C-allantoin uptake by PucI, of up to around 35 nmol (mg cells)$^{-1}$, are significantly higher than those for uptake of ^{14}C-labelled 5-aryl-substituted hydantoins by Mhp1, which are up to around 1 nmol (mg cells)$^{-1}$ (Simmons *et al.*, 2014). The concentration dependence of PucI-mediated ^{14}C-allantoin uptake conformed to Michaelis–Menten kinetics in both uninduced and induced cells containing the PucI-expressing plasmid, with a significantly higher uptake into induced cells (Fig. 5b). In induced cells, values for the apparent affinity of initial-rate transport ($K_m$app) and the maximum velocity (V_{max}) in a typical experiment were 24.4 ± 3 μM and 14.8 nmol (mg cells)$^{-1}$, respectively. These direct measurements of allantoin transport are the first to our knowledge using radiolabelled substrate for any putative allantoin transporter or NCS-1 family protein.

Specificity of ligand recognition by PucI

The specificity of ligand recognition by PucI was investigated by testing the effects of a tenfold excess of 20 unlabelled potential competing compounds on the PucI-mediated uptake of ^{14}C-allantoin into energized *E. coli* whole cells (Fig. 6). The compounds tested included a range of hydantoins, nucleobases and nucleosides and their catabolites (Fig. S9), which are substrates of NCS-1 transporters and NCS-2/NAT transporters (Goudela *et al.*, 2005) or closely related chemically to these, and therefore most likely to be recognized by PucI. The highest competitive effect was produced by allantoin, therefore confirming this as the principal compound recognized by PucI. All of the other compounds tested had some degree of competitive effect on the uptake of ^{14}C-allantoin. After allantoin, the compounds having the highest effect were

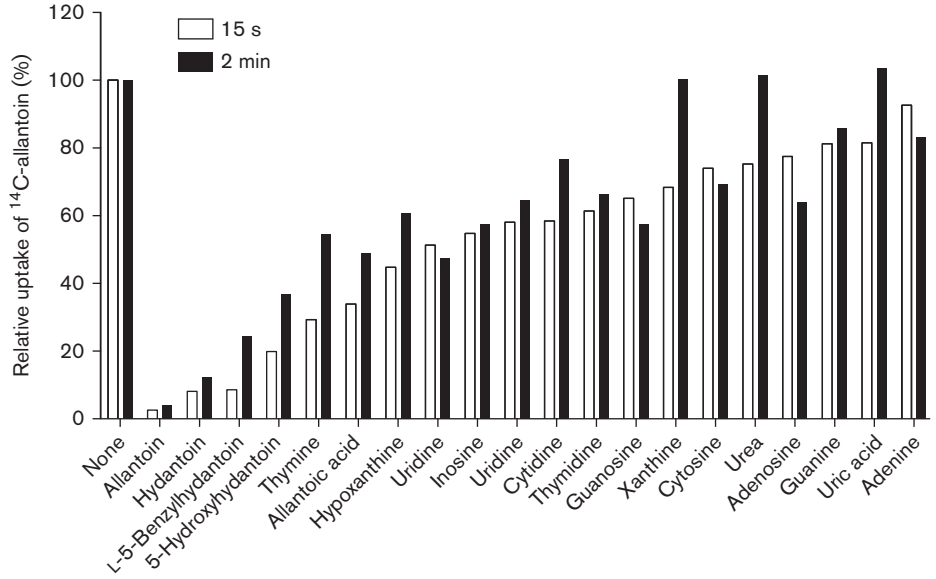

Fig. 6. Competition of PucI-mediated ^{14}C-allantoin uptake into energized whole cells by unlabelled compounds. Relative uptake (%) of ^{14}C-allantoin (50 μM) after 15 s and 2 min into energized *E. coli* BL21(DE3) cells containing the construct pTTQ18-pucI(His$_6$) that were induced with IPTG in the presence of a number of unlabelled competing compounds (500 μM). Cells were cultured in minimal medium with 20 mM glycerol and induced at OD_{680} 0.4–0.6 with 0.5 mM IPTG for 1 h. Harvested cells were washed three times with assay buffer (150 mM KCl, 5 mM MES, pH 6.6) and resuspended to an OD_{680} of 2.0. Cells were energized with 20 mM glycerol and bubbled air for 3 min followed by incubation with ^{14}C-allantoin (50 μM) and removal of aliquots for analysis after 15 s and 2 min. Unlabelled competing compounds (500 μM) were introduced prior to the energization period. The non-competed uptake rate was taken as 100 % corresponding to 14.6 and 30.4 nmol ^{14}C-allantoin (mg cells)$^{-1}$ for 15 s and 2 min post addition of ^{14}C-allantoin, respectively. Final uptake samples contain 2 % DMSO or no DMSO for the control sample (denoted by 'None' in the figure). The data points represent the mean of duplicate measurements. Structures of many of the compounds are shown in Fig. S9.

hydantoin, L-5-benzylhydantoin and 5-hydroxyhydantoin. PucI, therefore, has a significantly higher level of recognition for the hydantoin moiety alone than does Mhp1, since a tenfold excess of hydantoin had only a small competitive effect on ^{14}C-indolylmethylhydantoin transport by Mhp1 (Simmons *et al.*, 2014). PucI can also recognize and bind hydantoins substituted at the 5 position with both hydrophilic and hydrophobic groups, whilst Mhp1 has significant recognition only for hydantoins substituted at the 5 position with aromatic groups (Simmons *et al.*, 2014). The competitive uptake results also suggest that PucI is able to recognize and bind a range of pyrimidine and purine nucleobases and nucleosides, although to a lesser extent than the hydantoins. In order of decreasing competitive effect, these include thymine, hypoxanthine, uracil, inosine, uridine, cytidine, thymidine and guanosine. Based on comparisons with related fungal and plant NCS-1 transporters, residues in PucI that could possibly have an effect on the recognition and/or transport of these compounds are discussed further below. The compounds identified as competitors of ^{14}C-allantoin uptake would themselves need to be used in radiolabelled form in uptake assays to test whether they are transported substrates, rather than inhibitors, for PucI.

Purification of PucI

As a first step towards determination of the structure of PucI, solubilization trials of inner membrane preparations from cells with amplified expression of PucI(His$_6$) using eight different detergents at a concentration of 1 % identified DDM as the most successful for achieving the highest recovery of PucI in solubilized form (not shown). The C-terminal hexahistidine tag on PucI was then used to assist its purification from a larger scale solubilization of the inner membranes (shown in Fig. S10) using IMAC with a Ni-NTA resin. SDS-PAGE and Western blot analysis of fractions from a purification that used imidazole at a concentration of 20 mM for incubation and washing steps, and 200 mM for elution, revealed a good solubilization of PucI protein from the inner membranes with a smaller relative amount left in the insoluble fraction and a substantial purification from the IMAC column (Fig. S10a, b). The eluted protein was obtained in quantities of hundreds of micrograms and had a purity of at least 77 % as determined by densitometric analysis of the Coomassie-stained SDS-PAGE separation. The purified PucI(His$_6$) protein had a far-UV circular dichroism spectrum characteristic of a high content of α-helix

(Fig. S10c), therefore confirming the integrity of the secondary structure of the protein solubilized in DDM detergent. Attempts at improving the expression level and purification of PucI have not yet surpassed those that we have shown here. The purification yield of PucI may not be good enough to consider for traditional *in surfo* crystallization, but it should be sufficient for *in meso* crystallization trials.

3D model of PucI

The protein sequence alignment between PucI and Mhp1 revealed that a majority of the residues in Mhp1 known to be involved in direct binding interactions with the hydantoin substrate (Simmons *et al.*, 2014) are identically conserved in PucI (Figs 4c and S3). A structural explanation for the differences in ligand recognition between PucI and Mhp1 would, therefore, require a crystal structure of PucI with bound ligand, which we do not have, or a good homology model. While the overall sequence homology between PucI and Mhp1 is only 51.2 % (26.7 % identical, 24.5 % highly similar) and there are gaps in the alignment, the particularly high sequence conservation of the ligand binding site suggests that a homology model of PucI may be useful for explaining differences in ligand specificities between the two proteins. A reasonable homology model of PucI was obtained based on the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B; Simmons *et al.*, 2014) (Fig. S11) and achieved a root-mean-square deviation (RMSD) for Cα atoms from Mhp1 of 0.263 Å over 456 residues. The PucI model was validated by Ramachandran plot analysis, which showed only three residues in the outlier region, and 92.2 and 7.2 % of residues in the favoured and allowed regions, respectively. The limits of the 12 transmembrane helices in this model match those in the diagram of putative membrane topology for PucI based on the crystal structure of Mhp1 (Fig. 4c). Using this model we can hypothesize about the origins of differences in substrate selectivity and cation dependence between PucI and Mhp1 as follows. Mhp1 is not able to transport allantoin (Simmons *et al.*, 2014), but nevertheless in PucI there are two conserved tryptophan residues, 119 and 240, critical for ligand binding in Mhp1, the former of which could overlap pi orbitals with the aromatic hydantoin ring of allantoin (Fig. 7a, b; Weyand *et al.*, 2008; Simmons *et al.*, 2014). In addition, Asn329 of PucI can form hydrogen bonds with the hydantoin ring of allantoin (Fig. 7a, b), just as Asn318 does in Mhp1 (Simmons *et al.*, 2014). However, replacement of Gly219 in Mhp1 by the Ile239 in PucI does cause some restriction on the substrate binding space (Fig. 7a, b), so it is interesting that benzylhydantoin, the substrate for Mhp1, still binds to PucI, albeit with lower affinity than allantoin. Perhaps this is accounted for by the additional replacement of Gln42 in Mhp1 by Asn43 in PucI, which would relieve the restriction on total space imposed by Ile239 (Fig. 7a, b). The ureido substituent on the hydantoin ring of allantoin can be accommodated in the model of PucI by hydrogen bonds to Asn43 and Gln123. This is very different from the accommodation in Mhp1 of the hydrophobic and aromatic substituent in a hydrophobic cavity (Weyand *et al.*, 2008; Simmons *et al.*, 2014). In a

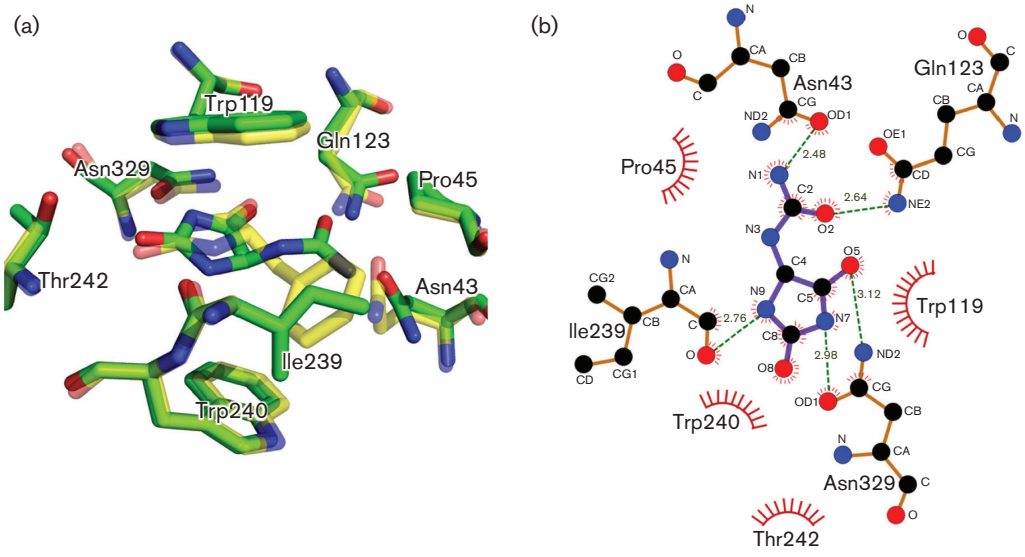

Fig. 7. Ligand binding site in a homology model of PucI. (a) Homology model of PucI (green carbon atoms) docked with (*S*)-allantoin (PDBeChem 3AL) superimposed with the crystal structure of the Mhp1-benzylhydantoin complex (PDB 4D1B; Simmons *et al.*, 2014; yellow carbon atoms) highlighting key differences in the ligand binding site of PucI. (b) PucI protein interactions with allantoin represented by LigPlot+ (http://www.ebi.ac.uk/thornton-srv/software/LigPlus/; Laskowski & Swindells, 2011). Residues are labelled accordingly to the PucI protein.

similar comparison of the residues composing the sodium ion binding site in Mhp1, Ala38 becomes Gly39, Ile41 becomes His42, Ala309 becomes Ala320, Ser312 becomes Ser323, and Thr313 becomes Val324. The loss of the threonine and gain of a histidine in PucI might, intriguingly, account for a change in cation specificity from Na^+ to H^+.

Despite the possible structural explanations for differences in substrate specificity between PucI and Mhp1 described above, other studies have demonstrated that substrate specificities of NCS-1 family proteins cannot be predicted by simple amino acid sequence comparisons, by phylogenetic analyses or from comparisons of amino acid residues in the major substrate binding site. Such comparisons performed with the Fur-type and Fcy-type fungal NCS-1 transporters showed how identical or highly similar residues can provide different substrate specificities, which is mainly due to convergent evolution within the major substrate binding site (Hamari *et al.*, 2009; Krypotou *et al.* 2015). This is upheld by the extremely high sequence homology shared between the plant proteins ZmNCS1 and SvNCS1, which have different specificity and affinity profiles for both transport and competitive binding. Indeed, none of the residue differences between ZmNCS1 and SvNCS1 corresponds to any of the functionally important residues identified by conservation or mutagenesis (Rapp *et al.*, 2016). These observations are also maintained for comparisons of putative major substrate binding site residues in PucI with those in other bacterial, fungal and plant NCS-1 proteins (Table 1). For example, residues in PucI are very similar to those in the fungal transporters FurD (uracil/uric acid), FurE (promiscuous), Dal4 (allantoin) and Fui1 (uridine). This comparison also shows that the positions corresponding to Trp119 (helix III) and Asn328 (helix VIII) in PucI are identically conserved in almost all of the NCS-1 proteins. Furthermore, the helix VI positions corresponding to Ile239 and Trp240 in PucI show more variability and these are critical for defining substrate specificity, as confirmed by experimental and modelling studies with the fungal NCS-1 homologues (Krypotou *et al.*, 2015).

The helix with the highest shared conservation in terms of identical residues between PucI, Mhp1 and all the presumed bacterial allantoin transporters is helix X (Fig. 4b, c). In Mhp1, helix X (residues 358–382) acts as an extracellular 'thin' gate that folds over the substrate as it binds to close the outward-facing cavity, which partially seals it from the exterior (Shimamura *et al.*, 2010), and this key event in the transport mechanism may underlie the evolutionary stability of this sequence. Also, mutational analyses of residues in helix X suggest that it makes an important contribution in defining the substrate specificities of NCS-1 proteins. The Mhp1 mutation L363A changes the compound naphthylmethylhydantoin from a competitive inhibitor to a substrate (Simmons *et al.*, 2014) and mutations in the FurD residues 386–389 result in changed substrate specificity (Krypotou *et al.*, 2015). For example, the FurD residue Met389 corresponds to Leu363 in Mhp1, and the mutation M389A enables FurD to transport allantoin, hypoxanthine and 8-azaguanine in addition to its natural substrates (uracil/uric acid). This position in the helix X gate corresponds with Leu377 in PucI, and is conserved as a leucine or methionine residue in all of the bacterial, fungal and plant NCS-1 proteins (Fig. S12). In addition to the major substrate binding site, residues in the putative outward-facing helix X gate therefore appear to be important for defining substrate specificity in NCS-1 family proteins. Mutational analysis in these regions of PucI may be useful for a further exploration of structure–substrate specificity and structure–transport activity relationships.

CONCLUSIONS

The putative 12-helix allantoin transport protein, PucI, from *B. subtilis* showed distinct evolutionary relationships with other putative bacterial allantoin transporters and with members of the NCS-1 family of secondary active transport proteins exemplified by the sodium-coupled hydantoin transport protein, Mhp1, from *M. liquefaciens*. Measurements of sequence homology and a phylogenetic analysis revealed wider relationships of PucI with NCS-1 transporters in bacteria, fungi (Fur-type and Fcy-type) and plants. Indeed, PucI has a similar value of overall sequence homology with the plant transporters AtNCS1 and CrNCS1 to that which it has with Mhp1. The PucI protein, cloned with a C-terminal hexahistidine tag and expressed in *E. coli*, demonstrated a high affinity for ^{14}C-allantoin uptake into energized whole cells, which was not dependent on sodium. PucI is, therefore, likely to be a proton-coupled symporter. These direct measurements of allantoin transport are the first to our knowledge using radiolabelled substrate for any putative allantoin transporter. Based on competitive uptake experiments, PucI showed highest recognition for binding of allantoin, and also good recognition for binding of hydantoin, L-5-benzylhydantoin and 5-hydroxyhydantoin. The second of these may have applications in biotechnology (Suzuki & Henderson, 2006). Recognition of some purine and pyrimidine nucleobases and nucleosides was evident to a lesser extent. Despite the differences in ligand recognition between PucI and Mhp1, a majority of the Mhp1 residues involved in direct binding interactions with the hydantoin substrate are identically conserved in PucI, and the sequence of PucI threaded comfortably into that of Mhp1, enabling construction of a hypothetical model with a distinct location for the substrate binding site. A matching of the majority of residues in the major putative substrate binding site of PucI with those in other NCS-1 proteins that have different substrate specificities is consistent with a theory of convergent evolution in this region. Based on comparisons with other NCS-1 transporters, it appears that the helix VI residues Ile239 and Trp240, and residues in a putative helix X outward-facing gate, are critical for defining substrate specificity in PucI and in the homologous proteins. A mutational analysis in these regions of PucI may be useful for further exploration

Table 1. Homology of substrate binding site residues between PucI and NCS-1 family transporters

Residues in the putative major substrate (allantoin) binding site in PucI from a 3D homology model based on the crystal structure of Mhp1 (Fig. 7) are compared with the corresponding residues from sequence alignments with characterized NCS-1 family proteins (Figs S3, S4 and S5). The NCS-1 proteins are: Mhp1 from *M. liquefaciens* (D6R8X8), CodB from *E. coli* (P0AA82), FurA from *Asp. nidulans* (Q5BFM0), FurD from *Asp. nidulans* (A6N844), FurE from *Asp. nidulans* (Q5ATG4), Fur4 from *Sac. cerevisiae* (P05316), Dal4 from *Sac. cerevisiae* (Q04895), Fui1 from *Sac. cerevisiae* (P38196), FcyB from *Asp. nidulans* (C8V329), Fcy2 from *Sac. cerevisiae* (P17064), Thi7 from *Sac. cerevisiae* (Q05998), Tpn1 from *Sac. cerevisiae* (P53099), Nrt1 from *Sac. cerevisiae* (Q08485), AtNCS1 (PLUTO) from *Ara. thaliana* (Q9LZD0), CrNCS1 from *C. reinhardtii* (A8J166), ZmNCS1 from *Z. mays* (B4FJ20), SvNCS1 from *Set. viridis* (V9SBV7). Residues are marked to highlight those that are identical (underlined) or highly similar (italic) to the corresponding residue in PucI. Also given are the experimentally characterized or putative substrate(s) for each protein and the transmembrane helix in which the corresponding residue is located based on the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B; Simmons *et al.*, 2014). x, No residue aligned; *, at the preceding residue position.

	Protein	Substrate(s)	Substrate binding site homology with PucI							
	PucI	Allantoin	*Asn43*	*Pro45*	*Trp119*	*Gln123*	*Ile239*	*Trp240*	*Thr242*	*Asn329*
Bacterial	Mhp1	5-Arylhydantoins	*Gln*	Ala	Trp	Gln	Gly	Trp	Ala	Asn
	CodB	Cytosine	x	Ser	Trp	Gly	Ser	*Phe*	*Ser*	Asn
Fungal										
Fur-type	FurA	Allantoin	Asn	Gly	x	x	Asn	*Phe*	Ala	Asn
	FurD	Uracil/uric acid	Asn	Ala	Trp	Gln	Asn	*Tyr*	Thr	Asn
	FurE	Uracil/uric acid/allantoin	Ser	Ser	Trp	Gln	Pro	Lys	Thr	Asn
	Fur4	Uracil	Asn	Asn	Trp	Gln	Asn	*Phe*	Thr	Asn
	Dal4	Allantoin	Asn	Asn	Trp	Gln	Asn	*Phe*	Ala	Asn
	Fui1	Uridine	Asn	Asn	Trp	Leu	Asn	*Phe*	Thr	Asn
Fcy-type	FcyB	Purines/cytosine	Val	Ser	Trp	*Asn*	Phe	Ala	Gly	Asn*
	Fcy2	Purines/cytosine	Val	Ala	Trp	*Asn*	Phe	Ala	Gly	Asn*
	Thi7	Thiamine	Ser	Gly	Asn	*Asn*	Ser	Val	Pro	Asn*
	Tpn1	Pyridoxine	Gly	Ser	Trp	*Asn*	Ile	Thr	Thr	Asn*
	Nrt1	Nicotinamide riboside	Thr	Gly	Asn	*Asn*	Ser	Ile	Pro	Asn*
Plant	AtNCS1	Adenine/guanine/uracil	Gly	Pro	Trp	Glu	Phe	Trp	Thr	Asn
	CrNCS1	Adenine/guanine/uracil/allantoin	Ser	Thr	Trp	Gln	Tyr	Trp	Thr	Asn
	ZmNCS1	Adenine/guanine/cytosine	Gly	Pro	Trp	Glu	Phe	Trp	Thr	Asn
	SvNCS1	Adenine/guanine/hypoxanthine/ cytosine/allantoin	Gly	Pro	Trp	Glu	Phe	Trp	Thr	Asn
Transmembrane helix in Mhp1			***I***		***III***			***VI***		***VIII***

of structure–substrate specificity and structure–transport activity relationships. While the tryptophan residues key for ligand binding in Mhp1, Trp117 and Trp220, were conserved in the residues predicted to be close to allantoin in PucI, there was not a sufficient change in fluorescence when PucI was exposed to either allantoin or L-5-benzylhydantoin to be useful for substrate binding assays. Nevertheless, a crystal structure of PucI with bound ligand is desirable to understand the structure–activity relationship of a class of proteins important in bacterial nitrogen metabolism, and so the PucI(His_6) protein was successfully solubilized in DDM detergent and purified in submilligram quantities, retaining its predicted high content of α-helix, which is suitable for future studies of its structure–activity relationship.

ACKNOWLEDGEMENTS

This work was supported by the EU EDICT consortium (contract 201924) and the BBSRC through grants numbers BB/C51725X/1 and BB/G020043/1. P. M. thanks the BBSRC, the EU E-MeP consortium and the University of Leeds for a PhD studentship, and P. J. F. H. acknowledges personal support from the Leverhulme Trust. The authors thank Przemyslaw Nogly for discussions on homology modelling and ligand docking.

REFERENCES

Adelman, J. L., Dale, A. L., Zwier, M. C., Bhatt, D., Chong, L. T., Zuckerman, D. M. & Grabe, M. (2011). Simulations of the alternating access mechanism of the sodium symporter Mhp1. *Biophys J* **101**, 2399–2407.

Arora, R. & Papaioannou, V. E. (2012). The murine allantois: a model system for the study of blood vessel formation. *Blood* **120**, 2562–2572.

Beier, L., Nygaard, P., Jarmer, H. & Saxild, H. H. (2002). Transcription analysis of the *Bacillus subtilis* PucR regulon and identification of a *cis*-acting sequence required for PucR-regulated expression of genes involved in purine catabolism. *J Bacteriol* **184**, 3232–3241.

Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* **37**, W465–W468.

Bettaney, K. E., Sukumar, P., Hussain, R., Siligardi, G., Henderson, P. J. & Patching, S. G. (2013). A systematic approach to the amplified expression, functional characterization and purification of inositol transporters from *Bacillus subtilis*. *Mol Membr Biol* **30**, 3–14.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M. & other authors (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252–W258.

Christiansen, L. C., Schou, S., Nygaard, P. & Saxild, H. H. (1997). Xanthine metabolism in *Bacillus subtilis*: characterization of the *xpt-pbuX* operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *J Bacteriol* **179**, 2540–2550.

Cruz-Ramos, H., Glaser, P., Wray, L. V., Jr. & Fisher, S. H. (1997). The *Bacillus subtilis ureABC* operon. *J Bacteriol* **179**, 3371–3373.

Danielsen, S., Boyd, D. & Neuhard, J. (1995). Membrane topology analysis of the *Escherichia coli* cytosine permease. *Microbiology* **141**, 2905–2913.

de Koning, H. & Diallinas, G. (2000). Nucleobase transporters. *Mol Membr Biol* **17**, 75–94.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.

Findlay, H. E., Rutherford, N. G., Henderson, P. J. & Booth, P. J. (2010). Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proc Natl Acad Sci U S A* **107**, 18451–18456.

Fisher, S. H. (1993). Utilization of amino acids and other nitrogen-containing compounds. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 221–228. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Fisher, S. H. (1999). Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la différence! *Mol Microbiol* **32**, 223–232.

Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessières, P., Aymerich, S. & Fromion, V. (2008). Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol* **2**, 20.

Goudela, S., Karatza, P., Koukaki, M., Frillingos, S. & Diallinas, G. (2005). Comparative substrate recognition by bacterial and fungal purine transporters of the NAT/NCS2 family. *Mol Membr Biol* **22**, 263–275.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.

Hamari, Z., Amillis, S., Drevet, C., Apostolaki, A., Vágvölgyi, C., Diallinas, G. & Scazzocchio, C. (2009). Convergent evolution and orphan genes in the Fur4p-like family and characterization of a general nucleoside transporter in *Aspergillus nidulans*. *Mol Microbiol* **73**, 43–57.

Henderson, P. J., Giddens, R. A. & Jones-Mortimer, M. C. (1977). Transport of galactose, glucose and their molecular analogues by *Escherichia coli* K12. *Biochem J* **162**, 309–320.

Inman, K. E. & Downs, K. M. (2007). The murine allantois: emerging paradigms in development of the mammalian umbilical cord and its relation to the fetus. *Genesis* **45**, 237–258.

Jackson, S. M., Patching, S. G., Ivanova, E., Simmons, K. J., Weyand, S., Shimamura, T., Brueckner, F., Suzuki, S., Iwata, S. & other authors (2013). Mhp1, the Na^+-hydantoin membrane transport protein. In *Encyclopedia of Biophysics*, pp. 1514–1521. Edited by G. C. K. Roberts. Berlin, Heidelberg: Springer.

Johnson, R. J., Sautin, Y. Y., Oliver, W. J., Roncal, C., Mu, W., Sanchez-Lozada, L. G., Rodriguez-Iturbe, B., Nakagawa, T. & Benner, S. A. (2009). Lessons from comparative physiology: could uric acid represent a physiologic alarm signal gone awry in western society? *J Comp Physiol B* **179**, 67–76.

Kazmier, K., Sharma, S., Islam, S. M., Roux, B. & Mchaourab, H. S. (2014). Conformational cycle and ion-coupling mechanism of the Na^+/hydantoin transporter Mhp1. *Proc Natl Acad Sci U S A* **111**, 14752–14757.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580.

Krypotou, E., Evangelidis, T., Bobonis, J., Pittis, A. A., Gabaldón, T., Scazzocchio, C., Mikros, E. & Diallinas, G. (2015). Origin, diversification and substrate specificity in the family of NCS1/FUR transporters. *Mol Microbiol* **96**, 927–950.

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

Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). Structure validation by Cα geometry: φ, ψ and Cβ deviation. *Proteins* **50**, 437–450.

Ma, P., Yuille, H. M., Blessie, V., Göhring, N., Iglói, Z., Nishiguchi, K., Nakayama, J., Henderson, P. J. F. & Phillips-Jones, M. K. (2008). Expression, purification and activities of the entire family of intact membrane sensor kinases from *Enterococcus faecalis*. *Mol Membr Biol* **25**, 449–473.

Ma, P., Varela, F., Magoch, M., Silva, A. R., Rosário, A. L., Brito, J., Oliveira, T. F., Nogly, P., Pessanha, M. & other authors (2013). An efficient strategy for small-scale screening and production of archaeal membrane transport proteins in *Escherichia coli*. *PLoS One* **8**, e76913.

Mourad, G. S., Tippmann-Crosby, J., Hunt, K. A., Gicheru, Y., Bade, K., Mansfield, T. A. & Schultes, N. P. (2012). Genetic and molecular characterization reveals a unique nucleobase cation symporter 1 in *Arabidopsis*. *FEBS Lett* **586**, 1370–1378.

Nygaard, P. & Saxild, H. H. (2005). The purine efflux pump PbuE in *Bacillus subtilis* modulates expression of the PurR and G-box (XptR) regulons by adjusting the purine base pool size. *J Bacteriol* **187**, 791–794.

Nygaard, P., Duckert, P. & Saxild, H. H. (1996). Role of adenine deaminase in purine salvage and nitrogen metabolism and characterization of the *ade* gene in *Bacillus subtilis*. *J Bacteriol* **178**, 846–853.

Nygaard, P., Bested, S. M., Andersen, K. A. K. & Saxild, H. H. (2000). *Bacillus subtilis* guanine deaminase is encoded by the *yknA* gene and is induced during growth with purines as the nitrogen source. *Microbiology* **146**, 3061–3069.

Pantazopoulou, A. & Diallinas, G. (2007). Fungal nucleobase transporters. *FEMS Microbiol Rev* **31**, 657–675.

Patching, S. G. (2009). Synthesis of highly pure ^{14}C-labelled *DL*-allantoin and ^{13}C NMR analysis of labelling integrity. *J Labelled Comp Radiopharm* **52**, 401–404.

Pizzichini, M., Pandolfi, M. L., Arezzini, L., Terzuoli, L., Fe, L., Bontemps, F., Van den Berghe, G. & Marinello, E. (1996). Labelling of uric acid and allantoin in different purine organs and urine of the rat. *Life Sci* **59**, 893–899.

Rapp, M., Schein, J., Hunt, K. A., Nalam, V., Mourad, G. S. & Schultes, N. P. (2016). The solute specificity profiles of nucleobase cation symporter 1 (NCS1) from *Zea mays* and *Setaria viridis* illustrate functional flexibility. *Protoplasma* **253**, 611–623.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A* **106**, 1760–1765.

Saidijam, M. & Patching, S. G. (2015). Amino acid composition analysis of secondary transport proteins from *Escherichia coli* with relation to functional classification, ligand specificity and structure. *J Biomol Struct Dyn* **33**, 2205–2220.

Saidijam, M., Psakis, G., Clough, J. L., Meuller, J., Suzuki, S., Hoyle, C. J., Palmer, S. L., Morrison, S. M., Pos, M. K. & other authors (2003). Collection and characterisation of bacterial membrane proteins. *FEBS Lett* **555**, 170–175.

Saxild, H. H., Brunstedt, K., Nielsen, K. I., Jarmer, H. & Nygaard, P. (2001). Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with *glyA*, *guaC*, *pbuG*, *xpt-pbuX*, *yqhZ-folD*, and *pbuO*. *J Bacteriol* **183**, 6175–6183.

Schein, J. R., Hunt, K. A., Minton, J. A., Schultes, N. P. & Mourad, G. S. (2013). The nucleobase cation symporter 1 of *Chlamydomonas reinhardtii* and that of the evolutionarily distant *Arabidopsis thaliana* display parallel function and establish a plant-specific solute transport profile. *Plant Physiol Biochem* **70**, 52–60.

Schultz, A. C., Nygaard, P. & Saxild, H. H. (2001). Functional analysis of 14 genes that constitute the purine catabolic pathway in *Bacillus subtilis* and evidence for a novel regulon controlled by the PucR transcription activator. *J Bacteriol* **183**, 3293–3302.

Shi, Y. (2013). Common folds and transport mechanisms of secondary active transporters. *Annu Rev Biophys* **42**, 51–72.

Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. (2010). Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. A., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Simmons, K. J., Jackson, S. M., Brueckner, F., Patching, S. G., Beckstein, O., Ivanova, E., Geng, T., Weyand, S., Drew, D. & other authors (2014). Molecular mechanism of ligand recognition by membrane transport protein, Mhp1. *EMBO J* **33**, 1831–1844.

Stark, M. J. R. (1987). Multicopy expression vectors carrying the *lac* repressor gene for regulated high-level expression of genes in *Escherichia coli*. *Gene* **51**, 255–267.

Suzuki, S. & Henderson, P. J. (2006). The hydantoin transport protein from *Microbacterium liquefaciens*. *J Bacteriol* **188**, 3329–3336.

Szakonyi, G., Leng, D., Ma, P., Bettaney, K. E., Saidijam, M., Ward, A., Zibaei, S., Gardiner, A. T., Cogdell, R. J. & other authors (2007). A genomic strategy for cloning, expressing and purifying efflux proteins of the major facilitator superfamily. *J Antimicrob Chemother* **59**, 1265–1270.

von Heijne, G. (1992). Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**, 487–494.

Ward, A., O'Reilly, J., Rutherford, N. G., Ferguson, S. M., Hoyle, C. K., Palmer, S. L., Clough, J. L., Venter, H., Xie, H. & other authors (1999). Expression of prokaryotic membrane transport proteins in *Escherichia coli*. *Biochem Soc Trans* **27**, 893–899.

Ward, A., Sanderson, N. M., O'Reilly, J., Rutherford, N. G., Poolman, B. & Henderson, P. J. F. (2000). The amplified expression, identification, purification, assay and properties of hexahistidine-tagged bacterial membrane transport proteins. In *Membrane transport – a Practical Approach*, pp. 141–166. Edited by S. A. Baldwin. Oxford: Blackwell.

Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M. & other authors (2008). Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709–713.

Weyand, S., Ma, P., Saidijam, M., Baldwin, J., Beckstein, O., Jackson, S., Suzuki, S., Patching, S. G., Shimamura, T. & other authors (2010). The nucleobase-cation-symport-1 family of membrane transport proteins. In *Handbook of Metalloproteins*, 11. Edited by A. Messerschmidt. Chichester: Wiley.

Weyand, S., Shimamura, T., Beckstein, O., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. (2011). The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *J Synchrotron Radiat* **18**, 20–23.

Witz, S., Panwar, P., Schober, M., Deppe, J., Pasha, F. A., Lemieux, M. J. & Möhlmann, T. (2014). Structure-function relationship of a plant NCS1 member – homology modeling and mutagenesis identified residues critical for substrate specificity of PLUTO, a nucleobase transporter from *Arabidopsis*. *PLoS One* **9**, e91343.

Wray, L. V., Jr, Ferson, A. E., Rohrer, K. & Fisher, S. H. (1996). TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **93**, 8841–8845.

Xi, H., Schneider, B. L. & Reitzer, L. (2000). Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J Bacteriol* **182**, 5332–5341.

Edited by: G. Unden