Supplementary Materials, Methods, Tables and Figures

The allantoin transport protein, Pucl, from *Bacillus subtilis*: evolutionary relationships, amplified expression, activity and specificity

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

General

Chemicals, reagents and media of the highest available quality were obtained from Sigma-Aldrich Co., Fisher Scientific UK Ltd, Melford Laboratories Ltd, BDH Chemical Supplies or Difco Laboratories, unless stated otherwise. All media, buffers and other solutions were prepared using either deionised water or MilliQ™ water. All media were sterilised by autoclaving or for thermally-sensitive solutions by passage through 0.2 µM Minisart® high-flow sterile syringe-driven filters (Sartorius) or using vacuum-driven 0.2 µM filters (Stericup®) from Millipore. Cellulose nitrate 25 mm ø filters (0.45 µM pore size) for radiolabelled substrate assays and cellulose ester GSTF 25 mm ø filters (0.22 µm pore size) (Whatman®) for protein determinations were from Millipore (UK) Ltd. DNA purification kits were from QIAGEN Ltd. Restriction endonucleases and T4 DNA ligase were from New England Biolabs, Pfu Turbo™ DNA polymerase was from Agilent Technologies UK, and 1 kb DNA ladder and SYBR Safe™ DNA gel stain was from Invitrogen. PCR amplification of DNA was performed using a Peltier Thermal cycler from MJ Research. Cell disruption was performed using a Constant Systems disruptor. Protein determinations used the method of Schaffner and Weissmann (1973) or a BCA assay using Pierce® BCA protein assay reagent A from Thermo Scientific. SDS-PAGE was performed by the method of Laemmli UK (1970), refined for membrane proteins as described by Henderson and Macpherson (1986) using 4% stacking gels and 15% resolving gels in a BioRad Mini PROTEAN 3 apparatus. Acrylamide (40%) and bisacrylamide (2%) solutions were from BioRad Laboratories and SDS-7 protein molecular weight markers were from Sigma-Aldrich Co. Western blotting was performed by semi-dry transfer using a BioRad TRANS-BLOT® SD apparatus; RGS-His antibody was from QIAGEN Ltd, SuperSignal®; West Pico luminal enhancer solution and stable peroxide solution were from Perbio Science UK; and Fluorotrans™ membrane was from Pall BioSupport, UK. High-range Rainbow molecular weight markers were from Amersham Biosciences UK Ltd.

Gene cloning and transformation of *E. coli*

Cloning was performed using the plasmid pTTQ18 (Stark, 1987), which is based on the pUC high expression series of plasmids with a polylinker/lacZα region flanked by the strong hybrid trp-lac
(tac) promoter, which was later modified to introduce an RGS(His$_6$) tag at the C-terminal end of the protein (Ward et al., 1999; Ward et al., 2000). The strategy is outlined below. PCR primers (forward: 5'-CCGGAATTCCGATATGAAATTAAAGAGAGTCAGCAGCAATCCA-3' and reverse: 5'-AAAACCTGCAGCTTCAGCCTGGCGGACCTGCGCATGGT-3') were designed to extract and amplify the pucl gene from B. subtilis 168 genomic DNA with introduction of EcoR1 and Pst1 restriction sites at the 5' and 3' ends, respectively, followed by digestion of the PCR product with these enzymes. The gene digests were ligated into the multi-cloning site of EcoR1/Pst1-digested plasmid pTTQ18 downstream from the IPTG-inducible tac promoter and immediately upstream from a RGS(His$_6$)-coding sequence that we had already engineered into the plasmid (Liang, 1994, unpublished). The ligation product was transformed into E. coli XL-1-Blue cells (Stratagene™) in the presence of carbenicillin (100 µg/ml) followed by PCR screening of colonies, extraction of plasmid DNA from positive clones and restriction digestion analysis using EcoR1 and Pst1 enzymes. Plasmid DNA from successful ligations was transformed into E. coli BL21(DE3) cells (Novagen™) followed by a test for inducible expression of the His-tagged protein by SDS-PAGE and western blot analysis of membranes prepared by the water lysis method (Witholt et al., 1976; Ward et al., 2000) from small-scale (50 ml) cell cultures that were uninduced or induced with IPTG. Clones of cells that showed successful amplified expression of the proteins were transferred into a freezing mixture (12.6 g/L K$_2$HPO$_4$, 0.9 g/L sodium citrate, 0.18 g/L MgSO$_4$, 1.8 g/L (NH$_4$)$_2$SO$_4$, 3.6 g/L KH$_2$PO$_4$, 96 g/L glycerol), frozen in liquid nitrogen and stored at -80 °C. Competent cells were prepared by the methods described by Inoue et al. (1990) or Chung et al. (1989) and transformations were performed based on the method described by Inoue et al. (1990). The optimum concentration of IPTG and length of time for induction were determined.

**Cell growth and membrane preparation**

Cells were grown in LB or 2TY liquid medium supplemented with glycerol (20 mM) and carbenicillin (100 µg/ml) in Falcon tubes (10 ml in 50 ml tubes) for starter cultures and in LB, 2TY or minimal medium in baffled flasks (50 ml in 250 ml flasks or 500 ml in 2 litre flasks for small-scale and large-scale cultures, respectively) at a temperature of 37 °C with shaking at 200 rpm. Cells were recovered from deep frozen stocks by streaking onto LB-agar plates with 100 µg/ml carbenicillin, using a single colony to inoculate LB medium in Falcon tubes, and then using a 2% (v/v) inoculum when transferring from one liquid culture to another. For expression tests and optimisation of induction conditions, small-scale cultures were grown to an A$_{680}$ of 0.4-0.6, then left uninduced or induced with the relevant concentration of IPTG and grown for the given further length of time before harvesting by centrifugation (3000 x g, 10 min, in Falcon tubes using a bench-top instrument), followed by preparation of membranes by the water lysis method (Witholt et al.,
1976; Ward et al., 2000). For large-scale membrane preparation, typically a total of 10 litres of cells were grown to an \(A_{680}\) of 0.4-0.6, then induced with IPTG (0.5 mM) and grown for a further 3 hours before harvesting by centrifugation (6000 x g, 15 min, 4 °C) and storage at -80 °C. At a later time the cells were thawed, suspended in Tris-EDTA buffer (20 mM Tris, pH 7.5 with 0.5 mM EDTA) and inner/outer membranes were separated by sucrose gradient ultracentrifugation and prepared as described in Ward et al. (2000), followed by washing and resuspension in Tris buffer (20 mM, pH 7.5), dispensing into aliquots, rapid freezing in liquid nitrogen and storage at -80 °C.

**Protein purification**

Inner membrane preparations were solubilised for up to 4 hours at 4 °C in a buffer containing 20 mM Tris (pH 8.0), 1% \(n\)-dodecyl-\(\beta\)-\(D\)-maltoside (DDM), 20% glycerol and 300 mM sodium chloride (Supplementary Table S1) at a protein concentration of 3 mg/ml followed by removal of insoluble material by ultracentrifugation (100,000 x g, 1 hour, 4 °C). Immobilised-metal affinity chromatography (IMAC) was performed by mixing the supernatant obtained above with Ni-NTA resin (QIAGEN) (1 ml per 30 mg of total protein) overnight at 4 °C, which was then packed into a column. Unbound material was collected followed by washing of the column with at least 40x column volumes of a buffer that contained imidazole at a concentration of 20 mM or 40 mM (Supplementary Table S1). The His-tagged protein was eluted from the column using ~ 7 ml (for a 1 ml column) of a buffer that contained 200 mM imidazole (Supplementary Table S1), which was then concentrated to a volume of ~ 300 µl by centrifugation using a concentrator with a MW cut off of 100 kDa (Vivaspin 20, Sartorius). Using the same column, the protein was washed a minimum of five times with at least 5 ml of a buffer containing 20 mM Tris (pH 8.0) or 10 mM KH\(_2\)PO\(_4\), (pH 7.6) and 0.05% DDM, before concentrating to a volume of 200-500 µl, dispensing into aliquots, rapid freezing in liquid nitrogen and storage at -80 °C.

**Circular dichroism spectroscopy**

Far-UV circular dichroism spectroscopy analysis of purified protein (0.05 mg/ml) 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. The sample was introduced in a Hellma quartz-glass cell of 1 mm path length and spectra were recorded over a wavelength range of 260-190 nm in steps of 1 nm at a scan rate of 10 nm/min. The response time was set at 1 second with a sensitivity of 20 mdeg.
Supplementary Table S1. Composition of buffers used for protein purification.

<table>
<thead>
<tr>
<th></th>
<th>Solubilisation buffer</th>
<th>Wash buffer</th>
<th>Elution buffer</th>
<th>Storage buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl (pH 7.5)</td>
<td>20 mM</td>
<td>20 mM</td>
<td>20 mM</td>
<td>--</td>
</tr>
<tr>
<td>imidazole</td>
<td>20 mM</td>
<td>20 or 40 mM</td>
<td>200 mM</td>
<td>--</td>
</tr>
<tr>
<td>Glycerol</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>NaCl</td>
<td>300 mM</td>
<td>150 mM</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>DDM</td>
<td>1%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
</tr>
<tr>
<td>KH₂PO₄ (pH 7.5)*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>10 mM</td>
</tr>
</tbody>
</table>

* when used instead of Tris in the storage buffer

Supplementary Table S2. Sequence homology between PucI and NCS-1 family transporters.

This table gives values of sequence homology for PucI from *B. subtilis* (P94575) with characterised 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 *A. nidulans* (Q5BFM0), FurD from *A. nidulans* (A6N844), FurE from *A. nidulans* (Q5ATG4), Fur4 from *S. cerevisiae* (P05316), Dal4 from *S. cerevisiae* (Q04895), Fui1 from *S. cerevisiae* (P38196), FcyB from *A. nidulans* (C8V329), Fcy2 from *S. cerevisiae* (P17064), Thi7 from *S. cerevisiae* (Q05998), Tpn1 from *S. cerevisiae* (P53099), Nrt1 from *S. cerevisiae* (Q08485), AtNCS1 (PLUTO) from *A. thaliana* (Q9LZD0), CrNCS1 from *C. reinhardtii* (A8J166), ZmNCS1 from *Zea mays* (B4FJ20), SvNCS1 from *Setaria viridis* (V9SBV7). Values are given for the number of residues (left) and the percentage of residues (right) in PucI that are identical, highly similar and a combined total of these from separate sequence alignments with Mhp1 or the given groups of proteins (Supplementary Figures S4, S5, S6, S7 and S8).
Supplementary Figure S1. Inner membrane preparation with amplified expression of the PucI(His$_6$) protein. SDS-PAGE analysis of inner (1), mixed (2) and outer (3) membranes prepared from a large-scale minimal medium culture of BL21(DE3) cells containing the construct pTTQ18-pucI(His$_6$). M = molecular weight markers, the arrow indicates the position of the amplified PucI(His$_6$) protein.
Supplementary Figure S2. Amino acid sequence and amino acid composition of the PucI protein from Bacillus subtilis. The amino acid sequence of the PucI protein (Bsu3645, P94575, ALLP_BACSU) from Bacillus subtilis (strain 168) in FASTA format (A) taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and the percentage content of each type of amino acid residue in the protein (B) determined using the ExPASy online tool ProtParam (http://web.expasy.org/protparam/, Gasteiger et al., 2005). Coloured single amino acids correspond with those in the topology diagram of PucI in Figure 4A of the main paper.

A

>sp|P94575|ALLP_BACSU Probable allantoin permease OS=Bacillus subtilis (strain 168) GN=pucI PE=2 SV=1
MKLKESSQQSNRLSNEDLVPVLGQEKRTWKAMNFASIMGCHNIPTYATVGGLIASPQVLAILIITASLI
LFGALALNHAGTGYLPPFVIRASYIGANIPALLRAFTAIMGMDIQTGFAGSTALNNILNNWPGWGEIG
GEWNLGIHSLGGSFVFFWAIHLLVHLHGEMIKRFEVWAGPLVLYLVFFGVMVAVDIAGGLGPIYSQFGKF
HTFSETFWFFAAGVTGIIGIATLILNIFDFTRALAEFTQKEQIKQFQYGGLPGTFALFAPASITVTSGQVAFGE
PiWVVDILARFDNYVIVLSVITLCLATISVNVAANIVSPAYDIANLPPKYNFKRGSFITALALLFTVPWK
LMESATSVFAGLGSGNLPGVAGVMADFYIIRKRELSDLYSETGRYVYYWKGYNYRAAATMLGALISLI
GMYFPVLKSLYDISWFVGLISFLFIVLMRVLLPASLAIETEHAQVRQAE

B

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>49</td>
<td>10.0%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>14</td>
<td>2.9%</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>16</td>
<td>3.3%</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>11</td>
<td>2.2%</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>13</td>
<td>2.7%</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>17</td>
<td>3.5%</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>46</td>
<td>9.4%</td>
</tr>
<tr>
<td>His (H)</td>
<td>9</td>
<td>1.8%</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>49</td>
<td>10.0%</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>55</td>
<td>11.2%</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>15</td>
<td>3.1%</td>
</tr>
<tr>
<td>Met (M)</td>
<td>14</td>
<td>2.9%</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>31</td>
<td>6.3%</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>22</td>
<td>4.5%</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>29</td>
<td>5.9%</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>25</td>
<td>5.1%</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>17</td>
<td>3.5%</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>20</td>
<td>4.1%</td>
</tr>
<tr>
<td>Val (V)</td>
<td>36</td>
<td>7.3%</td>
</tr>
</tbody>
</table>

Supplementary Figure S3. Protein sequence alignment between putative allantoin permeases from 24 different species of bacteria. Amino acid 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). PucI from Bacillus subtilis is shown at the top. Residues coloured red are identical and those coloured blue are highly similar. Details about the proteins are listed at the end of the alignment.
PucI

AllPBcereus  MKLKESQQQSNRLSDLLPVQLGQKRRTWKAMNFASIWM
AllPEfaeca  MKLKESQQQSNRLSDLLPVQLGQKRRTWKAINFASIWM
AllPLBact   MDNALEKYSRGYSDDLLPPTENKRTWTGFNYTLWM
AllPRaqu    MNESECTQRYRERGYSNDLLPKLKEKRNKFNYTLWM
AllPBagres  MERQEQQQRRYLARGYSDDLPKEEKQTRKAPNYTFLWM
AllPEColi    MEHRKLFQGQRYSEDLLPKTQOSRTKWTGFNYTLWM
AllPKibdel  MEHRELYGQRYSDLLPKTEAQRNKFNYTFLWM
AllPucI     MDHAESGMADGGFAGDGLFNA
AllPSarat   MISSQREKQYQRYHEDLLPKETDKRKTAINFYTLWM
AllPYinter  MDIENKREVRSRGYSDLPPKTSDKRNRFAPNYTFLWM
AllPArubri  MNSLENEDALPVPAGDWSWVNNSTVWM
AllPAsinuis MTKSLSNEVMDKTAKHNYTAFLWM

PucI

AllPBcereus  GC1HNIPTYATVGLIAIGLSWPQVTLAIIITAASLIVFAGALALNHGATGKYGFLPFVIIRA
AllPEfaeca  GCIHNIPTYATVGLIAIGLSWPQVTLAIIITAASLIVFAGALALNHGATGKYGFLPFVIIRA
AllPLBact    GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPRaqu     GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPBagres   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPEColi    GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPSafgha   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPKibdel   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPucI      MDSLENDALPVPAGDWSWVNNSTVWM
AllPSarat   MISSQREKQYQRYHEDLLPKETDKRKTAINFYTLWM

PucI

AllPBcereus  GCIHNIPTYATVGLIAIGLSWPQVTLAIIITAASLIVFAGALALNHGATGKYGFLPFVIIRA
AllPEfaeca  GCIHNIPTYATVGLIAIGLSWPQVTLAIIITAASLIVFAGALALNHGATGKYGFLPFVIIRA
AllPLBact    GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPRaqu     GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPBagres   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPEColi    GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPSafgha   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPKibdel   GSCHVNPYVAVGGFLIGLSPLQVLMLAVVLSSTFTAVAMNLNVGSKGYVPFAMILQS
AllPucI      MDSLENDALPVPAGDWSWVNNSTVWM
AllPSarat   MISSQREKQYQRYHEDLLPKETDKRKTAINFYTLWM
Current details about the bacterial putative allantoin permeases from the UniProt KnowledgeBase are listed below in alphabetical order of the bacterial species. Pucl from *Bacillus subtilis* is highlighted in blue.

AllPArubri

>tr|A0A0D6GP5P0|A0A0D6GP5P0_9PROT  
Cytosine/purines/uracil/thiamine/allantoin permease OS=Acidisphaera rubrifaciens  
HS-AP3 GN=Asru_0108_06 PE=4 SV=1

AllPAcido

>tr|T0BTV5|T0BTV5_9BACL  
Uncharacterized protein OS=Alicyclobacillus acidoterrestris ATCC 49025  
GN=N007_08025 PE=4 SV=1

AllPAjapon

>tr|A0A075UWS0|A0A075UWS0_9PSEU  
Cytosine/purines/uracil/thiamine/allantoin permease family protein OS=Amycolatopsis japonica GN=AJAP_29195 PE=4 SV=1

AllPBcereus

>tr|A0A0K6K4C4|A0A0K6K4C4_BACCE  
Putative allantoin permease OS=Bacillus cereus  
GN=pucI_2 PE=4 SV=1

PucI

>sp|P94575|ALLP_BACSU  
Probable allantoin permease OS=Bacillus subtilis (strain 168) GN=pucI PE=2 SV=1

AllPBagres

>tr|A0A085GIB7|A0A085GIB7_9ENTR  
Allantoin permease OS=Buttiauxella agrestis ATCC 33320 GN=ybbW PE=4 SV=1

AllPCapicu

>tr|A0A017SX37|A0A017SX37_9DELT  
Cytosine/purine/uracil/thiamine/allantoin permease family protein OS=Chondromyces apiculatus DSM 436 GN=CAP_8588 PE=4 SV=1

AllPCfreun

>tr|A0A064EDD5|A0A064EDD5_CITFR  
Uncharacterized protein OS=Citrobacter freundii MGH 56 GN=AF42_00326 PE=4 SV=1

AllPKkluyv

>tr|B9E3U4|B9E3U4_CLOK1  
Uncharacterized protein OS=Clostridium kluveri (strain NBRC 12016) GN=CKR_2118 PE=4 SV=1

AllPEfaeca

>tr|A0A0E1RIK8|A0A0E1RIK8_ENTFL  
Allantoin permease OS=Enterococcus faecalis str. Symbioflor 1 GN=allP PE=4 SV=1

AllPEcoli

>sp|P75712|ALLP_ECOLI  
Putative allantoin permease OS=Escherichia coli (strain K12) GN=ybbW PE=1 SV=2

AllPKibdel

>tr|A0A0B7CDN7|A0A0B7CDN7_9PSEU  
Cytosine/purine/uracil/thiamine/allantoin permease family protein OS=Kibdelosporangium sp. MJ126-NF4 PE=4 SV=1

AllPKflavi

>tr|D2PV18|D2PV18_KRIFD  
NCS1 nucleoside transporter family OS=Kribbella flavida (strain DSM 17836 / JCM 10339 / NBRC 14399) GN=Kfla_2410 PE=4 SV=1
**AllPKutzn**
>tr|W7SE62|W7SE62_9PSEU
NCS1 family nucleobase:cation symporter-1 OS=Kutzneria sp. 744 GN=KUTG_02746 PE=4 SV=1

**AllPLBact**
>tr|A0A099W9I6|A0A099W9I6_9LIST
Allantoin permease OS=Listeriaceae bacterium FSL A5-0209 GN=EP56_09325 PE=4 SV=1

**AllPPdurus**
>tr|A0A0F7F7F2|A0A0F7F7F2_PAEDU
Allantoin permease OS=Paenibacillus durus ATCC 35681 GN=VK70_04590 PE=4 SV=1

**AllPRagua**
>tr|H8NQW0|H8NQW0_RAHAQ
Allantoin permease OS=Rahnella aquatilis HX2 GN=Q7S_01470 PE=4 SV=1

**AllPPricke**
>tr|R0E5V8|R0E5V8_RALPI
NCS1 nucleoside transporter-like protein OS=Ralstonia pickettii OR214 GN=OR214_02516 PE=4 SV=1

**AllPStypfi**
>tr|A0A0F6AY07|A0A0F6AY07_SALT1
Allantoin permease OS=Salmonella typhimurium (strain 14028s / SGSC 2262) GN=allP PE=4 SV=1

**AllPSerrat**
>tr|A0A087L1Z7|A0A087L1Z7_9ENTR
Allantoin permease OS=Serratia sp. Ag1 GN=IV04_10740 PE=4 SV=1

**AllPSdysen**
>tr|F3V2W9|F3V2W9_SHIDY
NCS1 nucleoside transporter family protein OS=Shigella dysenteriae 155-74 GN=ncs1 PE=4 SV=1

**AllPSusita**
>tr|Q01P63|Q01P63_SOLUE
NCS1 nucleoside transporter family OS=Solibacter usitatus (strain Ellin6076) GN=Acid_7658 PE=4 SV=1

**AllPSafgha**
>tr|S4ME44|S4ME44_9ACTN
Putative allantoin permease OS=Streptomyces afghaniensis 772 GN=STAFG_8236 PE=4 SV=1

**AllPSacido**
>tr|G8TUQ4|G8TUQ4_SULAD
Uncharacterized protein OS=Sulfobacillus acidophilus (strain ATCC 700253 / DSM 10332 / NAL) GN=Sulac_2310 PE=4 SV=1

**AllPYinter**
>tr|C4SZI6|C4SZI6_YERIN
Allantoin permease OS=Yersinia intermedia ATCC 29909 GN=yinte0001_12410 PE=4 SV=1
Supplementary Figure S4. Protein sequence alignment between PucI from *Bacillus subtilis* and Mhp1 from *Microbacterium liquefaciens*. The amino acid sequences of the PucI protein from *Bacillus subtilis* strain 168 (Bsu3645, P94575, ALLP_BACSU) and the Mhp1 protein from *Microbacterium liquefaciens* (D6R8X8, D6R8X8_9MICO) taken from the UniProt KnowledgeBase (http://www.uniprot.org/) were aligned using the online multiple sequence alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/, Sievers et al., 2011). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting is used to show helical regions in Mhp1 based on the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B, Simmons et al., 2014) as follows: transmembrane helix (grey), break in transmembrane helix (yellow), internal helix (cyan), external helix (green). Helical regions correspond with those shown in the topology diagram of PucI in Figure 4C of the main paper.

PucI  MKLKESQQSNRLSNEDLVLPGQEKRTWKAHMFSIWMGCIHNIPTYAVTGGLIAMGLSP  60
Mhp1  MNS-TPIEARESSLPSNAPTRYARSVPGPSLAAIFAMACIGVAIFIAA-GQMTSSQV  58

PucI  WQVLAIITASLILFLGALANDHAKGYLPPVFIIIRASYGIGANIPALLRAFTAIMWL  120
Mhp1  WQVICIAAGCTAVILFQTSQIMRGINFTVAARMPFGIRGSLIPITL7ALLSLFWE  118

PucI  GIQTFAGSTALNILLNMPGWGEIGGEWNLGIHLSGLSFFWAIHLLVHLHGMESI  180
Mhp1  GFQTWLGALALDEITL-LLTGFILNP------------LVIVIFCAIQVVTFFEYIFI  164

PucI  KRFEWAGPLVYLFFGGMVWAVDI-AGGLGPIYSQPDKHIETFSPFPAAGVTCGIGI  239
Mhp1  RWMNVFAFVSLLTAMGVYMVYLMGDADVSGGEV-SMGE------NPFGSTAIIMIIVGC  219

PucI  WATLILNIDPFTRAPETQKE--------------IKQFYGLPGTALFASITVTSGSQ  287
Mhp1  WIAVVSIIHDIKECKVDPNASRFGQTKADARYATAGWLMVPASTIFGIGA--ASMLV  277

PucI  VAFGEPIWVDVDILARFDNPYTVILSVITLCAITISVNAANIVSPAYDIFANPKEYNF  347
Mhp1  VGWHLWIAITTEQCSSPMALLQFQV-FVLLATWSNPANLLSPAYTLCSTPPVRFTE  336

PucI  KGFSFITALLALFTPVWKLMMESATSVYAFGLGMMGPGVAGMMADYFIRKRELSVDD  407
Mhp1  KTGVVSAVGLDMPWFAQMG-----------LNTTLNLSASLGPAGIMISYELVRRRISLHD  393

PucI  LYESRYGVYYWKGYNRAAFATMLGALISI----------GMYVFVLKLYSWDSWFVGVI  459
Mhp1  LYTPIKGYITYRVRGWNLVALAYAVALAVSFLTPDLMFVTGLAALLHIPAMRWWAKTFP  453

PucI  SFLFIYI----VLMRVHPASLAIMTETVEAQQVRAB------------QAE---490
Mhp1  LFSAEHRENEDYRIFPGVPAPADESATANTKEQNNPAGGGRSHHHH  501

**Colour key:**
- **Red**    Identical
- **Blue**   Highly similar
- **Grey**   Transmembrane helix
- **Yellow** Break in transmembrane helix
- **Cyan**   Internal helix
- **Green**  External helix
**Supplementary Figure S5. Protein sequence alignment between PucI and bacterial NCS-1 family transporters.** The amino acid sequence of PucI from *B. subtilis* strain 168 (P94575) was aligned with those of Mhp1 from *M. liquefaciens* (D6R8X8) and CodB from *E. coli* (P0AAA82). Sequences were taken from the UniProt KnowledgeBase ([http://www.uniprot.org/](http://www.uniprot.org/)) and aligned using the multiple sequence alignment tool Clustal Omega ([http://www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/), Sievers et al., 2011). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting (cyan) is used to show residues in the putative substrate (allantoin) binding site of PucI (Figure 7).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PucI</td>
<td>MKLKEQQQSRLNEQDVPGLGKEK-TWKAAMFASIAWMGCHIHPFTYATGGLIAIGLS</td>
</tr>
<tr>
<td>Mhp1</td>
<td>MNS-TPIEEARSLLNPSNAPTYREA-SVGFPFLAAINFAMAIQVAFIAA-GCMSSFQ</td>
</tr>
<tr>
<td>CodB</td>
<td>--------MSQDNNF--------SQGFPQBSAKVLALTF--------VLMGLTF-FAJSMWVTGGLTGLF</td>
</tr>
<tr>
<td>PucI</td>
<td>PWQVLAIITASLILFGLAALNHAGTGYGLPPFVIIRASYGYGANIPALLRAFTAIMNH</td>
</tr>
<tr>
<td>Mhp1</td>
<td>WQVQVIAYAAGCTTAVILLFTQPSAARWHRGINTVARRMPFGPILSIPITLKLALSLFW</td>
</tr>
<tr>
<td>CodB</td>
<td>YHDFLAVLGNIILGDIYTSFLGYIAGKTGLTTHALLRFSGVKGWLSPLLGGTQVGW</td>
</tr>
<tr>
<td>PucI</td>
<td>LGMDTFASTLALLILLNMPWGEIGEWEHNIGHALSSLSSVFFWAIHLLVLHGIEMES</td>
</tr>
<tr>
<td>Mhp1</td>
<td>FGQVWGALALDEITRLGTFTNPQ-------------WLVIFGAIQVVTFFGIFT</td>
</tr>
<tr>
<td>CodB</td>
<td>FGQVWMAFAPVSGKAT----------GL----------DNLLLAVSGLMTVVFQGSA</td>
</tr>
<tr>
<td>PucI</td>
<td>IYAWLILNPDPDTRFAETQKEQ----------IKQFYGGLPCGAFISATVTSGS</td>
</tr>
<tr>
<td>Mhp1</td>
<td>GWAVVVSSH DATVKECKVDPNASREGQTKADARYATAQWLMPVAPFIFGIA--ASMV</td>
</tr>
<tr>
<td>CodB</td>
<td>SFISAGTTLTADFVRGNNKALAVLVA----------MVAPFGLN-SLIMIFGAAGAAPALGA</td>
</tr>
<tr>
<td>PucI</td>
<td>QVAFGEPIDVDVDSLARFNDPPYVIVLSVTLCATLSTVNAVINPAYSIDANALPKIN</td>
</tr>
<tr>
<td>Mhp1</td>
<td>LVGWEENPSIAITEEVTGAYSVTPMIALEVQVFV-VILLATWSTPAANLSPAYTLCSTPFRVFT</td>
</tr>
<tr>
<td>CodB</td>
<td>CINQGIFIDIATCDVPARI-------------IVLGQNIWATNDYALGSFAN------ITCMS</td>
</tr>
<tr>
<td>PucI</td>
<td>FTKVIGFITALLAFFTVWKLAMESATSAYQPFLGIGGMFLPGVAVGMDAYFIRKRERLSD</td>
</tr>
<tr>
<td>Mhp1</td>
<td>FKTVGIVSAYVGLMMPWFQAVG---LNTFPNLNASASLGPLAGIMSDFLYLVRRIRIISLH</td>
</tr>
<tr>
<td>CodB</td>
<td>SKTLSLVINGIIETVACNLWLYN---FVGLTFLSSAPIPFGVWIIALDIYLMRRRREYHFA</td>
</tr>
</tbody>
</table>
Supplementary Figure S6. Protein sequence alignment between PucI and fungal (Fur-type) NCS-1 family transporters. The amino acid sequence of PucI from *B. subtilis* strain 168 (P94575) was aligned with those of FurA from *A. nidulans* (Q5BFM0), FurD from *A. nidulans* (A6N844), FurE from *A. nidulans* (Q5ATG4), Fur4 from *S. cerevisiae* (P05316), Dal4 from *S. cerevisiae* (Q04895) and Fui1 from *S. cerevisiae* (P38196). 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). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting (cyan) is used to show residues in the putative substrate (allantoin) binding site of PucI (Figure 7).
Supplementary Figure S7. Protein sequence alignment between PucI and fungal (Fcy-type) NCS-1 family transporters. The amino acid sequence of PucI from *B. subtilis* strain 168 (P94575) was aligned with those of FcyB from *A. nidulans* (C8V329), Fcy2 from *S. cerevisiae* (P17064), Thi7 from *S. cerevisiae* (Q05998), Tpn1 from *S. cerevisiae* (P53099) and Nrt1 from *S. cerevisiae* (Q08485). 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). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting (cyan) is used to show residues in the putative substrate (allantoin) binding site of PucI (Figure 7).

<table>
<thead>
<tr>
<th>PucI</th>
<th>FcyB</th>
<th>Fcy2</th>
<th>Thi7</th>
<th>Tpn1</th>
<th>Nrt1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAGA</td>
<td>MLEEGN</td>
<td>FHKF</td>
<td>ALRF</td>
<td>MNGA</td>
<td>ALRF</td>
</tr>
<tr>
<td>MGCS</td>
<td>VYMI</td>
<td>MFF</td>
<td>LAMF</td>
<td>GAVS</td>
<td>MGCL</td>
</tr>
<tr>
<td>NF</td>
<td>LSFG</td>
<td>LSFG</td>
<td>IYMG</td>
<td>VGG</td>
<td>VGG</td>
</tr>
<tr>
<td>A</td>
<td>YVMI</td>
<td>LSFV</td>
<td>VUFF</td>
<td>VGG</td>
<td>VGG</td>
</tr>
<tr>
<td>G</td>
<td>MSFF</td>
<td>LSFG</td>
<td>IYMG</td>
<td>VGG</td>
<td>VGG</td>
</tr>
<tr>
<td>K</td>
<td>LSFG</td>
<td>LSFG</td>
<td>IYMG</td>
<td>VGG</td>
<td>VGG</td>
</tr>
<tr>
<td>E</td>
<td>LSFG</td>
<td>LSFG</td>
<td>IYMG</td>
<td>VGG</td>
<td>VGG</td>
</tr>
</tbody>
</table>

The amino acid sequence of PucI from *S. cerevisiae* strain 168 (P94575) and Nrt1 from *B. subtilis* type strain 168 (P17064) was aligned with those of FcyB from *A. nidulans* (C8V329), Fcy2 from *S. cerevisiae* (P17064), Thi7 from *S. cerevisiae* (Q05998), Tpn1 from *S. cerevisiae* (P53099) and Nrt1 from *S. cerevisiae* (Q08485). 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). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting (cyan) is used to show residues in the putative substrate (allantoin) binding site of PucI (Figure 7).
**Supplementary Figure S8.** Protein sequence alignment between PucI and plant NCS1 family transporters. The amino acid sequence of PucI from *B. subtilis* strain 168 (P94575) was aligned with those of AtNCS1 (PLUTO) from *A. thaliana* (Q9LZD0), CrNCS1 from *C. reinhardtii* (A8J166), ZmNCS1 from *Zea mays* (B4FJ20) and SvNCS1 from *Setaria viridis* (V9SBV7). 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). Residues are coloured to indicate those that are identical (red) and highly similar (blue). Coloured highlighting (cyan) is used to show residues in the putative substrate (allantoin) binding site of PucI (Figure 7).
Supplementary Figure S9. Membrane topology analyses of the PucI protein from *Bacillus subtilis*. The amino acid sequence of the PucI protein (Bsu3645, P94575, ALLP_BACSU) from *Bacillus subtilis* (strain 168) was analysed using the online topology prediction tools TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), which uses a hidden Markov model (Krogh et al., 2001), (A) and TOPCONS consensus prediction server (http://topcons.cbr.su.se/, Bernsel et al., 2009) (B). These predictions were in agreement of PucI having twelve putative transmembrane-spanning α-helices with both the N- and C-terminal ends of the protein at the cytoplasmic side of the membrane.
Supplementary Figure S10. Effect of induction time on PucI-mediated $^{14}$C-allantoin uptake into whole cells. Uptake of $^{14}$C-allantoin (50 µM) after 2 minutes into energised BL21(DE3) cells containing the construct pTTQ18-pucI(His$_6$) that were uninduced or induced with IPTG for a range of different lengths of time from 0.1 to 22 hours. Cells were cultured in minimal medium with 20 mM glycerol and induced at an $A_{680}$ of 0.4-0.6 with 0.5 mM IPTG for the given length of time. Uninduced cells were grown in the same way as induced cells except that no IPTG was added. Harvested cells were washed three-times with assay buffer (150 mM KCl, 5 mM MES, pH 6.6) and resuspended to an $A_{680}$ of 2.0. Cells were energised with 20 mM glycerol and bubbled air for 3 minutes followed by incubation with $^{14}$C-allantoin (50 µM) and removal of aliquots for analysis after 2 minutes. The data points represent the mean of triplicate measurements and the error bars represent the standard errors of the means.
Supplementary Figure S11. Effect of sodium ions on PuCl-mediated $^{14}$C-allantoin uptake into energised whole cells. Uptake of $^{14}$C-allantoin (50 µM) after 15 seconds and 2 minutes into energised BL21(DE3) cells containing the construct pTTQ18-pucl(His$_6$) that were induced with IPTG. Cells were cultured in minimal medium with 20 mM glycerol and induced at an $A_{680}$ of 0.4-0.6 with 0.5 mM IPTG for 1 hour. Harvested cells were washed three-times with assay buffer (150 mM KCl, 5 mM MES, pH 6.6) and resuspended to an $A_{680}$ of 2.0. Cells were energised with 20 mM glycerol, NaCl at a range of concentrations from 0-150 mM and bubbled air for 3 minutes followed by incubation with $^{14}$C-allantoin (50 µM) and removal of aliquots for analysis after 15 seconds and 2 minutes. The data points represent the average of duplicate measurements.
Supplementary Figure S12. Structures of compounds used as potential competitors of PucI-mediated \textsuperscript{14}C-allantoin uptake into whole cells. The structures 1-20 are arranged in order of decreasing competitive effect on PucI-mediated \textsuperscript{14}C allantoin uptake into whole cells as shown in Figure 6 of the main paper.

1. Allantoin

2. Hydantoin

3. L-5-Benzylhydantoin

4. 5-Hydroxyhydantoin

5. Thymine

6. Allantoic acid

7. Hypoxanthine

8. Uracil

9. Inosine

10. Uridine

11. Cytidine

12. Thymidine

13. Guanosine

14. Xanthine

15. Cytosine

16. Urea

17. Adenosine

18. Guanine

19. Uric acid

20. Adenine
Supplementary Figure S13. Detergent solubilisation and purification of the PucI(His$_6$) protein and integrity of its alpha-helical secondary structure. Protein solubilisation and purification were performed as described above in Materials and Methods, and analysed by SDS-PAGE (A) and Western blotting (B). Samples: 1. Inner membranes; 2. Insoluble fraction from solubilisation (pellet); 3. Soluble fraction from solubilisation (supernatant); 4. Unbound fraction from column; 5. Eluted proteins. $M =$ molecular weight markers. The arrows indicate the positions of the PucI(His$_6$) protein. A far-UV circular dichroism spectrum (C) of the purified PucI(His$_6$) protein (0.05 mg/ml) in potassium phosphate buffer (10 mM, pH 7.6) with 0.05% DDM was obtained as described in Materials and Methods. The spectrum represents an accumulation of ten scans from which a buffer control was subtracted. The blue line represents the voltage applied to the photomultiplier.
Supplementary Figure S14. Overlaid crystal structure of the Mhp1-benzylhydantoin complex (4DB1, red) with the predicted model of Pucl (green). See Materials and methods for derivation.
Supplementary Figure S15. Putative helix X outward-facing gate for substrate specificity of NCS-1 family transporters. Part of a complete sequence alignment between PucI and NCS-1 family transport proteins in the region of transmembrane helix X in Mhp1. The proteins are PucI from *B. subtilis* strain 168 (P94575), Mhp1 from *M. liquefaciens* (D6R8X8), CodB from *E. coli* (P0AA82), FurA from *A. nidulans* (Q5BFM0), FurD from *A. nidulans* (A6N844), FurE from *A. nidulans* (Q5ATG4), Fur4 from *S. cerevisiae* (P05316), Dal4 from *S. cerevisiae* (Q04895), Fui1 from *S. cerevisiae* (P38196), FcyB from *A. nidulans* (C8V329), Fcy2 from *S. cerevisiae* (P17064), Thi7 from *S. cerevisiae* (Q05998), Tpn1 from *S. cerevisiae* (P53099), Nrt1 from *S. cerevisiae* (Q08485), AtNCS1 (PLUTO) from *A. thaliana* (Q9LZD0), CrNCS1 from *C. reinhardtii* (A8J166), ZmNCS1 from *Zea mays* (B4FJ20) and SvNCS1 from *Setaria viridis* (V9SBV7). 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). Coloured highlighting is used to show transmembrane helix X in Mhp1 (grey) based on the crystal structure of Mhp1 with bound benzylhydantoin (PDB 4D1B, Simmons et al., 2014) and the position of a residue involved in substrate specificity (cyan). Coloured residues (red) are those that have been mutated in Mhp1 (Leu363; Simmons et al., 2014) and in FurD (Leu386, Asn387, Phe388, Met389; Krypotou et al., 2015) resulting in changed substrate specificity.

PucI  PWKLMESATS-VYAFLGLIGGMLGPVAGVMMADYFIIRKR  401
Mhp1  PWQPAGQVLNTF----LNLASALGPLAGIMISDYFLVRRR  387
CodB  LWLY-----NNF-VGWLTFLSAAIPPVGGVIIADYLMNRRR
FurA  PWNLVSDSNQF-TTYLSAYSIFLSAIAGVMICDYYVVRKG
FurD  PWKILEASANF-LNEKSAYAIIFGPIAAIMLDWEFLIKNR  413
FurE  PWYIQNSAASF-SSFLGGYSLFLGAIAGVIVVDYWVCRGR
Fur4  PWNLMATSSSF-TMALSAAYAIFLSIIAGVCSDYVFVVRGG
Dal4  PWNLMAASSKF-TSALGAYAIFLSIIAGVICADYFVVRGR
Fui1  PWDLLSSSSKF-TTALAYAVFLSIAAGVISADYFIVRKG
FcyB  -------SHFETVLENFMNFIAYWLAISAIAMDHFVFVKRG
Fcy2  -------YYFDGMENFMDSIGGYYLAIYIAISCEHFFYRRS
Thi7  PWNFYNNSSSTF-LTVSSFGVVMTPIIISVMCDNLIRKR
Tpn1  -------NHFSILGNFLPMIGYWISMYFILLFEENLVFRF
Nrt1  PWNFYNNSSSTF-LTVSSFGVVMTPIIAVMCDNLIRKR
AtNCS1 PWRLLKKSESFYTWLIGSYALLGPIGGIILVYLYLKKM
CrNCS1 PWNLVSTHGFPVNTWGLIGYALLGPVIGMDSYFPVRQR
ZmNCS1 PWRLLSSSESFYTWLGLYSALMGPIGGVVLADHYIVRRT
SvNCS1 PWRLLSSSESFYTWLGLYSALMGPIGGVVLADHYIVRRT
References


