

Article

Distantly Related Homologue of UhpT in *Pseudomonas aeruginosa*

Tommaso Orioli *  and Daniela Dolce

Department of Paediatric Medicine, Cystic Fibrosis Center, Meyer Children's University Hospital IRCCS, 50139 Florence, Italy

* Correspondence: tommaso.orioli@meyer.it

Abstract: *Pseudomonas aeruginosa* (PA) is an opportunistic Gram-negative bacteria that affects patients in intensive care units and chronic respiratory disease patients. Compared to other bacteria, it has a wide genome (around 6.3-Mb) that supports its metabolic versatility and antimicrobial resistance. Fosfomycin (FF) is primarily used as an oral treatment for urinary tract infections (UTIs). FF diffuses inside the cell via glycerol-3-phosphate transporter (GlpT) PA, as well as in other bacteria. In other bacteria, such as *E. coli*, glucose-6-phosphate transporter (UhpT) functions as FF transporter. Since mutant GlpT leads to FF resistant PA, it is assumed that GlpT is the only FF transporter. However, it is also assumed that PA uses glucose-6-phosphate and, thus, homologous proteins of UhpT may be present in its genome. Here, we present an attempt to find a distant related homologue of UhpT in PA. A Hidden Markov Model (HMM) was created to seek for Major facilitator family (MFS) domain in 21 PA genomes of 14 CF patients annotated with prokka and the statistical analysis was performed (MCC: 0.84, ACC: 0.99). Then, the HMM was applied to PA genomes. Besides the actual GlpT, annotated as glpt_1, one more GlpT protein was found in 21 out of 21 genomes, annotated as glpt_2. Since glpt_2 clusters closer to UhpT than GlpT, glpt_2 was selected to build a model. Computing a structural superimposition, the model and the template of UhpT have 0.6 Å of RMSD. The model of glpt_2 has some characteristics that are fundamental to UhpT functions. The binding site, consisting of 2 arginines (Arg46 and Arg275) and Lys45, is totally conserved, as well as the topology of the structure. Asp90 is also conserved in glpt_2 model. No studies aimed at searching for distant related homologous of UhpT. Since the high genetic exchange and high mutational rate in bacteria, it is likely that PA has a UhpT-like protein in the PA genome. The binding site is superimposable to UhpT protein as well as the overall topology. In fact, the 12 TMs are completely comparable, suggesting a well-defined folding of the protein across the bilayer lipid membrane. To enforce our hypothesis, in all 21 PA genomes, we also found a protein annotated as membrane sensor protein UhpC, important for expression and function of UhpT in *E. coli*. Since PA strains are wild-type, we can assume that most of the PA have proteins like this. The presence of a homologue of UhpT suggests that this protein is conserved in PA genome.

Keywords: *Pseudomonas aeruginosa*; GlpT; UhpT; MFS; HMM; homology-modelling; cystic fibrosis



Citation: Orioli, T.; Dolce, D.

Distantly Related Homologue of UhpT in *Pseudomonas aeruginosa*. *Bacteria* **2022**, *1*, 266–278. <https://doi.org/10.3390/bacteria1040020>

Academic Editor: Bart C. Weimer

Received: 7 October 2022

Accepted: 3 November 2022

Published: 7 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pseudomonas aeruginosa (PA) is an opportunistic Gram-negative bacteria that thrives soil and host-environment and affects patients in intensive care units and chronic respiratory disease patients, such as cystic fibrosis (CF) [1,2]. Compared to other bacteria, it has a wide genome (around 6.3-Mb) that supports its metabolic versatility. Therefore, a key point of its pathogenicity is a rapid adaptation to different and challenging environments. PA causes infections whose high mortality rate [3] is attributable to the organism's high resistance to many antimicrobials, particularly multidrug resistance in healthcare settings [4,5]. Generally, two mechanisms are responsible for antimicrobial resistance: acquisition of resistance genes (e.g., those encoding β -lactamases [6,7]) or aminoglycoside-modifying enzymes [8] from other bacteria or mutations of chromosomal genes [9].

Among all mentioned mechanisms, mutations of chromosomal genes play a crucial role in fosfomycin-resistant PA. Fosfomycin (FF) is primarily used as an oral treatment for urinary tract infections (UTIs). However, FF is under study in therapy of a variety of infections because it could be active against multidrug-resistant (MDR) bacteria [10]. For PA treatments, fosfomycin has shown its efficacy in combinations with other antibiotics [11].

FF is a broad-spectrum bactericidal antibiotic whose function is inhibiting peptidoglycan biosynthesis. In particular, it links UDP-GlcNAc enol-pyruvyltransferase (MurA) which catalyzes the reaction between UDP-N-acetylglucosamine and phosphoenolpyruvate [12,13]. In *E. coli* there are two carrier-dependent systems which can act as fosfomycin uptake, namely the glycerol-3-phosphate (glycerol-3-P) transporter (GlpT) and glucose-6-phosphate (glucose-6-P) transporter (UhpT). Both systems are up-regulated by their substrates, glycerol-3-P and glucose-6-P, respectively, and by cAMP, which is positively regulated by *ptsI* and *cyaA* genes [14,15]. Moreover, the UhpT system requires the regulatory genes *uhpA*, *uhpB*, and *uhpC* [16]. Both the mechanisms belong to organophosphate:phosphate antiporter family (OPA).

Resistance to FF has been revised recently [17,18]. Breakpoints for susceptibility have been set by Clinical and Laboratory Standards Institute (CLSI) and European Committee on Antimicrobial Susceptibility Testing (EUCAST). For wild-type PA, the minimum inhibitory concentrations (MICs) is ≤ 128 mg/L. In PA, the main mutational resistance is mutations in the GlpT channel. A study demonstrated that missense mutations on *glpT* gene result in FF resistance of PA with a good fitness of FF-resistant colonies [15] assuming that GlpT may be the only channel for FF in PA. Compared to *E. coli*, PA utilizes a wider range of carbon sources to overcome the lack of glucose-6-P in the cell medium. This suggests the presence of a protein homologue of *E. coli* UhpT system in PA. However, in the same study, they tried to find a homologue protein in PA strains, since it is generally assumed that a glucose-6-P system may exist. In fact, it has been stated that glucose-6-P should be added in growth medium for FF sensitivity to allow FF efficacy, confirming that the presence of glucose-6-P may activate its channel in some PA strains [19,20].

Here we present an attempt to find an UhpT homologous 21 PA strains found in 14 cystic fibrosis patients. Generally, therapies with solely FF are rarely used.

First, the Major facilitator family (MFS, PFAM id: PF07690) was studied. MFS is a family of membrane channel proteins whose function is the transport of small molecules in or out of the cell in response to chemiosmotic gradients [21]. The proteins in this family are widely distributed in all kingdoms and are the main responsible of the transportation of sugars. However, drugs, metabolites, oligosaccharides, amino acids and oxyanions were all transported by MFS family members [22]. LacY is one of the representative permeases of this family. MFSs can function by solute uniport, solute/cation symport, solute/cation antiport and/or solute/solute antiport with inwardly and/or outwardly directed polarity. Generally, MFSs contain 12 transmembrane (TM) helices, with two 6-helix bundles formed by the N and C terminal homologous domains [23] of the transporter which are connected by an extended cytoplasmic loop (among 30 to 100 residues) that may suggest a large degree of relative motion between the two domains. Both C and N terminals are located on the cytoplasmic side. The MFSs topology is organized as follows:

- 4 TM helices are positioned in the center of the transporter, which contains the majority of residues important for substrate binding site and creates a central pore;
- 4 TM helices are positioned outside of the previous TM helices and are important for structural integrity;
- 4 TM helices are positioned on the side of the proteins and are important for interdomain contacts.

Generally, the movement across the membrane through MFSs is mediated by the same mechanism. In steady state, MFS has an outward conformation that reveals the inside face of the channel and the binding site. When the ligand reaches the binding site, a sudden change in conformation happens and the MFS shifts in inward conformation. Now the binding site is in communication with cytoplasm, which is rich in inorganic phosphate

(Pi). Pi has a higher affinity with the binding site than the ligand. Therefore, the ligand is released in cytoplasm and the protein shifts back in outward conformation [23,24]. The function of the specific MFS lays in a specific binding site. In LacY is mainly coordinated to residues in the N domain, with Glu135, Arg144 and Trp151 [25]. In FucP, Glu135 and Gln162 are essential for galactose binding [26]. In PepT, residues Tyr29, Tyr 30 and Tyr68 are essential for peptide-binding affinity [27]. In GlpT, two arginines (Arg45 and Arg269) and a Lys46 participate at glycerol-3-P binding, which are conserved even in UhpT binding pocket [28]. Moreover, Asp388 and Lys391 are important for substrate recognition in UhpT [29].

In this study, we focused on organophosphate:phosphate antiporter family (OPA), a sub-family of MFS proteins whose function is to transport small carbohydrate molecules inside the cell to use them as carbon source. These proteins are critical for both bacteria and human. Particularly, mutations of GlpT in human cause glycogen storage disease type Ib [30]. In bacteria, many MFS proteins have been characterized and their function studied. In PA, MFS are also important. By searching for PFAM id (PF07690) and filter for *Pseudomonas aeruginosa* organism in UniProt database, we end up with over 3000 proteins and over 200 clusters, by using UniRef50. MFSs have an important role in many biological processes: xenobiotic detoxification as Bcr/CflA family efflux transporter, nitrate assimilation as nitrate/nitrite transporter or even as pharmacological resistance as the chloramphenicol resistance protein CmlA.

2. Results

The Hidden Markov Model (HMM, see Methods) was used to search for MFS proteins in 21 genomes of PA annotated with prokka. For each genome, we used the *hmmsearch* tool of HMMER using the HMM as input model and the PA genomes as query. The list of the proteins in output of each PA genome is in Supplemental Materials File S1. To validate our model, whole-genome sequences from NCBI of PA were downloaded, and the *hmmsearch* was computed. We compared the number of sequences which passed the HMM filter in PA genomes with the set of genomes from NCBI. The statistics are in Supplemental Materials File S2. On average, in the NCBI set 75.1 sequences out of 6138.6 total sequences that passed the HMM were found (1.23%). In our genomes, 74.4 sequences out of 6136.9 total sequences (1.21%) were found. The 21 PA genomes and NCBI genomes are comparable. In all 21 genomes, the GlpT protein annotated was found. Its product name is “Glycerol-3-phosphate transporter” and *glpt_1* was the gene name. A pairwise alignment with UniProt ID A0A2R3IQP3 was computed, and the similarity was 99–100%. Then, we can claim that either prokka or HMM were able to identify GlpT.

However, 21 out of 21 genomes also have a protein whose product name is still “Glycerol-3-phosphate transporter”, but the gene name is *glpt_2*. Its length is 4 residues longer. The pairwise alignment with the Needleman-Wunsch global alignment algorithm was used by using Needle from EMBOSS [31] and the sequence identity is very low (28.1%) as shown in Figure 1.

Even if the low similarity suggests that the two proteins are not the same, prokka still has annotated this sequence with the same product name. Then, we assumed that this protein may be a homologue of GlpT and may share the same function. As seen above, GlpT and UhpT belong to the same family (MFS), meaning that the protein function is conserved, although the protein sequence similarity is low. To enforce our assumption, a MSA with *glpt_1* and *glpt_2* and some GlpT and UhpT proteins of different organisms was computed. The list of the proteins collected is in Table 1 and the fasta sequences of *glpt_1*, *glpt_2* and UhpC are in Supplemental Materials File S3.

glpt_1	1	MFGIFRPAPHQAP--LPDAEVDSTYRRLRWQIFAGIFIGYAGYLLRKN	47
glpt_2	1	-----MNQRPGTALPSSPV-SSFQRWRLQIFAITWLAYAAFYFTRKA	41
glpt_1	48	FSLAMPYLIED-GYSRQLGVA-MSAIAI-AYGLSKFIMGLVSDRSNPRV	94
glpt_2	42	FSVAKLGIAEDPDFQLDKMAMADLDAIYLAAYAVGQFTWGLADRFGPRI	91
glpt_1	95	FLPFGLLVSAAMFIFG-FAHWATSSVAIMFVLLFINGWAQGMGWPPSGR	143
glpt_2	92	VVFGGLLISALAALVMGTFA----TLPIFATCMVVQGLAQSTGWSGLCK	136
glpt_1	144	TMVHWSQKERGSVVSVNVAHNVGGGLIGPLF-LLGLALFGDWRSFAYV	192
glpt_2	137	NIGSFFATHERGRVLGLWSTCYAFGGLVASPFAGWWAYSVFGTWHAAFYS	186
glpt_1	193	PAIVAVGVAVFAYFAMRDTQPSCGLPPIE-----KHKDDYPEGYDEH	234
glpt_2	187	SAAVVGGVALLFLLQNRPEDEVGHPPEVQAASEAVRYAGFRPEGGNGG	236
glpt_1	235	HERELSTKEIFVTYVLRNKLWYIALANVFVYLLRYGVLWDAPTYLKEAK	284
glpt_2	237	SLQVLAK-----VVRNRTVLCGLAYFLLKPARYAILLWGPVIIVYE-R	278
glpt_1	285	HFNVDKSSWAYF--FYEWAGIPGTLLCGWLSDKLFKGNRGATGVVFMILV	332
glpt_2	279	MPALGKVGAAIVPTAFEVAGLLGPILIGFASDKLFGARRIPACVISLLAL	328
glpt_1	333	TVGTLIYWLNPA--GNPAIDMAALIMIGFLIYGP-VMLIGLQALELAPK	378
glpt_2	329	TVCLALF--VPAMTTGSIPIVVGLLFMMGLTLYGPDMSICGSAADIFGTS	376
glpt_1	379	KAAGTAAGFTGLFGYLGGSVAANALVGYTVDFHFGWDGGFKLLIGACLLAI	428
glpt_2	377	EAAGTASGFVNGCGSV-GAILGGLPGY-----FD-TYTVFIGFTVAAL	418
glpt_1	429	LFLCLTLRKPAPVQAAPQQAGG-----	449
glpt_2	419	IATAVLL--PHWNSRPDGAVKSHRNIPSLGDFRPIRS	453

Figure 1. Pairwise global alignment between *glpt_1* and *glpt_2*. The identity is 28.1% and suggests that the two sequences represent two different proteins.

Table 1. UniProt identifiers and organisms of GlpT and UhpT proteins used in MSA.

UniProt ID of UhpT	Organism	UniProt ID of GlpT	Organism
P0AGC0	<i>E. coli</i>	P37948	<i>B. subtilis</i>
P0AGC2	<i>S. flexneri</i>	P96335	<i>H. influenzae</i>
P27670	<i>S. typhi</i>	P08194	<i>E. coli</i>
A0A485IC54 *	<i>P. aeruginosa</i>	A0A072Z *	<i>P. aeruginosa</i>

Four different proteins were used to compute the MSA. In order to maintain the heterogeneity of the species, four different microorganisms which have either UhpT or GlpT were chosen. The id marked with * are the GlpT and UhpT of PA. The sequence identity of these two sequences and *glpt_1* and *glpt_2* is 100%.

In Figure 2, the tree obtained from MSA is reported. Clustal Omega algorithm [32] was used. In the tree *glpt_2* clusters with UhpT proteins, suggesting that *glpt_2* is closer to UhpT than GlpT and *glpt_1*, which are the actual GlpT protein in PA. Since their function and structure is conserved, the alignments show one big cluster where either GlpT or UhpT proteins are packed together. However, in this sole cluster, UhpT forms a sub-cluster where *glpt_2* is included.

Therefore, *glpt_2* was selected as a plausible UhpT homologue in PA. To do so, MODELLER v10.2 was used to compute a homology modeling experiment. MODELLER is used for homology or comparative modeling of protein three-dimensional structures [33,34]. The user provides a well-defined protein structure and a protein sequence to compare to it. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints [35,36] and automatically calculates a model containing all non-hydrogen atoms. MODELLER has been used for homology modeling experiments and it is suitable

for identifying distantly related homologues. For building a new model, a defined crystal structure of UhpT from AlphaFold repository [37,38] was retrieved. AlphaFold is a machine learning system that predicts a protein's 3D structure from its amino acid sequence with an accuracy very close to structural experiments and it is, nowadays, the best secondary structure prediction method. It is possible to search for gene or protein names in AlphaFold database and, more importantly, it reports the level of confidence predicted for each residue of the protein. Therefore, the P0AGC0 was selected as crystal structure template. This protein has the best confidence score across all the TMs helices, which are the key points of the protein functions. The only part of the protein that has a very low confidence score is the cytoplasmic loop between the C and N domains, which is very unstructured [24] and may be the reason why it was poorly predicted by AlphaFold. MODELLER produces 5 outputs and, in order to choose which is the best one, the structural superimposition by using Chimera [39] was performed for every output and the Root Mean Square Deviation (RMSD) was calculated. RMSD is the measure of the average distance between the atoms of superimposed proteins (Supplemental Materials File S4). The less is the RMSD the better is the structural superimposition. The RMSD between P0AGC0 and *glpt_2* model was 0.614 Å. Both the crystal structures and the superimposition are shown in Figure 3. The sequence identity between the two structures is 24%.

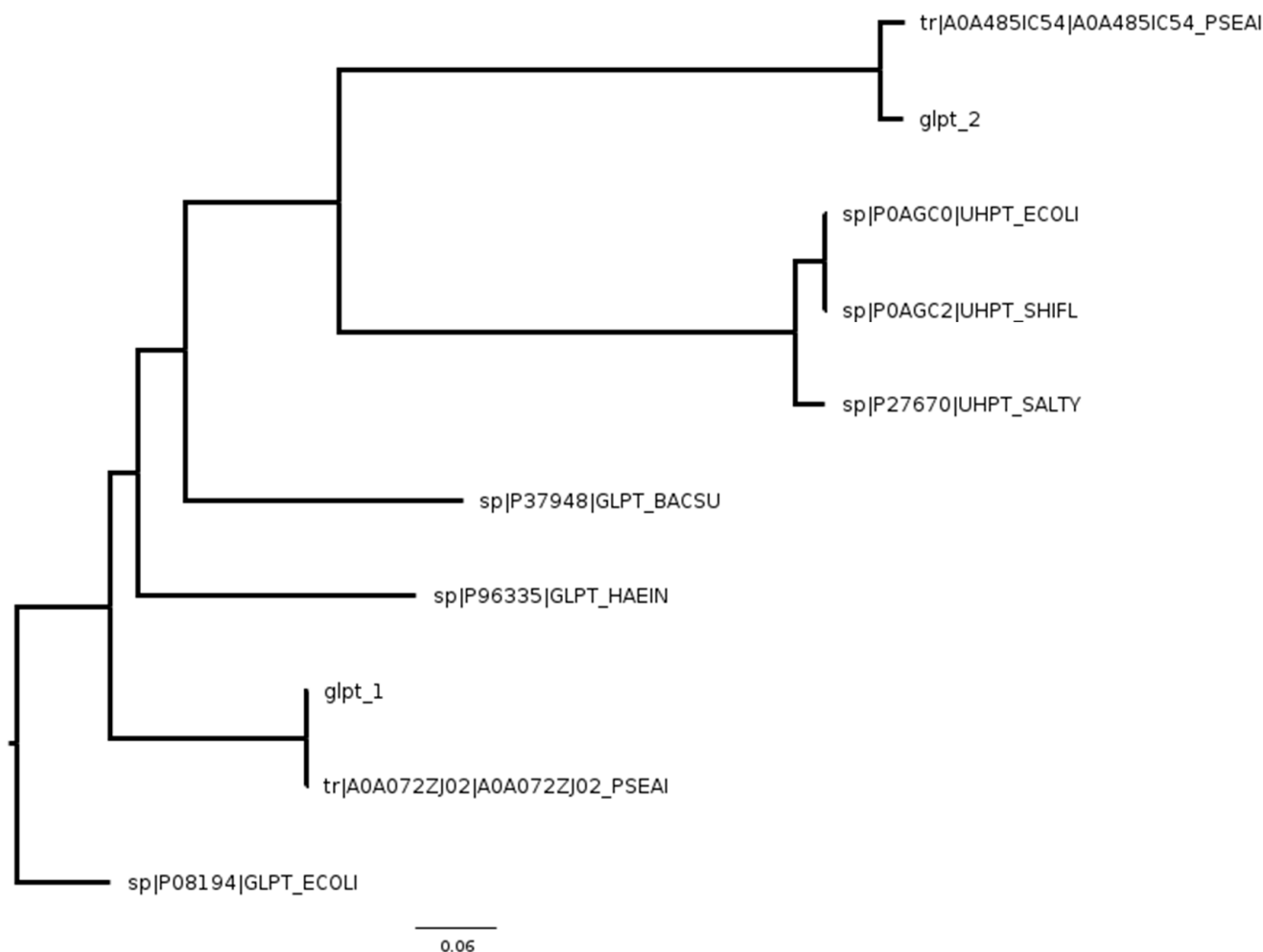


Figure 2. Distance tree from the MSA. The UniProt identifiers of the 8 proteins used for the MSA are in Table 1. We can see that *glpt_2* is closer to UhpT than GlpT.



Figure 3. The UhpT and the model of glpt_2 superimposed. In the superimposition the two structures are barely indistinguishable. The model topology is conserved in all 12 TM domains. In Supplemental Materials File S5 the UhpT structure is colored by pLDDT score of AlphaFold2.

3. Discussion

The topology of the model and the UhpT crystal structure are comparable. The model has all 12 TMs conserved as in MFS and both the C and N domains end in the cytoplasmic part of the membrane. As seen in Figures 4 and 5, the binding site is highly conserved. The two arginines (Arg45 and Arg269) responsible for either GlpT and UhpT proteins and perfectly superimposable: Arg46 and Arg275 in UhpT and Arg39 and Arg264 in the model have similar distances (9.70 Å vs. 9.88 Å, respectively) than crystal structure of GlpT, that is 9.9 Å [28]. Either the two lysine's are conserved as Lys45 and Lys38. Then, we can claim that binding site could function as well as in UhpT proteins.

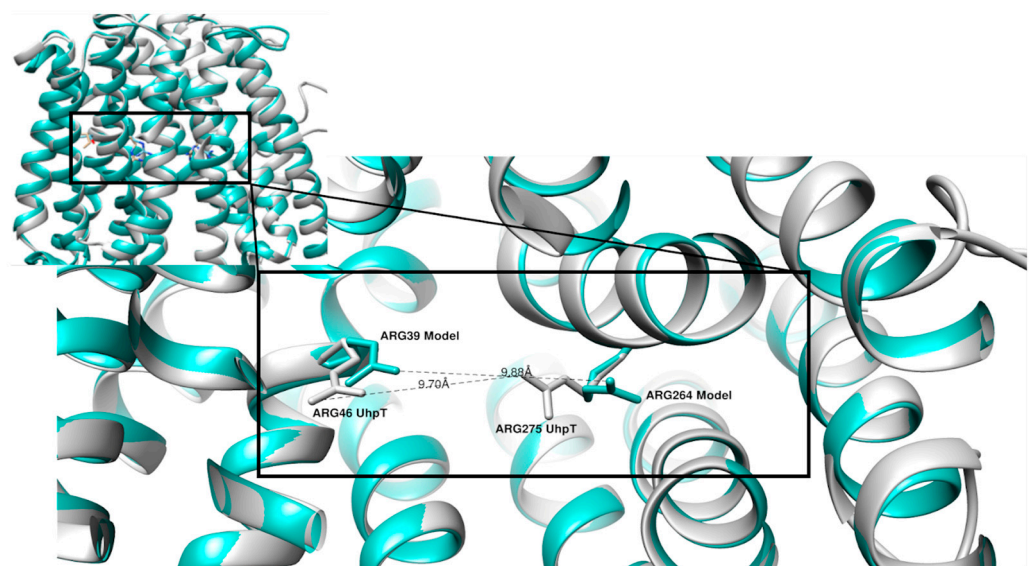


Figure 4. The binding site superimposed with the distances. The binding site is conserved in the zoomed black area. We can see the two arginines of UhpT protein (in gray) and the two arginines of the model (in light blue). The distance between the pair of arginines is comparable and fits with previous studies. The pLDDT score is 95.41 for both ARG46 and ARG275 (Supplemental Materials File S5).

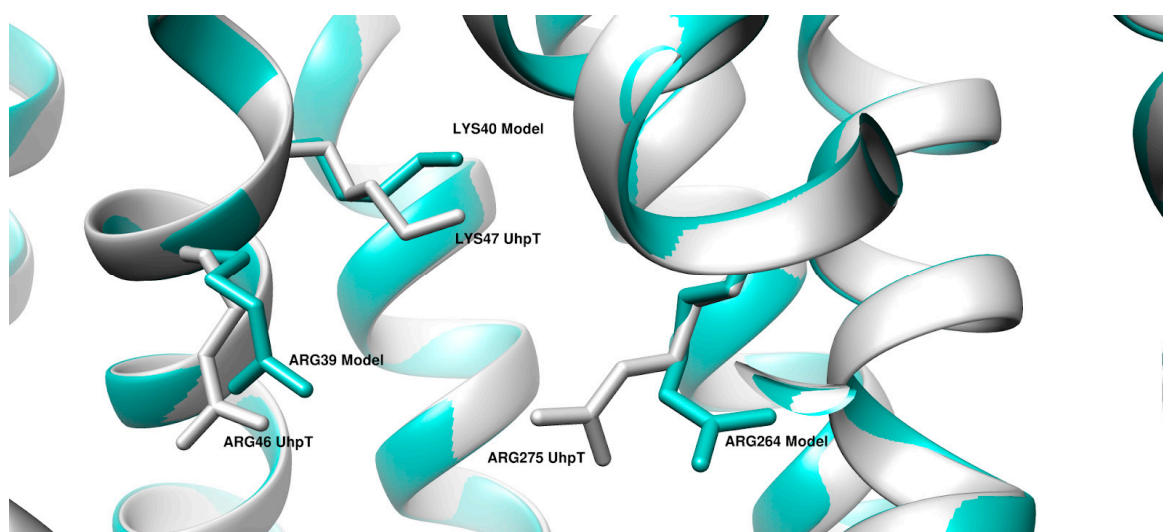


Figure 5. The binding site superimposed with the lysine. The pLDDT score of LYS47 is 94.65 (Supplemental Materials File S5).

Moving out from the substrate binding, it has shown that the equivalent of Asp88 in GlpT is important for the interaction between the second and the seventh helices during inward to outward interconversion [40]. In both UhpT and model, we found the asparagine, respectively, in position 90 and 85, although they are not superimposed and are 6.57 Å apart (Figure 6).

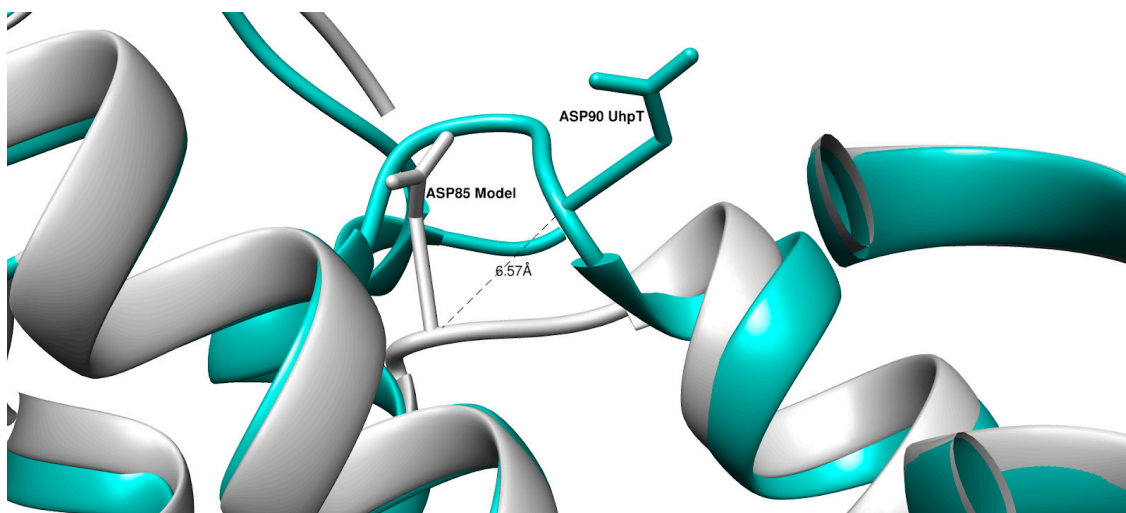


Figure 6. Distance between the Asp90 vs. Asp85. The distance between the two carbon atoms in the backbone of the aspartates (Asp90 vs. Asp85 in UhpT and model, respectively). The pLDDT score is 76.05 (Supplemental Materials File S5).

Moreover, we found the relatively high percentage of aromatic residues in the sequence, 13.6% vs. 11.9%, respectively, that tends to increase if counted only the TMs helices (15.4% vs. 12.9%), whereas in GlpT is even higher (15.2% and 18.9% in whole sequence and in TMs only, respectively).

However, this model presents some differences either with UhpT and with GlpT. In UhpT has been underlined the importance of Asn388 and Lys391 in substrate recognition [29]. It has been proved that these two residues form a salt bridge that is crucial in selecting the glucose-6-phosphate to the detriment of other organophosphate substrates. However, in the same paper, it is not excluded that glucose-6-phosphate still can be selected

and transported by the channel. Moreover, in the organophosphate transporter family, the motif $W_{173}NXXHN_{178}$ [41] is highly conserved (>95%). This motif is found in UhpT in the same position but is not present in our model in any position. Finally, in our model, there is one large undefined loop in the N terminal TM helix (Figure 7).

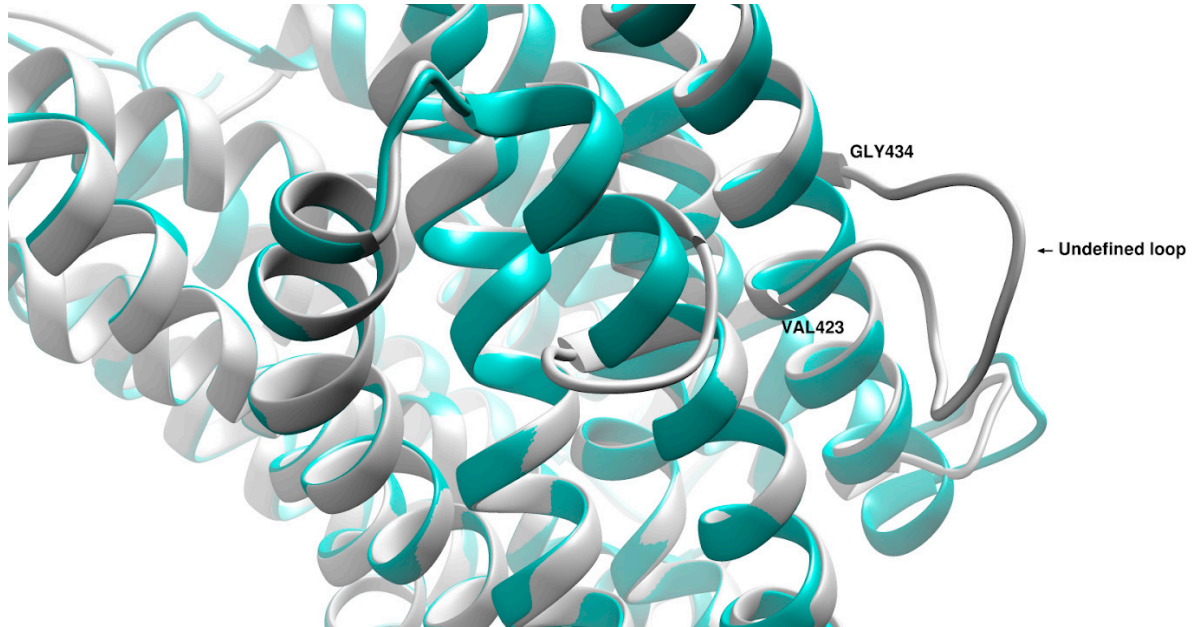


Figure 7. The loop in the model between Val423 and Gly434.

4. Materials and Methods

4.1. PA Collection and Sequencing Pipeline

21 first acquisition PA strains were selected from 13 different cystic fibrosis patients with initial infection of PA (5 male vs. 8 female; age 3–27 years old). These strains were chosen because they may be wild-type concerning genetic background and antimicrobial pressure. Especially for GlpT and UhpT, the main cause for FF resistance is mutations on *glpT* and *uhpT* (in *E. coli*). From the oro-pharyngeal swab, the PA colonies were isolated from Cetrimide agar or McConkey agar after 48-h incubation in controlled temperature (35–37 °C).

DNA extraction was performed from pure PA cultures after 24 h of incubation at 37 °C on Columbia agar + 5% sheep blood (bioMérieux) using QIAamp DNA Mini Kit (QIAGEN). Whole-DNA libraries were prepared with Illumina DNA Prep Library Preparation Kit (Illumina). Quality checks were performed with Qubit Fluorometric Quantification (ThermoFisher Scientific). Expecting 100× coverage on average, a pool of six libraries were run with MiSeq Reagent Kit v2, 300-cycles.

First, the quality of the reads was checked with FastQC v.0.11.9 [42]. Trimmomatic v.0.39 [43] was used to remove the reads with command *PE -phred33 -threads 4 SLIDING-WINDOW:4:20 MINLEN:70*. Reads shorter than 70 bp were removed. Trimmed reads were assembled using SPAdes v.3.15.4 [44] with default command and contigs shorter than 1 kb were removed. Genome assemblies were annotated with prokka v.1.14.6 [45] with default command. Prokka is a tool to annotate bacterial, archaeal and viral genomes quickly and produce standards-compliant output files. From the output files of prokka, the *name_of_sampl.faa* was selected for further analysis. This file contains all the proteins annotated with the predicted name, if available. Otherwise, the protein is annotated as a hypothetical protein.

4.2. Building Hidden Markov Model (HMM)

An HMM is a finite model that describes a probability distribution over a finite number of possible sequences [46] providing a tool for building complex models by drawing an intuitive picture [47]. HMMs have proven its efficacy in predicting protein domain from protein sequence. Generally, the steps to generate a reliable HMM are (i) the retrieval of a training set to build the protein domain model and the training of the model, (ii) the retrieval of the testing set and (iii) the validation of the model.

In order to collect the proteins for building the model, Pfam identifier of MFS (PF07690) was searched in Protein Data Bank (PDB) [48], with some filters:

- the proteins have to be an X-ray experiment;
- resolution must be ≤ 3.5 Å;
- the proteins have to be wild-type;
- no mutations in the sequence.

36 proteins were retrieved. Since most of these proteins contain more than one chain and may be identical or very similar, PDBeFold were run in order to reduce the similarity of the training set. PDBeFold is an on-line tool that allows pairwise or multiple comparison and 3D alignment of protein structures [49]. The chain A of 1PW4 was used as template [28] and the rest of the 36-protein dataset as a query. From the output, 18 proteins remained. To avoid redundancy, skipredundant from EMBOSS was run. This tool automatically clusters proteins based on pairwise alignment by using Needleman-Wunsch global alignment algorithm [50] with 95% of identity. As a result, 9 proteins were retrieved, which list is reported in Table 2. The structural multiple sequence alignment (MSA) of these 6 structures, based on PDBeFold tool, was used as an input to create an HMM model.

Table 2. List of the structures used to train the HMM.

PDB id and Chain	Name of the Protein	Organism
1pw4:A	Crystal Structure of the Glycerol-3-Phosphate Transporter	<i>Escherichia coli</i>
6e9n:A	D-galactonate:proton symporter in the inward open form	<i>Escherichia coli</i>
7cko:A	Human MCT1/Basigin-2 complex in the presence of anti-cancer drug candidate 7ACC2 in the inward-open conformation	<i>Homo sapiens</i>
6kki:A	Drug:Proton Antiporter-1 (DHA1) Family SotB, in the inward-occluded conformation	<i>Escherichia coli</i> K-12
6kkj:B	Drug:Proton Antiporter-1 (DHA1) Family SotB, in the inward open conformation	<i>Escherichia coli</i> K-12
4zp2:A	Multidrug transporter MdfA in complex with n-dodecyl-N, N-dimethylamine-N-oxide	<i>Escherichia coli</i> K-12
6oop:A	Protein B	<i>Escherichia coli</i>
4zow:A	Multidrug transporter MdfA in complex with chloramphenicol	<i>Escherichia coli</i> K-12
6vs0:A	Protein B	<i>Escherichia coli</i>

In the first column, there are the PDB identifiers and the chain, in the second there is the name of the protein and in the third the organism.

The HMM was created by using HMMER v3.3.2 [51]. An HMM was created with command *hmmbuild* using the MSA file from PDBeFold. In order to validate the model, a positive and a negative set of protein sequences was retrieved from UniProt [52]. As positive set (PS) all the sequences that are (i) manually annotated and reviewed, (ii) 300–500 residues sequence length and (iii) with PFam id PF07690 were filtered. As negative set (NS), the same characteristics were used but filtering out PFam id PF07690 proteins. For both sets, UniRef50 [53] was applied to avoid redundancy. Then, the *blastall* command of BLAST [54] against the dataset of the sequences of the training set was used to avoid the bias in the testing procedure that could happen if the testing set contains the sequences used to train the model. Finally, the PS and NS contain 341 and 55,258 proteins, respectively. The UniProt identifiers of both sets are in Supplemental Materials Files S6 and S7 as fasta files. The command *hmmsearch* (with the option *-E 0.05*) was used on PS and NS to test the model:

this command allows the user to search one or more profiles against a sequence database and it gives the score and the E-value for both the whole sequence and the best matching domain. From the results, a statistical evaluation of the HMM was performed (Table 3).

Table 3. Confusion matrix of predicted and the actual MFS proteins.

Protein with MFS Domain		Predicted MFS Domain	
		Positive	Negative
	Positive	288 (TP)	53 (FN)
	Negative	54 (FP)	55,204 (TN)

In the table the results are reported as true positives, true negative, false positive and false negatives.

In Table 3, the confusion matrix is reported. In the matrix the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were reported.

True positive rate (TPR), true negative rate (TNR), accuracy (ACC) and Matthew Correlation Coefficient (MCC) were calculated. The definition of the statistical measures is in Supplemental Materials File S4 and the results are in Table 4.

Table 4. Scoring indices calculated on numbers of Table 3.

Scoring Index	Value
TPR	0.84
TNR	0.99
ACC	0.99
MCC	0.84

5. Conclusions

Certainly, two of the most important MFS proteins are GlpT and UhpT. Whereas the first is ubiquitous and is found in all bacteria, the second was not found in PA. However, they are likewise important for the FF transportation inside the cell and then, crucial in PA acquisition of FF resistance. However, no studies aimed at searching for distant related homologous of UhpT have been carried out. Since the high genetic exchange and high mutational rate in bacteria, it might as well search for an UhpT-like protein in the PA genome. In this study, in addition to GlpT, we found a second GlpT protein in 21 out of 21 PA, called *glpt_2*, which has characteristics to be an MFS. We had a search in the literature and in the common databases (e.g., UniProt) and we did not find any studies about *glpt_2* in PA. However, if we use BLAST to search *glpt_2* sequence, we can easily find it, although the information about its specific function is lacking. We searched for *glpt_2* sequence in the NCBI set of 433 whole-genome sequences and we found it in all genomes. Also, we found it even in double hits in some genomes. The information are in Supplemental Materials File S2.

Therefore, we tried to build a new protein structure by using homology-modeling approach through MODELLER. A UhpT crystal protein was retrieved from AlphaFold databases. The model of UhpT has some characteristics that suggest the function conservation. The binding site, comprehending two arginines (Arg39 and Arg264) and Lys40, is perfectly superimposable to UhpT protein as well as the overall topology. In fact, the 12 TMs are completely comparable, suggesting a well-defined folding of the protein across the bilayer lipid membrane. To enforce our hypothesis, in all 21 PA genomes, we also found a protein annotated as membrane sensor protein UhpC. UhpC is lacking in PA reference genomes, but it is crucial for UhpT expression and function in *E. coli*, where UhpT is functioning and well characterized. We computed the alignment between the UhpC of *E. coli* and the sequence found in one of our genomes and they share low identity (27.9%). Unfortunately, we were not able to find the other two proteins of the Uhp complex, namely UhpA and UhpB.

The model we have built is an attempt to find a distantly related homologue of a glucose-6-P channel of *E. coli*. Since the PA strains we used are considered as wild-type, we can assume that most of the PA have proteins like these. Obviously, we cannot say whether this protein is functional or if it is transcript as well. However, having this protein in the PA genome may prove that even PA can use glucose as a carbon source.

The presence of a homologue of UhpT suggests that this protein is conserved in PA genome because it is generally found in all Gram-negative bacteria; however, PA prefers different carbon sources. Moreover, glucose is the less preferred carbon source by PA, preferring other substrates, such as succinate and citrate [55,56]. Therefore, having a well-functioning UhpT probably does not affect the fitness of the PA in its habitat. On the other hand, since the mutation rate in bacteria is usually high, we cannot exclude that, in the absence of all favorite carbon source substrates, UhpT-functioning PA can be selected.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bacteria1040020/s1>.

Author Contributions: Conceptualization, T.O. and D.D.; methodology, T.O.; validation, T.O. and D.D.; formal analysis, T.O.; investigation, T.O. and D.D.; resources, T.O. and D.D.; data curation, T.O. and D.D.; writing—original draft preparation, T.O.; writing—review and editing, T.O. and D.D.; visualization, T.O. and D.D.; supervision, T.O. and D.D.; project administration, D.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the local ethics committee (Meyer Children’s Hospital, 27/2020) and informed written consent was obtained from parents of involved subject for the use of anonymous clinical data for research purposes.

Data Availability Statement: All data that support this paper are stored in Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lyczak, J.B.; Cannon, C.L.; Pier, G.B. Lung Infections Associated with Cystic Fibrosis. *Clin. Microbiol. Rev.* **2002**, *15*, 194–222. [CrossRef] [PubMed]
2. Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **2000**, *406*, 959–964. [CrossRef] [PubMed]
3. Poole, K. Outer Membranes and Efflux: The Path to Multidrug Resistance in Gram- Negative Bacteria. *Curr. Pharm. Biotechnol.* **2002**, *3*, 77–98. [CrossRef] [PubMed]
4. Ferrara, A.M. Potentially multidrug-resistant non-fermentative Gram-negative pathogens causing nosocomial pneumonia. *Int. J. Antimicrob. Agents* **2006**, *27*, 183–195. [CrossRef] [PubMed]
5. Rossolini, G.M.; Mantengoli, E. Treatment and control of severe infections caused by multiresistant *Pseudomonas aeruginosa*. *Clin. Microbiol. Infect.* **2005**, *11*, 17–32. [CrossRef]
6. Gupta, V. Metallo beta lactamases in *Pseudomonas aeruginosa* and *Acinetobacter* species. *Expert Opin. Investig. Drugs* **2008**, *17*, 131–143. [CrossRef]
7. Zhao, W.-H.; Hu, Z.-Q. β -Lactamases identified in clinical isolates of *Pseudomonas aeruginosa*. *Crit. Rev. Microbiol.* **2010**, *36*, 245–258. [CrossRef]
8. Ramirez, M.S.; Tolmasky, M.E. Aminoglycoside modifying enzymes. *Drug Resist. Update* **2010**, *13*, 151–171. [CrossRef]
9. Strateva, T.; Yordanov, D. *Pseudomonas aeruginosa*—A phenomenon of bacterial resistance. *J. Med. Microbiol.* **2009**, *58*, 1133–1148. [CrossRef]
10. Kahan, F.M.; Kahan, J.S.; Cassidy, P.J.; Kropp, H. The mechanism of action of fosfomycin (phosphonomycin). *Ann. N. Y. Acad. Sci.* **1974**, *235*, 364–386. [CrossRef]
11. Falagas, M.E.; Giannopoulou, K.P.; Kokolakis, G.N.; Rafailidis, P.I. Fosfomycin: Use Beyond Urinary Tract and Gastrointestinal Infections. *Clin. Infect. Dis.* **2008**, *46*, 1069–1077. [CrossRef] [PubMed]
12. Oliver, A.; Levin, B.R.; Juan, C.; Baquero, F.; Blázquez, J. Hypermutation and the Preexistence of Antibiotic-Resistant *Pseudomonas aeruginosa* Mutants: Implications for Susceptibility Testing and Treatment of Chronic Infections. *Antimicrob. Agents Chemother.* **2004**, *48*, 4226–4233. [CrossRef] [PubMed]

13. Falagas, M.E.; Kastoris, A.C.; Karageorgopoulos, D.; Rafailidis, P.I. Fosfomycin for the treatment of infections caused by multidrug-resistant non-fermenting Gram-negative bacilli: A systematic review of microbiological, animal and clinical studies. *Int. J. Antimicrob. Agents* **2009**, *34*, 111–120. [\[CrossRef\]](#)
14. Alper, M.D.; Ames, B.N. Transport of antibiotics and metabolite analogs by systems under cyclic AMP control: Positive selection of *Salmonella typhimurium* cya and crp mutants. *J. Bacteriol.* **1978**, *133*, 149–157. [\[CrossRef\]](#)
15. Castañeda-García, A.; Rodríguez-Rojas, A.; Guelfo, J.R.; Blázquez, J. The Glycerol-3-Phosphate Permease GlpT Is the Only Fosfomycin Transporter in *Pseudomonas aeruginosa*. *J. Bacteriol.* **2009**, *191*, 6968–6974. [\[CrossRef\]](#)
16. Olekhovich, I.; Dahl, J.L.; Kadner, R.J. Separate contributions of UhpA and CAP to activation of transcription of the uhpT promoter of *Escherichia coli*. *J. Mol. Biol.* **1999**, *292*, 973–986. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Karageorgopoulos, D.; Wang, R.; Yu, X.-H.; Falagas, M.E. Fosfomycin: Evaluation of the published evidence on the emergence of antimicrobial resistance in Gram-negative pathogens. *J. Antimicrob. Chemother.* **2012**, *67*, 255–268. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Nikolaidis, I.; Favini-Stabile, S.; Dessen, A. Resistance to antibiotics targeted to the bacterial cell wall. *Protein Sci.* **2014**, *23*, 243–259. [\[CrossRef\]](#)
19. Mirakhur, A.; Gallagher, M.; Ledson, M.; Hart, C.; Walshaw, M. Fosfomycin therapy for multiresistant *Pseudomonas aeruginosa* in cystic fibrosis. *J. Cyst. Fibros.* **2003**, *2*, 19–24. [\[CrossRef\]](#)
20. Okazaki, M.; Suzuki, K.; Asano, N.; Araki, K.; Shukuya, N.; Egami, T.; Uchimura, H.; Watanabe, T.; Higurashi, Y.; Morita, K. Effectiveness of fosfomycin combined with other antimicrobial agents against multidrug-resistant *Pseudomonas aeruginosa* isolates using the efficacy time index assay. *J. Infect. Chemother.* **2002**, *8*, 37–42. [\[CrossRef\]](#)
21. Pao, S.S.; Paulsen, I.T.; Saier, M.H., Jr. Major Facilitator Superfamily. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1–34. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Marger, M.D.; Saier, M.H., Jr. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem. Sci.* **1993**, *18*, 13–20. [\[CrossRef\]](#)
23. Maiden, M.C.J.; Davis, E.O.; Baldwin, S.A.; Moore, D.C.M.; Henderson, P.J.F. Mammalian and bacterial sugar transport proteins are homologous. *Nature* **1987**, *325*, 641–643. [\[CrossRef\]](#)
24. Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **2013**, *38*, 151–159. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Zhou, Y.; Jiang, X.; Kaback, H.R. Role of the irreplaceable residues in the LacY alternating access mechanism. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12438–12442. [\[CrossRef\]](#)
26. Dang, S.; Sun, L.; Huang, Y.; Lu, F.; Liu, Y.; Gong, H.; Wang, J.; Yan, N. Structure of a fucose transporter in an outward-open conformation. *Nature* **2010**, *467*, 734–738. [\[CrossRef\]](#)
27. Solcan, N.; Kwok, J.; Fowler, P.W.; Cameron, A.D.; Drew, D.; Iwata, S.; Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **2012**, *31*, 3411–3421. [\[CrossRef\]](#)
28. Huang, Y.; Lemieux, M.J.; Song, J.; Auer, M.; Wang, D.N. Structure and Mechanism of the Glycerol-3-Phosphate Transporter from *Escherichia coli*. *Science* **2003**, *301*, 616–620. [\[CrossRef\]](#)
29. Hall, J.A.; Maloney, P.C. Altered Oxyanion Selectivity in Mutants of UhpT, the Pi-linked Sugar Phosphate Carrier of *Escherichia coli*. *J. Biol. Chem.* **2005**, *280*, 3376–3381. [\[CrossRef\]](#)
30. Chen, S.-Y.; Pan, C.-J.; Lee, S.; Peng, W.; Chou, J.Y. Functional analysis of mutations in the glucose-6-phosphate transporter that cause glycogen storage disease type Ib. *Mol. Genet. Metab.* **2008**, *95*, 220–223. [\[CrossRef\]](#)
31. Rice, P.; Longden, I.; Bleasby, A. EMBOS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [\[CrossRef\]](#)
32. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Martí-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sánchez, R.; Melo, F.; Šali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5–6. [\[CrossRef\]](#)
35. Fiser, A.; Do, R.K.; Šali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773. [\[CrossRef\]](#)
36. Šali, A.; Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [\[CrossRef\]](#)
37. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [\[CrossRef\]](#)
38. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [\[CrossRef\]](#)
39. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [\[CrossRef\]](#)
40. Jessen-Marshall, A.E.; Brooker, R.J. Evidence That Transmembrane Segment 2 of the Lactose Permease Is Part of a Conformationally Sensitive Interface between the Two Halves of the Protein. *J. Biol. Chem.* **1996**, *271*, 1400–1404. [\[CrossRef\]](#)
41. Fann, M.-C.; Busch, A.; Maloney, P.C. Functional Characterization of Cysteine Residues in GlpT, the Glycerol 3-Phosphate Transporter of *Escherichia coli*. *J. Bacteriol.* **2003**, *185*, 3863–3870. [\[CrossRef\]](#) [\[PubMed\]](#)

42. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 May 2022).
43. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
44. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
45. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
46. Eddy, S.R. Hidden Markov models. *Curr. Opin. Struct. Biol.* **1996**, *6*, 361–365. [[CrossRef](#)]
47. Eddy, S.R. What is a hidden Markov model? *Nat. Biotechnol.* **2004**, *22*, 1315–1316. [[CrossRef](#)]
48. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]
49. Krissinel, E.; Henrick, K. Multiple Alignment of Protein Structure in Three Dimensions. In *Computational Life Sciences*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 67–68.
50. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [[CrossRef](#)]
51. HMMER. Available online: <http://hmmer.org/> (accessed on 1 May 2022).
52. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [[CrossRef](#)]
53. Suzek, B.; Wang, Y.; Huang, H.; McGarvey, P.B.; Wu, C.H. The UniProt Consortium UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31*, 926–932. [[CrossRef](#)]
54. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
55. Frimmersdorf, E.; Horatzek, S.; Pelnikovich, A.; Wiehlmann, L.; Schomburg, D. How *Pseudomonas aeruginosa* adapts to various environments: A metabolomic approach. *Environ. Microbiol.* **2010**, *12*, 1734–1747. [[CrossRef](#)] [[PubMed](#)]
56. Ng, F.M.-W.; Dawes, E.A. Chemostat studies on the regulation of glucose metabolism in *Pseudomonas aeruginosa* by citrate. *Biochem. J.* **1973**, *132*, 129–140. [[CrossRef](#)] [[PubMed](#)]