Supplementary Materials, Methods, Tables and Figures

The 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 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'-CCGGAATTCGCATATGAAATTAAAAGAGAGTCAGCAGCAATCCA-3' and reverse: 5'-AAAACTGCAGCTTCAGCCTGGCGGACCTGCGCATGTT-3') were designed to extract and amplify the *pucI* 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^{TM}) 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^{TM}) 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_2HPO_4, 0.9 g/L sodium citrate, 0.18 g/L $MgSO_4$, 1.8 g/L $(NH_4)_2SO_4$, 3.6 g/L KH_2PO_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-β-*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 xg, 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_2PO_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 ºC 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.

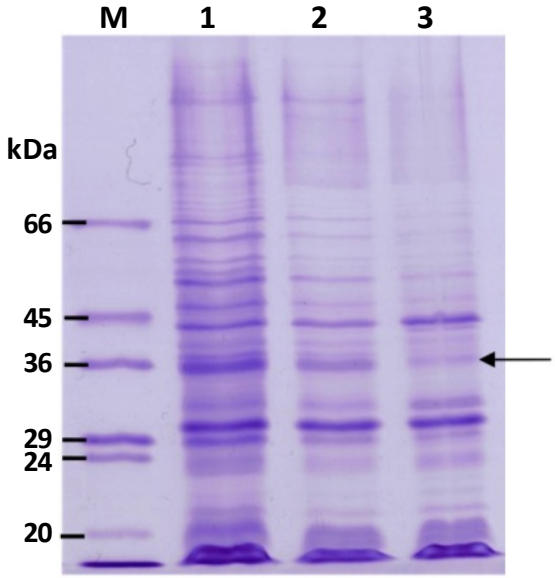
	Solubilisation buffer	Wash buffer	Elution buffer	Storage buffer
Tris-HCl (pH 7.5)	20 mM	20 mM	20 mM	--
Imidazole	20 mM	20 or 40 mM	200 mM	--
Glycerol	20%	10%	10%	5%
NaCl	300 mM	150 mM	--	--
DDM	1%	0.05%	0.05%	0.05%
KH_2PO_4 (pH 7.5)*	--	--	--	10 mM

* 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).

	Sequence homology with PucI					
NCS1 proteins	**Identical**		**Highly similar**		**Overall**	
Mhp1	123	25.1%	132	26.9%	255	52.0%
Bacterial (Mhp1, CodB)	40	8.2%	73	14.9%	113	23.1%
Fungal (Fur-type: FurA, FurD, FurE, Fur4, Dal4, Fui1)	31	6.3%	68	13.9%	99	20.2%
Fungal (Fcy-type: FcyB, Fcy2, Thi7, Tpn1, Nrt1)	10	2.0%	53	10.8%	63	12.8%
Plant (AtNCS1, CrNCS1, ZmNCS1, SvNCS1)	108	22.0%	124	25.3%	232	47.3%

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
MKLKESQQQSNRLSNEDLVPLGQEKRTWKAMNFASIWMGCIHNIPTYATVGGLIAIGLSPWQVLAIIITASLI
LFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMWLGIQTFAGSTALNILLLNMWPGWGEIG
GEWNILGIHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAGGLGPIYSQPGKF
HTFSETFWPFAAGVTGIIGIWATLILNIPDFTRFAETQKEQIKGQFYGLPGTFALFAFASITVTSGSQVAFGE
PIWDVVDILARFDNPYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYINFKRGSFITALLALFTVPWK
LMESATSVYAFLGLIGGMLGPVAGVMMADYFIIRKRELSVDDLYSETGRYVYWKGYNYRAFAATMLGALISLI
GMYVPVLKSLYDISWFVGVLISFLFYIVLMRVHPPASLAIETVEHAQVRQAE
```

B

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

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            ----------------------MKLKESQQQSNRLSNEDLVPLGQEKRTWKAMNFASIWM
AllPBcereus     ----------------------MKLKESQHQSNRLSNEDLLPLGQEKRTWKAINFASIWM
AllPEfaeca      --------------MEKNVSRVTAQEETAMKARGYNEDLLPSSPKQRTMGARNFFTLWM
AllPLBact       ----------------------MDNAQLEKYRSRGYSDDLLPKTENKRTWGTFNYFTLWM
AllPRaqua       -------------------MNESECTQQERYRERGYSNDLLPKLKEKRNWKGFNYFTLWM
AllPBagres      -------------------MERQEQQQRELYRARGYSDDLLPKEKEKQTWKAFNYFTLWM
AllPEcoli       ----------------------MEHQRKLFQQRGYSEDLLPKTQSQRTWKTFNYFTLWM
AllPSdysen      ----------------------MEHQRKLFQQRGYSEDLLPKTQSQRTWKTFNYFTLWM
AllPCfreun      ----------------------MEHQRELYQQRGYSDDLLPKTAEQRNWKTFNYFTLWM
AllPStyphi      ------------------------------------------------------------
AllPSerrat      -------------------MESISSKQREKYQQRGYHEDLLPKETDKKTWKAINYFTLWM
AllPYinter      -------------------MNDIEENKREVYRSRGYPEDLLPKTKDKKNWRAFNYFTLWM
AllPArubri      --------------MDHAESGTMAADGGFAGDTGLFNADLAPVPPAGRDWSWVNMSTVWM
AllPAacido      ------------------------------------------------------MNFTTVWM
AllPPdurus      ------------------------METNKLSPSLSNTDLLPVKPEERTWKAFNFASIWM
AllPSafgha      MTDTAPTAPPPTTQVTLADGRVEIAPGAPAPTGPYANEDLLPVPVEKRTWTTYNFSALWV
AllPKflavi      --------MTSTEQTYHPDGRVELTDPEAVATSRYGNAELAPTRLAERRWTTYNYAALWM
AllPKutzn       MTS-GAAMAHSPVPVTPTDGRVELADDAAIADSRFYNSELAPVPLEKRTWTTYNFFALWM
AllPAjapon      -----MEPTARGTQHVHPDGRVELGEVESLKDSRFYNEELAPVPVEKRTWTTYTYFALWM
AllPKibdel      ---------MDGTHLTHPDGRVDLVDSSGIAASRFYNPELAPVPVEGRRWSTYNYFALWM
AllPCkluyv      -----------MEQLVEKEIYELDKSDINVTESKLYNDDNAPVPVKERTWNTYNFTALWI
AllPSacido      ----------------------MTHPLSEPEVDIANRDLLPTTSSQRQWTLYNYLTLWI
AllPRpicke      -----------MSQ---TTSSAFSADAVGAPDPTLWNEDLNPTPPAARTWTATNYAALWV
AllPCapicu      ------------------------------MSLSNEDLAPTPAEKRTWTMWHYAALWV
AllPSusita      -------------------------MKPRYDLSLYNEDLAPVPPEKRTWGTYNYAALWI

PucI            GCIHNIPTYATVGGLIAIGLSPWQVLAIIITASLIVFGALALNGHAGTKYGLPFPVIIRA
AllPBcereus     GCIHNIPTYATVGGLIAIGLSPWQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRA
AllPEfaeca      GSIHNIPNYAAVGGFIFLGLSPLQVMLAVVLSSFIVATFMNLNGVAGSKYGIPFAMHLQS
AllPLBact       GSVHNVPNYVAVGGFLILGLSTVSIMAAIIVSAFIIAAVMVLNGAAGSKYGVPFAMILRA
AllPRaqua       GSVHNVPNYIAVGGFLILGLSTFSVMMAIIISALFIAAVMVLNGAAGSKYGVPFAMILRG
AllPBagres      GSVHNVPNYVMVGGFFILGLSTLSIMLAIILSAFFIAFVMVMNGAAGTKYGVPFAMILRA
AllPEcoli       GSVHNVPNYVMVGGFFILGLSTFSIMLAIILSAFFIAAVMVLNGAAGSKYGVPFAMILRA
AllPSdysen      GSVHNVPNYVMVGGFFILGLSTFSIMLAIILSAFFIAAVMVLNGAAGSKYGVPFAMILRA
AllPCfreun      GSVHNVPNYVMVGGFFILGLSTFSIMLAIIISALFIALVMVMNGAAGSKYGVPFAMILRG
AllPStyphi      ----------MVGGFFILGLSTFNIMLAIIISALFIAAAMVMNGAAGSKYGVPFAMILRG
AllPSerrat      GSVHNVPNYVAVGGFFILGLSTVSIMAAIILSAFVIAFVMVMNGAAGSKYGIPFAMLLRA
AllPYinter      GSVHNVPNYVAVGGFFILGLSTVSIMAAIILSALIIAFVMVMNGAAGSKYGIPFAMILRA
AllPArubri      GMVHNVVAYEAAAGLMQLGLSALQSLAAVAVAYFVLFVAMWFNARPGTAYGIPFCVLIRS
AllPAacido      GMVHNIVAYETAASLLSLGMSVWQALLTVIVANAVLIVAMCLNSVAGARYGLPFPVLVRA
AllPPdurus      GCIHNIPTYATVGGLIAIGMSPWQVLAVILVASLILYAALSLNGHAGAKYAIPFPVFIRS
AllPSafgha      GMAHNTASYTLASGLIAVGMDWKQAVFTIALANVIVLIPMLLTGHAGPKYGIPFPVFARA
AllPKflavi      GMAHNIPSYLLASGLVTLGMNWLQAFLTITLGNLIVLVPLLLNSHAGTKYGIPFPVFARA
AllPKutzn       GMAHNIPSYTLAASLIALGMDWVQAFLTITLGNLIVLVPMLLNSHAGTKYGIPFPVFARS
AllPAjapon      GMAHNIPSYALAASLIALGMDWVQALLTITIGNLIVLIPMLLNSHAGTKYGIPFPVFARA
AllPKibdel      GMAHNIPSYTLAASLIALGMDWVQAFMTITLGNLIVLAPMLLNSHAGTKYGIPFPVFARA
AllPCkluyv      GMAHCIPTYMLAGSLISLGMDWKQALFTITFGNLIVLIPILLNAHPGTKYGINFPVFSRA
AllPSacido      GMAHNVSTYMMAGGFIALGLSWWEAILTVLVGTLIVLVPILLNSHAGTQYGIPFPVYARA
AllPRpicke      SMVVSVPAYMLASGLMSEGMNWWQAVLTVFLGNLIVLVPMVLVGHAGTKYGIPFPVLVRA
AllPCapicu      GMSVCIPTYTMASGLIDQGMSWKEAIACVALGNVIVLAPMILNAHPGTRYGVPFPVLARA
AllPSusita      SMSVCVPTYMLASGLIAGGMNWWQAILTILLGNLIVLVPMVLNAHAGTKYGIPFPVLVRT
```

```
PucI            SYGIYGANIPALLRAFT-AIMWLGIQTFAGSTALNILLLNMW-PGWGE---IGGEWNILG
AllPBcereus     SYGIYGANIPALLRAFT-AIMWLGIQTFAGSTALNILLLNIW-PGWGE---IGGEWNILG
AllPEfaeca      TYGSLGAKLPGFLRGCVAAIAWFGLQTFTGSLALLIILGKFW-PNFLE---IGGSFQFFG
AllPLBact       SYGVRGALFPGILRGCVAAIMWFGLQCYAGSLAFLILIGKIW-PSFLE---IGGGVSILG
AllPRaqua       SYGIRGALFPGILRGCIAAIMWFGLQCYAGSLAFLILIGKIW-PEFLT---LGGDFNLLG
AllPBagres      SYGVRGSLFPGILRGGIAAIMWFGLQCYAGSLAFLILIGKIW-PGFLS---LGGDFNILG
AllPEcoli       SYGVRGALFPGLLRGGIAAIMWFGLQCYAGSLACLILIGKIW-PGFLT---LGGDFTLLG
AllPSdysen      SYGVRGALFPGLLRGGIAAIMWFGLQCYAGSLACLILIGKIW-PGFLT---LGGDFTLLG
AllPCfreun      SYGVRGALFPGLLRGGIAAIMWFGLQCYAGSLAFLILIGKIW-PGFLT---LGGDFTLLG
AllPStyphi      SYGVRGALFPGLLRGGIAAIMWFGLQCYAGSLAFLILIGKIW-PGFLT---LGGDFKLLG
AllPSerrat      SYGVRGALLPGILRGCVAAIMWFGLQCYAGSLAFLILLGKLW-PEFLT---LGGDFSLLG
AllPYinter      SYGVRGALFPGILRGCVAAIMWFGLQCYAGSLAFLILIGKLW-PEFLT---LGGDFNILG
AllPArubri      SFGPRGAQLPVVIRGFC-AIFWFAVQGYAGSLAIDAIIGTLI-PAWNT----LT-MPILG
AllPAacido      AFGHKGAQIPVFVRAFV-AIFWFSIQAYAGSEAVGAVFGALI-PGWAS----LGHYHIIG
AllPPdurus      SYGVLGANVPALLRGFV-AIMWFGIQAFAGSTALNILLLNVW-DGWGT---LGGDWNLLG
AllPSafgha      SFGIRGANLPAVVRALV-ACGWFGIQTWIGGEAIYFLAGKLVGSGWTD------AAKVGG
AllPKflavi      FYGVVGANFPALLRAFV-ACGWFGIQTWIGGQAIHVIVGELAGAGWRD------ATAIAG
AllPKutzn       FYGVRGANLPALLRAFI-ACGWFGIQTWVGGEALYVIVGKLFGGGWSN------AAAIGG
AllPAjapon      FFGMRGANLAALLRAFI-ACGWFGIQTWVGGEAIYIIVGRLAGSGWKD------SAVVLG
AllPKibdel      FYGVRGANLAALLRAFI-ACAWFGIQTWVGGEALYVIVGKLTGSGWIN------AAEVGG
AllPCkluyv      AFGVFGANIPAVLRAVV-ACGWFGINTYIGGSALNVLFSAVI-PGWKT---LGGSFEIAG
AllPSacido      SFGGVGAGVPALLRALV-AAGWFGINAAIGGQAVQMFLSMLI-PGWAH---LSTAFTFVG
AllPRpicke      SFGVRGAQLPAILRAIV-ACGWFGIQTWLGSQAIYTILNVVT-DNML----VGSNIPGLG
AllPCapicu      SFGVLGANIPALLRALV-ACGWFGIQTWIGGQALYQLFAAAV-PAIVEPLSSATFKAAVG
AllPSusita      SFGVRGANLPAVLRALV-ACGWFGIQAWIGGQAIYSMLKIIW-PP------------AGD

PucI            IHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAGGLGPIY
AllPBcereus     IHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAGGLGPIY
AllPEfaeca      LRLPELMAFTLFWLLNVAIGFGGSKILNRFTAILSPLIYVVIIGLTIWAIRAGGGLTPIL
AllPLBact       ITIPGLIAFLLFWAVNVAIGFGGGGVLNKFTAILNPCIYIVFGGMAIWAIYLA-GFGNIV
AllPRaqua       ISMPGLIAFMIFWAINVMIGFGGGGVLNKFTAILNPCIYVVFGGMAVWAISLA-GLENIL
AllPBagres      LNLPGLIAFLFFWAINVAIGFGGGGVLNKFTAVLNPCIYIVFGGMAIWAISLV-GIQPIL
AllPEcoli       LSLPGLITFLIFWLVNVGIGFGGGGKVLNKFTAILNPCIYIVFGGMAIWAISLV-GIGPIF
AllPSdysen      LSLPGLITFLLFWLVNVGIGFGGGGKVLNKFTAILNPCIYIVFGGMAIWAISLV-GLGPIF
AllPCfreun      LSLPGLITFLIFWLINVGIGFGGGGKVLNKFTAILNPCIYIVFGGMAIWAISLV-GIGPIL
AllPStyphi      LSLPGLITFLIFWIINVGIGFGGGGKVLNKFTAILNPCIYIVFGGMAIWAISLV-GIGPIL
AllPSerrat      ISLPGLIAFLIFWAVNVAIGLGGGGSILNKFTAILNPCIYVVFGGMAIWAISLA-GLSNII
AllPYinter      LSLPGLIAFIIFWAVNVAIGLGGGGSILNKFTAVLNPCIYIVFGGMAIWAISLA-GFDNII
AllPArubri      MALKGWVAVALFWALHAWIVSHGVHRIRNFELIAGPLVILVGLLATAWGLTVAHGVGPLF
AllPAacido      MGLNTAIAVALFWLLHIWVVSHGINRVKYFELWAGPLVIVLGLCLVVWSITVAHGFGPAF
AllPPdurus      LHLPGLLSFLLFWGLNVLVLHHGMESIKKFEVWAGPLVYVVFGGMVWWAIDIAGGLGPIY
AllPSafgha      YAWTMWLSFAIFWALQVVIIYRGMETIRRFENWAAPFVLVGAFVMLWWMSDKAGGFGPLF
AllPKflavi      HPWTLWLSFAVFWALQMWLIWRGIEGLRRFENWAAPLVTVAFLALMIAILVKAGGPGPIL
AllPKutzn       QPWTLWLSFAAFWVVQMLIIWRGMDAIRRFENWTAPLVSVGFLILLVYVLVKAGGFGPIL
AllPAjapon      QHWTLWLSFGLFWLFQMLIIWRGMEAVRRFENWTAPLVSVGFLILLGYVLVKAGGLGPIL
AllPKibdel      QPWTLWLCFAAFWVVQMLIIWRGMEAIRRFENWTAPLVSIGFLILLAYVVIKAGGFGPIL
AllPCkluyv      LSLPAAITFMIFWGIQMFIIFKGMEQLKKFENWAAPAVIILAVFLVIWAVSSAHGFGPLL
AllPSacido      LDFGGWVSFLLFWFLNIWIIYHGIDAVRRFEAWAGPLVLLLGIGLLLWAYNAAHGFGPML
AllPRpicke      INPGQGFCFLLFWALHIWLITKGLESIKRFQALATPLLILAALGLVWWAYTNAGGFGPML
AllPCapicu      IQPGELLGFAVFWAITLRFILRGTESIKFLESWAAPFLIVMGLVLLGWAYTRAGGFGPML
AllPSusita      FAAGVWICFFAFWALNMWVVWRGIDTIKFLEGVGAPFLLGIGLLLLLWITGKAGGLGPVL
```

```
PucI            SQ-PGKFHT----FSETFWPFAAGVTGIIGIWATLILNIPDFTRFAETQKEQIKGQFYGL
AllPBcereus     SQ-PGRFHT----FSETFWPFAAGVTGIIGIWATLILNIPDFTRFAETQKEQIKGQFYGL
AllPEfaeca      SYQVSGAIR----SVNPLVAYLIIFNSVVAVWSAPGASVADFTKNARSTRAQVVGQTAGL
AllPLBact       NYVPANVQT----GGNSIFLFLVVINAVVAVWAAPAVSASDFTQNAKSFKAQATGQTFGL
AllPRaqua       NYVPANAEQ----SGNPLFLFLVVINAVVAVWAAPAVSASDFTQNASSFKQQAWGQTLGL
AllPBagres      DYVPAGVQK----AENSGFLFLVVINAVVAVWAAPAVSASDFTQNAQSFRQQALGQTLGL
AllPEcoli       DYIPSGIQK----AENGGFLFLVVINAVVAVWAAPAVSASDFTQNAHSFREQALGQTLGL
AllPSdysen      DYIPSGIQK----AENSGFLFLVVINAVVAVWAAPVVSASDFTQNAHSFREQALGQTLGL
AllPCfreun      DYIPGGVQK----AGNSGFLFLVVINAVVAVWAAPAVSASDFTQNAHSFREQALGQTLGL
AllPStyphi      DYLPSGVQK----AEHSGFLFLVVINAVVAVWAAPAVSASDFTQNAHSFRAQALGQTLGL
AllPSerrat      AYVPANTDI----TSNSGFMFLVVINAVVAVWAAPAVSASDFTQYAKSFQQQAVGQTLGL
AllPYinter      SYVPANVVM----AEHSGFMFLVVINAVVAVWAAPAVSASDFTQNASSFRQQAFGQTAGL
AllPArubri      DQ-PSRLTG-----TDAWLTFCVGVTGMIGIWSTFAVNIPDLSRFVRSERDQVIGQLIGL
AllPAacido      TQ-PSKLHG-----VAFWQAFGLSVTGLVGTWSTLVLNIPDLTRFSRSQKDQIVGQAIGL
AllPPdurus      AQ-ASKFQS----FGDLFWVFVASVTGIIGIWATLILNIPDFTRFAKSQKEQIKGQFWGL
AllPSafgha      DQ-PSKLGW----GPDFWKLFAPALMGMIGFWSTLSLNIPDFTRYGRSQKAQTWGQALGL
AllPKflavi      SQ-PSTLGW----DADFWKIFAPSLMGMIAFWATLSLNMPDFTRFGQGQRQQVLGQIIGL
AllPKutzn       SE-PSKLGW----GSGFWAVFAPSLMAMIAFWSTLSLNMPDFTRFGGSQRKQFWGQILGL
AllPAjapon      SE-PGKLGW----GPDFWKVFAPSLMAMIAFWSTLSLNMPDFTRFGGSQGKQVRGQILGL
AllPKibdel      SE-PSKLGW----GGDFWKVFAPALMGMIAFWSTLSLNMPDFTRFGGSQRKQVTGQILGL
AllPCkluyv      SE-ESKLKT----MGDFMKVFPAALTSMVGFWATLSLNIPDFTRFAKGQKEQMVGQSLGL
AllPSacido      HQ-PAKVHG-----AALWAVEIPALTSVVGNWATLSLNIPDFTRFAKSQKAQIWGQTLGL
AllPRpicke      SA-PSAFAAGGKRAGEFWGFFWPSLTAMVGYWATLALNIPDFTRFARSQRDQLVGQAVGL
AllPCapicu      AQ-PSKLE------GRFWKVFGPGLTAMVGFWATLSLNIPDFTRYAKSQRDQALGQAIGL
AllPSusita      KT-PSKFHT----TAEFARFFIPSLTGMVGFWATVALNIPDFTRYAKSQKAQIWGQVLGL

PucI            PGTFALFAFASITVTSGSQVAFGEPIWDVVDILARFDNPYVIVLSVITLCIATISVNVAA
AllPBcereus     PGTFALFAFASITVTSGSQVAFGEPIWDVVDILARFDNPYVIVLSVITLCIATISVNVAA
AllPEfaeca      VVGYGIFAFSSVVILLGGSLYFGIQEWNILNIIDRLDNVAVVVLAMSVFLLTTISTNATG
AllPLBact       AVAYVLFAIASVCILAGASIHYGTETWNVLDIVQKWDSLFASIFAVLVILMTTISTNATG
AllPRaqua       IVAYVLFAVASVCILAGASIHYGVDTWNVLDIVQKWDSLFASVFAVLVILMTTISTNATG
AllPBagres      LVAYILFAVAGVCIIAGASIHYGEDTWNVLDIVQKWDSLFASFFAVLVILMTTISTNATG
AllPEcoli       VVAYILFAVAGVCIIAGASIHYGADTWNVLDIVQRWDSLFASFFAVLVILMTTISTNATG
AllPSdysen      VVAYILFAVAGVCIIAGASIHYGADTWNVLDIVQRWDSLFASFFAVLVILMTTISTNATG
AllPCfreun      IVAYVLFAIASVCIIAGASIHYGVDTWNVLDIVQRWDSLFASFFAVLVILMTTISTNATG
AllPStyphi      IVAYILFAVASVCIIAGASIHYGMDTWNVLDIVQRWDSLFASFFAVLVILMTTISTNATG
AllPSerrat      VVAYLLFAVASVCILAGASIHYGVDTWNVLDIVQKWDSVFASVFAVLVILMTTISTNATG
AllPYinter      VVAYILFAVASVCILAGASIHYGVDTWNVLDIVQKWDSLFASVFAVLVILMTTISTNATG
AllPArubri      PLTAIVFTAMSVVTTSATILVFGHPIWDPVQILLALHEPWVLLLGGVTIIVATLSVNVAA
AllPAacido      PGTAILFSVMSIVITSGTLIAFGTAVTDPVQILGKFNNSIVLMFGAFALLIATLSVNVAA
AllPPdurus      PGTFILFAFASITVTSGSQVAFGTPIWDVVEILKYFNHPFIIAVSVITLCMASVSVNVAA
AllPSafgha      PTTMTLFAFLSVMVTSGSQAVYGEAIWDPVQLAAKTDNTVGLLFALVTVLVATLSVNVAA
AllPKflavi      PTTMSFIALVSIVTTSGTVVVYGSAIWDPVELTRRFENPLVVTIGLVMAILATMSCNVAA
AllPKutzn       PTTMSFIAIVAILTTSGAVALYGEAIWDPAQLAARFDSPVLVVVALIALVLATISANLAA
AllPAjapon      PTTMTFIAIVAILTTSGGSVLYGEQIWDPAKLADRFDSPVVVVVALVALVLATVSANLAA
AllPKibdel      PTTMSFIALVAILTTSGAMSLYGEAIWDPAQLASRFDSPLLVVIALIALVLATVSANLAA
AllPCkluyv      PITMTIFSAMGIIITSATVVIYGKAMWDPVDIIAKFTNPVALLIGFFGIVVASLSVNIAA
AllPSacido      PTTMTVFSAIGVLVTSATIVVFHQAIADPVTLLGHFHNVLLLLISLGAVVVATLSVNVAA
AllPRpicke      PLPMGLLALVAVLVTSSTVVIYGQAIWDPVTLAGKMTGPSV-IVALLALITATLMTNIAA
AllPCapicu      PGTMVLFSFIGVAVTSATPIIFGETIWDPVKLLGRIGGALILIVAMFGLGVATLSTNLAA
AllPSusita      PTTMTFYSFIGVAVTSASVVLFGRPIWDPVELLGKFNQPLVAFIAMIALLLATLSTNVAA
```

```
PucI            NIVSPAYDIANALPKYINFKRGSFITALLALFTVPWKLMESA-TSVYAFLGLIGGMLGPV
AllPBcereus     NIVSPAYDIANALPKYINFKRGSFITALLALFTVPWKLMESA-TSVYAFLGLIGGMLGPV
AllPEfaeca      NIIPAGYQLAALFPKKMTYKKGVMIASVISFLIMPWKLMENA-DSIFIFLNAIGAVLGPV
AllPLBact       NIIPAGYQIAAIFPKKLTYKHGVMIASIISVLICPWKLMENQ-ASIYLFLDIIGGILGPV
AllPRaqua       NIIPAGFQIAAIAPKKLTYKKGVLIASLISVVICPWKLMENQ-ESIYLFLDIIGGMLGPV
AllPBagres      NIIPAGYQIAAIAPTKLTYKNGVLIASIISLLICPWKLMENQ-SSIYLFLDIIGGMLGPV
AllPEcoli       NIIPAGYQIAAIAPTKLTYKNGVLIASIISLLICPWKLMENQ-DSIYLFLDIIGGMLGPV
AllPSdysen      NIIPAGYQIAAIAPTKLTYKNGVLIASIISLLICPWKLMENQ-DSIYLFLDIIGGMLGPV
AllPCfreun      NIIPAGYQIAAIAPTKLTYKKGVLIASIISLLICPWKLMENQ-SSIYLFLDIIGGMLGPV
AllPStyphi      NIIPAGYQIAALAPTKLNYKNGVMIASIISLLICPWKLMENQ-DSIYLFLDIIGGMLGPV
AllPSerrat      NIIPAGYQIAAIAPKKLTYKNGVVIASIISLIICPWKLMENQ-ESIYLFLDVIGGILGPV
AllPYinter      NIIPAGYQIAAIAPKKLTYKNGVIIASLISLIICPWKLMENQ-ESIYLFLDVIGGILGPV
AllPArubri      NIMPAAYDLVNLMPRRLGFNSASMLVLVIGLFFAPWLWFHNA-NSIFAVLGGIGGLLGPV
AllPAacido      NVVSPAYDLVNLFPKKLNFVRAGVISVVIGLCFAPWLWYDNG-GVIFSVLNAIGGGLGPV
AllPPdurus      NIVSPAYDLANLFPKWITFKRGGYIAAILSLLTVPWKMMEQS-TSIFAFLGTIGGALGPV
AllPSafgha      NLVSPAFDFSNIAPRKISFRAGALATCVLGVLIFPWKLYSDPQGYIFTWLGLVGGLLGTV
AllPKflavi      NVVSPSYDFANALPRWLNFRTAGLLTGVIGVLIQPWRLISDPDIYIFAWLSFYGGLLASV
AllPKutzn       NVVSPSYDFSNAVPKRITFATGGLITGVLGVLIQPWRLISDPHIYIFTWLGFYGGVLAAV
AllPAjapon      NVVSPSYDFSNAFPKKITFAVGGLITGIIGIVIQPWRLYSDPNIYIFAWLGFYGGLLGAV
AllPKibdel      NVVSPSYDFSNAFPKKITFATGGLITGVVGILIQPWRLISDPSIYIFAWLGFYGGLLAAI
AllPCkluyv      NIVSPANDFSNMAPKHISFKMGSLITGIIGILIMPWKLLSDPSGYIYAWLGTYSGILGPV
AllPSacido      NVVSPAYDFIQLFPKHLNFSRAGLLTGILGIVMVPWLLISNPHIYIFSWLNVYGGFLGPI
AllPRpicke      NVVSPAYDFSNLAPHRISFRTGGYITAGIGLAMMPWKILETTKGYIFTWLVGYGALLGPV
AllPCapicu      NVVSPANDFSNLSPSRISYRMGGVITAVIGALIMPWKLIESSQGYIFVWLVGYSALLGPI
AllPSusita      NVVSPSNDFANLNPQRISFRTGGMITGVIGVLMMPWKLLSDLSAYVFGWLVGYSGLLGPI

PucI            AGVMMADYFIIRKRELSVDDLYSE-TGRYVY---WKGYNYRAFAATMLGALISLIGM---
AllPBcereus     AGVMMADYFIIRKRELSVDDLYSE-TGRYVY---CKGYNYRAFAATILGALISLIGM---
AllPEfaeca      AGVMIANYYFVQKQQINLNALYVD-KHKKEEANPFYGLNKPAYVATILALVLSLSGQ---
AllPLBact       IGVMLAHYFIIMRRQINLDSLYTE-PGQFSY--YKNGFNSLAFVVTIVAVIISLSGK---
AllPRaqua       IGVMMAHYFIVVRSELDLDTLYTA-PGNYHY--YDRGFNTVAFAVTLIAVVLSLGGK---
AllPBagres      IGVMMAHYFIVMRSQIDLDTLYTK-AGDYKF--YDNGFNVTAFSVTLIAVVLSLGGK---
AllPEcoli       IGVMMAHYFVVMRGQINLDELYTA-PGDYKY--YDNGFNLTAFSVTLVAVILSLGGK---
AllPSdysen      IGVMMAHYFVVMRGKINLDELYTE-PSDYKY--YDNGFNLTAFSVTLVAVILSLGGK---
AllPCfreun      IGVMLAHYFIVMRGKINLDELYTA-SGDYQY--YDNGFNLTAFSVTLVAVILSLGGK---
AllPStyphi      IGVMLAHYFVVMRGKINLDELYTA-SGDYKY--YDNGFNLTAFSVTLVAVILSLGGK---
AllPSerrat      IGVMMAHYFIVIRSDINLDTLYTE-PGNYKY--YENGFNSVAFIVTLVAVVLSLGGK---
AllPYinter      IGVMMAHYFVIMRRDIDLDTLYTE-DGNYKY--YDNGFNTTAFVVTLISVILSLGGK---
AllPArubri      TGIMLTDYYLIRRQRLSVPELYRY-EGRYAG---RGGWNPAGVWAFLIGGTCALIGA---
AllPAacido      AGIMLADFFMIKRRKYDVLSFYRS-DSEYRY---TNGWNLRAIGALVIGLIAAFIGL---
AllPPdurus      AGVMFADYFIIRKRTLEVDELYKL-NGKYTY---YKGYNYRAFVATAIGAFVSLIGQ---
AllPSafgha      AGILIADYWILRRSRLDLADLYRT-GGRYWY---EGGWNWRAVVAFAVGGVLAVGGASF-
AllPKflavi      AGVLIAGYWFVDRTNLFLADLYLV-NGRYWY---SAGWNWRAVVATLVGSVLAVGGAYG-
AllPKutzn       AGVLVAGYWLIDRTQLSLPDLYQE-NGKYWF---TGGWNWRALVATVVGAVIAVGGAYSA
AllPAjapon      AGVLVAGYWVVNRTKLRLRDLYTE-RGIYWF---NGGWNWRALVATLAGAVLAVGGAYG-
AllPKibdel      AGVFVAGYWTLAKTRLNLADLYKDGEGAYWF---HGGWNWRAVVATLLAGVLAVGGAYG-
AllPCkluyv      AAIIICDYWIIKKKNLVLKDLYLT-KGKYTY---NKGFNLRAVISLAVGIFAALIGK---
AllPSacido      AGILIADYWVFRKTTLAVVELYAK-GGRYYY---ANGYNWRAIAALAGGVVIALIGK---
AllPRpicke      AGIMMVDYFLVRRTVLKTGELFRV-DGPYGY---GNGWNGRAIAALVIGVLPNLPGFFKQ
AllPCapicu      GGIMIVDYFLVRKKRLDVDDLYRR-GGIYEY---SNGVNWKAIIAMAIAVAVNLPGFLAE
AllPSusita      AGVMIADYFLVRHAHLDIDDLYRR-NGIYEY---DNGINRRAVVALAAGIGVALIGL---
```

```
PucI           ----------YVPVLKSLYDISWFVGVLISFLFYIVLMRVHPPASLAIETVEHAQVRQAE
AllPBcereus    ----------YVPALKSLYDISWFVGVLISFLFYIVLMRVHPPASSAIEPFESRQVRQAE
AllPEfaeca     ----------FIPQVKIIADISWFVGFATGFVLYLVLKKWTWDSKKVKETAY----QEGK
AllPLBact      ----------FIPVLEPVSRLSWFVGVIVAFGAYALFASLHRKKNPSFYDEN----TEVQ
AllPRaqua      ----------FIPVLEPLSRVSWFVGVIVAFVLYSVFMKREPSLQPQNV----------
AllPBagres     ----------FIPLFEPLSRVSWFVGVITAFVLYVLLKKRDAPGISTEHKAA--------
AllPEcoli      ----------FIHFMEPLSRVSWFVGVIVAFAAYALLKKRTTAEKTGEQKTI----G---
AllPSdysen     ----------FIPFMEPLSRVSWFVGVIVAFAAYALLKKRTTAEKTGEQKTI----G---
AllPCfreun     ----------FIPFLEPLSRVSWFVGVIVAFVAYALLKKRTGAQSAGVQKVT----GQM-
AllPStyphi     ----------FIPFMEPLSRVSWFVGVIVAFVAYALLKKRTGFENTGEKKLA----G---
AllPSerrat     ----------FISILEPLSRVSWFVGVISAFCLYALIKSKVAASGKNIPDVD----IITK
AllPYinter     ----------FIPLLEPLSRISWFVGVITAFVLYVLIKRRTIANKTEYA----------
AllPArubri     ----------VVPALHTIYAFAWFIGIAVGAAVYGVLATRRRAVEGLSPARA--------
AllPAacido     ----------VVPVLSILYTYSWFIGVIVGGVAYVLLMRSSMSVEAIEPVAVGMFEEN--
AllPPdurus     ----------FVPSLKYLYDISWFVGVLFAFVTYIALMRLHPPAAIAINESKESLIEKTV
AllPSafgha     ---KPLIDGRPIPALADLADYGWAVGLGTSMLLYLVLMAARGGNRATV------------
AllPKflavi     ---GPFPTEGLIPFLQPLYDYSWVVGLLAGFLGYVGLTVAFPHRTDKAVHAAPTF-----
AllPKutzn      PGTGPFPADGLIPFLKPVYDYSWVAGLIAAFLLYLVLTPRTSATSAVTVATN--------
AllPAjapon     ---GPFPADGLIPFLKPLYDYNWVVGLAGAFVVYLLLSLPERKRTTDIEEDASERRGPSR
AllPKibdel     ---GPFPADGLIPLLKPLYDYSWVVGAVVGYVVYLVLSVSTKH----TEEEASAADRSRR
AllPCkluyv     ----------IIPSLNGLANYAWFVGFAVSFVIYYLLSASSKEPESAAELANESS-NS--
AllPSacido     ----------VVPALAWLFNYSWFVGFIVAFIVYLGLMQTAESPDVRLAGSR--------
AllPRpicke     A----GFVASVPGVFEALYTYAWFVGLAISAVVYVILMRGRR------------------
AllPCapicu     A--IPSLKDAVPPLLKTLYTYAWFVGVLVAGSIYYLLMVRSVREPSEPGAAPASP-G---
AllPSusita     ----------IVPPLKFLYDYAWFVGFAVAGGVYVCLMRGTGV-PARAIR----------

PucI           ----------
AllPBcereus    ----------
AllPEfaeca     ----------
AllPLBact      IKKIGVEGNEL
AllPRaqua      ----------
AllPBagres     ----------
AllPEcoli      ----------
AllPSdysen     ----------
AllPCfreun     ----------
AllPStyphi     ----------
AllPSerrat     ----------
AllPYinter     ----------
AllPArubri     ----------
AllPAacido     ----------
AllPPdurus     ----------
AllPSafgha     ----------
AllPKflavi     ----------
AllPKutzn      ----------
AllPAjapon     IDPAAVDG---
AllPKibdel     IDPAAVDG---
AllPCkluyv     ----------
AllPSacido     ----------
AllPRpicke     ----------
AllPCapicu     ----------
AllPSusita     ----------
```

Current details about the bacterial putative allantoin permeases from the UniProt KnowledgeBase are listed below in alphabetical order of the bacterial species. PucI from *Bacillus subtilis* is highlighted in blue.

AllPArubri
>tr|A0A0D6P5P0|A0A0D6P5P0_9PROT
Cytosine/purines uracil thiamine allantoin permease OS=Acidisphaera rubrifaciens HS-AP3 GN=Asru_0108_06 PE=4 SV=1

AllPAacido
>tr|T0BNV5|T0BNV5_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

AllPCkluyv
>tr|B9E3U4|B9E3U4_CLOK1
Uncharacterized protein OS=Clostridium kluyveri (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

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

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

AllPStyphi
>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    MKLKESQQQSNRLSNEDLVPLGQEKRTWKAMNFASIWMGCIHNIPTYATVGGLIAIGLSP   60
Mhp1    MNS-TPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAA-GQMTSSFQV   58

PucI    WQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMWL   120
Mhp1    WQVIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWF   118

PucI    GIQTFAGSTALNILLLNMWPGWGEIGGEWNILGIHLSGLLSFVFFWAIHLLVLHHGMESI   180
Mhp1    GFQTWLGALALDEITR-LLTGFTNLP-------------LWIVIFGAIQVVTTFYGITFI   164

PucI    KRFEVWAGPLVYLVFGGMVWWAVDI-AGGLGPIYSQPGKFHTFSETFWPFAAGVTGIIGI   239
Mhp1    RWMNVFASPVLLAMGVYMVYLMLDGADVSLGEVMSMGGE-----NPGMPFSTAIMIFVGG   219

PucI    WATLILNIPDFTRFAETQKEQ------------IKGQFYGLPGTFALFAFASITVTSGSQ   287
Mhp1    WIAVVVSIHDIVKECKVDPNASREGQTKADARYATAQWLGMVPASIIFGFIGA--ASMVL   277

PucI    VAFGEPIWDVVDILARFDNPYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYINF   347
Mhp1    VGEWNPVIAITEVVGGVSIPMAILFQV-FVLLATWSTNPAANLLSPAYTLCSTFPRVFTF   336

PucI    KRGSFITALLALFTVPWKLMESATSVYAFLGLIGGMLGPVAGVMMADYFIIRKRELSVDD   407
Mhp1    KTGVIVSAVVGLLMMPWQFAGV---LNTFLNLLASALGPLAGIMISDYFLVRRRRISLHD   393

PucI    LYSETGRYVYWKGYNYRAFAATMLGALISLI--------GMYVPVLKSLYDISWFVGVLI   459
Mhp1    LYRTKGIYTYWRGVNWVALAVYAVALAVSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFP   453

PucI    SFLFYI----VLMRVHPPASLAIETVEHAQVRQAE-------------               490
Mhp1    LFSEAAESRNEDYLRPIGPVAPADESATANTKEQNQPAGGRGSHHHHHH               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* (P0AA82). 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).

```
PucI    MKLKESQQQSNRLSNEDLVPLGQEKR-TWKAMNFASIWMGCIHNIPTYATVGGLIAIGLS   59
Mhp1    MNS-TPIEEARSLLNPSNAPTRYAER-SVGPFSLAAIWFAMAIQVAIFIAA-GQMTSSFQ
CodB    ------MSQDNNF---SQGPVPQSARKGVLALTF--VMLGLTF-FSASMWTGGTLGTGLS

PucI    PWQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMW   119
Mhp1    VWQVIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFW
CodB    YHDFFLAVLIGNLLLGIYTSFLGYIGAKTGLTTHLLARFSFGVKGSWLPSLLLGGTQVGW

PucI    LGIQTFAGSTALNILLLNMWPGWGEIGGEWNILGIHLSGLLSFVFFWAIHLLVLHHGMES   179
Mhp1    FGFQTWLGALALDEITR-LLTGFTNLP-------------LWIVIFGAIQVVTTFYGITF
CodB    FGVGVAMFAIPVGKAT-----GL--------------DINLLIAVSGLLMTVTVFFGISA

PucI    IKRFEVWAGPLVYLVFGGMVWWAVDI-AGGLGPIYSQPGKFHTFSETFWPFAAGVTGIIG   238
Mhp1    IRWMNVFASPVLLAMGVYMVYLMLDGADVSLGEVMSMGGE-----NPGMPFSTAIMIFVG
CodB    LTVLSVIAVPAIACLGGYSVWLAVNGMG-GLDALKAV------VPAQPLDFNVALALVVG

PucI    IWATLILNIPDFTRFAETQKEQ-----------IKGQFYGLPGTFALFAFASITVTSGS   286
Mhp1    GWIAVVVSIHDIVKECKVDPNASREGQTKADARYATAQWLGMVPASIIFGFIGA--ASMV
CodB    SFISAGTLTADFVRFGRNAKLAVLVA--------MVAFFLGN-SLMFIFGAAGAAALGMA

PucI    QVAFGEPIWDVVDILARFDNPYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYIN   346
Mhp1    LVGEWNPVIAITEVVGGVSIPMAILFQV-FVLLATWSTNPAANLLSPAYTLCSTFPRVFT
CodB    -------DISDVMIAQGLLLPA-----IVVLGLNIWTTNDNALYASG-LGFAN--ITGMS

PucI    FKRGSFITALLALFTVPWKLMESATSVYAFLGLIGGMLGPVAGVMMADYFIIRKRELSVD   406
Mhp1    FKTGVIVSAVVGLLMMPWQFAGV---LNTFLNLLASALGPLAGIMISDYFLVRRRRISLH
CodB    SKTLSVINGIIGTVCALWLYNN----FVGWLTFLSAAIPPVGGVIIADYLMNRRRYEHFA

PucI    DLYSETGRYVYWKGYNYRAFAATMLGALISLI--------GMYVPVLKSLYDISWFVGVL   458
Mhp1    DLYRTKGIYTYWRGVNWVALAVYAVALAVSFLTPDLMFVTGLIAALLLHIPAMRWVAKTF
CodB    T--------TRMMSVNWVAILAVALGIAAGHWLPGIVPVNAVLGGA--------------

PucI    ISFLFYI----VLMRVHPPASLAIETVEHAQVRQAE-------------  490
Mhp1    PLFSEAESRNEDYLRPIGPVAPADESATANTKEQNQPAGGRGSHHHHHH
CodB    ---LSYLILNPILNRKT---TAAMTHVEANSVE----------------
```

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).

```
PucI   ------------------------------------------------------------
FurA   ------------------------------------------------------------
FurD   ------------------------------------------------------------
FurE   ------------------------------------------------------------
Fur4   -MPDNLSLHLSGSSKRL-NS-RQLMESSNETFAPNNVDLEKEYKSSQSNITTEVY-E-AS
Dal4   MANDALSAIFSNPSRKGVQPSTSIVSY--TNNEDDIIDVENGKFNKNKNINTNVYVD-NS
Fui1   -MPVS-DSGFDNSSKTMKDDTIPTEDYEEITKESEMGDATK----ITSKIDANVIEKKDT

PucI   ----------------------------------------MKLKESQ----QQSNRLSN  15
FurA   ----------MSAIKRWIK----K-----------------LEVESDPGLTNTQLMLTN
FurD   ---------------------------------MRFGRFHLRVEQSRSAFASGNARWTN
FurE   ---------------------------------MGL-RERLQVKQGDASLA-TEAVASN
Fur4   SFEEKVSSEKPQYSSFWKKIYYEYVV-----------------VDKSILGVSILDSFMYN
Dal4   SIEESEVVPLPETKSIWSKIYYDFIV-----------------LDKTTLNVSLKESFLYN
Fui1   DSENNITIAQDDEKVSWLQRVVEFFEVKNDSTDLADHKPENPIRTFKDLQESLRSTYLYN

PucI   EDLV--PLGQEKRTWKAMNFASIWMGCIHNIPTYATVGGLIAIGLSPWQVLAIIITASLI  73
FurA   HDLRP--VEPDRRQWRWYNFIFFWIADSLNIG--------------------------
FurD   LDLDP--VPRAGRVWGPLSFISYWISDAFNAATWQFASSIIAVGLSWRESLGIVALSFFI
FurE   KDLDPIPLDSPKRTWRWPSLLGFWVAEAFSISMYQVTSTSVSKGLSAPMAIAAVVVGHIL
Fur4   QDLKP--VEKERRVWSWYNYCYFWLAECFNINTWQIAATGLQLGLNWWQCWITIWIGYGF
Dal4   RDLKP--VEEERRCWSWFNYLYFWLADCFNINTWQIAGTGLQLGLNWWQCWLTVWIGYTF
Fui1   TDLRP--VEAKRRTWTWKQYIFFWISGSFNVNTWQISATGLQLGLNWWQTWICIWVGYTF

PucI   LFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMWLGIQTFAGSTALNI  133
FurA   ----------------------------------------------------YIGGQCITL
FurD   ISFVIAANGAVGSIYHIPFPVIARASWGFWGSYIAIISRVILAIFWFAIQNVNGANAVKA
FurE   VCIPAMLDGYVGAIFGINFPVYTRASFGMKGSYFAVFVRGIVAIIWFGTQTYQAGQCVST
Fur4   VGAFVVLASRVGSAYHLSFPISSRASFGIFFSLWPVINRVVMAIVWYSVQAYIAATPVSL
Dal4   AGIFVVLNSRFGSAYHLSFPITVRASFGIFFSMWPIINRVVMAIVWYAVQAWLGATPVAL
Fui1   VAFFLILGSKVGNNYHISFPISSRVSFGIYFSIWIVINRVVMACVWNSTLAYIGSQCVQL

PucI   LLLNMWPGWGEIGGEWNI----LGIHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGP  189
FurA   MIRAIWPSYESLPNGIPE---SSGVDTKNFLSFFLFWLLSLPALWFPVHQIRHLFTVKSI
FurD   MISAIWPSFLSMKNTIPQ---DQGIETNTMIAYMIFWIVQMPFLCIHPNKVRWLFATKSV
FurE   MLSAIWPSFNHFPNHLPS---SGPITSAELLCFFLAIILQAPLLWLKVSKLRYLFIVKTC
Fur4   MLKSIFGKD--LQDKIPDHFGSPNATTYEFMCFFIFWAASLPFLLVPPHKIRHLFTVKAV
Dal4   MLKSIFGKN—LEDRIPNHFGSPNSTTFEFMCFFIFWVVSIPFVLVAPHKIRHLFTVKAA
Fui1   MLKAIFGTN--LNTRIKDTIKNPNLTNFEFMCFMVFWVACLPFLWFPPDKLRHIFALKSA
```

```
PucI   LVYLVFGGMVWWAVDIAGG---LGPIYSQPGKFHTFSETFWPFAAGVTGIIGIWATLILN  246
FurA   YSPIAAIAFFAWAISRANG---LGPIVHQSHT-VHGSTLAWAVVKALMSCLGNFAALIMN
FurD   LVPAAWIAILIWAFVA-EG---KGALFEQRAT-VSGSQYSWVWLASMTSVLGNYATLSVN
FurE   IMPIFGIVLFAWAVKAANG---FGPVFSKPSKITDGTPVAVVFLQCVTSAIGPKATLALN
Fur4   LVPFASFGFLIWAIRRAHGRIALGSLTDVQPH---GSAFSWAFLRSLMGCMANFSTMVIN
Dal4   LIPFAAFGFLIWALKKSHGKIELGTLNDYSPH---GSEFSWIFVRSLMACVANFAALIIN
Fui1   ITPFAAFGFLIWTLCKAKGHLALGSLNDNGGA-ISKTVLAWSVIRAIMSALDNFSTLILN

PucI   IPDFTRFAETQKEQIKGQFYGLPGTFALFAFASITVTSGSQVAF--GEPIWDVVDILARF  304
FurA   DPDFSRFARKPKDALWAQLLTIPIGFGITSFIGIIASSSSAVIFG-GDAIWNPLDLLGRF
FurD   QSDFSRYSRVSAKWQLLYIPLLPVIFTFISFIGIAASSAGWTRYNTPSIPWDPIELISHW
FurE   MPDFTRYAKTPREVFWTQAVGLVVLVSLCGVLGATVSSASEVIY--GVQTWNPLEVAVLW
Fur4   APDFSRFSKNPNSALWSQLVCIPFLFSITCLIGILVTAAGYEIYG--INYWSPLDVLEKF
Dal4   APDFGRFAKNPQASLWPQLVAIPLFFAITCLIGIIVTAAGYHLYG--VNYWSPLDVLGQF
Fui1   APDFTRFGKTYKSSVYSQLIALPVCYAIISLIGILSVSAAYTLYG--VNYWSPLDILNRY

PucI   D-------NPYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYINFKRGSFITALL  357
FurA   LE-GASSAERFGVFIIALGFALAQLGTNISANSVSAGTDMTALLPRYITIRRGSYICAAI
FurD   D-------SRAARFFGAFSFALASLGVNISANSISAANDLMALFPTYVDLRRGQIICGVI
FurE   N-------NRAAQFFAAFCWCLAAIGTNISANSVSFSNDLALWFPKYVDTRRGAYICALL
Fur4   LQTTYNKGTRAGVFLISFVFAVAQLGTNISANSLSCGTDMSAIFPKFINIKRGSLFCAAM
Dal4   LETTYTRGTRAGVFLISFVFALAQLGTNISANSLACGADMTALFPRYINIRRGSLFCVAM
Fui1   LD-NYTSGNRAGVFLISFIFAFDQLGANLSGNSIPAGTDLTALLPKFINIRRGSYICALI

PucI   ALFTVPWKLMESATSVYAFLGLIGGMLGPVAGVMMADYFIIRKRELSVDDLYSETGR-YV  416
FurA   GLAMCPWNLVSDSNQFTTYLSAYSIFLSAIAGVMICDYYVVRKGYLIVKDLYSGEKDSAY
FurD   SWALVPWKILESASNFLNFMSAYAIFLGPIAAIMLWDFWLIKNRKYDTVALYQPDTP-IY
FurE   SILSMPWYIQNSAASFSSFLGGYSLFLGAIAGVIVVDYWVCRGRRLRLRSLYEAHGT-HY
Fur4   ALCICPWNLMATSSKFTMALSAYAIFLSSIAGVVCSDYFVVRRGYIKLTHIYSHQKGSFY
Dal4   ALCICPWNLMASSSKFTSALGAYAIFLSSIAGVICADYFVVRRGYVKLTHLFLAQKGSFY
Fui1   SLAICPWDLLSSSSKFTTALAAYAVFLSAIAGVISADYFIVRKGYVNIFHCYTDKPGSYY

PucI   YW---KGYNYRAFAATMLGALISLIGMYV----------PVLKSLYDISWFVGVLISFLF  463
FurA   RF--NYGFSWQAYASYLSGLLINIVGFAGAVGR---DVPVGAQYIYNVNYLSGFIVSFVM
FurD   RF-NAWLVNWRAVVAFLVGVIPSLPGLSNSVNSR--IQVGVGIHPYQFGWLLGFVGTSLV
FurE   FT---KGVNIRAMISFVCGIAPNLPGLAAVTGQD--GVPKGANYLYSCSWLVSIVVSGMV
Fur4   MYGNRFGINWRALAAYLCGVAPCLPGFIAEVGAPAIKVSDGAMKLYYLSYWVGYGLSFSS
Dal4   MFGNKFGANWRAFVAYICGIAPNLPGFIGDVGAPKITVSEGAMRLYYLGYPVGFFISAVI
Fui1   MY-NKYGTNWRAVVAYIFGIAPNFAGFLGSVGV---SVPIGAMKVYYLNYFVGYLLAALS

PucI   YIVLMRVHPPASLAIETVEHA--------QVR--------QA-E----------------  490
FurA   YFIITRLCPIAATSD-------TWNEVNTDLE-LDTEGHD-IDAEDIHTGKPIGFETSEP
FurD   YIALSYGFPVREALIERAVLSDEVYEGR-EVE---GEGVEEG-REELGE----------S
FurE   YYLLFFVWPFDVE--EKVIVLEGMEEGD-RVV--------RV-EE-AVV----------Q
Fur4   YTALCYFFPVPGCPVNNIIKDKGWFQRWANVDDFEEEWKDTIERDDLVDDNISVYEHEHE
Dal4   YLILCYFFPVPGTPVTNFLTEKGWFQRWAYVEDFEQDWKNELRRDDLCDDTVSIYDGTEE
Fui1   YCILVYFYPIKGIPGDAKITDRKWLEEWVEVEEFGTEREAFEEYGGVST-------G-YE

PucI   --------------
FurA   REDYKGAKAGSASV
FurD   KREGVGKEKGFAVY
FurE   KKEAVSA-------
Fur4   KTFI----------
Dal4   KIVY----------
Fui1   KIRYI---------
```

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).

```
PucI   ------------------------------------------------------------
FcyB   --MAGA---FDFDLEKNPPVVQSTADNSSDGAVPGETFTY----------G---DSTYAK
Fcy2   MLEEGNNVYEIQDLEKRSPVIGSSLENEKKVA-ASETFTATSEDDQQYIVESSEATKLSW
Thi7   ----------------------------------------------------MSFGSKVSR
Tpn1   --MNRDNMDTTKRKEDHTKHTTDVIEFYEEGTAASSLNIATEKANSSPSILRRIINRAAW
Nrt1   ----------------------------------------------------MSFSSIVSK

PucI   ----MK-LKESQQQSNRLSNEDLVPLGQEK---RTWKAMNFASIWMGCIHNIPTYAT-VG  51
FcyB   IQRLAAELN--------IEQRGIERVPAAEQ--TDTSVFNIGSMWLAANMVVSSFAIGVL
Fcy2   FHKFFASLN--------AETKGVEPVTEDEK--TDDSILNAASMWFSANMVIASYALGAL
Thi7   ALRFLEIPVKDRASVSFLKNPDLQPIKSAN---QTWGFWSNFAYWGVMSFSVGTWMS-AS
Tpn1   LSKKVDAMG--------VESTGIQRISPYERGTSKKQFLHVAGLWLSATGGLSSMSSFLL
Nrt1   FLRYLEIPAKNRTAVNFLRNPDLQPIKSAN---QTWGFWSNLAYWGAVSFTAGTWMS-GS

PucI   GLIAIGLSPWQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALL  111
FcyB   GKSVYSLGFVDAILTVLFFNLLGIMTVCFFSCFGP-FGLRQMVFSRLWFGWYVTKGFAVL
Fcy2   GPMVFGLNFGQSVLVIIFFNIMGLIFVAFFSVFGAELGLRQMILSRYLVGNVTARIFSLI
Thi7   SALGVGLSYPETIGTFIVGDVLTIIFTLANSCPGYDWKVGFTLAQRFVFGIYGSAFGIII
Tpn1   GPLLFGLSFRESVASSLISVTIGCLIAAYCSIMGPQSGCRQMVTARYLFGWWFVKLVALA
Nrt1   AALSVGLSYPETIVSFLLGNVLTIIFTMANSYPGYDWKIGFTLAQRFVFGIYGSAFGIII

PucI   RAFTAIMWLGIQTFAGSTALNILLLNMWPG-WGEIGGEW-NILGIHLSGLLSFVFFWAIH  169
FcyB   NILACLGWSAANAIVGAQMLHAVNSD--------------------VPGFAAILIISICT
Fcy2   NVIACVGWGIVNTSVSAQLLNMVNEGS-----GHV-----------CPIWAGCLIIIGGT
Thi7   RILMSIVNYGSNAWVGGLCINMILD-SWSHHYLHLPNTLSSKVAMTTKELIGFIIFHVLT
Tpn1   SIIGVMGWSVVNSVVGGEMLAAISND--------K-----------VPLWVGIVIVTVCS
Nrt1   RILMSIVNYGSNAWLGGLSINMILD-SWSHHYLHLPNTLSPSVAMTTKQLVGFIIFHVLT

PucI   LLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAGGLGPIYSQPGKFHTFSETFWPF  229
FcyB   LLVTFAGYKVVHLYEYWSWIPTFIVFMIILGTFAHSGDFQNIPM------GVGTSEMGSV
Fcy2   VLVTFFGYSVIHAYEKWSWVPNFAVFLVIIAQLSRSGKFKGGEW------VGGATTAGSV
Thi7   AFCYLMKPYHMNYILIWSCVATFFSMLGMVIYLAKQAHGVGELFTSTKSTATGSTKAWAW
Tpn1   FLVAIFGIKQVIKVETYLSVPVLTAFLLLYISSSDKYSFVNAYVS--KGNLDSSTRKGNW
Nrt1   ALCYFMKPYHMNYLLIWSCVATCFAMLGIVIYLTKNAHGVGELFTSTKSTVTGSKRAWAW

PucI   AAGVTGIIGIWATLILNIPDFTRFAE---TQKEQIKGQFYGLPGTFALFAFASITVTSGS  286
FcyB   LSFGSAVYGFATGWTSYAADYTVYQPANRSKRKIFLSTWLGLIVPLLFVEMLGVAVMTAT
Fcy2   LSFGSSIFGFAAGWTTYAADYTVYMPKSTNKYKIFFSLVAGLAFPLFFTMILGAASAM-A
Thi7   VYMISYWFGSVSPGSTNQSDYSRFGS---SNWAIWAGTICALLIPTTLIPVFGVIGASTC
Tpn1   MSFFSLCYSITATWGSITADYYILFPEDTPYIQIFCLTFFGTFLPTCFVGILGLLLAS-V
Nrt1   VYMISYWFGSISPGSTNQSDYSRFGS---SNLAIWTGSVCALLIPATLVPIFGVISASTC
```

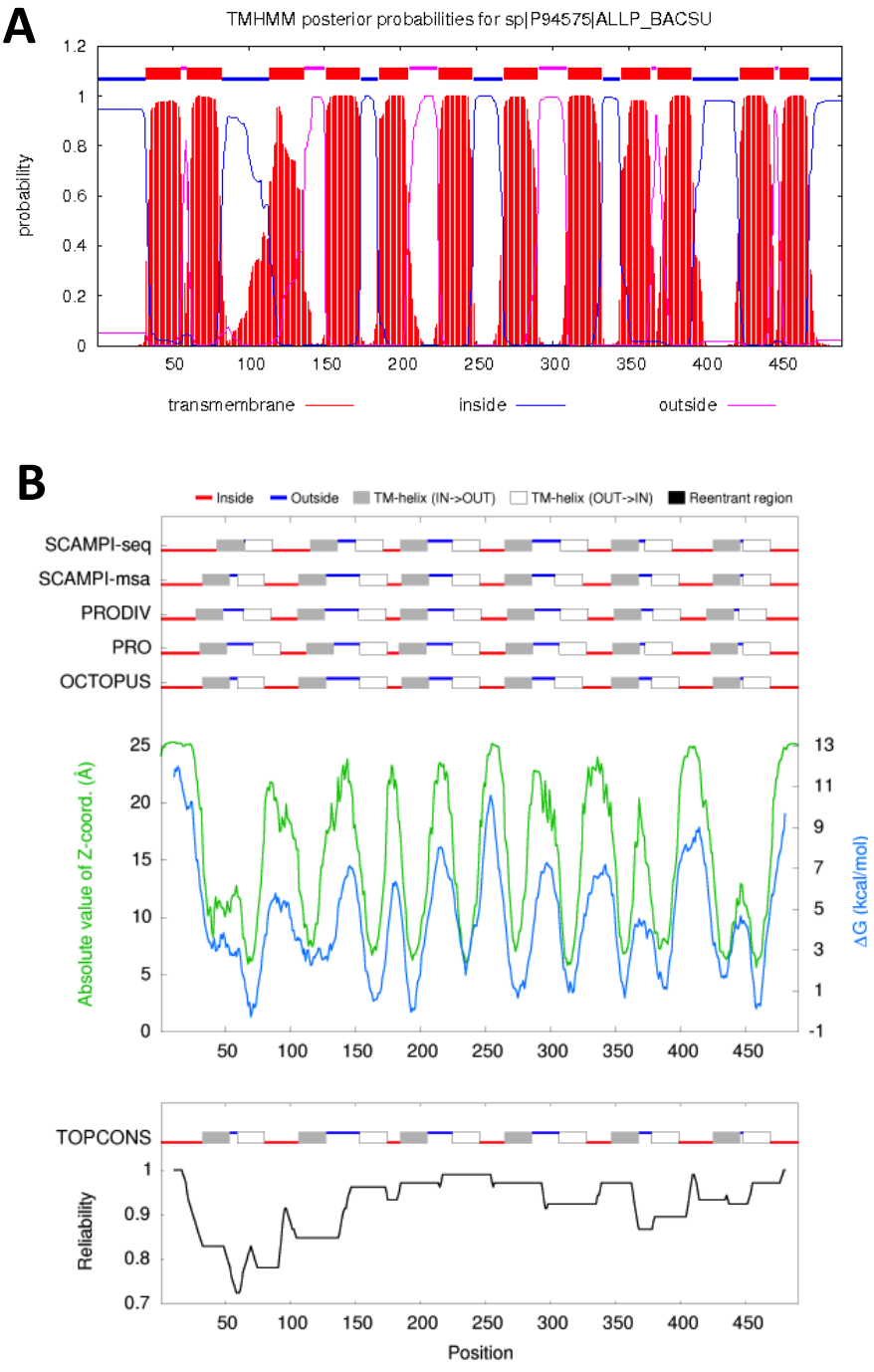
```
PucI   QVAFGEPIWDVVD--------ILARFDNPYVI-------VLSVITLCIATISVNVAANIV  331
FcyB   DIK--GSKYDVGYATSGNGGLIAAV-LQ--PL---GGFGDFCLVILALSIVANNCPNFY-
Fcy2   ALN--DPTWKAYYDKNAMGGVIYAI-LVPNSL---NGFGQFCCVLLALSTIANNIPNMY-
Thi7   DKLYGEQYWMPMD--------IFNHWLTTNYSAGARAGAFFCGLSFVLSQMSYTISNCGF
Tpn1   AMS--YKPWSVEYDSHGMGGLLWAG-FQ--RW---NGFGKFCVVVLVFSLVSNNIINTY-
Nrt1   DKLYGKQFWMPMD--------IFDYWLTNNYSAGARAGAFFCGLCFTMSQMSSTISNCGF

PucI   SPAYDI------ANALPKYINFKRGSFITALLALFTVPWKLMESATSVYAFLGLIGGMLG  385
FcyB   SVALTVQVLSRYAQRVPRFIWT----LFGTGVSIAIAIPGYSHFETVLENFMNFIAYWLA
Fcy2   TVALSAQALWAPLAKIPRVVWT----MAGNAATLGISIPATYYFDGFMENFMDSIGYYLA
Thi7   ASGMDL------AGLLPKYVDIKRGALFAACVSWACLPWNFYNSSSTFLTVMSSFGVVMT
Tpn1   SAAFSIQLSSVFCAKIPRWFWS----IVCTIICLVCALIGRNHFSTILGNFLPMIGYWIS
Nrt1   ATGMDM------AGLLPKYVDIKRGALFCACISWACLPWNFYNSSSTFLTVMSSFGVVMT

PucI   PVAGVMMADYFIIRKRELS----------------------------VDDLYSETG-RY  415
FcyB   IYSAIAIMDHFVFKRGFS----------------------------------------
Fcy2   IYIAISCSEHFFYRRSFS----------------------------------------
Thi7   PIISVMICDNFLIRKRQYS----------------------------ITNAFILKG-EY
Tpn1   MYFILLFEENLVFRRFFLHLYTKEFPTVTGEINGPELVGSSKEVEKDAVTNIHLLKRKHK
Nrt1   PIIAVMICDNFLIRKRQYS----------------------------ITNAFILKG-EY

PucI   VYWKGYNYRAFA-------------AT---MLGALISLIGMYVP--VLKSLYDI------  451
FcyB   ----GYVVENFDKREKLPVGIAATIAFGFGVAGMITGMSQPWYVGPIARH--AAGGDVGF
Fcy2   ----AYNIDDWDNWEHLPIGIAGTAALIVGAFGVALGMCQTYWVGEIGRLIGKYGGDIGF
Thi7   YFTKGVNWRAIV-------------AW---VCGMTPGLPGIAWE--VNNDYFHNTGIVNF
Tpn1   VTKHRYNWDKWEDYEVLTHGYAATFAFIVGVAGVVVGMAQAYWIGPIAAKFGEYGGDVAM
Nrt1   YFTKGVNWRAIV-------------AW---VCGMAPGLPGIAWE--VNNNYFHDSGIVKF

PucI   ---SWFVGVLISFLFYIVLMRVHPPASLAIETVE--------HAQVRQAE----------  490
FcyB   ELGFAFA-----AFSYLC-LR----------PFEIKFFGR--------------------
Fcy2   ELGASWA-----FIIYNI-LR----------PLELKYFGR--------------------
Thi7   FYGDSFFSFLISFFVYWGLCLLFPFK-ITVKHDDKDYYGAFTDEEARKKGMVPYSEISEE
Tpn1   WLSMAFS-----GVVYPP-CR----------YLELRKFGR--------------------
Nrt1   FYGDSFFSFLISFFVYWGLCVFFPFK-ITVRHDDKDYYGAFTDEEARKKGMIPYSEISEE

PucI   ------------------------------------------------------------
FcyB   ------------------------------------------------------------
Fcy2   ------------------------------------------------------------
Thi7   EIRAYTLGEGYTTGHEYRPEGSDDEIPELVKTSSENTNEFEIVHHKNNEKQSSTASEKAA
Tpn1   ------------------------------------------------------------
Nrt1   EIRAYTLGECYTTGHEYKPESSDNESPELIKTSSENTNVFEIVHQKDDEKHSFSTTQQVV
```

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).

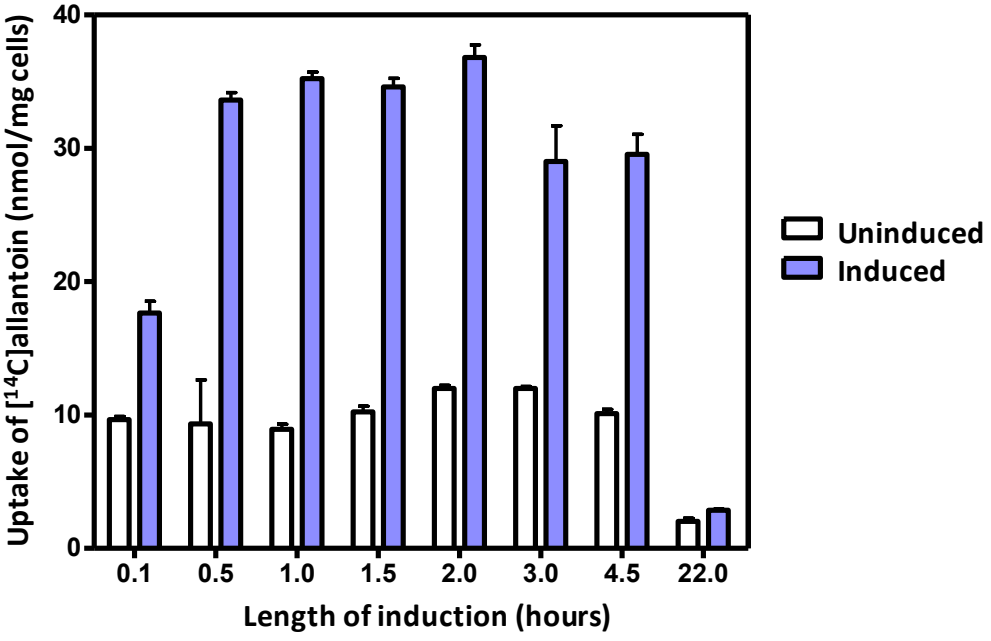
```
PucI    ------------------------------------------------------------
AtNCS1  MVSNCLSLSLHLNLHPHKHNRHSLSSLRSRTKAKLYQHVSFTDSSHKSSYTSCVSTFDIQ
CrNCS1  ------------------------------------------------------------
ZMNCS1  -----MAMS--MAM-----------------S--------------------KVFTSR
SvNCS1  -----MAMS--MAM-----------------S--------------------KAITAR

PucI    ----------------------------------------------------MKLKESQQQS  10
AtNCS1  RKSSKHYEL------GKHSFSPILPGDNLVLSRSGVIRPRLSAMTGSEINDHGYDESQFD
CrNCS1  -------------------------MGMFSD----------------PITARPPTNPD
ZMNCS1  HSEHLHHRLVAASSQAAAPRLPLLPRSPGLAAVTVAYRPRLRP-------------ASPR
SvNCS1  HATHLQHRLVASSSQ-AAPRLPLLPRRPSLALTVASPPRRLLP-------------ASPR

PucI    NRLSNEDLVPLGQEKRTWKAMNFASIWMGCIHNIPTYATVGGLIAIGLSPWQVLAIIITA  70
AtNCS1  PSLTNDDLKPTTPSQRTFSWLDMSSLWIGLVVGVPTYYLAGSLVDLGMAWWQGIATVVTA
CrNCS1  PSLINEDFSPTTQDKRTFDTTDYATFWITLVISITTYYLAASLVDLGMSWWQGILTVFFG
ZMNCS1  STSSESDLSPTPPSERTMTAWDLASLWVGLVVGVPSYYLAGSLVDLGMSALQGVATVAFA
SvNCS1  SSSSESDLAPTPPSERTMTAWDLASLWVGLVVGVPSYYLAGSLVDLGMSALQGVATVAFA

PucI    SLILFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMWLGIQTFAGSTA  130
AtNCS1  NLILLVPLVLTAQPGTLYGISFPVLARSSFGIRGAHIPTLLRALVGCGWYGIETWIGGEA
CrNCS1  NLITLLPMVLNAHPGTKYGVPFPVLARASFGIQGANLPSLSRAIVACGWFGIQTWIGGSS
ZMNCS1  NLIVLVTLVLTAAPAVTHGLPFPVLARAAFGVRGAHVPAVIRALIGCGWFGIESWIGGRA
SvNCS1  NLIVLVTLVLTAAPAVTHGLPFPVLARAAFGVRGAHVPAVIRALVGCGWFGIESWIGGRA

PucI    LNILLLNMWPGWGEIGGEWNILGIHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGPL  190
AtNCS1  IFLLLPGHIKKS-ALSHTLPWLGTSPLEFSCFIVFWLAQLCIVWRGMDGIRKLEKYSAPI
CrNCS1  IFQMLMAVTGG-AVAAAPIAWLGISLPELLCFLGFWAAQVWIVVRGMESIRILEKYSAPI
ZMNCS1  IFLLLPSRLKSYQPLLAPVPGLGVAPLEFACFLAFWAAQLGVIMHGMEGIRKLEKLSAPV
SvNCS1  IFLLLPSRLKSYQPLLAPVPGLGAAPLEFACFLAFWAAQLGVIMRGMEGIRKLEKFAAPV

PucI    VYLVFGGMVWWAVDIAGGLGPIYSQPGKFHT---FSETFWP-FAAGVTGIIGIWATLILN  246
AtNCS1  LISLTSCLLAWSYLKAGGFGHMLSLSSKL-----TSAQFWTLFFPSLTANISFWATLALN
CrNCS1  LIGLSLALMGWAVTTAGGFGPMLSTPSQFGVGMPKEGQFWSVFWPAVTANVGYWATLSLN
ZMNCS1  LIVLTSALLAWAYTSAGGFGRILSLPPRL-----TGAEFWKVFFPSLTANISFWATVAIN
SvNCS1  LFVLTSALLAWAYTSAGGFGRILSLPPRL-----TGAEFRKVFFPSLTANISFWATVAIN

PucI    IPDFTRFAETQKEQIKGQFYGLPGTFALFAFASITVTSGSQVAFGEPIWDVVDILARFDN  306
AtNCS1  IPDFSRFAKSQTDQIIGQ-VGLPVFMGLFTFVGVAVTSSTSIIFGRVISNPIELLGQIGG
CrNCS1  IPDFTRYAKSQKDQVMGQAIGLPLFMALFTFLGLAVTSATVVIYGEAIIDPVQLLGRMEG
ZMNCS1  IPDFARYARSQADQVLGQ-AGLPVFMGMFTFAGLAITSATEAIFGHVISDPIELLGRIGG
SvNCS1  IPDFARYARSQADQVLGQ-AGLPVFMGMFTFAGLAITSATEAIFGHVVSDPIELLGRIGG
```

```
PucI     PYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYINFKRGSFITALLALFTVPWKL  366
AtNCS1   LATTLLAIVGISLATLTTNIAANVVAPANALVNLNPKFFTFGRGAFLTAVLGIVFQPWRL
CrNCS1   LVPICISLFGLMWATLTTNIAANVVAPANAFVNCAPKWISFEAGGILTAVLGLLMCPWNL
ZMNCS1   PATTFLAIFGIGLATITTNIAANVVAPANALVSMSPRRFTFAKGAFVTALLGIAFQPWRL
SvNCS1   PVTTFLAIFGIGLATITTNIAANVVAPANALVSMSPRRFTFAKGALVTALLGIAFQPWRL

PucI     MESATS-VYAFLGLIGGMLGPVAGVMMADYFIIRKRELSVDDLYSETGRYVYW--KGYNY  423
AtNCS1   LKSSESFVYTWLIGYSALLGPIGGIILVDYYLIKKMKLNIGDLYSLSPSGEYYFSKGYNV
CrNCS1   VSSTHGFVNTWLIGYSALLGPVIGIVMSDYFIVRQRQLDIDSLYSKGDKSIYWYKGGWNP
ZMNCS1   LSSSESFVYTWLLGYSALMGPIGGVVLADHYIVRRTALDVDALYSEDSGSPYYFQGGFNV
SvNCS1   LSSSESFVYTWLLGYSALMGPIGGVILADHYIVRRTALDVDALYSEDSGSPYYFQNGFNV

PucI     RAFAATMLGALISLIG---------MYVPVLKSLYDISWFVGVLISFLFYIVLMRVHPPA  474
AtNCS1   AAVVALVAGIIPVVPGFLHKISALSKISNGFVVVYDNALFFSFIIAGFVYWIIMSRLGRK
CrNCS1   AALWAILIGVLPTLPGFLSTIGVLSGLPPIFGQLYDLAWFVGVAVSSVVYCLLMRGAPGA
ZMNCS1   ASMVAMAAGVAPIVPGFLHKVGVLPSVPSAFVTSYNNAWFVSFFVAGAVYCLLCNRRGKQ
SvNCS1   AAMAAMAAGVAPIVPGFLQKVGVLPSVSKAFATAYNNAWFVSFFVAGAVYCLLCGRGGVQ

PucI     SLAIETVEHAQVRQAE------------------  490
AtNCS1   QSSLSSSSHPLL----------------------
CrNCS1   YKS---GGDPSFNGVGGGLDTEPPGDMTIDTILVF
ZMNCS1   EREHYS-----------------------------
SvNCS1   AKQHSN-----------------------------
```

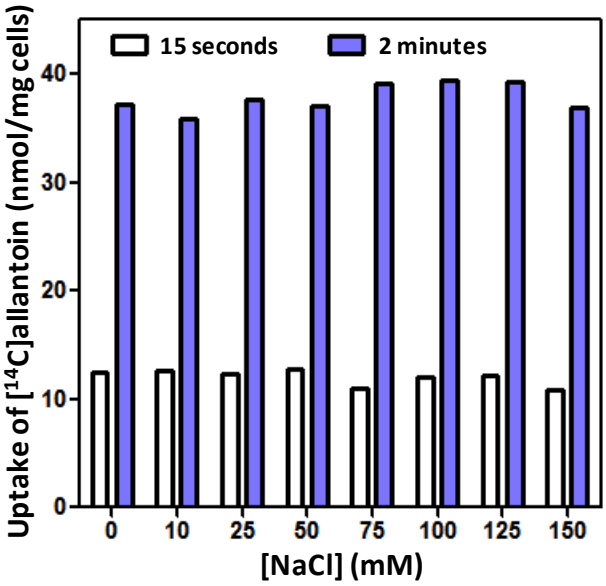
Supplementary Figure S9. Membrane topology analyses of the PucI protein from *Bacillus subtilis*. The amino acid sequence of the PucI protein (Bsu3645, P94575, ALLP_BACSU) from *Bacillus subtilis* (strain 168) was analysed using the online topology prediction tools TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), 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.



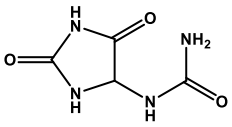
Supplementary Figure S10. Effect of induction time on PucI-mediated ^{14}C-allantoin uptake into whole cells. Uptake of ^{14}C-allantoin (50 μM) after 2 minutes into energised BL21(DE3) cells containing the construct pTTQ18-pucI(His_6) that were uninduced or induced with IPTG for a range of different 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.



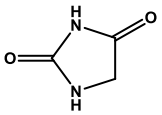
Supplementary Figure S11. Effect of sodium ions on PucI-mediated ^{14}C-allantoin uptake into energised whole cells. Uptake of ^{14}C-allantoin (50 μM) after 15 seconds and 2 minutes into energised BL21(DE3) cells containing the construct pTTQ18-pucI(His$_6$) that were induced with IPTG. Cells were cultured 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.



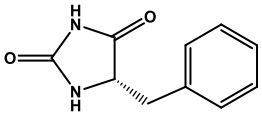
Supplementary Figure S12. Structures of compounds used as potential competitors of PucI-mediated ^{14}C-allantoin uptake into whole cells. The structures 1-20 are arranged in order of decreasing competitive effect on PucI-mediated ^{14}C allantoin uptake into whole cells as shown in Figure 6 of the main paper.



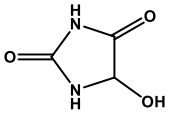
1. Allantoin



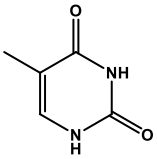
2. Hydantoin



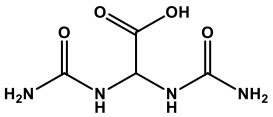
3. *L*-5-Benzylhydantoin



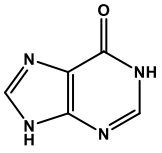
4. 5-Hydroxyhydantoin



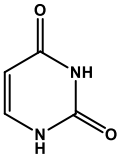
5. Thymine



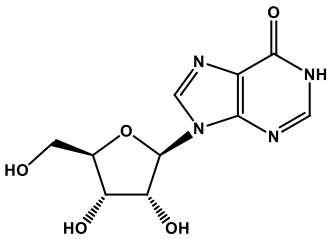
6. Allantoic acid



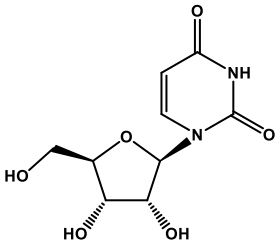
7. Hypoxanthine



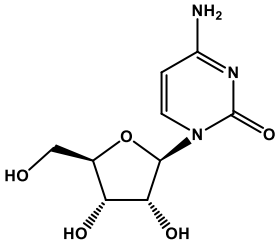
8. Uracil



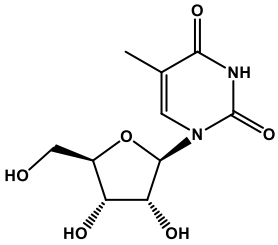
9. Inosine



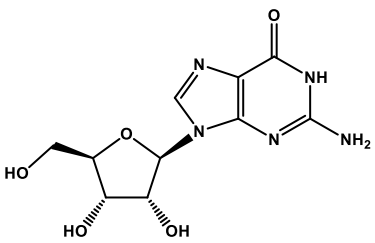
10. Uridine



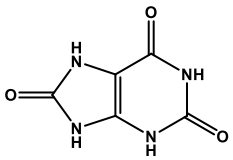
11. Cytidine



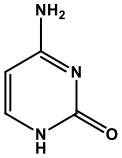
12. Thymidine



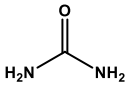
13. Guanosine



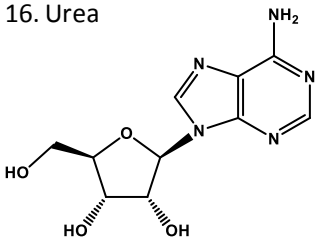
14. Xanthine



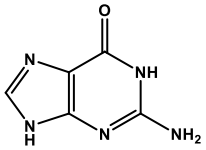
15. Cytosine



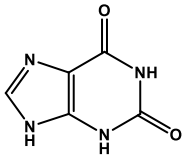
16. Urea



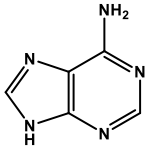
17. Adenosine



18. Guanine

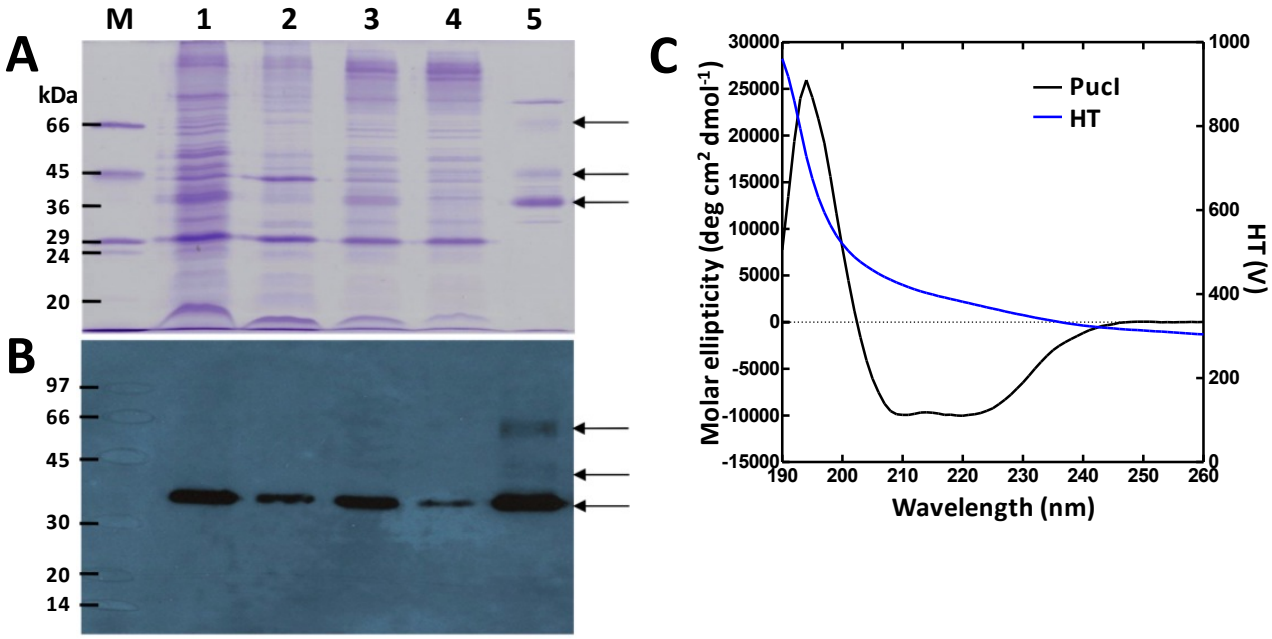


19. Uric acid

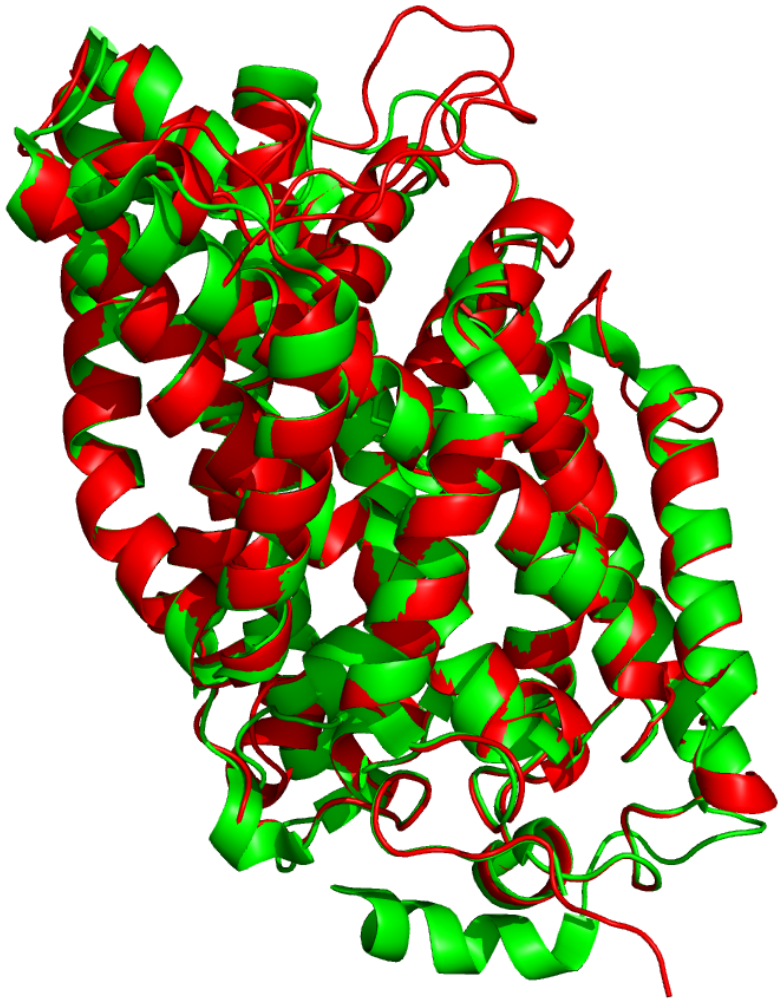


20. Adenine

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 PucI (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     PWQFAGVLNTF----LNLLASALGPLAGIMISDYFLVRRR  387
CodB     LWLY----NNF-VGWLTFLSAAIPPVGGVIIADYLMNRRR
FurA     PWNLVSDSNQF-TTYLSAYSIFLSAIAGVMICDYYVVRKG
FurD     PWKILESASNF-LNFMSAYAIFLGPIAAIMLWDFWLIKNR  413
FurE     PWYIQNSAASF-SSFLGGYSLFLGAIAGVIVVDYWVCRGR
Fur4     PWNLMATSSKF-TMALSAYAIFLSSIAGVVCSDYFVVRRG
Dal4     PWNLMASSSKF-TSALGAYAIFLSSIAGVICADYFVVRRG
Fui1     PWDLLSSSSKF-TTALAAYAVFLSAIAGVISADYFIVRKG
FcyB     -----SHFETVLENFMNFIAYWLAIYSAIAIMDHFVFKRG
Fcy2     -----YYFDGFMENFMDSIGYYLAIYIAISCSEHFFYRRS
Thi7     PWNFYNSSSTF-LTVMSSFGVVMTPIISVMICDNFLIRKR
Tpn1     -----NHFSTILGNFLPMIGYWISMYFILLFEENLVFRRF
Nrt1     PWNFYNSSSTF-LTVMSSFGVVMTPIIAVMICDNFLIRKR
AtNCS1   PWRLLKSSESFVYTWLIGYSALLGPIGGIILVDYYLIKKM
CrNCS1   PWNLVSSTHGFVNTWLIGYSALLGPVIGIVMSDYFIVRQR
ZmNCS1   PWRLLSSSESFVYTWLLGYSALMGPIGGVVLADHYIVRRT
SvNCS1   PWRLLSSSESFVYTWLLGYSALMGPIGGVILADHYIVRRT
```

References

Bernsel A, Viklund H, Hennerdal A, Elofsson A. 2009. TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 37:W465-W468.

Chung CT, Niemela SL, Miller RH. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci 86:2172-2175.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, pp. 571-607.

Henderson PJF, Macpherson AJ. 1986. Assay, genetics, proteins and reconstitution of proton-linked galactose, arabinose and xylose transport systems of *Escherichia coli*. Methods Enzymol 125:387-429.

Inoue H, Nojimab H, Okayama H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23-28.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 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 AA, 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.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Schaffner W, Weissmann C. 1973. A rapid, sensitive, and specific method for the determination of protein in dilute solution. Anal Biochem 56:502-514.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

Simmons KJ, Jackson SM, Brueckner F, Patching SG, Beckstein O, Ivanova E, Geng T, Weyand S, Drew D, Lanigan J, Sharples DJ, Sansom MS, Iwata S, Fishwick CW, Johnson AP, Cameron AD, Henderson PJ. 2014. Molecular mechanism of ligand recognition by membrane transport protein, Mhp1. EMBO J 33:1831-1844.

Stark MJR. 1987. Multicopy expression vectors carrying the lac repressor gene for regulated high-level expression of genes in *Escherichia coli*. Gene 51:255-267.

Ward A, O'Reilly J, Rutherford NG, Ferguson SM, Hoyle CK, Palmer SL, Clough JL, Venter H, Xie H, Litherland GJ, Martin GE, Wood JM, Roberts PE, Groves MA, Liang WJ, Steel A, McKeown BJ, Henderson PJ. 1999. Expression of prokaryotic membrane transport proteins in *Escherichia coli*. Biochem Soc Trans 27:893-899.

Ward A, Sanderson NM, O'Reilly J, Rutherford NG, Poolman B, Henderson PJF. 2000. The amplified expression, identification, purification, assay and properties of hexahistidine-tagged bacterial membrane transport proteins. In Baldwin SA (ed), Membrane transport - a practical approach. Oxford: Blackwell. pp. 141-166.

Witholt B, Boekhout M, Brock M, Kingma J, van Heerikhuizen H, de Leij L. 1976. An efficient and reproducible procedure for the formation of spheroplasts from variously grown *Escherichia coli*. Anal Biochem 74:160-170.