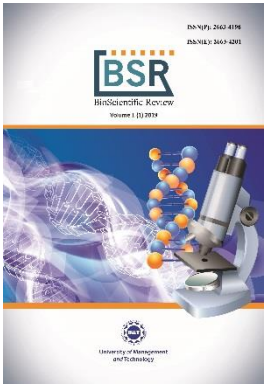

BioScientific Review (BSR)

Volume 3 Issue 4, December 2021

$ISSN_{(P)}$: 2663-4198 $ISSN_{(E)}$: 2663-4201

Journal DOI: https://doi.org/10.32350/BSR

Issue DOI: https://doi.org/10.32350/BSR.0304

Homepage: https://journals.umt.edu.pk/index.php/BSR

Journal QR Code:

Article: **Cloning, Amplified Expression and Bioinformatics Analysis of a Putative Nucleobase Cation Symporter-1 (NCS-1) Protein Obtained from *Rhodococcus erythropolis***

Author(s): Irshad Ahmad[1], Youri Lee[2], Nighat Nawaz[3], Rizwan Elahi[4], Israr Ali Khan[1], Muhammad Zahid Mustafa[5], Simon G. Patching[6]

Affiliation:
[1]Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan
[2]College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, China
[3]Department of Chemistry, Islamia College Peshawar, Peshawar, Pakistan
[4]Professional College of Medical Sciences, Peshawar, Pakistan
[5]Centre for Advanced Studies in Vaccinology and Biotechnology, University of Balochistan, Quetta, Pakistan
[6]School of Biomedical Sciences, University of Leeds, Leeds, LS2 9JT, UK

Article DOI: https://doi.org/10.32350/BSR.0304.05

Article QR:

Irshad Ahmad

Citation: Ahmad I, Lee Y, Nawaz N, et al. Cloning, amplified expression and bioinformatics analysis of a putative nucleobase cation symporter-1 (NCS-1) Protein Obtained from *Rhodococcus erythropolis*. *BioSci Rev.* 2021;3(4):54–71.

Indexing

Crossref

HEC

EuroPub

INTERNET ARCHIVE

HJRS HEC Journal Recognition System

J-Gate INDEXED

LOCKSS

Estd. 1990

A publication of the
Department of Life Sciences, School of Science
University of Management and Technology, Lahore, Pakistan

Cloning, Amplified Expression and Bioinformatics Analysis of a Putative Nucleobase Cation Symporter-1 (NCS-1) Protein Obtained from *Rhodococcus erythropolis*

Irshad Ahmad[1*], Youri Lee[2], Nighat Nawaz[3], Rizwan Elahi[4], Israr Ali Khan[1], Muhammad Zahid Mustafa[5], Simon G. Patching[6]

[1]Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan
[2]College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, China
[3]Department of Chemistry, Islamia College Peshawar, Peshawar, Pakistan
[4]Professional College of Medical Sciences, Peshawar, Pakistan
[5]Centre for Advanced Studies in Vaccinology and Biotechnology, University of Balochistan, Quetta, Pakistan
[6]School of Biomedical Sciences, University of Leeds, Leeds, LS2 9JT, UK

*Corresponding Author: irshadibms@kmu.edu.pk

Article Info

Received: September 16, 2021
Revised: November 19, 2021
Accepted: November 19, 2021

Keywords

bioinformatics analysis, gene cloning, membrane topology, NCS-1 family, protein expression, transport protein

Abstract

The *Rhodococcus erythropolis* gene *DYC18_RS18060* (1437 bp) putatively codes for a secondary transporter of the Nucleobase Cation Symporter-1 (NCS-1) protein family (478 amino acids). The *DYC18_RS18060* gene was successfully cloned from *R. erythropolis* genomic DNA with the addition of *Eco*RI and *Pst*I restriction sites at the 5′ and 3′ ends, respectively, using PCR technology. The amplified gene was introduced into IPTG-inducible plasmid pTTQ18, immediately upstream of the sequence coding for a His_6-tag. The construct was transformed into *Escherichia coli BL21*(DE3). Then, the amplified expression of the DYC18_RS18060-His_6 protein was achieved with detection through SDS-PAGE and western blotting. Computational methods predicted that DYC18_RS18060 has a molecular weight of 51.1 kDa and isoelectric point of 6.58. The protein was predicted to be hydrophobic in nature (aliphatic index 113.24, grand average of hydropathicity 0.728). It was also predicted to form twelve transmembrane spanning α-helices, with both N- and C-terminal ends at the cytoplasmic side of the membrane. Database sequence similarity searches and phylogenetic analysis suggested that the substrate of DYC18_RS18060 could be cytosine; however, this was uncertain based on the comparison of residues involved in substrate binding in experimentally characterised NCS-1 proteins. The current study lays the foundations for further structural and functional studies of DYC18_RS18060 and other NCS-1 proteins.

1. Introduction

The Nucleobase Cation Symporter-1 (NCS-1) family of secondary active transport proteins is widespread in bacteria, archaea, fungi and plants [1-9]. The principal function of NCS-1 proteins is the uptake of nucleobases, nucleosides, hydantoins and other similar compounds from the environment in salvage pathways. This requires a symport mechanism driven by a gradient of protons or sodium ions [6]. NCS-1 proteins typically comprise 419-635 amino acids and putatively form twelve transmembrane spanning α-helices [3, 5].

The structural organisation of NCS-1 proteins was determined by using high-resolution crystal structures of the sodium-driven 5-arylhydantoin transporter Mhp1 from *Microbacterium liquefaciens* [9-16]. Mhp1 is pivotal for explaining the alternating access mechanism of membrane transport and its ion-coupling [14, 17-21]. Moreover, it is used as a model to develop free energy calculations for protein conformational changes [22]. Whilst Mhp1 is the only NCS-1 protein with high-resolution structures, 27 other NCS-1 proteins (5 bacterial, 16 fungal, 6 plant) have been characterised experimentally [6, 9, 23-27]. NCS-1 proteins characterised from bacteria include an allantoin transporter from *Bacillus subtilis* [7] and cytosine transporters from both *Escherichia coli* [28] and *Vibrio parahaemolyticus* [27]. There is a considerable lack of information available about NCS-1 proteins determined by experimental studies. So, we have undertaken the current study of a bacterial NCS-1 protein.

A crucial step in the structural and functional characterisation of a membrane protein is overcoming the challenge of achieving amplified expression [29]. This is necessary to ensure that sufficient quantities of the protein can be made available for crystallisation trials and for applying various chemical, biochemical and biophysical techniques. In the current work, we cloned the *Rhodococcus erythropolis* gene *DYC18_RS18060*. It putatively codes for an NCS-1 transporter and we achieved amplified recombinant protein expression in *E. coli*. We also performed a bioinformatics analysis of the chemical and physical properties, predicted the structural and functional characteristics, as well as the evolutionary relationships of the DYC18_RS18060 protein. Whilst DYC18_RS18060 is not itself a drug target, bacterial NCS-1 proteins are close homologues of human LeuT-fold solute carrier transporters [30-33], which are drug targets in the treatment of diseases [34-36].

2. Methodology

2.1. Design of PCR Primers Used for Cloning of the *DYC18_RS18060* Gene

The sequence of the *R. erythropolis DYC18_RS18060* gene (1437 bp) was obtained from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/gene/61556622). Restriction sites in the desired gene were mapped using Webcutter 2 (https://users.unimi.it/camelot/tools/cut2.html). The mapping was done to check for the presence of any internal *Eco*RI or *Pst*I sites that would be cut by the enzymes intended to be used for gene cloning with plasmid pTTQ18 [37]. PCR primers for amplifying the *DYC18_RS18060* gene with an in-frame *Eco*RI site (GAATTC) at the 5′ end and a *Pst*I site (CTGCAG) at the 3′ end were designed through GeneLink

(https://www.genelink.com/). The properties and the quality of the designed primers were predicted using OligoAnalyzer 3.1 software. The forward primer 5′-CCGGAATTCGCATATGACTCACGATGG-3′ and the reverse primer 5′-AAAACTGCAGTCAGACGCGAGAGTCG-3′ were synthesised commercially (Thermo Fisher Scientific).

2.2. Gene Cloning and Amplification from Genomic DNA

Genomic DNA was extracted from a culture of *R. erythropolis* using a GenElute bacterial DNA kit (Sigma), according to the manufacturer's instructions. The amplification of the desired *DYC18_ RS18060* gene was carried out via PCR, using the primers described above on a Bio-rad thermocycler (Waltham, USA) in total reaction volumes of 50 µL. PCR samples contained the following components: 1 µL genomic DNA (50 ng/µL), 1.5 µL forward primer (10 µM), 1.5 µL reverse primer (10 µM), 2.5 µL pfu Turbo polymerase (2500 units/mL) (Agilent Technologies, UK), 5 µL pfu Turbo buffer (10x), 1 µL dNTPs (10 mM each), and 37.5 µL sterile water. The following PCR conditions were used: 1x cycle of 95°C for 3 minutes, then 30x cycles of 95°C for 30 seconds, 58°C for 30 seconds and 72°C for 2 minutes, then 1x cycle of 72°C for 2 minutes, and finally held at 4°C. The PCR was analysed through agarose gel electrophoresis to confirm successful amplification. The product obtained from the PCR was run on and extracted from an agarose gel and both the PCR product and plasmid pTTQ18 were cut using *Eco*RI and *Pst*I restriction enzymes (New England Biolabs). Then, the gene was ligated into pTTQ18 using T4 DNA ligase (New England Biolabs) and transformed into the Omnimax strain of *E. coli* (Thermo Fisher Scientific). The plasmid construct was analysed through agarose gel electrophoresis and subjected to automated DNA sequencing to confirm that the desired gene had been cloned without mutation and inserted into pTTQ18 in the correct orientation. Positive clones were transformed into BL21(DE3) *E. coli* cells (Invitrogen) for the optimisation of DYC18_RS18060-His_6 expression.

2.3. Recombinant Protein Expression

A clone of *E. coli* BL21(DE3) cells transformed with pTTQ18/DYC18_ RS18060-His_6 was streaked onto an LB-agar plate (1.5%) containing carbenicillin (Melford Laboratories, UK) (100 µg/mL) and incubated at 37°C, overnight. The expression of DYC18_RS18060-His_6 was tested from small-scale cultures grown in LB medium (50 mL) supplemented with carbenicillin (100 µg/mL). A single colony was used to inoculate the LB medium and the culture was incubated (37°C, 220 rpm) up to an A_{600} of 0.6. Induction was initiated by adding isopropyl-β-d-1-thiogalactopyranoside (IPTG) (Melford Laboratories, UK) (0.5 mM). Growth was continued for 2 hours before harvesting the cells by centrifugation (12000 x g, 4 °C, 10 minutes). Mixed (inner and outer) membranes were isolated from the cells using a water lysis procedure. The successful amplified expression of the DYC18_RS18060-His_6 protein was checked through sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting.

2.4. Large-Scale Cultures and Membrane Preparation

For large-scale membrane preparation, a total of 10 litres of cells in 2-litre flasks were grown to an A_{600} of 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 stored 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 disrupted by passing twice through a cell disrupter (Constant Systems) at 30 kpsi. Undisrupted cells and cell debris were removed by centrifugation at 12000 x g for 45 minutes at 4°C. The supernatant containing all (inner and outer) membranes was collected. The inner and outer membranes were separated by sucrose gradient ultracentrifugation and prepared as described in Ward et al. [38], followed by washing and resuspension in Tris buffer (20 mM, pH 7.5), dispensing into aliquots, freezing in liquid nitrogen and storage at -80°C.

2.5. SDS-PAGE and Western Blotting

SDS-PAGE comprised the use of 4% stacking gels and 15% resolving gels made from acrylamide (40%) and bisacrylamide (2%) solutions (BioRad Laboratories). The samples contained 10 µg protein and gels were stained with Coomassie Brilliant Blue R-250 (Thermo Fisher Scientific). For western blotting, the samples containing 5 µg protein were first separated by SDS-PAGE. They were then transferred from the gel to a Fluorotrans™ membrane (Pall BioSupport, UK) using a Trans-Blot semi-dry transfer cell (BioRad) operating at 18 volts for 35 minutes. This process involved the pre-soaking of four pieces of filter paper in 0.5x SDS-PAGE running buffer. Then membrane was layered on two of these soaked filter papers followed by the respective layering of the polyacrylamide gel and two more pieces of filter paper. Following transfer, the membrane was incubated with bovine serum albumin (3%) in TBST (20 mM Tris-HCl pH 7.6, 0.05% v/v Tween-20, 0.5M NaCl) for 3 hours at 4°C to block the non-specific binding sites. It was washed twice with TBST (20 mL) at room temperature for 10 minutes and then incubated for 1 hour with HisProbe-HRP antibody (QIAGEN Ltd) (10 mL) diluted to 1:5000 with TBST, followed by three washes with TBST (20 mL) for 10 minutes each. A 6-mL SuperSignal West Pico chemiluminescent solution was prepared by mixing 3 mL West Pico luminol/enhancer solution (Perbio Science, UK) (3 mL) and West Pico stable peroxide solution (Perbio Science, UK) (3 mL). Finally, the membrane was incubated with this solution for 3 minutes before being wrapped in acetate for exposure (Syngene G:Box).

2.6. Computational Methods

Gene and protein sequence information was obtained from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov) and from the UniProt KnowledgeBase (https://www.uniprot.org). Similar sequences were identified using the Basic Local Alignment Search Tool available at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or UniProt. The protein's chemical and physical parameters were calculated using the ExPASy tool ProtParam (https://web.expasy.org/protparam/) [39]. Putative transmembrane regions in the protein were identified using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [40] and TOPCONS

(https://topcons.cbr. su.se) [41]. Homology modelling was performed using the ExPASy SWISS-MODEL tool (https://swissmodel.expasy. org) [42]. The alignment of protein sequences was achieved using Clustal Omega at EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) [43], from which the nearest-neighbour phylogenetic results were extracted in Newick format and displayed using iTOL (Interactive Tree of Life, http://itol.embl.de/index.shtml) [44].

3. Results

3.1. Cloning and Amplification of the *DYC18_RS18060* Gene

PCR primers designed for cloning and amplifying the *DYC18_RS18060* gene from *R. erythropolis* with a His_6-tag were predicted to be free of dimers or other secondary structures. They were also predicted to have other ideal properties, including a melting temperature of ≥65 °C, a GC content of less than 40% and termination with a G or C base. The analysis of the PCR product through agarose gel electrophoresis showed that the *DYC18_RS18060* gene was successfully cloned and amplified. According to the restriction digestion analysis of the plasmid construct, the *DYC18_RS18060* gene was successfully ligated into pTTQ18 at the *Eco*RI and *Pst*I restriction sites (Figure 1). The gene insert ran on the agarose gel at a position consistent with a predicted length of 1437 bp, as given by the database entries for *DYC18_RS18060*. DNA sequencing confirmed that the *DYC18_RS18060* gene was cloned without mutation and inserted into pTTQ18 in the correct orientation.

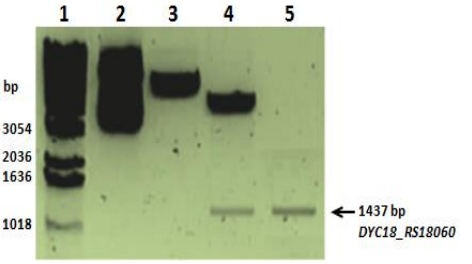

Figure 1. Agarose gel showing the restriction digestion analysis of the plasmid construct containing gene *DYC18_RS18060*

The product from the PCR for amplifying the *DYC18_RS18060* gene was run on and extracted from an agarose gel and digested by *Eco*RI and *Pst*I restriction enzymes. A 1.5% agarose gel was loaded with the following samples: (1) 1 kb DNA ladder size markers; (2) undigested plasmid pTTQ18; (3) *Eco*RI-digested plasmid pTTQ18; (4) *Eco*RI/*Pst*I-digested constructed plasmid pTTQ18/DYC18_RS18060-His_6; (5) *Eco*RI/*Pst*I-digested PCR product. The arrow indicates the band for the *DYC18_RS18060*-His_6 gene with a size of 1437 bp.

3.2. Detection of Amplified DYC18_RS18060-His_6 Expression

The constructed plasmid pTTQ18/DYC18_RS18060-His_6 was introduced into BL21(DE3) *E. coli* cells for expression studies. Recombinant DYC18_RS18060-His_6 expression was tested from 50-mL cultures of cells grown in LB medium and induced with IPTG. Both SDS-PAGE and western blotting detected an amplified protein band with an apparent size of ~37 kDa during membrane preparations from induced cultures (Figure 2).

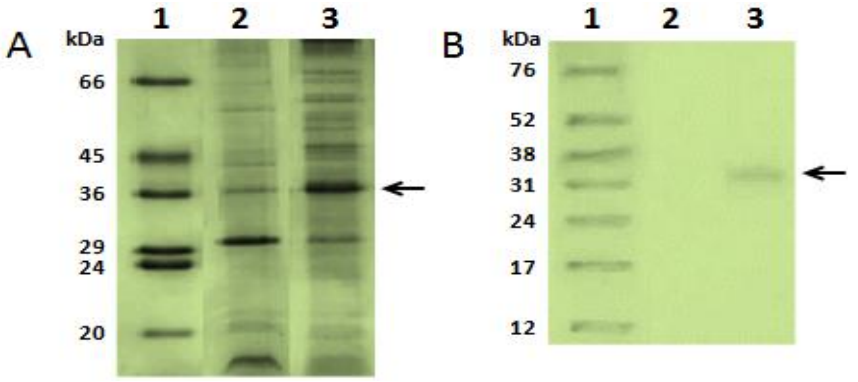

Figure 2. Test for the Amplified Expression of the DYC18_RS18060-His_6 Protein

Coomassie-stained SDS-PAGE (A) and western blot (B) analysis of the inner and outer membrane preparations from the cultures of induced (0.5 mM IPTG) BL21(DE3)/DYC18_RS18060-His_6 cells. Cells were grown in LB medium at 37°C with shaking at 220 rpm until the A_{600} reached 0.6, then they were induced with IPTG (0.5 mM) and grown for a further 3 hours. Cells were harvested by centrifugation (6000 x g, 15 min, 4°C), were disrupted using a cell disrupter and then the inner and outer membranes were separated by sucrose density gradient ultracentrifugation. The samples were loaded on the gel as follows: (1) molecular weight markers (kDa) (gel: SDS-7, Sigma-Aldrich; blot: Rainbow™, Amersham Biosciences) (10 μg), (2) outer membranes (10 μg), and (3) inner membranes (10 μg). The blot was probed using a RGS-His_6 antibody. The arrow indicates the position of the amplified DYC18_RS18060-His_6 protein migrating at a size of ~37 kDa.

3.3. Bioinformatics Analysis of DYC18_RS18060

Several databases and computational methods were used to obtain and analyse the chemical and physical properties, predicted structural and functional characteristics, as well as the evolutionary relationships of the DYC18_RS18060 protein.

The 478 amino acids in DYC18_RS18060 were predicted to have a molecular weight of 51087.74 Da and a theoretical pI of 6.58. The protein was found to have a high aliphatic index of 113.24 and a high grand average of hydropathicity (GRAVY), that is, 0.728. It was predicted to form twelve transmembrane-spanning α-helices with both the N- and C-terminal ends at the cytoplasmic side of the membrane, as illustrated by tools that predict the positions of transmembrane helices (Figure 3A) and by homology modelling based on the X-ray crystal structure of Mhp1 (Figure 3B), with which it shares 22.6% sequence identity. The protein was found to contain a slight excess of negatively charged residues (26x Asp/Glu) over positively charged residues (25x Arg/Lys). A large majority of the positively charged residues (18 out of 25) were located in the predicted loop regions at the cytoplasmic side of the membrane (Figure 3).

A

```
MTHDGPAEVILTPERRTIDVVPDAERHGTPRSQFTLWFGANMQITAIVDGALAVVFGADA  60
IWAIVGLLIGNIFGGAVMALHSAQGPRMGLPQMISSRAQFGVKGAVVPLVLVILMYLGFA 120
ATGTVLAGQAVNKILHIDSPTVGIVVFGLLTAFVAVTGYKLIHIVGRIATVVGIVGFSYL 180
AVRLFLEYDVASYVGIKGFDIVTFLLAISLGAGWQLTFGPYVADYSRYLPRSTSESTTFW 240
STFLGSVIGSQWSMTFGALVAACAGDAFLGNQVGFMGDLAGPAAIAFLIYFVILVGKLTV 300
NVLNAYGGFMSILTTVTAFNGQSRISSTARTLYILGFTAVSVLIAIAASADFLDNFKNFV 360
LVLLMVFTPWSAINLIDYYLISKERIDIPALYDVNGRYGAWNFTALACYAAGVLAQIPFL 420
AQKMYTGPVTDMLGGADISWIVGIVFTGLIYYPLAKRTSNPPSSMIYPDHTAMTDSRV   478
```

B

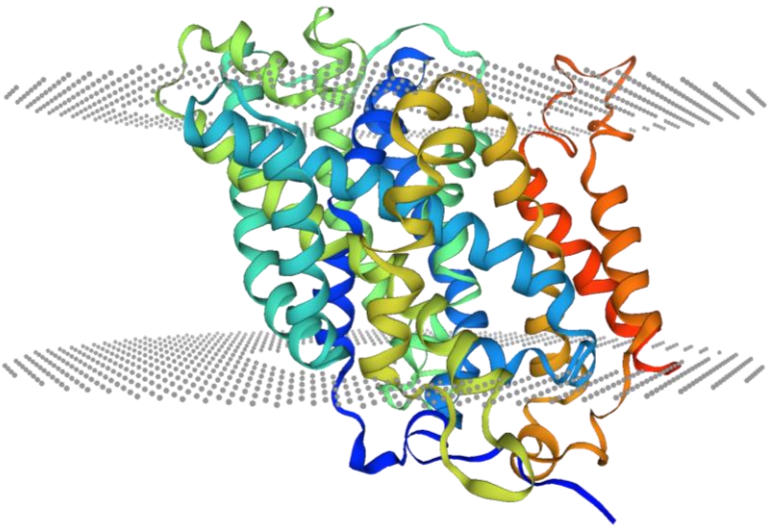

Figure 3. Predicted structure of *R. erythropolis* protein DYC18_RS18060

(A) Predicted transmembrane helices in DYC18_RS18060. The amino acid sequence of DYC18_RS18060 (478 residues) was analysed by the membrane topology prediction tools TMHMM and TOPCONS. The predicted positions of twelve transmembrane helices in the sequence of DYC18_RS18060 are highlighted (*grey*). Positively charged (Arg/Lys) (*red*) and negatively charged (Asp/Glu) (*blue*) residues are coloured. Positions corresponding to residues involved in substrate binding in Mhp1 based on sequence alignment between DYC18_RS18060 and Mhp1 (Figure 5 and Table 1) are also highlighted (*green*). (B) Three-dimensional homology model of DYC18_RS18060 based on the 2.85-Å X-ray crystal structure of Mhp1 (PDB 2JLN) [13], with which it shares 22.6% sequence identity, is generated using the ExPASy SWISS-MODEL tool . The model is coloured in rainbow effect with the N-terminus in blue and the C-terminus in red. The grey dots represent the predicted outer barriers of the membrane.

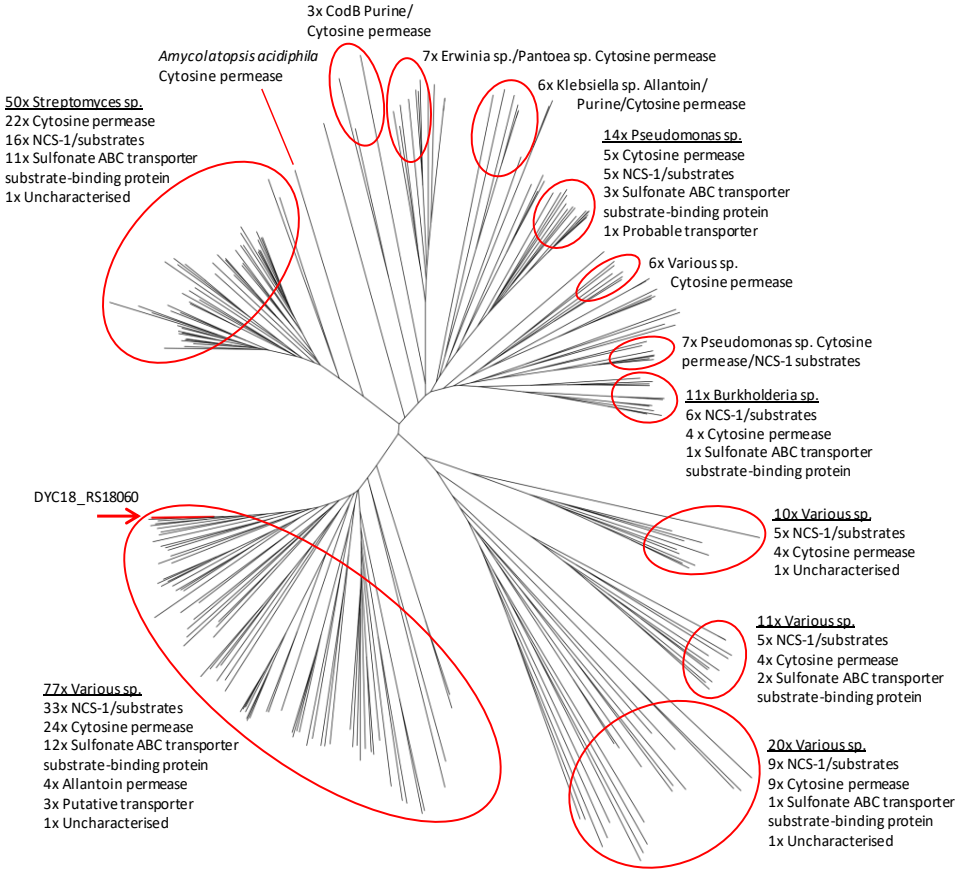

Figure 4. Closest evolutionary relationships of *R. erythropolis* protein DYC18_RS18060

The sequence of the DYC18_RS18060 protein was subjected to a BLAST search against proteins in the UniProt database. The sequences of the top 250 results were aligned using Clustal Omega. The nearest-neighbour phylogenetic results were extracted in Newick format and displayed as an unrooted phylogenetic tree using iTOL. The DYC18_RS18060 protein is indicated (*red arrow*). Some of the proteins are grouped (*red ellipses*) and some details are given about the host bacterial species and the putative function of the protein, as listed in the UniProt database.

The NCBI entry for the *DYC18_RS18060* gene (https://www.ncbi.nlm.nih.gov/gene/61556622) lists it as coding for a cytosine permease, as per the NCS-1 family. A BLASTP search of the NCBI database produced a top 100 results with sequence identities ranging from 100.0% to 79.9%. They were all listed as cytosine permeases, albeit all from *Rhodococcus* or *Norcadia* species. The NCBI entry for

DYC18_RS18060 also refers to an identical *R. erythropolis* protein in the UniProt KnowledgeBase, listed as an allantoin permease (https://www.uniprot.org/uniprot/A0A0E4AAD2). When the sequence of the protein DYC18_RS18060 was subjected to a BLAST search against all proteins in the UNIPROT database, the top 250 results with sequence identities of 89.7-39.7% were mostly (202 out of 250) predicted to be NCS-1-type proteins (Figure 4), with 99 of these listed as "Cytosine permease". Forty-nine were listed as "Nucleobase:cation symporter-1, NCS-1 family", 33 as "Permease for cytosine/purines, uracil, thiamine, allantoin", nine as "Allantoin permease", six as "Putative transporter", five as "NCS-1 nucleoside transporter family protein" and one as "Permease for cytosine allantoin". Interestingly, 38 were listed as "Sulfonate ABC transporter substrate-binding protein".

When the sequence of DYC18_RS18060 was aligned with those of experimentally characterised NCS-1 transporters, it was found to share only 25.0%, 24.4%, and 22.6% overall sequence identities with CodB (cytosine), PucI (allantoin) and Mhp1 (hydantoins), respectively (Figure 5). From the sequence alignment between DYC18_RS18060 and crystographically defined Mhp1, it was found that four out of the nine residues involved in substrate interactions in Mhp1 were identical at the corresponding positions in DYC18_RS18060 and a further two were similar (Table 1).

Table 1. Conservation of Residues Involved in Substrate Binding in Characterised Bacterial NCS-1 Proteins

	Mhp1	**PucI**	**CodB**	**Rhod**
TMI	Gln42	Asn43	Phe33	Gln43
	Ala44	Pro45	Ala35	Thr45
TMIII	Trp117	Trp119	Trp108	Phe119
	Gln121	Gln123	Gly112	Gly123
TMVI	Gly219	Ile239	Ser203	Gly213
	Trp220	Trp240	Phe204	Trp214
	Ala222	Thr242	Ser206	Leu216
TMVIII	Asn318	Asn329	Leu284	Asn301
TMX	Leu363	Leu377	Leu325	Val360

*Residues in the substrate binding site of crystallographically defined Mhp1 (5-arylhydantoins) are compared with those at the corresponding positions in PucI (allantoin), CodB and *R. erythropolis* DYC18_RS18060 based on sequence alignments. Colouring indicates residues that are identical (*red*) or highly similar (*blue)* to residues at the same positions in Mhp1.

```
Rhod  MTHDGPAEVILTPERRTIDVVPDAERH-GTPRSQFTLWFGANMQITAIVDGALAVVFGAD  59
Mhp1  MNS-TPIEEARSLLNPSNAPTRYAER-SVGPFSLAAIWFAMAIQVAIFIAA-GQMTSSFQ  57
PucI  MKLKESQQQSNRLSNEDLVPLGQEKR-TWKAMNFASIWMGCIHNIPTYATVGGLIAIGLS  59
CodB  ------MSQDNNF---SQGPVPQSARKGVLALTF--VMLGLTFFSA-SMWTGGTLGTGLS  48

Rhod  AIWAIVGLLIGNIFGGAVMALHSAQGPRMGLPQMISSRAQFGVKGAVVPLVLVILMYLGF  119
Mhp1  VWQVIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFW  117
PucI  PWQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRASYGIYGANIPALLRAFTAIMW  119
CodB  YHDFFLAVLIGNLLLGIYTSFLGYIGAKTGLTTHLLARFSFGVKGSWLPSLLLGGTQVGW  108

Rhod  AATGTVLAGQAVNKILHIDSP------------------TVGIVVFGLLTAFVAVTGYKL  161
Mhp1  FGFQTWLGALALDEITR-LLTGFTNLP-------------LWIVIFGAIQVVTTFYGITF  163
PucI  LGIQTFAGSTALNILLLNMWPGWGEIGGEWNILGIHLSGLLSFVFFWAIHLLVLHHGMES  179
CodB  FGVGVAMFAIPVGKAT-----GL---------------DINLLIAVSGLLMTVTVFFGISA  149

Rhod  IHIVGRIATVVGIV---GFSYLAVRLFLEYDVASYVGIKG------FDIVTFLLAISLGA  212
Mhp1  IRWMNVFASPVLLAMGVYMVYLMLDGAD-VSLGEVMSMGGE-----NPGMPFSTAIMIFV  217
PucI  IKRFEVWAGPLVYLVFGGMVWWAVDI-A-GGLGPIYSQPGKFHTFSETFWPFAAGVTGII  237
CodB  LTVLSVIAVPAIACLGGYSVWLAVNGMG--GLDALKAV------VPAQPLDFNVALALVV  201

Rhod  GGWLTFGPYVADYSRYLPRSTSESTTFWSTFLGSVIGSQWS------MTFGALVAACAGD  266
Mhp1  GGWIAVVVSIHDIVKECKVDPNASREGQTKADARYATAQWLGMVPASIIFGFIGA--ASM  275
PucI  GIWATLILNIPDFTRFAETQKEQ------------IKGQFYGLPGTFALFAFASITVTSG  285
CodB  GSEISAGTLTADFVRFGRNAKLAVLVA--------MVAFFLGN-SLMFIFGAAGAAALGM  252

Rhod  AFLG------NQVGFMGDLAGPAAIAFLIYFVILVGKLTVNV----LNAYGGFMSILTTV  316
Mhp1  VLVGEWNPVIAITEVVGGVSIPMAILFQV-FVLLA-TWSTNPAANLLSPAYTLCSTFPRV  333
PucI  SQVAFGEPIWDVVDILARFDNPYVIVLSVITLCIA-TISVNVAANIVSPAYDIANALPKY  344
CodB  A-------DISDVMIAQGLLLPA-----IVVLGLN-IWTTNDNALYASG-LGFAN--ITG  296

Rhod  TAFNGQSRISSTARTLYILGFTAVSVLIAIAASADFLDNFKNFVLVLLMVFTPWSAINLI  376
Mhp1  FTFKTGVIVSAVV------GL-----LMMPWQFAGV---LNTFLNLLASALGPLAGIMIS  379
PucI  INFKRGSFITALL------AL-----FTVPWKLMESATSVYAFLGLIGGMLGPVAGVMMA  393
CodB  MSSKTLSVINGII------GT-----VCALWLYNN----FVGWLTFLSAAIPPVGGVIIA  341

Rhod  DYYLISKERIDIPALYDVNGRYGAWN---FTALACYAAGVLAQIPFLAQKMYTGPV-TDM  432
Mhp1  DYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALAVSFLTPDLMFVTGLIAALL  439
PucI  DYFIIRKRELSVDDLYSETGRYVYWKGYNYRAFAATMLGALISLI--------GMYVPVL  445
CodB  DYLMNRRRYEHFAT--------TRMMSVNWVAILAVALGIAAGHWLPGIVPVNAVLGGA-  392

Rhod  LGGADISWIVGIVFTGLIYYPLA---KRTSNPPSSMIYP-DHTAMTDSRV  478
Mhp1  LHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESATANTKEQNQR  489
PucI  KSLYDISWFVGVLISFLFYI----VLMRVHPPASLAIETVEHAQVRQAE-  490
CodB  ----------------LSYLILNPILNRKT---TAAMTHVEANSVE----  419
```

Figure 5. Multiple sequence alignment between *R. erythropolis* protein DYC18_RS18060 and characterised bacterial NCS-1 proteins

The sequences of DYC18_RS18060, Mhp1 from *M. liquefaciens* (D6R8X8), PucI from *B. subtilis* (P94575) and CodB from *E. coli* (P0AA82) were aligned using Clustal Omega. Residues that are identically conserved in three or four of the proteins are highlighted (*red text*). Residues involved in substrate binding in Mhp1 are

also highlighted (*green*). Transmembrane helices in Mhp1 based on its crystal structure (PDB 4D1B) [13] are highlighted as follows: transmembrane helix (*grey*), internal helix (*cyan*), external helix (*pink*). Putative transmembrane helices in DYC18_RS18060 based on TMHMM prediction (Figure 3) are also highlighted (*grey*).

4. Discussion

Structural and functional studies of a membrane protein require sufficient quantities of the respective protein in native membranes, or purified protein that is detergent-solubilised or reconstituted in a native-like environment [45-48], especially for applying techniques such as X-ray diffraction (XRD), cryo-electron microscopy (cryo-EM), mass spectrometry (MS), surface plasmon resonance (SPR), and nuclear magnetic resonance (NMR) spectroscopy [49-56]. Since the natural expression level of membrane proteins is usually too low, amplified expression must be achieved [29, 57]. In the current research, we demonstrated the successful amplified expression of the *R. erythropolis* protein DYC18_RS18060 with a C-terminal His_6-tag, as observed by an amplified band at ~37 kDa by SDS-PAGE and western blotting. Whilst the predicted molecular weight of DYC18_RS18060-His_6 is 51.1 kDa, it is well known that membrane proteins migrate anomalously through SDS-PAGE at lower molecular weight positions than their actual weight/position [58]. Applying the correction factor to the observed molecular weight (divide by 0.82) for fast migrating proteins gives a corrected apparent molecular weight of 45.1 kDa, which reduces the margin of error from 27.6% to 11.7%. Culture volumes can now be further scaled up to produce sufficient material for purifying milligram quantities of DYC18_RS18060-His_6.

In the computational analysis of DYC18_RS18060, the high aliphatic index and GRAVY values reflected the high contents of aliphatic residues in the protein (11.5% alanine, 10.0% glycine, 10.3% leucine, 10.0% valine), which are similar to the average contents of the aliphatic residues in secondary transporters from *E. coli* [59]. The observation that a large majority of the positively charged residues are located in the predicted loop regions at the cytoplasmic side of the membrane agrees with the positive-inside rule of von Heijne [60].

Gene and protein database entries for DYC18_RS18060 and for the nearest-neighbour proteins identified through BLAST searches currently list them as cytosine permease, allantoin permease or as a transporter of other NCS-1 family substrates; however, none of these proteins have been characterised by laboratory experiments yet. Combined with the observation that DYC18_RS18060 shares a relatively low overall sequence identity with any of the experimentally characterised NCS-1 transporters, it was also found that the residues in DYC18_RS18060 correspondng to the positions involved in interations with substrates in Mhp1 are not highly conserved with any of the experimentally characterised NCS-1 transporters. For example, when comparing DYC18_RS18060 and CodB at the same positions, only one glycine residue (Gly123) is identical and three residues (Phe119, Trp214, Val360) are similar. In

contrast, the experimentally charactersised cytosine transporter VPA1242 from *V. parahaemolyticus* [27] shares 75.2% sequence identity with CodB and at the Mhp1-defined substrate binding positions, all nine residues are identical in VPA1242 and CodB. Overall, there is no strong evidence for DYC18_RS18060 having the same substrate specificity as any of CodB (cytosine), PucI (allantoin) or Mhp1 (hydantoins). So, transport measurements using radiolabelled potential substrates [7, 12, 15] need to be performed for defining the substrate specificity of DYC18_RS18060.

5. Conclusion

In order to obtain further information about the structure, function and evolutionary relationships of bacterial NCS-1 family transport proteins, we have to overcome the challenge of cloning the *R. erythropolis* gene *DYC18_RS18060* with the introduction of a His_6-tag and amplifying expression of the translated protein in *E. coli* inner membranes. Large-scale flask or fermentor cultures can be used to produce sufficient quantities of membrane preparations to purify and reconstitute the DYC18_RS18060-His_6 protein and to assess its purity, yield and thermal stability. The protein can then be analysed using a multitude of chemical, biochemical and biophysical techniques. Bioinformatics analysis of DYC18_RS18060 was consistent with the protein having an overall structural organisation of an NCS-1 protein; however, its predicted role as a cytosine permease (currently given in databases) was not certain based on the comparison of the residues involved in substrate binding in experimentally characterised bacterial NCS-1 proteins. The substrate specificity of DYC18_RS18060 needs to be determined by transport measurements using radiolabelled potential substrates. This work laid the foundations for further structural and functional studies of DYC18_RS18060 and other NCS-1 proteins.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] de Koning H, Diallinas G. Nucleobase transporters (review). *Mol Membr Biol*. 2000;17(2):75-94. https://doi.org/10.1080/09687680050117101

[2] Pantazopoulou A, Diallinas G. Fungal nucleobase transporters. *FEMS Microbiol Rev*. 2007;31(6):657-675. https://doi.org/10.1111/j.1574-6976.2007.00083.x

[3] Saier MH Jr, Yen MR, Noto K, Tamang DG, Elkan C. The Transporter Classification Database: recent advances. *Nucleic Acids Res*. 2009;37:D274-D278. https://doi.org/10.1093/nar/gkn862

[4] Weyand S, Ma P, Saidijam M, Baldwin J, et al. The Nucleobase-Cation-Symport-1 family of membrane transport proteins. In: *Handbook of Metalloproteins*. John Wiley and Sons; 2010. https://doi.org/10.1002/0470028637.met268

[5] Witz S, Panwar P, Schober M, Deppe J, Pasha FA, Lemieux MJ, and Möhlmann T. Structure-function relationship of a plant NCS1 member - 0homology modeling and mutagenesis identified residues critical for substrate specificity of PLUTO, a nucleobase

transporter from Arabidopsis. *PLoS One*. 2014;9(3):e91343. https://doi.org/10.1371/journal.pone.0091343

[6] Krypotou E, Evangelidis T, Bobonis J, Pittis AA, Gabaldón T, Scazzocchio C, Mikros E, Diallinas G. Origin, diversification and substrate specificity in the family of NCS1/FUR transporters. *Mol Microbiol*. 2015;96(5):927-950. https://doi.org/10.1111/mmi.12982

[7] Ma P, Patching SG, Ivanova E, et al. Allantoin transport protein, PucI, from *Bacillus subtilis*: evolutionary relationships, amplified expression, activity and specificity. *Microbiol*. 2016;162(5):823-836.

[8] Sioupouli G, Lambrinidis G, Mikros E, Amillis S, Diallinas G. Cryptic purine transporters in *Aspergillus nidulans* reveal the role of specific residues in the evolution of specificity in the NCS1 family. *Mol Microbiol*. 2017;103(2):319-32. https://doi.org/10.1111/mmi.13559

[9] Patching SG. Recent developments in Nucleobase Cation Symporter-1 (NCS1) family transport proteins from bacteria, archaea, fungi and plants. *J Biosci*. 2018;43(4):797-815. https://doi.org/10.1007/s12038-018-9780-3

[10] Suzuki S, Henderson PJ. The hydantoin transport protein from *Microbacterium liquefaciens*. *J Bacteriol*. 2006;188(9):3329-3336.

[11] Jackson SM, Ivanova E, Calabrese AN, et al. Structure, Substrate Recognition, and Mechanism of the Na+-Hydantoin Membrane Transport Protein, Mhp1. In*Encyclopedia of Biophysics* 2018 (pp. 1-12). Springer.

[12] Patching SG. Synthesis, NMR analysis and applications of isotope-labelled hydantoins. *J Diagnos Imaging in Therapy*. 2017;4(1):3-26. http://dx.doi.org/10.17229/jdit.2017-0225-026

[13] Weyand S, Shimamura T, Yajima S, et al. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Sci*. 2008;322 (5902):709-713.

[14] Shimamura T, Weyand S, Beckstein O, et al. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Sci*. 2010;328(5977):470-473.

[15] Simmons KJ, Jackson SM, Brueckner F, et al. Molecular mechanism of ligand recognition by membrane transport protein, Mhp1. *EMBO J*. 2014;33(16):1831-1844. https://doi.org/10.15252/embj.201387557

[16] Calabrese AN, Jackson SM, Jones LN, et al. Topological dissection of the membrane transport protein Mhp1 derived from cysteine accessibility and mass spectrometry. *Anal Chem*. 2017;89(17):8844-8852. https://doi.org/10.1021/acs.analchem.7b01310

[17] Weyand S, Shimamura T, Beckstein O, et al. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *J Synchrotron Radiat*. 2011;18(1):20-23.

https://doi.org/10.1107/S0909049510032449

[18] Adelman JL, Dale AL, Zwier MC, Bhatt D, Chong LT, Zuckerman DM, Grabe M. Simulations of the alternating access mechanism of the sodium symporter Mhp1. *Biophys J*. 2011;101(10):2399-4207. https://doi.org/10.1016/j.bpj.2011.09.061

[19] Shi Y. Common folds and transport mechanisms of secondary active transporters. *Annu Rev Biophys*. 2013;42:51-72.

[20] Kazmier K, Sharma S, Islam SM, Roux B, Mchaourab HS. Conformational cycle and ion-coupling mechanism of the Na^+/hydantoin transporter Mhp1. *Proc Natl Acad Sci U S A*. 2014;111(41):14752-14757. https://doi.org/10.1073/pnas.1410431111

[21] Li J, Zhao Z, Tajkhorshid E. Locking two rigid-body bundles in an outward-facing conformation: The ion-coupling mechanism in a LeuT-fold transporter. *Sci Rep*. 2019;9(1):19479. https://doi.org/10.1038/s41598-019-55722-6

[22] Meshkin H, Zhu F. Toward convergence in free energy calculations for protein conformational Changes: A case study on the thin gate of Mhp1 transporter. *J Chem Theory Comput*. 2021;17(10):6583-6596. https://doi.org/10.1021/acs.jctc.1c00585

[23] Mourad GS, Tippmann-Crosby J, Hunt KA, et al. Genetic and molecular characterization reveals a unique nucleobase cation symporter 1 in Arabidopsis. *FEBS Lett*. 2012;86(9):1370-1378. https://doi.org/10.1016/j.febslet.2012.03.058

[24] Schein JR, Hunt KA, Minton JA, Schultes NP, Mourad GS. The nucleobase cation symporter 1 of *Chlamydomonas reinhardtii* and that of the evolutionarily distant *Arabidopsis thaliana* display parallel function and establish a plant-specific solute transport profile. *Plant Physiol Biochem*. 2013;70:52-60. https://doi.org/10.1016/j.plaphy.2013.05.015

[25] Minton JA, Rapp M, Stoffer AJ, Schultes NP, Mourad GS. Heterologous complementation studies reveal the solute transport profiles of a two-member nucleobase cation symporter 1 (NCS1) family in *Physcomitrella patens*. *Plant Physiol Biochem*. 2016;100:12-17. https://doi.org/10.1016/j.plaphy.2015.12.014

[26] Rapp M, Schein J, Hunt KA, Nalam V, Mourad GS, Schultes NP. The solute specificity profiles of nucleobase cation symporter 1 (NCS1) from *Zea mays* and *Setaria viridis* illustrate functional flexibility. *Protoplasma*. 2016;253(2):611-623.

[27] Ahmad I, Ma P, Nawaz N, Sharples DJ, Henderson PJF, Patching SG. Cloning, amplified expression, functional characterisation and purification of *Vibrio parahaemolyticus* NCS1 cytosine transporter VPA1242 (Chapter 8). In:

Patching SG, editor. A Closer Look at Membrane Proteins. Independent Publishing Network. 2020;241-67.

[28] Danielsen S, Boyd D, Neuhard J. 1995. Membrane topology analysis of the *Escherichia coli* cytosine permease. *Microbiol*. 1995;141(11): 2905-2913. https://doi.org/10.1099/13500872-141-11-2905

[29] Ahmad I, Nawaz N, Darwesh NM, ur Rahman S, Mustafa, MZ, Khan SB, Patching SG. Overcoming challenges for amplified expression of recombinant proteins using *Escherichia coli*. *Prot Expr Purif*. 2018;144:12-18. https://doi.org/10.1016/j.pep.2017.11.005

[30] Singh SK, Pal A. Biophysical approaches to the study of LeuT, a prokaryotic homolog of neurotransmitter sodium symporters. *Methods Enzymol*. 2015;557:167-198. https://doi.org/10.1016/bs.mie.2015.01.002

[31] Bai X, Moraes TF, Reithmeier RAF. Structural biology of solute carrier (SLC) membrane transport proteins. *Mol Membr Biol*. 2017;34(1-2):1-32. https://doi.org/10.1080/09687688.2018.1448123

[32] Kazmier K, Claxton DP, Mchaourab HS. Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes. Curr Opin Struct Biol. 2017;45:100-108. https://doi.org/10.1016/j.sbi.2016.12.006

[33] Razavi AM, Khelashvili G, Weinstein H. How structural elements evolving from bacterial to human SLC6 transporters enabled new functional properties. *BMC Biol*. 2018;16(1):31. https://doi.org/10.1186/s12915-018-0495-6

[34] Schumann T, König J, Henke C, et al. Solute carrier transporters as potential targets for the treatment of metabolic disease. *Pharmacol Rev*. 2020;72(1):343-379. https://doi.org/10.1124/pr.118.015735

[35] Wang WW, Gallo L, Jadhav A, Hawkins R, Parker CG. The druggability of solute carriers. *J Med Chem*. 2020;63(8):3834-3867. https://doi.org/10.1021/acs.jmedchem.9b01237

[36] Pizzagalli MD, Bensimon A, Superti-Furga G. A guide to plasma membrane solute carrier proteins. *FEBS J*. 2021;288(9):2784-2835. https://doi.org/10.1111/febs.15531

[37] Stark MJ. Multicopy expression vectors carrying the lac repressor gene for regulated high-level expression of genes in *Escherichia coli*. *Gene*. 1987;51:255-267. https://doi.org/10.1016/0378-1119(87)90314-3

[38] Ward A, et al. (2000) The amplified expression, identification, purification, assay, and properties of hexahistidine-tagged bacterial membrane transport proteins. Membrane Transport: A Practical Approach, ed Baldwin SA (Oxford Univ Press, Oxford) 141–166.

[39] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and

analysis tools on the ExPASy Server. In: Walker JM (ed) *The proteomics protocols handbook*. Humana Press. 2005; 571-607.

[40] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567-580. https://doi.org/10.1006/jmbi.2000.4315

[41] Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res*. 2009;37(Web Server issue):W465-W468. https://doi.org/10.1093/nar/gkp363

[42] Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. https://doi.org/10.1093/nar/gky427

[43] Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539. https://doi.org/10.1038/msb.2011.75

[44] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44:W242-W245. https://doi.org/10.1093/nar/gkw290

[45] Brown CJ, Trieber C, Overduin M. Structural biology of endogenous membrane protein assemblies in native nanodiscs. *Curr Opin Struct Biol*. 2021;69:70-77. https://doi.org/10.1016/j.sbi.2021.03.008

[46] Lemieux MJ, Overduin M. Structure and function of proteins in membranes and nanodiscs. *Biochim Biophys Acta Biomembr*. 2021;1863(1):183445. https://doi.org/10.1016/j.bbamem.2020.183445

[47] Mouhib M, Benediktsdottir A, Nilsson CS, Chi CN. Influence of detergent and lipid composition on reconstituted membrane proteins for structural studies. *ACS Omega*. 2021;6(38):24377-24381. https://doi.org/10.1021/acsomega.1c02542

[48] Strickland KM, Neselu K, Grant AJ, Espy CL, McCarty NA, Schmidt-Krey I. Reconstitution of detergent-solubilized membrane proteins into proteoliposomes and nanodiscs for functional and structural studies. *Methods Mol Biol*. 2021;2302:21-35.

[49] Hammerschmid D, van Dyck JF, Sobott F, Calabrese AN. Interrogating membrane protein structure and lipid interactions by native mass spectrometry. *Methods Mol Biol*. 2020;2168:233-261.

[50] Wu M, Lander GC. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr Opin Struct Biol*. 2020;64:9-16. https://doi.org/10.1016/j.sbi.2020.05.007

[51] Yeh V, Goode A, Bonev BB. Membrane protein structure determination and characterisation by solution and solid-state NMR. *Bio*.

2020;9(11):396. https://doi.org/10.3390/biology9110396

[52] Kwan TOC, Reis R, Siligardi G, Hussain R, Cheruvara H, Moraes I. Selection of Biophysical Methods for Characterisation of Membrane Proteins. *Int J Mol Sci*. 2019;20(10):2605. https://doi.org/10.3390/ijms20102605

[53] Beriashvili D, Schellevis RD, Napoli F, Weingarth M, Baldus M. High-resolution studies of proteins in natural membranes by solid-state NMR. *J Vis Exp*. 2021;(169). https://doi.org/10.3791/62197.PMID:33749679

[54] Günsel U, Hagn F. Lipid Nanodiscs for High-Resolution NMR Studies of Membrane Proteins. *Chem Rev*. 2021. https://doi.org/10.1021/acs.chemrev.1c00702

[55] Januliene D, Moeller A. Single-Particle Cryo-EM of Membrane Proteins. InStructure and Function of Membrane Proteins 2021 (pp. 153-178). Humana, New York, NY.

[56] Weissenberger G, Henderikx RJM, Peters PJ. Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods*. 2021;18(5):463-471. https://doi.org/10.1038/s41592-021-01130-6

[57] Gordon E, Horsefield R, Swarts HG, de Pont JJ, Neutze R, Snijder A. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr Purif*. 2008;62(1):1-8. https://doi.org/10.1016/j.pep.2008.07.005

[58] Rath A, Deber CM. Correction factors for membrane protein molecular weight readouts on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal Biochem*. 2013;434(1):67-72. https://doi.org/10.1016/j.ab.2012.11.007

[59] Saidijam M, Patching SG. Amino acid composition analysis of secondary transport proteins from *Escherichia coli* with relation to functional classification, ligand specificity and structure. *J Biomol Struct Dyn*. 2015;33(10):2205-20. https://doi.org/10.1080/07391102.2014.998283

[60] von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol*. 1992;225(2):487-494. https://doi.org/10.1016/0022-2836(92)90934-C