Article:

MICROBIOLOGY

RESEARCH ARTICLE
Drousiotis *et al.*, *Microbiology* 2023;169:001308
DOI 10.1099/mic.0.001308

MICROBIOLOGY SOCIETY

OPEN ACCESS

Characterization of the L-arabinofuranose-specific GafABCD ABC transporter essential for L-arabinose-dependent growth of the lignocellulose-degrading bacterium *Shewanella* sp. ANA-3

Konstantinos Drousiotis[1]†, Reyme Herman[1]†, Judith Hawkhead[1], Andrew Leech[2], Anthony Wilkinson[3] and Gavin H. Thomas[1,*]

Abstract

Microbes that have evolved to live on lignocellulosic biomass face unique challenges in the effective and efficient use of this material as food. The bacterium *Shewanella* sp. ANA-3 has the potential to utilize arabinan and arabinoxylan, and uptake of the monosaccharide, L-arabinose, derived from these polymers, is known to be mediated by a single ABC transporter. We demonstrate that the substrate binding protein of this system, GafA$_{Sw}$, binds specifically to L-arabinofuranose, which is the rare furanose form of L-arabinose found in lignocellulosic biomass. The structure of GafA$_{Sw}$ was resolved to 1.7 Å and comparison to *Escherichia coli* YtfQ (GafA$_{Ec}$) revealed binding site adaptations that confer specificity for furanose over pyranose forms of monosaccharides, while selecting arabinose over another related monosaccharide, galactose. The discovery of a bacterium with a natural predilection for a sugar found abundantly in certain lignocellulosic materials suggests an intimate connection in the enzymatic release and uptake of the sugar, perhaps to prevent other microbes scavenging this nutrient before it mutarotates to L-arabinopyranose. This biological discovery also provides a clear route to engineer more efficient utilization of plant biomass components in industrial biotechnology.

INTRODUCTION

The acquisition of carbon and energy sources is essential for all microbial cells to grow. In bacteria, sugars are known to be important sources of these nutrients, which must be first transported from the environment into the bacterium. These processes have been studied for decades in bacterial systems and many systems have been characterized in detail, demonstrating that a diverse range of systems can be used for the same substrates [1–3]. In the model bacterium *Escherichia coli* transporters are known for a wide range of different monosaccharides, with glucose being the preferred carbon source and then a range of others being used such as D-galactose, L-arabinose and D-xylose [4]. High-affinity transport can be mediated by ATP-binding cassette (ABC) transporters that use a substrate binding protein (SBP) to recognize the sugar with usually low or sub-micromolar affinity [5, 6]. Structures of these SBPs bound to different sugars have been known for many years and in every case the sugar adopts the pyranose (six-membered ring) form that is the most abundant in solution [6]. This includes structures of the D-galactose and L-arabinose SBPs, MglB and AraF, bound to D-galactopyranose and L-arabinopyranose, respectively [7–10].

A surprising discovery was made in 2009 with the finding that *E. coli* contains a second ABC transporter that handles both D-galactose and L-arabinose [11], although with lower affinity than either of the two well-known ABC systems. The crystal

structure of the SBP, YtfQ, resolved this apparent paradox with the discovery that YtfQ bound D-galactofuranose (PDB: 2VK2) – the rare furanose form of the sugar that constitutes about 8% of the D-galactose pool in solution (the other 92% being D-galactopyranose that is recognized by MglB) [11]. This unique binding feature was confirmed by NMR experiments, and the coordinated expression of the *ytfQRTyjfF* and *mglABCD* operons by the galactose responsive transcription factors GalR and GalS suggested that *E. coli* expresses both transporters to capture all the free D-galactose present in the environment [11]. That study was hence the first to recognize that for a particular monosaccharide, there could be transporters specific for either the pyranose or the furanose forms, which due to their very different chemical shapes would dictate separate and distinct systems for each form. Shortly afterwards the discovery of a D-ribofuranose transporter [12] in bacteria already known to have a D-ribopyranose transporter supported this hypothesis, prompting a broader consideration of this question for other sugars found in nature [6].

One of the most abundant uses of furanose forms of sugars is in lignocellulose, specifically components of hemicellulose that contain L-arabinose, such as arabinoxylan, arabinan and rhamnogalacturonan-II [13–16]. In these complex polymers the L-arabinose is exclusively in the L-arabinofuranose (L-Ara*f*) form [17, 18] and after release from these glucans by the action of L-arabinofuranosidases [19] will be present in the L-Ara*f* form for a short time before spontaneous chemical interconversion to the pyranose form, which forms the vast majority (around 90%) at chemical equilibrium [20].

YtfQ is also able to bind L-arabinose, in fact with an apparent higher affinity than D-galactose, and later work has demonstrated that the operon *ytfQRTyjfF* is regulated by the L-arabinose responsive transcription factor AraC [21]. This led us to speculate that the uptake of L-Ara*f* in bacteria could be of physiological significance and we noted that the YtfQ homologue in *Mycobacterium smegmatis*, MSMEG_1712, was encoded with L-arabinose-degrading genes [11]. This hypothesis has been supported in an elegant study that demonstrated that MSMEG_1712 is an L-Ara*f* and D-Gal*f* binding protein [22]. The absence of this protein in the related pathogen *Mycobacterium tuberculosis* but its presence in the soil-dwelling *M. smegmatis* suggests a physiological function outside of the host in a soil environment rich in plant and fungal biomass-derived sugars [23]. Given the confirmed functions of YtfQ proteins we now propose to rename YtfQ and its homologues as GafA for their galacto- and arabinofuranose binding function and the *ytfQRTyjfF* operon to be *gafABCD* following nomenclature for other ABC transporters. In this study we identify GafA homologues in a range of bacteria and focus on the function of these systems in the use of lignocellulosic sugars. Specifically, we focus on two uncharacterized GafA proteins from soil bacteria, namely *Sinorhizobium meliloti* (GafA$_{Sm}$) and the lignocellulosic degrader *Shewanella* sp. ANA 3 (GafA$_{Sw}$), as published genetic evidence demonstrates that the latter bacterium is dependent on this uncharacterized transporter for growth with L-arabinose as a sole carbon source [24, 25].

METHODS

Identification of *gafAs* in bacterial phyla

The repertoire of GafA proteins was explored initially using SEED viewer and MicrobesOnline [24], and organisms with interesting linked genes were identified, including some from biotechnologically important genomes (included in Fig. 1a). A larger dataset using the protein sequences of these YtfQ/GafA proteins was collected using BLASTP analysis. The lowest scoring GafAs from α, β, γ and δ proteobacteria as well as *Firmicutes*, *Spirochaetes* and *Actinobacteria*, which still presented at least 45% or higher identity to the query protein sequence, were inputted on new BLASTP searches. The fresh searches were intended to collect more GafAs and expand the collection by restricting the search for each bacterial orthologue to its respective phylum or class. The process was repeated until previous results were observed again so that a non-redundant set of GafAs was produced. To avoid creating large and impractical clades, only one bacterium from each species was included in the downstream phylogenetic analysis. The identified protein sequences, which amounted to 110, were used to produce a phylogram as described below.

Phylogenetic tree reconstruction

Multiple sequence alignments and homology searches were performed using the online sequence analysis software MAFFT [26] using the L-iNS-I refinement method or Clustal Omega [27]. The sequence alignments were curated in the Gblocks 0.91b tool of the phylogeny.fr [28]. The alignments were inputted into PhyML 3.0 [29] for automatic model selection using the Akaike information criterion (AIC). The AIC estimates the relative quality of statistical models and chooses one based on the quantity of information lost when a given method is used to represent the process that generated the data. The substitution model parameters as calculated by the AIC are described in Table 1 and their use in the phylotrees constructed is designated. Reconstruction of the phylograms was performed in PhyML 3.0 using the maximum-likelihood method based on the substitution model and the parameters calculated by the AIC. The confidence of the branches was inferred using 500 bootstrap replications. The resulting newick files were inputted into the Interactive Tree Of Life (iTOL) to display and annotate the trees (http://itol.embl.de) [30]. Further annotations were made in the scalable vector graphics editor, Boxy SVG.

Table 1. Data collection and refinement statistics

Gaf$_{SW}$-L-Ara*f* (5OCP)			
Diffraction Source, λ (Å)	DLS i03, 0.9795	**Refinement***	
Space group	P2$_1$2$_1$2$_1$	Refinement programme	REFMAC 5.8.0158
a, b, c (Å)	73.92, 86.33, 87.28	Resolution range (Å)	43.45–1.70
α, β, γ (°)	90.00, 90.00, 90.00	No. of reflections (working set)	58 975 (4272)
Resolution range (Å)*	43.64–1.70 (1.73–1.70)	No. of reflections (test set)	3081 (231)
Total Reflections	62122	Completeness	99.9 (99.7)
R_{merge}	0.082 (0.597)	Final R_{work}	0.167 (0.218)
I/σ(I)	3.77	Final R_{free}	0.204 (0.261)
Data redundancy	7.4 (7.1)	Wilson B-factor (A²)	18.0
		Number non-H ofatoms	
		Protein	4618
		Ligand	92
		Solvent	461
		RMS deviations	
		Bond lengths (Å)	0.02
		Bond angles (°)	2.03
		Average B factors (Å²)	
		Protein	23.36
		Ligand	33.05
		Solvent	34.85
		Ramachandran plot (favoured/ allowed/outlier)	93.0/6.6/0.4

*Values in parentheses refer to the outer resolution shell 1.73–1.70 Å.

Functional annotations of the genes involved in the arabinose, galactose and xylose metabolism and related pathways were derived using Microbes Online (http://www.microbesonline.org/); PATRIC 3. 4. 2 (https://www.patricbrc.org) and Biocyc (https://biocyc.org/).

Cloning of substrate binding proteins

The gene sequence for *shewana3_2073* was codon optimized in JCat and synthetically produced by IDT. The synthetic gene was cloned into pET20b vector using Gibson assembly. The *smb21587* gene sequence was amplified from genomic DNA of *S. meliloti* 1021 using Q5 DNA polymerase (NEB) with the following primers: GafASmF CATGCCATGGCCGAACTCGT CGTCGGCTTT and GafASmR CCGCTCGAGGTAGCCGAGGCCTTTCTTTTCTTCG. The PCR product was digested with *Nco*I and *Xho*I to allow cloning into pET20b. The sequence of the cloned genes was confirmed by DNA sequencing (Sigma Oligo) using the T7 and T7 terminator primers.

Expression and purification of SBPs

Expression of the proteins was optimal at 20 °C after 20 h of growth following induction with 1 mM IPTG. Expression was performed with cells grown either in Lysogeny broth (LB) or Terrific broth (TB). The periplasmic fraction was isolated by osmotic shock [31] and loaded in the Ni-nitrilotriacetic acid column for purification. Fractions were pooled and dialysed in PBS. Protein levels were quantified on an Epoch Microplate Spectrophotometer.

For preparation of ligand-free protein, the filtered periplasmic extract was injected into a pre-equilibrated His-tag column. The proteins were washed with a decreasing gradient of guanidine hydrochloride (4–0.5 M) in washing buffer (i.e. 50 mM Tris, 200 mM NaCl and 25 mM imidazole, pH 7.5). The column was further treated with refolding buffer (i.e. washing buffer mixed with

500 mM arginine monohydrochloride, 4 mM reduced glutathione and 0.4 mM oxidized glutathione, pH 7.5). Finally, the protein was eluted using elution buffer (i.e. washing buffer mixed with 20% glycerol and 500 mM instead of 25 mM imidazole, pH 7.5).

Circular dichroism (CD)

Spectra were obtained using a J-810 spectropolarimeter (Jasco) controlled by the supplied software SpectraManager version 1.53.00 (Jasco). Proteins were dialysed into 50 mM NaF and 20 mM Tris-HCl pH 7.5 and diluted to a concentration of 10 µM. The spectrum was recorded at 20 °C (Peltier temperature controller) in a 1 mm pathlength quartz cuvette (Starna) between 180 and 240 nm at 100 nm min^{-1} with 1 nm pitch. The molar elipticity (θ) data obtained were corrected by subtracting the buffer control and were plotted against the wavelength (nm) in GraphPad Prism 7.0.

Crystallization and structural determination of GafA$_{Sw}$

The Hydra-96 Microdispenser (Robbins Scientific) and Mosquito Crystal (TTP Labtech) robots were used to dispense commercially sourced crystallization solutions and the protein–ligand solution containing GafA$_{Sw}$ with L-arabinose. The crystallization screens were performed in the vapour-diffusion sitting–drop format with a mix of 150 nl crystallization solution and 150 nl GafA$_{Sw}$ and L-arabinose at final concentrations of 8 mg ml^{-1} and 1.25 mM, respectively. The crystallization conditions that produced the crystal used to resolve the structure of this protein was 0.2 M ammonium nitrate, pH 6.2, and 20% PEG 3,350 (B7 of the PEG/ION HT tray; Hampton Research). This crystal was harvested from the sitting–drop and coated in a solution of the aforementioned crystallization solution supplemented with glycerol as the cryo-protectant, to a final concentration of 20% (v/v).

Diffraction data were collected to a resolution of 1.7 Å using the macromolecular crystallography (MX) Beamline I03 at the Diamond Light Source, Hartwell Science and Innovation Campus, and processed with DIALS. The space group and cell dimensions (Table 1) were consistent with two GafA$_{Sw}$ chains in the asymmetric unit giving a Matthews coefficient of 2.1 Å^3 Da^{-1} and a solvent content of 42.04%. The CCP4i2 suite of programs was used to scale the data and resolve the structure. The processed DIALS data were scaled using AIMLESS, POINTLESS, Ctruncate and FreeRflag [32, 33]. The structure was resolved by molecular replacement using MOLREP [34, 35] using the coordinates for an *E. coli* D-galactofuranose binding protein, YtfQ (PDB: 2VK2), as the search model. As expected two solutions were obtained. The initial model built by MOLREP was then refined using iterative cycles of REFMAC5 [35] followed by manual model building in COOT [36] until convergence was reached. The final structure was then deposited with the PDB ID 5OCP via wwPDB [37].

For structural comparisons between the furanose binding proteins we included the GafA$_{Ms}$ bound to L-Ara*f* (PDB: 6HBM). This has two different binding positions for L-Ara*f* in the two monomers in the asymmetric unit. The terminal carbon of the ligand in one monomer is next to Glu28 and the other is next to Asp105, suggesting that the ligand is rotated 180° about the plane of the furan ring. Using our structure of GafA$_{Sw}$ bound to L-Ara*f* (PDB: 5OCP), we determined that the probable orientation of L-Ara*f* in GafA$_{Ms}$ is in chain A of the PDB structure with the terminal carbon being coordinated by Asp105, which is more consistent with the L-fucofuranose (PDB: 6HYH) and L-galactofuranose (PDB: 6HBD) bound structures. Hereafter, we used chain A from 6HBM for structural comparisons.

Differential scanning fluorimetry (DSF)

Protein–ligand solutions were dispensed into the wells of a 96-well thin wall PCR plate (ie. Genomic Fast Optical 0.1 ml plates). Each well contained 3–5 µM protein, 0.1× (2.5 µl of 0.8×) SYPRO orange and appropriate volumes of ligands. Different stocks of the ligands were prepared, so that 2 µl was transferred in each well to reach desired concentrations of 0.6, 6, 60, 600 and 1200 µM with four repeats of each ligand concentration for each protein. A final well volume of 20 µl was filled up by the protein's buffer. For each run, there were eight reference wells where ligands were excluded. Also, four reference wells were included that contained only buffer. The plates were sealed with optical sealing tape (Bio-Rad). The instrument run was set as instructed in the manual (Applied Biosystems). The plate was heated to 99 °C in increments of 0.5 °C. Data analysis was undertaken using Microsoft Excel. ΔT_m calculations were made by subtracting the T_m of the mean protein-only well with the data with added ligands. We note that the ratio of the furanose/pyranose forms of the monosaccharides will probably change slightly during the temperature ramp, although this should not impact on the conclusion of the experiment profiling potential ligands.

Fluorescence spectroscopy

Tryptophan and tyrosine fluorescence spectroscopy was performed using a FluoroMax 4 fluorescence spectrometer (Horiba Jobin-Yvon) with a water bath for temperature control. The maximum emission for each protein was determined by spectral analysis, i.e. excitation at wavelengths of 280, 295 and 297 nm at slit widths equal to 3 nm. Kinetic experiments for quantification of the binding affinities were performed with purified protein at concentrations ranging from 0.5 to 1.5 µM in PBS, pH 7.5. Total volume of the sample was 3 ml, which was excited at 280 nm for GafASm and 297 nm for GafASw with slit widths of 3 nm. Emission was monitored at 330 (GafASw) or 342 nm (GafASm) with slit widths of 3 nm. The fluorometer operated in a time-based acquisition mode with a run time of 360–500 s and an integration time of 1 s. Increasing concentrations of ligand were added to the protein
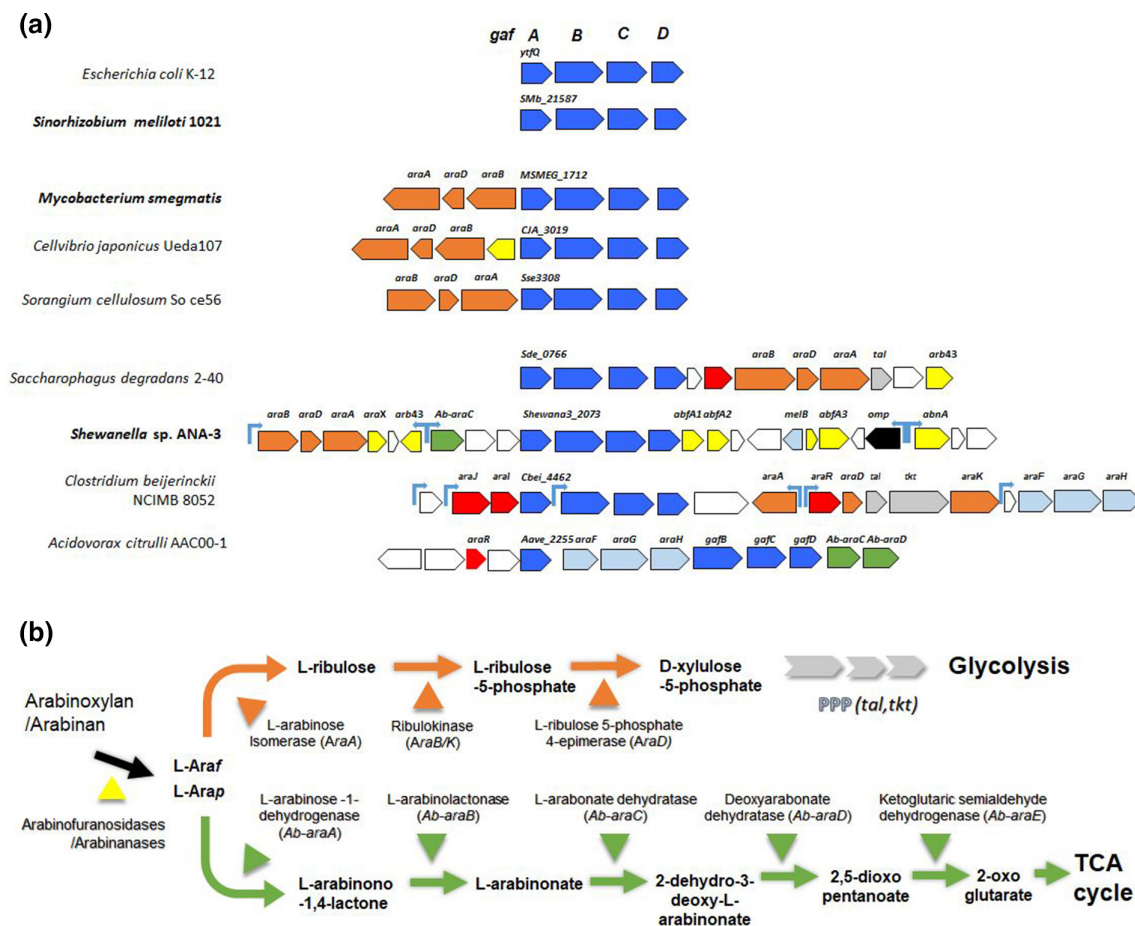
Fig. 1. (a) Genetic context of *gafABCD* genes encoding known and candidate galacto- and arabino-furanose transporters, highlighting linkage to arabinose catabolic gene clusters in soil-dwelling microbes. The locus tag for the Gaf SBP subunit is indicated. Genes encoding Gaf ABC transporters are in blue, while those for other sugar ABC transporters are in light blue and a possibly related porin in black. Catabolic genes in the AraBAD pathway are in orange and the alternative *A. brasiliense* system in green. Known and probable transcriptional regulators are in red. Arabinases and L-arabinofuranosidases are in yellow, while genes encoding pentose phosphate pathway components are in grey. Genes in the clusters with no strong functional predictions are in white. The light blue arrows indicate promoters regulated by AraJ/R-like regulators (from RegPrecise). (b) Simple schematic showing the steps in the AraBAD (orange) and *A. brasiliense* (green) pathways for L-arabinose catabolism.

solution and fluorescence change was monitored. The cumulative fluorescence change for each timepoint was plotted against the cumulative concentration of ligand in SigmaPlot 11 and the K_D was calculated from the hyperbolic fit of the binding curve.

Isothermal titration calorimetry

Calorimetry experiments were performed in the VP-ITC instrument (MicroCal, GE Health Sciences). Proteins and ligands were dissolved in the same buffer in each experiment (i.e. PBS or NaCl). The concentration of the protein in the cell ranged from 50 to 120 μM, according to the *c* value, where $c=[\text{protein}]/(\text{predicted})\ K_D$. The ligand concentration in the syringe was in 10- or 7-molar excess compared to the protein. Experiments were carried out in PBS or Tris-HCl (NaCl), pH 7.5 at 25 °C. The solutions in the cell and syringe were both degassed at 20 °C for 5 min before use. A typical run included 26 titrations, each one delivering 10 μl s^{-1} ligand with 240 s delay between injections. The acquired raw titration data were analysed in MicroCal Origin 7 software where binding isotherms were fitted by an iteration process using the one-set of sites model.

RESULTS

Gaf ABC transporters are present in diverse bacteria and often co-occur with arabinan/L-arabinose catabolic genes

To extend the study of Gaf ABC transport systems beyond the characterization of GafA$_{Ec}$ (YtfQ), we identified around 100 homologues in other bacteria using SEED and MicrobesOnline [24, 38]. These were found in several biotechnologically important
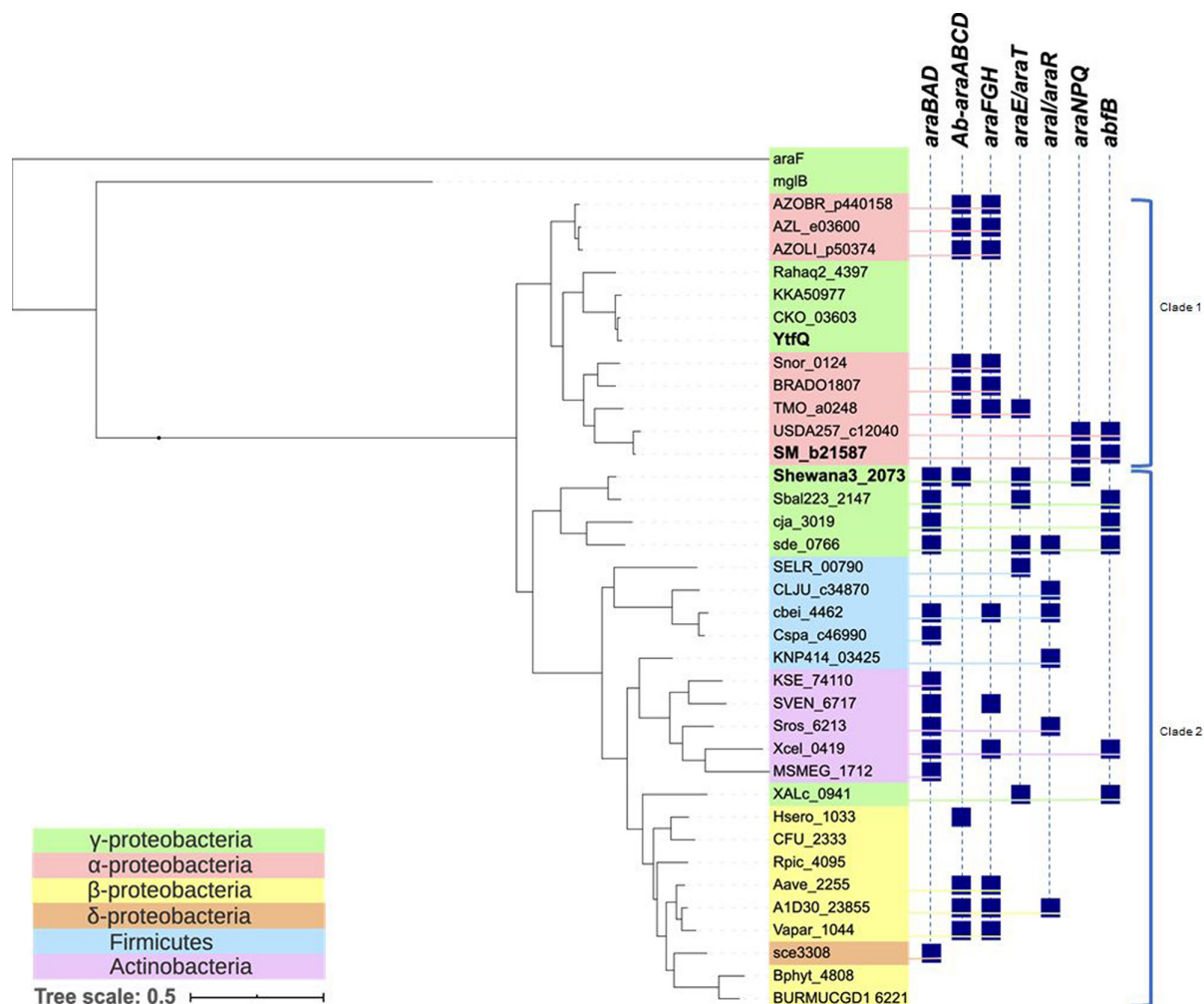
Fig. 2. Phylogeny of GafA proteins. Maximum likelihood analysis of 36 selected GafA proteins, with AraF and MglB as outgroups, with the taxonomic position of the organisms coloured according to the inset key. To the right is a mapping of presence/absence of genes related to arabinose transport and utilization (±20 kb of the *gafA* gene). Genes/operons code for: *araBAD*=primary arabinose catabolism, *Ab-araABCD*=arabinose catabolic pathway III from *A. brasiliense*, *araFGH*=arabinopyranose ABC transport, *araE/araT*=arabinose secondary transport, *araI/araR*=arabinose, *araNPQ*=arabinosides ABC transport, *abfB*=α-ʟ-arabinofuranosidase. Protein names are based on locus tags and can be identified in UniProt.

bacteria and plant symbionts with representatives from diverse phyla (Fig. 1a). While some resembled the *E. coli* system, in not being linked to any related catabolic genes, it was notable that many orthologues were encoded within larger clusters containing genes involved in ʟ-arabinose catabolism, most often the classical *araBAD* encoded route via ʟ-ribulose (Fig. 1b). This included the previously noted *M. smegmatis* system and other bacteria that are likely to encounter ʟ-arabinose in their soil environment, such as *Cellvibrio japonicus* Ueda 107 (γ-proteobacteria) and *Sorangium cellulosum* So ce56 (δ-proteobacteria). More complex examples were also seen in the γ-proteobacteria *Saccharophagus degradans* and *Shewanella* sp. ANA-3, as well as *Clostridium beijerincki* NCIMB 8052 (*Firmicutes*) and *Acidovorax citrulli* AAC00-1 (β-proteobacteria). These bacteria contain extended catabolic clusters including genes encoding secreted ʟ-arabinofuranosidases (yellow symbols in Fig. 1) and in some cases genes for an alternative oxidative non-phosphorylative arabinose catabolic pathway discovered in *Azospirillum brasiliense* that produces 2-oxo-glutarate [39] (Fig. 1b). Intriguingly the *Clostridium beijerinckii* cluster also contains genes encoding a typical ʟ-arabinopyranose type ABC transporter (*araFGH*) as well as the putative GafABCD system suggesting an evolutionary advantage to expressing transporters together to access all available ʟ-arabinose, and in *Acidovorax citrulli* AAC00-1 the pyranose transporter genes interpose with the genes for the predicted furanose transporter (Fig. 1a). It is notable that there is no direct link to any ᴅ-galactose catabolic genes, while linkage to ʟ-arabinose catabolism is found in many cases in strongly suggestive biological contexts.

Examining the GafA protein sequences themselves and their phylogenetic relationships (Fig. 2), the sequences fall into two nominal clades, when using the related AraF and MglB proteins, which bind ʟ-arabinopyranose and ᴅ-galactopyranose, respectively, as outgroups. When the presence of other genes in the same cluster are added, the split into Clade I and Clade II is reinforced,

Table 2. DSF data for GafA$_{Sm}$ and GafA$_{Sw}$

NC indicates no apparent change in ΔT_m upon ligand addition. Ratios of furanose/pyranose are taken from literature values at the indicated temperatures [20].

Monosaccharide	Furanose/pyranose ratio (%) at indicated temp.	ΔT_m (°C) at 1.2 mM [ligand]	
		GafA$_{Sm}$	GafA$_{Sw}$
L-Arabinose	8.5/91.5 (31 °C)	14.4±0.2	6.5±0.2
D-Galactose	6/94 (31 °C)	11.6±0.2	NC
D-Fucose	5/95 (31 °C)	11.8±0.2	2.1±0.2
D-Allose	8.5/91.5 (31 °C)	1.0±0.1	5.0±0.1
D-Talose	29/71 (28 °C)	3.4±0.2	NC
D-Glucose	0.3/99.7 (31 °C)	NC	NC

particularly by the *araBAD* genes, encoding the classical pathway to L-arabinose catabolism, which are exclusively seen linked to Clade II GafA systems (Fig. 2). While other genes relating to L-arabinose catabolism are more prevalent in Clade II clusters, this grouping is not exclusive, with some linkage to Clade 1 clusters. Also, when the phylogeny of the organisms is overlaid on the tree (Fig. 2), clade I contains only sequences from proteobacteria (α and γ proteobacteria) and includes GafA$_{EC}$ (YtfQ) that binds D-galactofuranose in addition to L-Ara*f*. Together these *in silico* data suggest that in some bacteria GafA might have evolved to confer not just specificity for furanose forms of common monosaccharides, but also selectively for L-Ara*f*.

Characterization of GafA proteins previously implicated in L-arabinose uptake

The genome context analysis revealed multiple Gaf systems encoded along with other genes related to L-arabinose catabolism. To help direct the experimental study, we sought further bioinformatic and experimental support before deciding which systems to characterize biochemically. The RegPrecise tool that predicts regulons in diverse bacteria was used to investigate predicted arabinose-induced regulons in the aforementioned bacterial candidates [40]. In *Clostrdium beijerinki* NCIMB 8052 there is a predicted AraR regulon with binding sites at six locations, all contained within the large gene cluster identified in Fig. 1(a). A similar prediction for the *Shewanella* sp. ANA-3 AraR regulon identifies five locations, again all within the large cluster indicated in Fig. 1(a), strongly supporting the hypothesis that these Gaf transporters are encoded as part of L-arabinosespecific regulons.

We then used the Fitness browser of MicrobesOnline to investigate the effect of *gafA* disruption caused by transposon insertion [24]. The deletion of *shewana3_2073* caused a notable decrease in fitness of *Shewanella* sp. ANA-3 when grown on L-arabinose as the sole carbon source [25]. Related data from Rodionov *et al.* [41] showed that *Shewanella* sp. ANA3 is unable to use galactose as the sole carbon source, consistent with a lack of a likely secondary transporter for galactose uptake, i.e. GalP. Together these data suggest a role for the Gaf system in L-arabinose utilization. Another example examined was the *gaf* system from *Sinorhizobium meliloti*. There is independent evidence that expression of the genes encoding this uncharacterized system are induced by the presence of L-arabinose, L-fucose and D-talose in the growth media [42]. Based on the evidence above, we chose Smb_21587 (GafA$_{Sm}$) from *Sinorhizobium meliloti* and Shewana3_2073 (GafA$_{Sw}$) from *Shewanella* sp. ANA-3 to analyse further, representing examples from Clade I and Clade II (Fig. 2), both of which have published supporting experimental evidence.

Coding sequences were cloned into pET20b and the recombinant proteins were expressed in *E. coli* as fusion proteins with a C-terminal hexahistidine tag, which were then purified using nickel affinity chromatography (Fig. S1A, available in the online version of this article). To remove pre-bound ligand and hence make the protein suitable for measuring ligand binding, the proteins were treated to an on-column unfold/refold step as used previously for GafA$_{Ec}$ and other SBPs [11, 43]. To confirm the protein had refolded, the overall structure was assessed using CD analysis, demonstrating a similar pattern of secondary structure for the refolded and native proteins (Fig. S1B).

GafA$_{Sm}$ is a typical GafA protein binding both D-galactose and L-arabinose

To study the ligand binding range of a Clade I GafA with a simple gene layout (Figs 1a and 2), we used a DSF method[44], where the binding of a ligand is reflected by the thermal stabilization of the protein. The addition of L-arabinose produced the greatest increase in melting temperature (ΔT_m of +14.4 °C at 1.2 mM ligand) while D-galactose and D-fucose also produced pronounced shifts (ΔT_m of +11.6 and +11.8 °C respectively, at 1.2 mM ligand) (Table 2, Fig. S3A), a pattern very similar to that observed previously with GafA$_{Ec}$ [11]. Consistent with the binding profile of GafA$_{Ec}$, the addition of D-talose and D-allose to GafA$_{Sm}$ produced small ΔT_m values and D-glucose failed to cause a change in thermal stability, even at higher concentrations (data not shown). The latter could be attributed to the negligible concentrations of the furanose form of glucose (<0.5%) due to its unstable nature, which differs from the other sugars that can be up to around 30% in the furanose form at equilibrium (Table 2 and Fig. S2). We
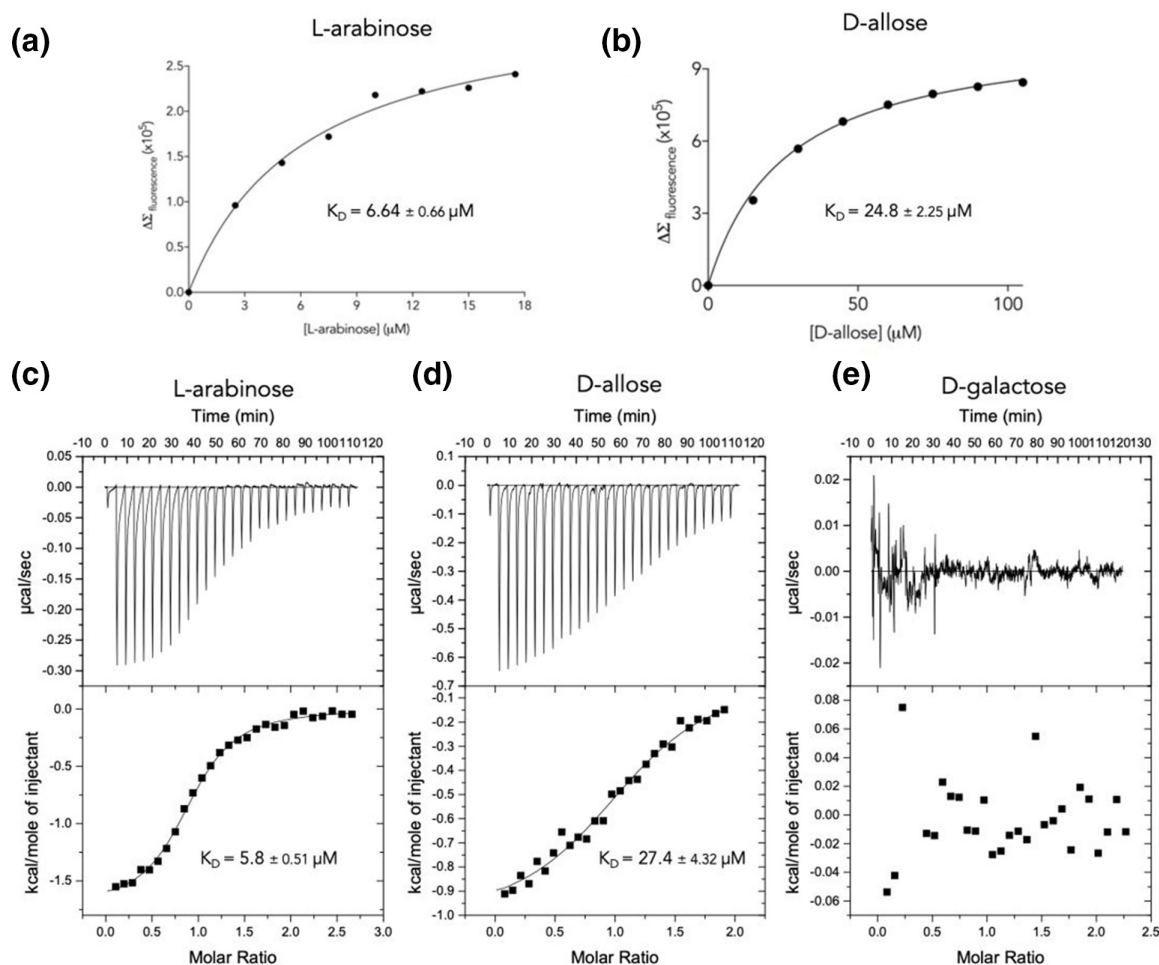
Fig. 3. Intrinsic fluorescence and ITC analysis of ligand binding to GafA$_{Sw}$. Plot of fluorescence changes against cumulative concentration of (a) L-arabinose and (b) D-allose . The binding affinity (K_D) was calculated using the one-site specific binding model. ITC experimental data for the binding of (c) L-arabinose, (d) D-allose and (e) D-galactose to GafA$_{Sw}$. The top panel (thermogram) represents the heat differences upon each injection of ligand and the lower panels (isotherm) show integrated heats of injection (■). The best fit (solid line) was calculated using the one-site binding model using the Microcal Origin software. Affinity was calculated using data from three different replicates for each ligand.

conclude the GafA$_{Sm}$ has similar properties to GafA$_{Ec}$, i.e. it binds to both D-galactose and L-arabinose and probably recognizes both of these sugars during its normal function, which is probably a conserved feature of Clade I GafA proteins that are not usually genetically linked to L-arabinose catabolic genes.

GafA$_{Sw}$ binds L-arabinose selectively over D-galactose

The GafA protein from Shewanella sp. ANA-3 sits in Clade II and its ligand binding specificity was similarly investigated using DSF analysis (Fig. S3 and Table 2). Again, L-arabinose gave the greatest increase in thermal stability of 6.5 °C, suggesting that it binds with the highest affinity. D-Allose caused the second highest change in thermal stability (ΔT_m of +4.9 °C). The addition of D-fucose, another binder of GafA$_{Ec}$ and GafA$_{Sm}$, resulted in a weaker increase in T_m (2.1 °C). We also found that D-talose, which could be weakly bound to GafA$_{Ec}$ and GafA$_{Sm}$, did not result in any changes in thermal stability of GafA$_{Sw}$. However, most significantly, D-galactose did not cause any detectable thermal shift, a unique feature of GafA proteins examined to date.

To validate these initial binding data using a more quantitative method, we assessed the binding of D-allose, L-arabinose and D-galactose to GafA$_{Sw}$ using intrinsic fluorescence spectroscopy and isothermal titration calorimetry (ITC). First, the changes in the intrinsic fluorescence of GafA$_{Sw}$ in the presence of a ligand were titrated with increasing concentrations of the various ligands. Through this assay, the K_D value for L-arabinose binding was determined to be 6.64±0.66 µM, with that for D-allose being 24.8±2.25 µM (Fig. 3a), consistent with the magnitude of the changes seen by DSF. These findings were further validated using ITC with the K_D calculated for L-arabinose and D-allose binding being 5.8±0.51 and 27.4±4.32 µM, respectively (Fig. 3b). Again, we were unable to detect any changes in intrinsic fluorescence when D-galactose was added to the protein at any concentrations,
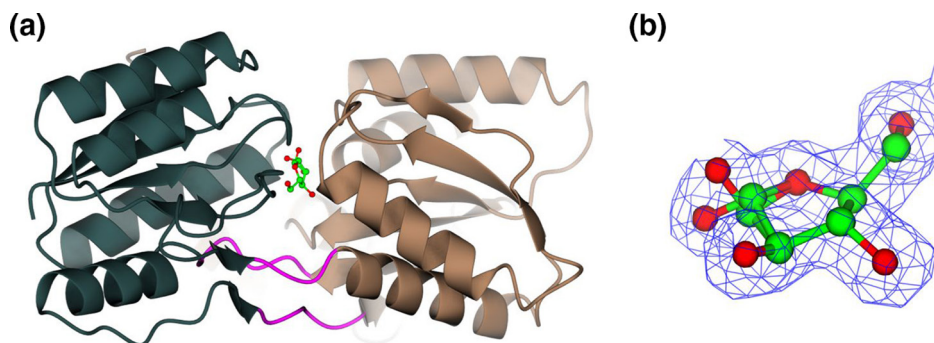
Fig. 4. GafA$_{Sw}$ is capable of binding both anomers of L-arabinofuranose. (a) GafA$_{Sw}$ bound to L-arabinofuranose with domain I in slate grey, domain II in dark brown and the hinge region in magenta. (b) The combined ligands of α-L-arabinofuranose and β-L-arabinofuranose. The blue chickenwire represents the experimentally derived composite electron density map (2Fo-Fc) of the ligand contoured to 1.0 RMSD.

nor detect binding by ITC (Fig. 3e). Hence, GafA$_{Sw}$ appears to have unique features in selecting L-arabinose over D-galactose that might relate to its biological function.

The structure of GafA$_{Sw}$ reveals bound L-arabinofuranose

To explore further the molecular basis of L-arabinose selectivity in GafA$_{Sw}$, we determined its 3D structure using X-ray crystallography. Purified GafA$_{Sw}$ was co-crystallized with L-arabinose and the structure of the 'closed' ligand bound form was resolved to 1.7 Å (Fig. 4a, Table 1). Two monomers with a root-mean-square deviation (RMSD) of 0.29 Å were observed in the asymmetric unit. The structure has been deposited in the PDB as 5OCP. Like other Cluster B SBPs, the protein contains two domains, each made up of a β-sheet comprising six strands surrounded by four or six α-helices, which are held together with a typical hinge region [6, 45, 46]. The most similar proteins identified at the structural level using the DALI server were the D-galactofuranose bound GafA$_{Ec}$ and the L-arabinofuranose and D-galactofuranose bound forms of GafA$_{Ms}$ (Fig. S4), which using the secondary-structure matching (SSM) algorithm [37] on CCP4mg [47] displayed calculated RMSDs of 1.01, 1.08 and 1.11 Å respectively.

Upon completing the refinement of the protein structure, we identified a significant residual feature in the electron density maps in an enclosed pocket sandwiched between the two protein domains. All possible forms of L-arabinose that were present in the crystallization mix were tested but the density present in the binding site only matches to the L-Ara*f* form (Fig. 4b). Further refinement revealed the presence of both the alpha- and beta-anomers of the sugar in approximately 50% occupancy. The accommodation of alpha- and beta-anomers of sugars is a recurring feature in SBPs first described for L-arabinopyranose binding to arabinose binding protein of *E. coli* [48] and more recently observed for D-galactose binding to GafA$_{Ec}$ [11].

Extensive ligand binding site similarities in GafA proteins

With the structures of GafA$_{Ec}$ and GafA$_{Ms}$ having high levels of overall similarity to GafA$_{Sw}$, we examined the ligand binding sites in more detail to identify commonalities and differences (Figs 5 and S6). There is remarkable conservation of all the amino acids in the binding sites of all three GafA proteins. Both the bottom and the top of the sugar ring are held through two aromatic residues, a Phe on the bottom side and a Trp on the top side, a common feature seen in other monosaccharide binding proteins [6]. Notably while the residues on the top binding surface are absolutely conserved at the sequence level, they adopt subtly different positions in the binding site and in how they coordinate the sugar.

The positioning of a conserved aspartate appears to dictate sugar selectivity in GafA proteins

We examined the role of the Asp residue (Asp110 in GafA$_{Sw}$), and its equivalents in our structure and the other GafA proteins. This positioning of the residue has previously been recognized as a key determinant of sugar recognition in GafA$_{Ec}$ as the salt bridge formed between Asp111 and Arg38 increases the size of the binding site allowing the more extended D-galactofuranose to bind over the D-galactopyranose [11]. By now having additional GafA proteins that bind furanoses of different sizes, the role of this Asp can be further assessed. It is notable that the position of this residue is the variable across the four structures shown in Fig. 6. As the GafA$_{Ms}$ protein has been crystallized with both D-galactofuranose and L-Ara*f* it provides a missing link to help us interpret selectivity of sugar binding to the other two proteins [22]. Both the GafA$_{Ec}$ and the GafA$_{Ms}$ bind D-galactofuranose with the Asp in an 'up' position making space in the binding site to accommodate the longer sugar molecule (Fig. 7a). By comparing the two structures of GafA$_{Ms}$ with either sugar, the coordination of the shorter L-Ara*f* requires the equivalent Asp105 to adopt a different rotamer to that seen with D-galactofuranose to preserve its charge-dipole H-bond to the sugar (Fig. 7b). This is then consistent with the position of the equivalent Asp in GafA$_{Sw}$ (Asp110) as this also adopts a similar 'down' position to coordinate
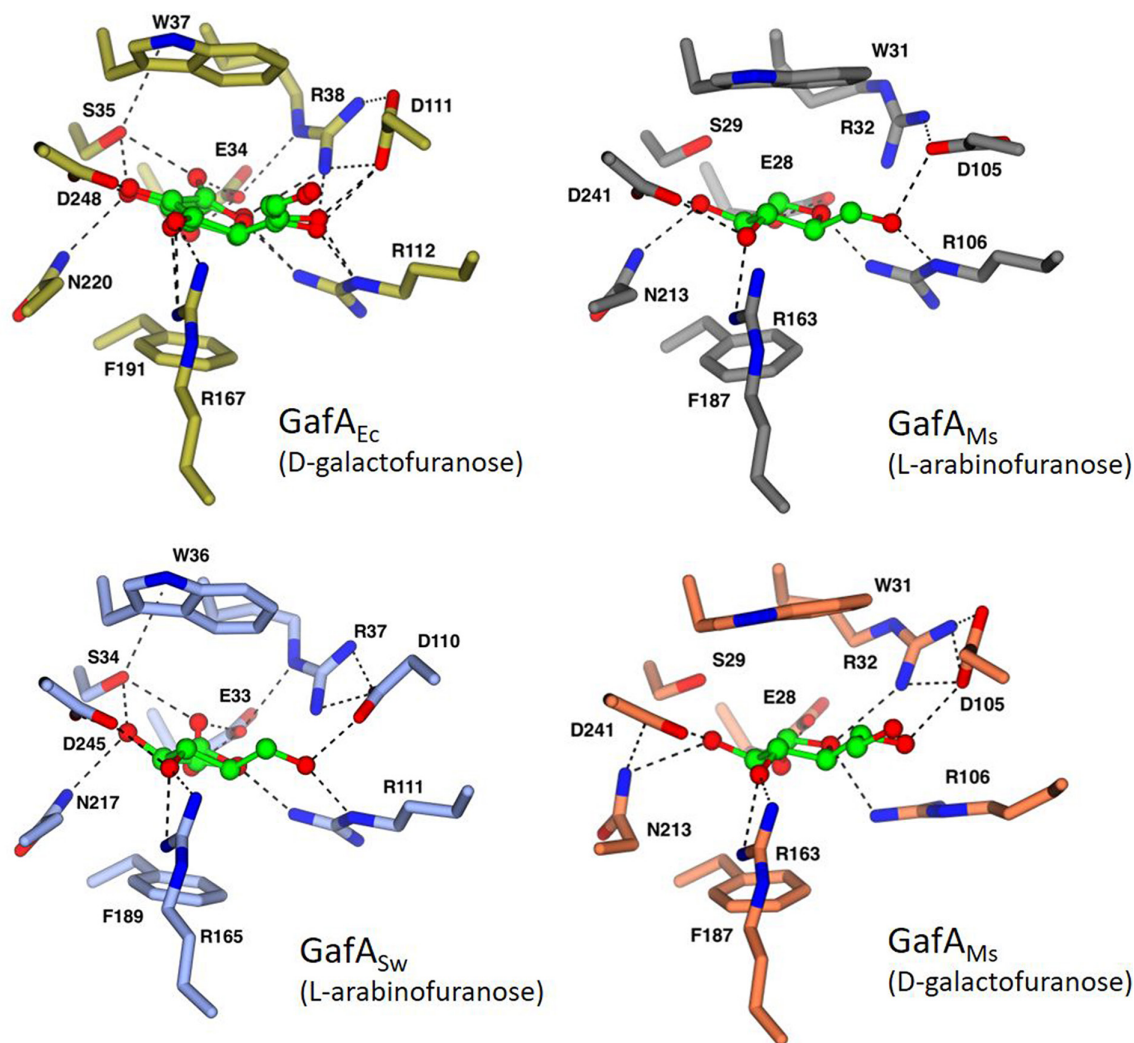
Fig. 5. Binding site residues are identical in GafA$_{Ec}$ (PDB: 2VK2, gold), GafA$_{Sw}$ (PDB: 5OCP, light blue) and GafA$_{Ms}$. Some amino acids adopt different positions in the presence of different ligands, as seen with GafA$_{Ms}$ bound to either L-arabinofuranose (PDB: 6HBM, grey) or D-galactofuranose (PDB: 6HBD, orange).

the shorter L-Ara*f* (Fig. 7c). While GafA$_{Ec}$ and GafA$_{Ms}$ can bind both sugars, GafA$_{Sw}$ can only recognize L-Ara*f* (Fig. 3), which is possibly caused by the bulky Phe40 variant seen in GafA$_{Sw}$ that would prevent the Asp110 from adopting multiple conformations through steric hindrance (Fig. S5). Hence, very subtle changes in the binding site are seen, which are driven by changes in the secondary shell of amino acids surrounding the binding site.

DISCUSSION

Bacteria have evolved a prodigious ability to scavenge nutrients from their environment, with high-affinity transporters enabling growth in environments with low (micromolar or below) concentrations of nutrients. Bacteria living in complex environments such as soil or seawater have evolved large repertoires of binding-protein-dependent transporters of the ABC, TTT and TRAP families [42, 49–51]. Predicting the specificity of binding-protein-dependent transporters is easier than for other classical secondary and other primary transporters, due to our detailed structural knowledge of the SBP subunit, and proteins have been structurally classified into families that largely sorts them into groups that bind generally similar ligands [45, 52, 53]. However, within these larger groupings there is much diversity and some classical families such as TRAP, which are commonly assumed to only bind dicarboxylates, do in fact bind a diverse range of organic acids including sugar acids, amino acids and sulphonates [44, 49, 54]. The Cluster B family of SBPs generally recognize sugars of different types and many monosaccharide-specific examples are known [45, 52]. Beyond predicting the general class of ligand, more specific predictions become more difficult [6, 53]. The discovery of GafA$_{Ec}$ was a surprise as the protein was clearly in the monosaccharide
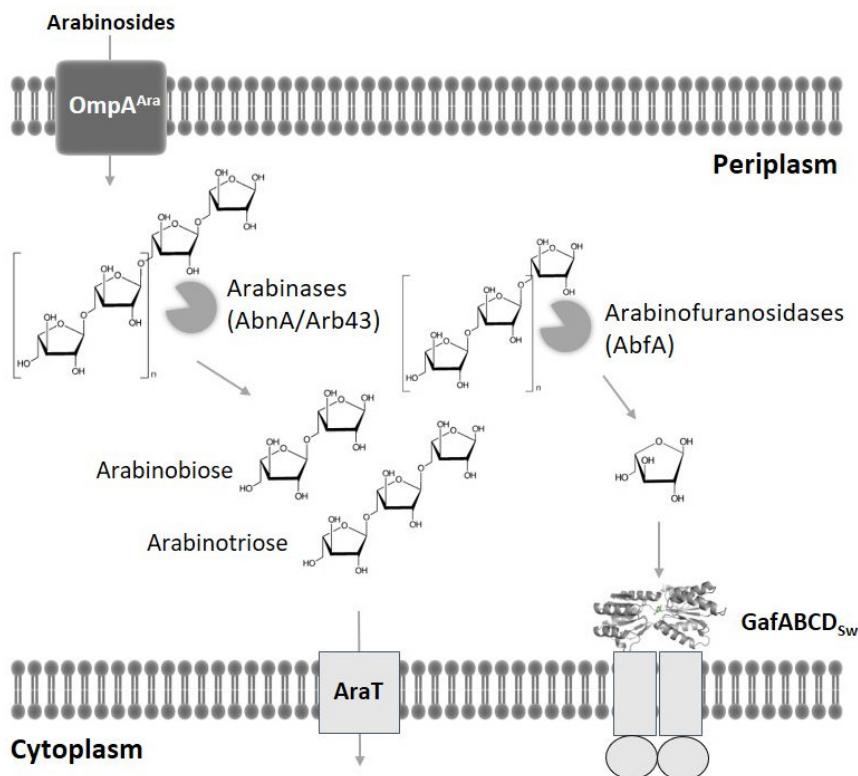
Fig. 6. Model for arabinan utilization by *Shewanella* sp. ANA-3. The model is based on a refinement of the predictions of Rodionoc *et al.* [41] and the genes linked to each are described in the main text.

cluster B class of SBPs and appears to be co-regulated with the known ᴅ-galactose transporter. However, while this system, the Mgl system, transports ᴅ-galactopyranose it then transpired that the bacterium makes at the same time a second ABC transporter to capture the ᴅ-galactofuranose [11]. In this study the structure of GafA$_{Sw}$ revealed how a GafA protein, while still recognizing a furanose form of a sugar, is selective for ʟ-Ara*f* over ᴅ-galactofuranose. At the same time the binding site residues are identical in all the GafA proteins and very highly conserved in all the other monosaccharide SBPs. Hence, this study confirms the ease of a general prediction of ligand specificity as a sugar, but reinforces that extremely subtle changes in the binding site architecture dictate the sugar selectivity. The experimentally measured K_D of 5.8±0.51 µM now needs to be corrected for the 12.5% of substrate present in the sample of ʟ-arabinose, giving an effective K_D of ~0.73 µM, which now being sub-micromolar is consistent with the affinity of other monosaccharide SBPs for their native ligands.
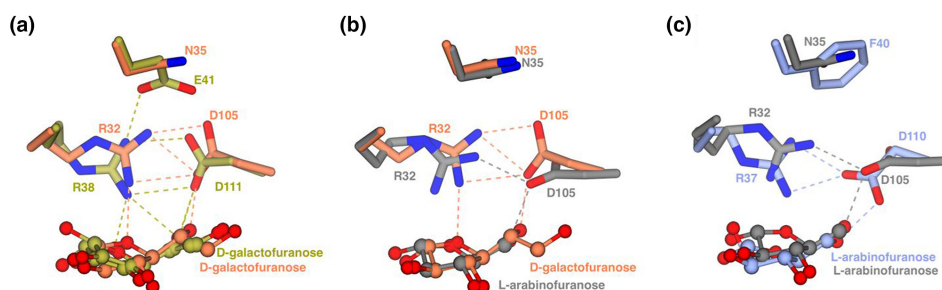
Fig. 7. The flexibility of a ligand interacting aspartate is required for the coordination of ᴅ-galactofuranose and not ʟ-arabinofuranose. (a) GafA$_{Ms}$ (orange) and GafA$_{Ec}$ (gold) can similarly accommodate ᴅ-galactofuranose with D105/D111 pointing away from the binding site. (b) GafA$_{Ms}$ (grey: ʟ-arabinofuranose bound, orange: ᴅ-galactofuranose bound) can also accommodate ʟ-arabinofuranose with D105 shifting closer to the binding site. (c) GafA$_{Sw}$ (light blue) can also accommodate ʟ-arabinofuranose with D105 shifting closer to the binding site but unlike GafA$_{Ms}$ (grey), it is unable to accommodate the larger ᴅ-galactofuranose due to the neighbouring F40.

The GafA$_{Sm}$ protein, which exhibited very similar overall binding characteristics to GafA$_{Ec}$, is similarly not encoded with other linked catabolic genes, although its expression is known to be induced by L-arabinose and D-fucose [42]. Our study corroborated the above results, as both of these sugars increased the thermal stability of GafA$_{Sm}$ but we also demonstrated D-galactose binding. Interestingy, the GafA$_{Sm}$ orthologue from *Rhizobium leguminosarum* bv. *viciae* 3841, namely RL2376, was found to bind D-galactose and its use as a biosensor was proposed [55], although it is highly likely to also bind L-arabinose and might not be particularly selective. Another study found that the GafA$_{Sm}$-encoding operon was expresed in response to desiccation induced by high NaCl in *Sinorhizobium meliloti* 1021 [56], and a known response to this is the production of galactose-containing oligosaccharides [57], consistent with a function of scavenging D-galactose to provide precursors for these glycans.

The use of the powerful bioinformatics tools MicrobesOnline and RegPrecise [24, 40], combined with the experimental fitness data included in MicrobesOnline [25], led us to a strong hypothesis that the GafA$_{Sw}$ protein was probably functioning in the context of L-Ara*f* uptake, which was borne out in our data and even more so in the unique specificity of this GafA protein for L-arabinose over D-galactose. Given this finding, a closer examination of the other genes in this extensive gene cluster enables a reconstruction of the possible function of the genes contained within it in the use of arabinan as a carbon source in the soil environment (Fig. 6). This includes a possible outer membrane porin for arabinosides (Shewana3_2085), a series of secreted and periplasmic arabinases and periplasmic and cytoplasmically located arabinofuranosidases (Shewana3_2067, 2069, 2077, 2078, 2082, 2086), the Gaf transporter (Shewana3_2077–76) along with other putative secondary carriers (Shewana3_2081 and Shewana3_2084), followed by the intracellular catabolic pathways shown in Fig. 1(b). In fact, 13 genes in this cluster are essential for growth on L-arabinose [25].

It is of interest that unlike *Sinorhizobium meliloti* and *E. coli*, an L-arabinopyranose (AraFGH) transporter is not present in the cluster or in the entire genome of *Shewanella*, suggesting that *gafABCD* encodes the sole L-arabinose transporter in this organism, which would be consistent with the strong growth phenotype. As far as we are aware, this would be unique in any bacterium and probably suggests that selective advantage of the immediate uptake of periplasmically liberated L-Ara*f* in the oligosaccharides that are being consumed by the bacterium. While the rate of spontaneous conversion of the released L-Ara*f* to L-arabinopyranose is likely to be in the range of minutes, data on this in physiological conditions are not available to our knowledge. From our previous work on GafA$_{Ec}$ when we released the D-galactofuranose bound to the protein in the presence of DMSO, which slows the rate of interconvertion, we were able to show significant furanose form in solution by NMR immediately after release, but a few hours later this was at equilibrium levels [11]. In summary, using a combination of published phenotypes, bioinformatics, biochemistry and structural biology we have expanded our knowledge of bacterial Gaf-type ABC transporters and discovered a system that has evolved selectivity for L-Ara*f* over D-galactofuranose. For the design and engineering of bacteria with enhanced plant biomass-degrading capabilities, this gene cluster, with its unique Gaf system, could be considered as useful targets for improving the capabilities of chassis strains for use in industrial biotechnology.

DATA SUMMARY

The data that support the findings of this study are publicly available in the RCSB PDB database under the PDB identifier = 5OCP (i.e. crystal structure of Shewana3_2073). Raw data that support the findings of this study are available from the corresponding author, upon request.

References
1. Henderson PJF. Proton-linked sugar transport systems in bacteria. *J Bioenerg Biomembr* 1990;22:525–569.

2. Saier MH. Families of transmembrane sugar transport proteins. *Mol Microbiol* 2000;35:699–710.

3. Thomas GH. Sialic acid acquisition in bacteria-one substrate, many transporters. *Biochem Soc Trans* 2016;44:760–765.

4. Henderson PJF. Sugar transport proteins. *Curr Opin Struct Biol* 1991;1:590–601.

5. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* 2008;72:317–364.

6. Maqbool A, Horler RSP, Muller A, Wilkinson AJ, Wilson KS, *et al*. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem Soc Trans* 2015;43:1011–1017.

7. Mowbray SL, Petsko GA. The X-ray structure of the periplasmic galactose binding protein from *Salmonella typhimurium* at 3.0-Å resolution. *J Biol Chem* 1983;258:7991–7997.

8. Gilliland GL, Quiocho FA. Structure of the L-arabinose-binding protein from *Escherichia coli* at 2.4 Å resolution. *J Mol Biol* 1981;146:341–362.

9. Mao B, Pear MR, McCammon JA, Quiocho FA. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem* 1982;257:1131–1133.

10. Quiocho FA, Pflugrath JW. The structure of D-galactose-binding protein at 4.1 Å resolution looks like L-arabinose-binding protein. *J Biol Chem* 1980;255:6559–51.

11. Horler RSP, Müller A, Williamson DC, Potts JR, Wilson KS, *et al*. Furanose-specific sugar transport: characterization of a bacterial galactofuranose-binding protein. *J Biol Chem* 2009;284:31156–31163.

12. Bagaria A, Kumaran D, Burley SK, Swaminathan S. Structural basis for a ribofuranosyl binding protein: insights into the furanose specific transport. *Proteins* 2011;79:1352–1357.

13. Dodd D, Cann IKO. Enzymatic deconstruction of xylan for biofuel production. *Glob Change Biol Bioenergy* 2009;1:2–17.

14. Shulami S, Raz-Pasteur A, Tabachnikov O, Gilead-Gropper S, Shner I, *et al*. The L-Arabinan utilization system of *Geobacillus stearothermophilus*. *J Bacteriol* 2011;193:2838–2850.

15. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, *et al*. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 2017;544:65–70.

16. Adelsberger H, Hertel C, Glawischnig E, Zverlov VV, Schwarz WH. Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the in vivo system from recombinant enzymes. *Microbiology* 2004;150:2257–2266.

17. Konishi T, Takeda T, Miyazaki Y, Ohnishi-Kameyama M, Hayashi T, *et al*. A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 2007;17:345–354.

18. Qaseem MF, Shaheen H, Wu AM. Cell wall hemicellulose for sustainable industrial utilization. *Renew Sustain Energy Rev* 2021;144:110996.

19. Cartmell A, McKee LS, Peña MJ, Larsbrink J, Brumer H, *et al*. The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J Biol Chem* 2011;286:15483–15495.

20. Franks F. Physical chemistry of small carbohydrates - equilibrium solution properties. *Pure and Appl Chemist* 1987;59:1189–1202.

21. Stringer AM, Currenti S, Bonocora RP, Baranowski C, Petrone BL, *et al*. Genome-scale analyses of *Escherichia coli* and *Salmonella enterica* AraC reveal noncanonical targets and an expanded core regulon. *J Bacteriol* 2014;196:660–671.

22. Li M, Müller C, Fröhlich K, Gorka O, Zhang L, *et al*. Detection and characterization of a mycobacterial L-arabinofuranose ABC transporter identified with a rapid lipoproteomics protocol. *Cell Chem Biol* 2019;26:852–862.

23. Gunina A, Kuzyakov Y. Sugars in soil and sweets for microorganisms: review of origin, content, composition and fate. *Soil Biol Biochem* 2015;90:87–100.

24. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, *et al*. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 2010;38:D396–400.

25. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, *et al*. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 2018;557:503–509.

26. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.

27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.

28. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, *et al*. Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 2008;36:W465–W469.

29. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--A web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* 2005;33:W557–W559.

30. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.

31. Quan S, Hiniker A, Collet JF, Bardwell JCA. Isolation of bacteria envelope proteins. *Methods Mol Biol* 2013;966:359–366.

32. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 2011;67:282–292.

33. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 2010;66:22–25.

34. Lebedev AA, Vagin AA, Murshudov GN. Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr D Biol Crystallogr* 2008;64:33–39.

35. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, *et al*. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011;67:355–367.

36. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486–501.

37. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 2004;60:2256–2268.

38. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, *et al*. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 2014;42:D206–14.

39. Watanabe S, Shimada N, Tajima K, Kodaki T, Makino K. Identification and characterization of L-arabonate dehydratase, L-2-keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism. Novel evolutionary insight into sugar metabolism. *J Biol Chem* 2006;281:33521–33536.

40. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, *et al*. RegPrecise 3.0--A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 2013;14:745.

41. Rodionov DA, Yang C, Li X, Rodionova IA, Wang Y, *et al*. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 2010;11:1–19.

42. Mauchline TH, Fowler JE, East AK, Sartor AL, Zaheer R, *et al*. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc Natl Acad Sci* 2006;103:17933–17938.

43. Maqbool A, Levdikov VM, Blagova EV, Hervé M, Horler RSP, *et al*. Compensating stereochemical changes allow murein tripeptide to be accommodated in a conventional peptide-binding protein. *J Biol Chem* 2011;286:31512–31521.

44. Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, *et al*. Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* 2015;54:909–931.

45. Berntsson RP-A, Smits SHJ, Schmitt L, Slotboom D-J, Poolman B. A structural classification of substrate-binding proteins. *FEBS Lett* 2010;584:2606–2617.

46. Scheepers GH, Lycklama a Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett* 2016;590:4393–4401.

47. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* 2011;67:386–394.

48. Quiocho FA, Vyas NK. Novel stereospecificity of the L-arabinose-binding protein. *Nature* 1984;310:381–386.

49. Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. Tripartite ATP-Independent Periplasmic (TRAP) transporters and Tripartite Tricarboxylate Transporters (TTT): from uptake to pathogenicity. *Front Cell Infect Microbiol* 2018;8.

50. Rosa LT, Springthorpe V, Bianconi ME, Thomas GH, Kelly DJ. Massive over-representation of solute-binding proteins (SBPs) from the tripartite tricarboxylate transporter (TTT) family in the genome of the α-proteobacterium *Rhodoplanes* sp. Z2-YC6860. *Microb Genom* 2018;4:e000176.

51. Ortega Á, Matilla MA, Krell T, Gao B. The repertoire of solute-binding proteins of model bacteria reveals large differences in number, type, and Ligand range. *Microbiol Spectr* 2022;10:e0205422.

52. Scheepers GH, Lycklama A Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett* 2016;590:4393–4401.

53. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res* 2017;45:D320–D324.

54. Mulligan C, Kelly DJ, Thomas GH. Tripartite ATP-independent periplasmic transporters: application of a relational database for genome-wide analysis of transporter gene frequency and organization. *J Mol Microbiol Biotechnol* 2007;12:218–226.

55. Bourdès A, Rudder S, East AK, Poole PS. Mining the *Sinorhizobium meliloti* transportome to develop FRET biosensors for sugars, dicarboxylates and cyclic polyols. *PLoS One* 2012;7:e43578.

56. Vriezen JAC, de Bruijn FJ, Nüsslein K. Identification and characterization of a NaCl-responsive genetic locus involved in survival during desiccation in *Sinorhizobium meliloti. Appl Environ Microbiol* 2013;79:5693–5700.

57. Zevenhuizen L, Faleschini P. Effect of the concentration of sodium chloride in the medium on the relative proportions of poly- and oligo-saccharides excreted by *Rhizobium meliloti* strain YE-2SL. *Carbohydr Res* 1991;209:203–209.

Edited by: T. Palmer and G. M. Seibold