ATP-binding Cassette (ABC) Transport System Solute-binding Protein-guided Identification of Novel d-Altritol and Galactitol Catabolic Pathways in Agrobacterium tumefaciens C58

Received for publication, August 20, 2015, and in revised form, October 12, 2015 Published, JBC Papers in Press, October 15, 2015, DOI 10.1074/jbc.M115.686857

Daniel J. Wichelecki1, Matthew W. Vetting1, Liyushang Chou1, Nawar Al-Obaidi1, Jason T. Bouvier1, Steven C. Almo2, and John A. Gerlt1,2

From the 1Departments of Biochemistry and Chemistry and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801 and the 2Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461

Background: Solute-binding proteins for ABC transport systems can be used to guide discovery of novel enzymes and metabolic pathways.

Results: Novel metabolic pathways were discovered for d-altritol and galactitol catabolism.

Conclusion: More than 2,000 proteins with previously unknown functions were annotated using this strategy.

Significance: Strategies are described that can be used to address misannotation in protein databases.

Innovations in the discovery of the functions of uncharacterized proteins/enzymes have become increasingly important as advances in sequencing technology flood protein databases with an exponentially growing number of open reading frames. This study documents one such innovation developed by the Enzyme Function Initiative (EFI; U54GM093342), the use of solute-binding proteins for transport systems to identify novel metabolic pathways. In a previous study, this strategy was applied to the tripartite ATP-independent periplasmic transporters. Here, we apply this strategy to the ATP-binding cassette transporters and report the discovery of novel catabolic pathways for d-altritol and galactitol in Agrobacterium tumefaciens C58. These efforts resulted in the description of three novel enzymatic reactions as follows: 1) oxidation of d-altritol to d-tagatose via a dehydrogenase in Pfam family PF00107, a previously unknown reaction; 2) phosphorylation of d-tagatose to d-tagatose 6-phosphate via a kinase in Pfam family PF00294, a previously orphan EC number; and 3) epimerization of d-tagatose 6-phosphate C-4 to d-fructose 6-phosphate via a member of Pfam family PF08013, another previously unknown reaction. The epimerization reaction catalyzed by a member of PF08013 is especially noteworthy, because the functions of members of PF08013 have been unknown. These discoveries were assisted by the following two synergistic bioinformatics web tools made available by the Enzyme Function Initiative: the EFI-Enzyme Similarity Tool and the EFI-Genome Neighborhood Tool.

The rate at which protein sequences are added to the UniProt Knowledgebase (UniProtKB)2 has created a daunting gap between the number of manually curated protein annotations (SwissProt database, ~550,000 in Release 2015_06) and the total number of proteins (SwissProt and TrEMBL databases, ~48 million in release 15_06). Approximately 50% of the automated annotations made by sequence similarity (TrEMBL database) are thought to be misleading or erroneous (1–4). The presence of a large fraction of entries in UniProtKB that are functionally ambiguous or incorrect undermines efforts to 1) determine drug targets in pathogens affecting human health or the agricultural industry, 2) develop in vitro enzymatic pathways for biomanufacturing, 3) further investigate the relationship between human health and the gut microbiota, and 4) understand critical environmental issues such as global nutrient cycles. The Enzyme Function Initiative (EFI) (5) worked to address these challenges by the development of robust strategies for accurate and efficient assignment of functions to unknown or uncharacterized proteins discovered in genome projects via bioinformatics, computational modeling, in vitro experimentation, and in vivo verification (6–11).

The most recent contribution from the EFI was the demonstration of the utility of solute-binding proteins (SBPs) for tripartite ATP-independent periplasmic transporters to guide

2 The abbreviations used are: UniProtKB, UniProt Knowledgebase; ABC, ATP-binding cassette transporter; AhSBP, d-altritol binding ABC SBP; AtuFK, putative A. tumefaciens C58 fructokinase; AtuSorD, putative A. tumefaciens C58 sorbit dehydrogenase; AtuTag6PK, putative A. tumefaciens C58 tagatose-6-phosphate kinase; AtuZnD, putative A. tumefaciens C58 zinc-binding dehydrogenase; DHA, dihydroxyacetone; DHAP, dihydroxyacetone phosphate; EC, Enzyme Commission; EFI, Enzyme Function Initiative; EFI-EST, Enzyme Function Initiative-Enzyme Similarity Tool; EFI-GNT, Enzyme Function Initiative-Genome Neighborhood Networks Tool; GNN, genome neighborhood network; Pfam, protein family; PPP, pentose phosphate Pathway; SBP, solute-binding protein; SSN, sequence similarity network; qRT, quantitative RT; DSF, differential scanning fluorimetry; FOR, forward; REV, reverse; PEP, P-enolpyruvate.
Novel D-Altritol and Galactitol Catabolic Pathways

functional discovery (12). The premise was that the small molecule ligands of SBPs represent the chemical “starting points” of metabolic pathways (i.e. the substrate for the first enzyme in a metabolic pathway). This information imposes significant constraints on the regions of chemical space and types of enzymatic transformations that must be considered for pathway deconvolution. In combination with annotated functions for the proteins encoded by the genes proximal to that encoding the SBP, these insights allow the facilitated prediction of new metabolic pathways and the discovery of novel enzymatic functions. In this study, we apply this strategy to an alternative class of transporters, the ATP-binding cassette (ABC) transporters.

Most ABC transporters are composed of three components as follows: a transmembrane domain, a nucleotide binding domain, and an SBP (13, 14). Hydrolysis of ATP provides the energy needed to transport the solute across the cell membrane (15). ABC transporters are widely spread throughout all phyla of life.

Applying the SBP-based strategy for pathway discovery to the ABC transporters enabled the discovery of a novel polyol metabolic pathway for catabolism of D-altritol (also known as D-talitol). For this pathway, the SBP-proximal genome neighborhood was found to contain three genes encoding enzymes that convert D-altritol to the central metabolite fructose 6-phosphate. Two enzymes perform novel reactions, and the third enzyme is responsible for a previously orphan Enzyme Commission (EC) number. Notably, one of these is a member of a family of ~6,000 previously unannotated proteins. Together, these results demonstrate the effectiveness of this strategy for facilitating functional discovery and correcting errors in the protein databases.

Rare sugars, such as D-altritol, have attracted attention because of their importance to the food, pharmaceutical, artificial sweetener, and nutrition industries (16, 17). From a biomedical perspective, rare sugars and their nucleoside derivatives are candidates for antibiotic drugs and antiviral drugs (18). Specifically, siRNA made from D-altritol-modified nucleic acids has been used to inhibit hepatitis B virus replication (19) and the expression of the multidrug-resistant efflux pump MDR1 (20). Defining metabolic pathways involving D-altritol may enable the development of new low cost synthetic routes, which will further its impact in medicine and industry.

D-Altritol has been isolated from brown algae of the order Fucales (21). Although D-altritol is not known to be a widespread plant metabolite, polyols are primary biosynthetic products; therefore, the existence of D-altritol may be phylogenetically widespread, albeit at low abundance (22). Notably, Izumori and co-workers (23–26) have shown that various organisms, including fungi, yeast, and bacteria, have the capability to synthesize D-altritol, although the pathway has not been studied.

In this study, D-altritol catabolism was predicted and verified in the tumor-producing plant pathogen Agrobacterium tumefaciens C58. Tumor production proceeds through insertion into the plant genome of a small segment of DNA from Agrobacterium’s Ti plasmid (27). This mode-of-action has made Agrobacterium an important tool in the genetic engineering of plants (28). Additionally, A. tumefaciens can infect a wide variety of plants, making it of great interest to the agricultural industry (it has been identified as the third most important plant pathogen) (29, 30). The fact that sugars are known to induce virulence factors in Agrobacterium (31) has resulted in many studies on carbohydrate metabolism in A. tumefaciens (32–35) but none involving D-altritol.

Experimental Procedures

Generation of Sequence Similarity Networks—Sequence similarity networks (SSNs) for Pfam families PF01547 (bacterial extracellular solute-binding protein), PF01116 (fructose-bisphosphate aldolase), PF00294 (PfkB family carbohydrate kinase), and PF08013 (tagatose-6-phosphate kinase) were generated with Option B of the EFI-Enzyme Similarity Tool (EFI-EST), which uses one or more Pfam families as input (11). The SSNs of PF00106 (short-chain dehydrogenase) and PF00107 (zinc-binding dehydrogenase) could not be generated via Option B because of the large number of sequences in these Pfam families; these SSNs were generated via Option A of EFI-EST, which uses a protein sequence as input and generates an SSN from the top 5,000 BLAST hits identified by the query sequence. All SSNs were visualized as 100% representative node networks generated with a minimum edge alignment score threshold corresponding to ~35% sequence identity selected from the percent identity-alignment score quartile plot provided by EFI-EST. The alignment score thresholds for edge inclusion were then made increasingly stringent using Cytoscape to generate the networks used in pathway prediction.

Generation of Genome Neighborhood Networks—The genome neighborhood networks (GNNs) for PF01547 (bacterial extracellular solute-binding protein) and PF00294 (PfkB family carbohydrate kinase) were generated using the EFI-Genome Neighborhood Tool (EFI-GNT). The input was the xgmml file for the SSN cluster containing the D-altritol binding ABC SBPs (AltSBPs). The output was generated using the default parameters (neighborhood size = 10, co-occurrence lower limit = 20%).

Cloning, Expression, and Purification of Polyol-binding ABC SBPs from PF01547—The genes encoding ABC SBPs were amplified from genomic DNAs by PCR using KOD Hot Start DNA polymerase (Novagen) and the primers listed in Table 1. The conditions were as follows: 2 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 30 s at 66 °C, and 30 s at 72 °C. The amplified fragments were cloned into the N-terminal tobacco etch virus-cleavable His6 tag containing vector, pNIC28-Bsa4 (36), by ligation-independent cloning (37). The periplasmic signal sequences, as predicted by SignalP (38), were not included in the cloned products. SBPs (selenomethionine-substituted) were expressed by autoinduction in a LEX48 airlift fermenter and purified by nickel-nitriolotriacetic acid and size exclusion chromatography. Details are as described previously (12). ABC SBPs were concentrated to 10–20 mg/ml, flash-frozen using liquid nitrogen, and stored at −20 °C.

Differential Scanning Fluorimetry (DSF) of PF01547 ABC SBPs—Ligand mixtures (1–6 compounds) were prepared as concentrated stock solutions in water (10 mM) and stored at −20 °C. In a 10-μl volume (100 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM DTT), the ligand concentration was 1 mM; the
protein concentration was 10 μM, and Sypro-Orange was at 5× (Thermo Fisher Scientific). The ligand library consists of 405 ligands and eight controls (protein without compound) in 192 duplicated wells (384 wells total) (supplemental Table S1). DSF ligands and eight controls (protein without compound) in 192 (Thermo Fisher Scientific). The ligand library consists of 405 ligands and eight controls (protein without compound) in 192 duplicated wells (384 wells total) (supplemental Table S1). DSF ligands and eight controls (protein without compound) in 192 (Thermo Fisher Scientific).
Novel D-Altritol and Galactitol Catabolic Pathways

from the cells using an RNasy mini kit (Qiagen). The RNA was further processed using RNase-free DNase (Qiagen) following the manufacturer’s protocol. Purity was verified with 30-μl PCRs containing 50 ng of mRNA, 1 mM MgCl₂, 1X Pfx amplification buffer, 2X PCR enhancer, 0.33 mM dNTP, 0.33 μM FOR/REV primers (AtuSorbD_RTPCR_FOR and AtuSorbD_RTPCR_REV), and 1.25 units of Pfx polymerase (Invitrogen Platinum Pfx DNA polymerase kit). The PCRs were analyzed on agarose gels to confirm amplification. Concentrations were determined spectrophotometrically at 280 nm. cDNA was generated using Protoscript First Strand (New England Biolabs) and 1 μg of mRNA following the manufacturer’s protocol. The qRT-PCRs were performed using the Light Cycler 480 SYBR Green I Master kit (Roche Applied Science) and a Light Cycler 480 II (Roche Applied Science) using the manufacturer’s protocol. Table 1 contains the sequences of the qRT-PCR primers (labeled as RT-PCR). The RpoD (RNA polymerase σ factor) primers were included to normalize the results based on the amount of cDNA added.

Discovery and Kinetic Characterization of AtuSorbD Activity—AtuSorbD was screened for oxidation activity in a reaction containing 50 mM Tris, pH 9, 10 mM MgCl₂, 1.5 mM NAD⁺, 1 μM AtuSorbD, and 5 mM polyol (as well as blanks without enzyme). Polysols tested include D-arabitol (Sigma), galactitol, D-mannitol (Sigma), D-sorbitol (Sigma), D-altritol and L-fucitol (Sigma), and volemitol (D-glycero-D-mannoheptitol, Sigma). The reactions were incubated at 25 °C for 60 min. The absorbance was measured at 340 nm (ε = 6220 M⁻¹ cm⁻¹) to determine activity. Active polysols (D-sorbitol and galactitol) were used to kinetically characterize AtuSorbD. Reaction conditions for the kinetics were the same as the screen except 400 mM enzyme was used, and the concentration of substrate was varied from 100 μM to 10 mM.

Discovery and Kinetic Characterization of AtuZnD Activity—AtuZnD was screened for oxidation activity in a reaction containing 50 mM Tris, pH 9, 10 mM MgCl₂, 1.5 mM NAD⁺, 1 μM AtuZnD, and 5 mM polyol (blanks without enzyme). Polysols tested include D-arabitol (Sigma), galactitol, D-mannitol (Sigma), D-sorbitol (Sigma), D-altritol and L-fucitol (Sigma), and volemitol (D-glycero-D-mannoheptitol, Sigma). The reactions were incubated at 25 °C for 60 min. The absorbance was measured at 340 nm (ε = 6220 M⁻¹ cm⁻¹) to determine activity. The most active polysol (D-altritol) was used to kinetically characterize AtuZnD. Reaction conditions for kinetics were the same as the screen except 400 mM enzyme was used and the concentration of substrate was varied from 250 μM to 25 mM.

Discovery and Kinetic Characterization of AtuFK Activity—AtuFK was tested for kinase activity with D-fructose (Sigma) and D-tagatose (Carbosynth) using a coupled-enzyme assay containing 50 mM HEPES, pH 7.9, 10 mM MgCl₂, 200 μM NADH, 1.5 mM ATP, 1.5 mM PEP, 9 units of lactate dehydrogenase, 9 units of pyruvate kinase, 1 μM AtuFK, and 5 mM substrate (blanks without enzyme). The reactions were incubated at 25 °C for 60 min. The absorbance was measured at 340 nm (ε = 6220 M⁻¹ cm⁻¹) to determine activity. Both D-fructose and D-tagatose were used to kinetically characterize AtuFK. Reaction conditions for kinetics were the same as the screen except 400 mM enzyme was used, and the concentration of substrate was varied from 100 μM to 10 mM.

Discovery and Kinetic Characterization of AtuTag6PK Activity—AtuTag6PK was tested for kinase activity with D-tagatose 6-phosphate (Sigma) and D-fructose 6-phosphate (Sigma) using a coupled enzyme assay containing 50 mM HEPES, pH 7.9, 10 mM MgCl₂, 200 μM NADH, 1.5 mM ATP, 1.5 mM PEP, 9 units of lactate dehydrogenase, 9 units of pyruvate kinase, 1 μM AtuTag6PK, and 5 mM substrate (blanks without enzyme). The reaction was incubated at 25 °C for 60 min, and the absorbance was measured at 340 nm (ε = 6,220 M⁻¹ cm⁻¹) to determine activity.

AtuTag6PK was tested for aldolase activity with D-tagatose 6-phosphate and D-fructose 6-phosphate using a coupled enzyme assay containing 50 mM HEPES, pH 7.9, 10 mM MgCl₂, 200 μM NADH, 1.5 mM ATP, 1.5 mM PEP, 9 units of lactate dehydrogenase, 9 units of pyruvate kinase, 1 unit of glyceraldehyde-3-phosphate dehydrogenase, 1 mM inorganic phosphate, 1 μM AtuTag6PK, and 5 mM substrate (blanks without enzyme). The reaction was incubated at 25 °C for 60 min, and the absorbance was measured at 340 nm (ε = 6,220 M⁻¹ cm⁻¹) to determine activity.

AtuTag6PK was tested for epimerase activity with D-tagatose 6-phosphate using a coupled enzyme assay containing 50 mM HEPES, pH 7.9, 10 mM MgCl₂, 200 μM NADH, 1.5 mM ATP, 1.5 mM PEP, 9 units of lactate dehydrogenase, 9 units of pyruvate kinase, 3 units of fructose 6-phosphate kinase, 1 μM AtuTag6PK, and 5 mM D-tagatose 6-phosphate (blanks without enzyme). The reaction was incubated at 25 °C for 60 min, and the absorbance was measured at 340 nm (ε = 6220 M⁻¹ cm⁻¹) to determine activity. Reaction conditions for kinetics were the same as the screen except 400 mM enzyme was used, and the concentration of substrate was varied from 100 μM to 5 mM.

NMR Verification of D-Tagatose Conversion to D-Fructose 6-Phosphate—Three reactions were performed in H₂O using 50 mM deuterated Tris, pH 7.9, 10 mM MgCl₂, 1 mM ATP, 60 mM PEP, 1 unit of pyruvate kinase, and 50 mM D-tagatose. One reaction was as described previously; one reaction contained 1 μM AtuFK; and one reaction contained both 1 μM AtuFK and AtuTag6PK. Reactions were incubated at 30 °C for 24 h followed by lyophilization and resuspension in D₂O at pH 7.9. The NMR spectra were acquired on a Varian UNITY INOVA 500NB (64 acquisitions).

Generation of Competent A. tumefaciens C58 Cells—Competent cells were generated by growing a specific strain to an OD of ~0.4 in LB or LB containing 50 μg/ml kanamycin for knock-out strains transformed with plasmids containing the wild-type gene. The cells were pelleted at 5,000 rpm for 5 min and resuspended in 15 ml of 10% glycerol (repeated twice). The cells were then pelleted and resuspended in 1 ml of 10% glycerol followed by another pelleting step and final resuspension in 300 μl of 10% glycerol. Competent cells were flash-frozen in liquid nitrogen and stored at −80 °C.

Genetic Knockouts of AtuSorbD, AtuZnD, AtuFK, and AtuTag6PK—The knock-outs of AtuSorbD, AtuZnD, AtuFK, and AtuTag6PK in A. tumefaciens C58 were constructed using overlap extension and a suicide vector (pK19mobSacB, derived from ATCC 87097) (39). Six hundred bp upstream and downstream of the start and stop codons of the genes encoding AtuSorbD, AtuZnD, AtuFK, and AtuTag6PK were amplified in
Novel d-Altritol and Galactitol Catabolic Pathways

TABLE 2

| UniProt ID | ΔTm (°C) of polyol DSF hits (°C) |
|------------|---------------------------------|
| A6X5Q5    | d-Arabinol (l-arabinol) Galactitol d-Mannitol (l-mannitol) d-Sorbitol (l-sorbitol) d-Altritol l-Fucitol Volemitol |
| B9JRF8    | 11.2 (1.5) 7.7 14.1 (1.1) 6.1 (2.3) 7.9 6.0 10.9 |
| 3PTPTN0   | 5.5 (0.6) 0.6 8.5 (0.4) 5.2 (2.0) 2.6 2.2 5.7 |
| 2K3Z9     | 9.5 (0.2) 0.9 14.3 (0.4) 8.0 (3.2) 5.0 6.6 10.5 |

Differential Scanning Fluorimetry of Polyol Binding ABC SBPs—ABC SBPs A6X5Q5, B9JRF8, 3PTPTN0, and Q2K3Z9 (UniProt IDs) were screened by DSF as part of a larger study of ABC SBPs from Pfam family PF01547. The library consisted of 405 compounds containing various aldoses, acid sugars, amino acids, dipeptides, phenolic acids, oligosaccharides, nucleotides, and various other potential metabolites (supplemental Table S1). A6X5Q5, B9JRF8, 3PTPTN0, and Q2K3Z9 were stabilized by a number of single polyols and polyol mixtures (Table 2). Deconvolution of those wells screened with d/l-ligand mixtures indicated a preference for the d-isomers (Fig. 1). The polyols range from 5 to 7 carbons and share the same configurations at three stereocenters (Fig. 2). d-Altritol was chosen for functional discovery efforts from the seven polyol hits (d-arabitol, d-galactitol, d-mannitol, d-sorbitol, d-altritol (d-talitol), l-fucitol, and volemitol) due to its unknown catabolism and use in modified siRNA-based therapies (19, 20).
Bioinformatic Analysis of Polyol-binding DSF Hits—The polyol-binding SBPs exhibit sequence identities of 58–83% and are located in the same cluster in the PF01547 SSN at an alignment score threshold of 140 (Fig. 3A). A6X5Q5 had the greatest stabilization by d-altritol (7.9 °C) and was the only protein to be significantly stabilized by d-galactitol (7.7 °C) (Fig. 1); therefore, the alignment score stringency was increased to 180 so that A6X5Q5 was located in a cluster separated from the other screened SBPs (Fig. 3B). Because the median sequence identity within the d-altritol-binding SBP (AltSBP) cluster is ~65%, and the sequences in the cluster are highly interconnected, all homologous SBPs in the cluster containing A6X5Q5 were hypothesized to be capable of binding d-altritol. This cluster was then submitted to the EFI-Genome Neighborhood Tool (EFI-GNT) to generate a GNN. EFI-GNT is used to identify Pfam families for enzymes (default to generate a GNN. EFI-GNT is used to identify Pfam families (default to generate a GNN). All of the sequences in the cluster containing A6X5Q5 (Fig. 3A) have 100% sequence identity (default to generate a GNN). All of the sequences in the cluster containing A6X5Q5 were used to generate the GNN.

For the AltSBP SSN cluster, we identified four frequently co-occurring enzyme Pfam families as follows: 1) PF00106 at 94% co-occurrence; 2) PF00107 at 83% co-occurrence; 3) PF00294 at 67% co-occurrence; and 4) PF08013 at 61% co-occurrence. PF00106 is a family of short-chain NAD derivatives (406,662 sequences in UniProt Release 2015_06). PF00107 is a family of zinc-binding alcohol dehydrogenases (195,604 sequences). PF00294 is the PfkB family of carbohydrate and pyrimidine kinases (118,351 sequences). PF08013 is annotated as a family of d-tagatose-6-phosphate kinases (5,909 sequences), although this annotation has been shown to be inaccurate (40).

PF08013 was of interest because no experimentally determined reactions had been associated with this family. Previously, GatZ and AgaZ/KbaZ that are members of this family were shown to be encoded by operons associated with galactitol and d-galactosamine catabolism, respectively (41, 42). Initial work led to the hypothesis that AgaZ and GatZ were d-tagatose-6-phosphate kinases based on the following: 1) the observation that the proposed catabolic pathways were missing a tagatose-6-phosphate kinase, and 2) AgaZ and GatZ had no significant similarity with any proteins with assigned functions (43). However, no kinase activity was observed for AgaZ or GatZ. Later, it was found that AgaY and AgaZ are the subunits of the heterodimeric d-tagatose-1,6-bisphosphate aldolase, AgaYZ (42). Further study revealed that the Z subunit was required for catalytic activity of the Y subunit, but it had no aldolase activity (40). Therefore, no enzymatic activity has been associated with the Z subunit (from PF08013); it has been thought to be a catalytically inactive subunit of d-tagatose-1,6-bisphosphate aldolase. In summary, the GNN of the AltSBPs shows that the following four enzymes are highly conserved in the genome neighborhood: an alcohol dehydrogenase; a short chain dehydrogenase; a kinase; and a protein associated with galactitol and N-galactosamine catabolism through tagatose 1,6-bisphosphate.

Prediction of the d-Altritol Catabolic Pathway—SSNs were generated for PF00106, PF00107, PF00294, and PF08013. Alignment score thresholds were chosen based on the values at which neighbors to the AltSBPs were located in highly interconnected clusters. The SSN for PF00106 was generated using an alignment score of 105 (~67% sequence identity) using option A of EFI-EST (Fig. 4A). Proteins in the cluster containing neighbors of AltSBP are generically annotated as sorbitol dehydrogenases; none have been reviewed by SwissProt. This indicates that the dehydrogenase from PF00106 likely oxidizes a polyol. Given that sorbitol dehydrogenases generally oxidize C-2, this protein is unlikely to oxidize d-altritol due to the inverse stereochemistry at C-2 (or C-5, if d-altritol is bound in the d-talitol orientation). Therefore, we hypothesized that the dehydrogenase from PF00106 oxidizes galactitol to d-tagatose (C-2 or C-5 oxidation) based on the following: 1) a highly conserved neighbor is from PF08013 and is indicated to be involved in galactitol metabolism through d-tagatose 1,6-bisphosphate; 2) galactitol shares the same stereochemistry at C-2 with d-sorbitol; 3) galactitol was a DSF hit; and 4) the SSN for PF00106 suggests the substrate is a polyol.

The SSN for PF00107 was generated at an alignment score threshold of 70 (~35% sequence identity) (Fig. 4B). The neighbors of AltSBP are collocated within a single cluster in this SSN. Proteins in this cluster are generically annotated as zinc-binding dehydrogenases; none have been reviewed by SwissProt. Many different polyol dehydrogenase annotations are present in PF00107, e.g. L-arabitol 4-dehydrogenase, d-galactitol 1-phosphate 5-dehydrogenase, and L-iditol 2-dehydrogenase. This suggests that the dehydrogenase from PF00107 also oxidizes a polyol, although it is difficult to predict the substrate specificity. Therefore, it is hypothesized that the dehydrogenase oxidizes d-altritol based on the following: 1) the DSF hit; 2) suggestions from the SSN that the substrate is a polyol; and 3) the stereochemistry-based hypothesis that the dehydrogenase in PF00106 cannot oxidize d-altritol.
The SSN for PF00294 was generated at an alignment score threshold of 55 (~35% identity) (Fig. 4C). The neighbors of AltSBP are either fructokinases or annotated as sugar kinases; none have been reviewed by SwissProt. The SwissProt-reviewed carbohydrate kinases present in this SSN are fructokinase, 2-keto-3-deoxy-glucokinase, 6-phosphofructokinase, ribokinase, and tagatose-6-phosphate kinase. On the basis of the association with members of PF08013 with tagatose 1,6-bisphosphate and the likelihood that the dehydrogenases from PF00106 and PF00107 oxidize galactitol/D-altritol to D-tagatose, we hypothesized that the neighbors in PF00294 phosphorylate D-tagatose to D-tagatose 6-phosphate.

The SSN for PF08013 was generated at an alignment score threshold of 145 (~60% sequence identity) (Fig. 4D). Proteins within the cluster containing neighbors of the AltSBPs have multiple annotations as follows: tagatose-6-phosphate kinase, tagatose-1,6-bisphosphate aldolase, and inactive subunit of tagatose 1,6-bisphosphate aldolase; none of the proteins in PF08013 have been reviewed by SwissProt. The members of PF08013 contain an aldolase-type TIM barrel domain (IPR013785), so it is unlikely these are kinases. Therefore, the initial hypothesis was that the neighbors in PF08013 are aldolases acting on the D-tagatose 6-phosphate, the product of the reaction catalyzed by the genome neighbors in PF00294, to form dihydroxyacetone (DHA) and glyceraldehyde 3-phosphate.

The predicted pathway is shown in Fig. 5A. The pathway begins with oxidation of the DSF hits galactitol (PF00106) and D-altritol (PF00107) to D-tagatose. D-Tagatose is phosphorylated to D-tagatose 6-phosphate (PF00294). Then D-tagatose 6-phosphate is cleaved by an aldolase (PF08013) into DHA and glyceraldehyde 3-phosphate, both of which can enter the central metabolism, after DHA is phosphorylated to dihydroxyacetone phosphate (DHAP) via EC 2.7.1.29.

**FIGURE 3.** SSN of PF01547 at an alignment score of 140 (~55% identity) is shown in A. The EFI-EST was utilized to perform a pairwise blast of all current sequences of PF01547. In the SSN, each sequence is a node, and those sequences with blast E-value smaller (higher similarity) than defined cutoffs are connected by an edge whose length is weighted by the blast E-value. The cluster containing polyol-binding SBPs is boxed in black. The polyol-binding cluster of PF01547 at a minimum alignment score threshold of 140 (~55% identity) is shown in B. The alignment score threshold is then increased to 180 (~65% identity) for use as an input for EFI-GNT. The green cluster (5) contains the D-altritol-binding SBP, UniProt ID A6X5Q5. UniProt IDs B9JRF8 and Q2K3Z9 cluster in the magenta cluster (1). UniProt ID D3PTN0 is located in the salmon cluster (2). C shows the most highly co-occurring catalytic Pfam families associated with the polyol-binding ABC SBPs.
qRT-PCR of Potential d-Altritol Catabolic Genes—Organisms that have an ABC SBP in the same cluster as A6X5Q5 with an alignment score threshold of 180 are from plant pathogens, plant-associated bacteria, or soil-dwelling bacteria. The genera include Agrobacterium, Rhizobium, and Mesorhizobium. The predominance of plant-associated bacteria is not surprising because polyols are primary photosynthetic products. A. tumefaciens C58 was chosen for in vivo validation of the predicted pathway because of its genetic tractability and ease in laboratory manipulations (34).

The genome neighborhood of the ABC SBP from A. tumefaciens C58 (AtuSBP, UniProt ID Q7CRS2) contains each of the conserved Pfam members identified by EFI-GNT as follows: an annotated sorbitol dehydrogenase (AtuSorbD, PF00106, UniProt ID A9CES4); a zinc-binding dehydrogenase (AtuZnD, PF00107, UniProt ID A9CES3); an annotated fructokinase (AtuFK, PF00294, Uniprot ID A9CES5); and an annotated tagatose-6-phosphate kinase (AtuTag6PK, PF08013, UniProt ID A9CES6) (Fig. 6A). Transcript levels for each of the genes were analyzed to establish involvement in d-altritol growth. Both MicrobesOnline (44) and the Database of prOkaryotic Operons (45, 46) predict that AtuSorbD and AtuZnD are present on one transcript, although AtuFK and AtuTag6PK are present on a separate transcript. Transcript levels were up-regulated several thousand-fold for all four genes when A. tumefaciens C58 was grown on d-altritol compared with d-glucose (Fig. 6B). Similar up-regulation was also observed when A. tumefaciens C58 was grown on galactitol compared with d-glucose (Fig. 6C). The transcriptional regulator and dehydrogenases that are located downstream of AtuZnD were not predicted to be present in an operon with AtuSorbD/AtuZnD nor were they up-regulated during growth on d-altritol or galactitol. Therefore, further characterization of this pathway focused on the enzymes in the Pfam families identified by the GNN analysis.

In Vitro Characterization of d-Altritol Catabolic Enzymes—AtuSorbD and AtuZnD were purified and screened for oxidation activity on all the DSF polyol hits as follows: d-arabitol, galactitol, d-mannitol, d-sorbitol, d-altritol, l-fucitol, and volemitol. AtuSorbD oxidized galactitol, d-sorbitol, and d-mannitol; d-altritol was oxidized at a significantly greater rate than d-mannitol and d-arabitol. Therefore, only the oxidations of galactitol to d-tagatose and of d-sorbitol to d-fructose were kinetically characterized (Table 3). AtuZnD oxidized d-altritol, d-mannitol, and d-arabitol; d-altritol was oxidized at a significantly greater rate than d-mannitol and d-arabitol. Therefore, kinetic constants were determined only for the conversion of d-altritol to d-tagatose.

AtuFK was tested for activity on the dehydrogenase products d-tagatose and d-fructose. Both substrates showed activity and
were kinetically characterized, with \(\delta\)-tagatose being significantly more active (Table 3).

Given its prediction as a class II aldolase, \(\text{AtuTag6PK}\) was tested for aldolase activity with \(\delta\)-tagatose 6-phosphate and \(\delta\)-fructose 6-phosphate via a coupled-enzyme assay; however, no activity was observed. Because of the annotation of its Pfam family, \(\text{AtuTag6PK}\) was tested for kinase activity on \(\delta\)-tagatose 6-phosphate and \(\delta\)-fructose 6-phosphate; again, no activity was observed. Some enzymes in Pfam family PF00596 (class II aldolases) are, in fact, epimerases, e.g. \(\delta\)-ribulose-5-phosphate 4-epimerase (47). Therefore, \(\text{AtuTag16PK}\) was tested for epimerization of \(\delta\)-tagatose 6-phosphate to \(\delta\)-fructose 6-phosphate (C-4 epimerization) using a coupled enzyme assay. Significant activity was observed, and kinetic constants were determined (Table 3). This last step in the proposed pathway was validated via \(^1\text{H}\) NMR by the conversion of \(\delta\)-tagatose to \(\delta\)-fructose 6-phosphate in a one-pot reaction containing \(\text{AtuFK}\) and \(\text{AtuTag6PK}\) (Fig. 7). Thus, the complete in vitro pathway (Fig. 5B) begins with the DSF hits galactitol (by PF00106) and \(\delta\)-altritol (by PF00107) being oxidized to \(\delta\)-tagatose. \(\delta\)-Tagatose is phosphorylated to \(\delta\)-tagatose 6-phosphate (by PF00294). Then \(\delta\)-tagatose 6-phosphate is epimerized to \(\delta\)-fructose 6-phosphate (by PF08013), which is a central metabolite in \(A.\) tumefaciens (32).

\textbf{In Vivo Verification of \(\delta\)-Altritol Catabolism in \(A.\) tumefaciens—}To verify the involvement of the genome neighborhood of the AltSBP in galactitol and \(\delta\)-altritol catabolism, we constructed knock-outs (deletions) of each of the following four genes encoding the four catabolic enzymes: \(\text{AtuSorbD}\), \(\text{AtuZnD}\), \(\text{AtuFK}\), and \(\text{AtuTag6PK}\). When grown on \(\delta\)-galactitol, the \(\Delta\text{AtuSorbD}\) strain showed a no-growth phenotype; the \(\Delta\text{AtuZnD}\) strain did not exhibit a growth defect (Fig. 8). In contrast, when grown on \(\delta\)-altritol, the \(\Delta\text{AtuSorbD}\) strain showed wild-type growth; the \(\Delta\text{AtuZnD}\) strain showed a no-growth phenotype (Fig. 8). Because of the strong no-growth phenotypes for both \(\Delta\text{AtuSorbD}\) and \(\Delta\text{AtuZnD}\), we constructed complementation plasmids carrying wild-type copies of each of these genes. Growth was restored to the levels of wild type for both knock-out strains with their respective carbon sources (Fig. 9). This validates, \textit{in vivo}, the predicted pathway for galactitol and \(\delta\)-altritol catabolism. Deletion of \(\text{AtuFK}\) did not result in growth defects for either \(\delta\)-galactitol or \(\delta\)-altritol (Fig. 8). Presumably, a promiscuous kinase located elsewhere in the genome is able to phosphorylate \(\delta\)-tagatose. The \(\Delta\text{AtuTag6PK}\) strain showed slow-growth phenotypes for both galactitol and \(\delta\)-altritol (Fig. 8); again, this suggests promiscuity or functional overlap encoded elsewhere in the genome, albeit at a lower efficiency. Further study is needed to identify the genes responsible for these activities.

\textbf{Discussion}

The use of transport system SBPs to guide functional discovery of novel metabolic pathways is a recent strategy developed by the EFI. Previous studies focused on the tripartite ATP-inde-
dependent periplasmic SBPs, where novel pathways for ethanolamine and d-Ala-d-Ala catabolism were discovered (12). The research presented here utilizes results from an analogous study of the ABC SBPs to elucidate novel pathways for catabolism of both galactitol and d-altritol, along with the discovery of novel enzymatic functions.

A pathway for d-altritol catabolism was previously unknown. Consequently, the elucidation of this pathway resulted in the annotation of multiple proteins of unknown function. *AtuZnD* catalyzes the NAD\(^{+}\)/H-dependent oxidation of d-altritol to d-tagatose; previously, this reaction was unknown (no EC number). *AtuZnD* is a zinc-binding dehydrogenase (PF00107). In the SSN for PF00107 generated with an alignment score threshold of 70, the cluster containing *AtuZnD* is highly interconnected, so it is hypothesized to be isofunctional (>50% sequence identity, Fig. 4). This cluster contains the sequences of 29 other dehydrogenases presumed to oxidize d-altritol to d-tagatose. Ninety percent of the proteins are from Agrobacterium, Rhizobium, or Ochrobactrum (all \(\alpha\)-proteobacteria). These genera comprise soil-dwelling bacteria with the ability to grow on, or near, plants either in symbiotic or pathogenic relationships. Therefore, the fact that these organisms possess the genetic capacity to catabolize polyols (primary photosynthetic products) for cellular carbon is reasonable.

*AtuFK* catalyzes the phosphorylation of d-tagatose to d-tagatose 6-phosphate. Previously, this activity had been identified only in the extracts of galactitol-grown *Mycobacterium butyricum*, but no gene or protein had been associated with the activity (EC 2.7.1.101) (48). The SSN of PF00294 was generated at an alignment score threshold of 55 (35% sequence identity). The cluster containing *AtuFK* is hypothesized to be isofunctional (40% sequence identity, Fig. 5) and contains 1,956 kinases that, therefore, are hypothesized to phosphorylate d-tagatose to d-tagatose 6-phosphate. These kinases are found in diverse organisms (Clostridia, \(\alpha\), \(\beta\), and \(\gamma\)-proteobacteria), indicating that additional metabolic pathways involve the phosphorylation of d-tagatose.

### TABLE 3

| Neighbor   | Substrate            | \(k_{\text{cat}}\) | \(K_m\) | \(k_{\text{cat}}/K_m\) |
|------------|----------------------|---------------------|---------|------------------------|
| *AtuSorbD* | Galactitol           | 3.5 ± 0.10          | 0.10    | 1.8 \(\times\) \(10^3\) |
| *AtuSorbD* | d-Sorbitol           | 4.4 ± 0.10          | 0.10    | 2.3 \(\times\) \(10^3\) |
| *AtuZnD*   | d-Altritol           | 1.9 ± 0.03          | 0.10    | 1.4 \(\times\) \(10^3\) |
| *AtuFK*    | d-Tagatose           | 3.0 ± 0.10          | 0.08    | 2.3 \(\times\) \(10^3\) |
| *AtuFK*    | d-Fructose           | 1.5 ± 0.10          |         | 1.9 \(\times\) \(10^3\) |
| *AtuTag6PK*| d-Tagatose 6-phosphate | 8.7 ± 0.40          | 0.98    | 8.9 \(\times\) \(10^3\) |

FIGURE 6. Genome neighborhood of the AltSBP homologue in *A. tumefaciens* C58 (†) is shown in A. Transporter proteins are colored red; transcriptional regulators are colored gold; and enzymes are colored green. The conserved catalytic Pfam family members are marked with *. B shows the significant up-regulation of the conserved genes when cells are grown on d-altritol. C shows the significant up-regulation of the conserved genes when cells are grown on galactitol. Protein identifiers in B and C correspond to the following KEGG locus annotations in A: *AtuSorbD* (sorbitol dehydrogenase); *AtuZnD* (zinc-binding dehydrogenase); *AtuFK* (fructokinase); *AtuTag6PK* (tagatose-6-P kinase); *AtuLacI* (transcriptional regulator); *AtuHomoD* (homoserine dehydrogenase); and *AtuHisOD* (3-hydroxyisobutyryl dehydrogenase).
AtuTag6PK catalyzes the epimerization of D-tagatose 6-phosphate to D-fructose 6-phosphate. Previously, this reaction had not been identified (no EC number). Furthermore, no protein in PF08013 had an assigned catalytic activity, although members were thought to be involved in galactitol and N-galactosamine catabolism via tagatose 1,6-bisphosphate (42). The SSN of PF08013 was generated with an alignment score threshold of 145 (\( \geq 60\% \) sequence identity); the cluster containing AtuTag6PK is hypothesized to be isofunctional (\( \geq 65\% \) sequence identity, Fig. 6). The sequences originate from Agrobacterium, Rhizobium, or Mesorhizobium (all \( \alpha \)-proteobacteria). This cluster contains 55 presumed orthologues that epimerize D-tagatose 6-phosphate to D-fructose 6-phosphate. Other members of PF08013 may perform different mechanistically related reactions (as is the case in PF00596, class II aldolases).

AtuSorbD catalyzes the oxidation of galactitol to D-tagatose (EC 1.1.1.16) and D-sorbitol to D-fructose (EC 1.1.1.14). The SSN of PF00106 was generated with an alignment score threshold of 105 (\( \geq 67\% \) sequence identity). The cluster containing AtuSorbD is hypothesized to be isofunctional (\( \geq 80\% \) identity, Fig. 7) and contains 27 likely orthologues that oxidize galactitol or D-sorbitol to D-tagatose. As with AtuZnD, \( \sim 90\% \) of sequences originate from plant-associated, soil-dwelling bacteria as follows: Agrobacterium, Rhizobium, or Ochrobactrum.

Galactitol 2-dehydrogenase is a known reaction (49), but only recently has a specific protein been associated with this activity (Uniprot Q1MLL4) (50). Q1MLL4 is also a member of PF00106, but it shares 32\% sequence identity with AtuSorbD. Consequently, the characterization of AtuSorbD assigns the same function to two different clusters in PF00106. The role of the enzymatic activity of Q1MLL4 in the metabolism of galactitol was unexplored in the earlier study.

Catabolism of galactitol is known to occur via two pathways. 1) Transport of galactitol via a phosphotransferase system (41). The resulting D-galactitol 1-phosphate is then oxidized to...
D-tagatose 6-phosphate, which is phosphorylated to D-tagatose 1,6-bisphosphate, and cleaved by an aldolase into the central metabolites glyceraldehyde 3-phosphate and DHAP. 2) The oxidoreductive D-galactose catabolic pathway in which D-galactose is reduced to galactitol (51). The galactitol then is oxidized and reduced at carbon 3 to epimerize galactitol to D-sorbitol. D-Sorbitol is oxidized at carbon 2 to form D-fructose.

Previously, no catabolic pathways had been identified by which galactitol is metabolized via D-tagatose, as found in this study. In total, the experimental characterizations conducted here produced new or corrected functional annotations for at least 2,067 protein sequences via confident annotation transfer.

When a GNN is generated for the D-tagatose kinase-containing cluster of PF00294, the most highly co-occurring Pfam families are found in a somewhat different genome context compared with that of the AltSBPs. In addition to PF00106 (52.6% co-occurrence), PF00107 (52.5% co-occurrence), and PF08013 (22.1% co-occurrence), genome neighbors also include the fructose-bisphosphate aldolase Pfam family (PF01116, 63.4% co-occurrence) and an additional member of the PfkB Pfam family (PF00294, 60.5% co-occurrence). The absence of these Pfam families in the AltSBP GNN indicates that the catabolic pathways containing D-tagatose kinase homologues that have conserved neighbors from PF01116 and PF00294 either do not use ABC transporters for D-altritol/galactitol or metabolize a different carbohydrate with D-tagatose as an intermediate.

Interestingly, when the PF01116 neighbors of D-tagatose kinase are mapped onto a SSN of PF01116 at an alignment score threshold of 100 (≈60% identity) (Fig. 10), they cluster with SwissProt-reviewed D-tagatose-1,6-bisphosphate aldolases. Therefore, the PF00294 neighbors of D-tagatose kinase likely are D-tagatose-6-phosphate kinases. This is supported by their similar co-occurrence with the proposed D-tagatose-1,6-bis-
Novel d-Altritol and Galactitol Catabolic Pathways

phosphate aldolases, and these proteins are not significantly similar to any proteins of reviewed function. They may represent yet another novel function within PF00294, which will require further study.

Considering the percent co-occurrences in the GNN, the pathway in which d-tagatose 6-phosphate is epimerized to d-fructose 6-phosphate is the least prevalent. The GNN suggests that ~60% of the putative pathways result in the conversion of d-tagatose to d-tagatose 1,6-bisphosphate, followed by cleavage by an aldolase (PF00294 and PF01116); in only ~40% of the pathways is d-tagatose epimerized to d-fructose 6-phosphate. Of the genera that contain AtuSorbD, AtuZnD, and AtuTag6PK (Agrobacterium, Rhizobium, Ochrobactrum, and Mesorhizobium), Ochrobactrum is the only genus predicted to utilize glycolysis for glucose metabolism. Agrobacterium, Rhizobium, and Mesorhizobium do not utilize glycolysis, but rather employ the Entner-Doudoroff pathway and the pentose phosphate pathway (PPP) (32, 52). Therefore, it follows that these organisms metabolize d-altritol to d-fructose 6-phosphate (a PPP intermediate) rather than glyceraldehyde 3-phosphate and DHAP (glycolysis intermediates and products of d-tagatose-1,6-bisphosphate aldolase). Further study is required, but organisms in the d-tagatose kinase cluster of PF00294, which utilize glycolysis, likely catabolize D-altritol/galactitol via d-tagatose 1,6-bisphosphate, although those that employ the Entner-Doudoroff pathway and PPP catabolize D-altritol/galactitol via d-fructose 6-phosphate.

Three other ligands (d-mannititol, d-arabinitol, and volemitol) provided greater thermal stabilization (14.1, 11.2, and 10.9 °C) than d-tagatose (7.7 °C) and d-altritol (7.9 °C), despite the data presented here that the collocated dehydrogenases utilize galactitol and d-altritol as substrates. Therefore, when possible, the largest number of DSF hits and related compounds should be used to screen enzymes predicted to be catabolically downstream of the transporter. Despite this caveat, the ability to define the correct chemotype and stereochemical centers by DSF drastically reduces the number of ligands required for biochemical analysis of associated enzymes. Further work is required to determine whether A6X5Q5 or its homolog from A. tumefaciens CS8 (AtuSBP, UniProt ID Q7CR52) can transport these alternative substrates for catabolic transformation by non-collocated genes.

The characterization of two novel catabolic pathways and the identification of three novel enzymatic activities demonstrate the utility of screening the ligand specificities SBPs to enable functional discovery. These efforts underscore the considerable impact that ligand screening of SBPs can have on correcting the functional annotations of mis- and unannotated proteins in UniProtKB.

Author Contributions—D. J. W., S. C. A., and J. A. G. designed the study. D. J. W. and J. A. G. wrote the paper. D. J. W. and L. C. performed the research. M. W. V. selected the ABC SBP targets, aided in the ABC SBP ligand library design, and curated the DSF results. N. A. purified the ABC SBPs and subjected them to DSF. J. T. B. aided in the ABC SBP ligand library design. S. C. A. and J. A. G. were the primary investigators of this study.

Acknowledgments—We thank Drs. Katie Whalen and Brian San Francisco for critical reading of the manuscript.

References

1. Devos, D., and Valencia, A. (2001) Intrinsic errors in genome annotation. Trends Genet. 17, 429–431.
2. Jones, C. E., Brown, A. L., and Baumann U. (2007) Estimating the annotation error rate of the curated GO database sequence annotations. BMC Bioinformatics 8, 170.
3. Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput. Biol. 5, e1000605.
4. Gills, W. R., Audit, B., De Angelis, D., Tsoka, S., and Ouzounis, C. A. (2002) Modeling the percolation of annotation errors in a database of protein sequences. Bioinformatics 18, 1641–1649.
5. Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poultier, C. D., Rausch, F. M., Sali, A., Shoichet, B. K., and Swedler, J. V. (2011) The Enzyme Function Initiative. Biochemistry 50, 9950–9962.
6. Erb, T. J., Evans, B. S., Cho, K., Warlick, B. P., Siriam, J., Wood, B. M., Imker, H. J., Swedler, J. V., Tabita, F. R., and Gerlt, J. A. (2012) A Rubisco-like protein links SAM metabolism with isoprenoid biosynthesis. Nat. Chem. Biol. 8, 926–932.
7. Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Bonnano, J. B., Hillerich, B. S., Seidel, R. D., Babbitt, P. C., Almo, S. C., Swedler, J. V., Gerlt, J. A., Cronan, J. E., and Jacobson, M. P. (2013) Discovery of new enzymes and metabolic pathways by structure and genome context guided analyses. Nature 502, 698–702.
8. Bouvier, J. T., Groninger-Poe, F. P., Vetting, M., Almo, S. C., and Gerlt, J. A. (2014) Galactarol β-lactone isomerase: lactone isomerization by a member of the amidohydrolase superfamily. Biochemistry 53, 614–616.
9. Wiechekii, D. J., Balkhazor, B. M., Chau, A. C., Vetting, M. W., Fedoroff, A. A., Fedorov, E. V., Lukk, T., Patkowsky, Y. V., Stead, M. B., Hillerich, B. S., Seidel, R. D., Almo, S. C., and Gerlt, J. A. (2014) Discovery of function in the enolase superfamily: d-mannitohydrolase and d-gluculactonohydrolases in the d-mannitohydrolase subgroup. Biochemistry 53, 2722–2731.
10. Zhao, S., Sakai, A., Zhang, X., Vetting, M. W., Kumar, R., Hillerich, B., San Francisco, B., Solbiati, I., Steves, A., Brown, S., Akiva, E., Barber, A., Seidel, R., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2014) Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. eLife 3, e03275.
11. Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., and Whalen, K. L. (2015) Enzyme function initiative–enzyme simi-
larity tool (EF-I-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037.

12. Vetting, M. W., Al-Obaidi, N., Zhao, S., San Francisco, B., Kim, J., Wichelecki, J. D., Bouvier, I. T., Solbiati, J. O., Yu, H., Zhang, X., Rodionov, D. A., Love, J. D., Hillerich, B. S., Seidel, R. D., Quinn, R. I., et al. (2015) Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute-binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**, 909–931.

13. Jones, P. M., and George, A. M. (2004) The ABC transporter structure and specificity of DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell. Mol. Life Sci.* **61**, 682–699.

14. Davidson, A. L., Dassa, E., Oreille, C., and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72**, 317–346.

15. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: the ABC transporters. *J. Biol. Chem.* **28976**, of rare hexoses using microorganisms and related enzymes. *Beilstein J. Org. Chem.* **9**, 2434–2445.

16. Alajarín, R., Garcia-Junceda, E., and Wong, C. (1995) A short enzymic synthesis of l-glucose from dihydroxyacetone phosphate and l-glycerol-dehyde. *Beilstein J. Org. Chem.* **9**, 2434–2445.

17. Zijie, L., Yahui, G., Hideki, N., Xiaodong, G., and Li C. (2013) Biosynthesis of rare hexoses using microorganisms and related enzymes. *Beilstein J. Org. Chem.* **9**, 2434–2445.

18. Zijie, L., Yahui, G., Hideki, N., Xiaodong, G., and Li C. (2013) Biosynthesis of rare hexoses using microorganisms and related enzymes. *Beilstein J. Org. Chem.* **9**, 2434–2445.

19. Fisher, M., Abramov, M., Van Aerschot, A., Xu, D., Juliano, R. L., and Herdewijn, P. (2007) Inhibition of MDR1 expression with altritol-modified antiviral siRNAs. *Artif. DNA PNA XNA* **1**, 1–26.

20. Fisher, M., Abramov, M., Van Aerschot, A., Xu, D., Juliano, R. L., and Herdewijn, P. (2007) Inhibition of MDR1 expression with altritol-modified siRNAs. *Nucleic Acids Res.* **35**, 1064–1074.

21. Raven, J. A., Beardall, J., Chudek, J. A., Scrimgeour, C. M., Clayton, M. N., and McNicol, S. G. (2001) Altritol synthesis by *Notothea anomala*. *Phytochemistry* **58**, 389–394.

22. Fox, T. C., Kennedy, R. A., and Loescher, W. H. (1986) Developmental changes in photosynthetic gas exchange in the polyol-synthesizing species. *Triticum aestivum*. *Plant Physiol.* **82**, 307–311.

23. Muniruzzaman, S., Koyama, T., Iizuka, T., and Umezawa, H. (1994) Development of α-talitol from D-psicose by *Bacillus pallidus*. *Biosci. Biotechnol. Biochem.* **58**, 393–394.

24. Muniruzzaman, S., Koyama, T., Iizuka, T., and Umezawa, H. (1994) Development of α-talitol from D-psicose by *Bacillus pallidus*. *Biosci. Biotechnol. Biochem.* **58**, 393–394.

25. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *J. Bacteriol.* **172**, 3–13.

26. Toporowski, M. C., Nomellini, J. F., Awram, P., and Smit, J. (2004) Two outer membrane proteins are required for maximal type I secretion of the *A. tumefaciens* type IV secretion system. *J. Bacteriol.* **186**, 8000–8009.

27. Brinkkötter, A., Shakeri-Garakani, A., and Lengeler, J. W. (2002) Two-class II α-talotase-bisphosphate aldolases from enteric bacteria. *Arch. Microbiol.* **177**, 410–419.

28. Nobelmann, B., and Lengeler, J. W. (1996) Molecular analysis of the genes coding for *Escherichia coli* and of their roles in galactitol transport and metabolism. *J. Bacteriol.* **178**, 6790–6795.

29. Brinkkötter, A., Klöss, H., Alpert, C., and Lengeler, J. W. (2000) Pathways for the utilization of N-acetylgalactosamine and galactosamines in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **64**, 125–135.

30. Reizer, J., Ramseier, T. M., Reizer, A., Charbit, A., and Saier, M. H., Jr. (1996) Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylglactosamine metabolic pathway in *Escherichia coli*. *Microbiology* **142**, 231–250.

31. Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J., and Arkin, A. P. (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38**, D396–D400.

32. Dam, P., Olman, V., Harris, K., Su, Z., and Xu, Y. (2007) Operon Prediction using both genome-specific and general genomic information. *Nucleic Acids Res.* **35**, 288–298.

33. Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009) DOOR: a Database for prOkaryotic OpeRons. *Bioinformatics* **25**, 4192–4203.

34. Arch. Microbiol. **177**, 410–419.

35. Nobelmann, B., and Lengeler, J. W. (1996) Molecular analysis of the genes coding for *Escherichia coli* and of their roles in galactitol transport and metabolism. *J. Bacteriol.* **178**, 6790–6795.

36. Nobelmann, B., and Lengeler, J. W. (1996) Molecular analysis of the genes coding for *Escherichia coli* and of their roles in galactitol transport and metabolism. *J. Bacteriol.* **178**, 6790–6795.