



(51) International Patent Classification:

C07K 14/40 (2006.01) **C12N 15/09** (2006.01)
C12N 1/19 (2006.01) **C12P 7/10** (2006.01)

(21) International Application Number:

PCT/US2014/064168

(22) International Filing Date:

5 November 2014 (05.11.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/900,115 5 November 2013 (05.11.2013) US

(71) Applicant: **BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 West 7th Street, Austin, Texas 78701 (US).(72) Inventors: **ALPER, Hal**; 6108 Colina Lane, Austin, Texas 78759 (US). **YOUNG, Eric**; 29 Acton Street #1, Arlington, Massachusetts 02476 (US).(74) Agents: **CLARK, Kevin M.** et al.; Kilpatrick Townsend & Stockton LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111 (US).

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

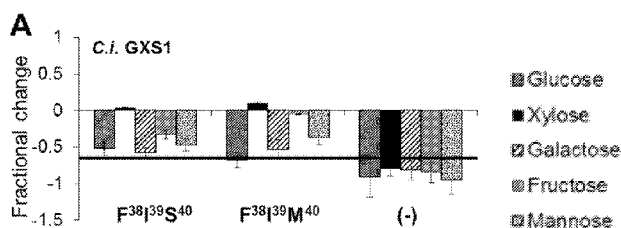
(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHODS FOR ENGINEERING SUGAR TRANSPORTER PREFERENCES

Figure 4



(57) **Abstract:** Provided herein are compositions and methods useful for transporting xylose, arabinose and other monosaccharides, into a yeast cell.

METHODS FOR ENGINEERING SUGAR TRANSPORTER PREFERENCES

CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 [0001] This application claims the benefit of U.S. Provisional Application No. 61/900,115, filed November 5, 2013, which is hereby incorporated by reference in its entirety and for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

- 10 [0002] This invention was made with Government support under grant number CBET-1067506 awarded by the National Science Foundation. The Government has certain rights in this invention.

REFERENCE TO A "SEQUENCE LISTING," A TABLE, OR A COMPUTER PROGRAM LISTING APPENDIX SUBMITTED ON A COMPACT DISK

- 15 [0003] The Sequence Listing written in file 93331-920858_ST25.TXT, created on November 3, 2014, 12,049 bytes, machine format IBM-PC, MS-Windows operating system, is hereby incorporated by reference.

BACKGROUND OF THE INVENTION

- [0004] The quest for an optimal xylose pathway in yeast is of utmost importance along the way to realizing the potential of lignocellulosic biomass conversion into fuels and chemicals. An often overlooked aspect of this catabolic pathway is the molecular transport of this sugar. Molecular transporter proteins facilitate monosaccharide uptake and serve as the first step in catabolic metabolism. In this capacity, the preferences, regulation, and kinetics of these transporters ultimately dictate total carbon flux. Optimization of intracellular catabolic pathways only increases the degree to which transport exerts control over metabolic flux. Thus, monosaccharide transport profiles and rates are important design criteria and a driving force to enable metabolic engineering advances. Among possible host organisms, *Saccharomyces cerevisiae* is an emerging industrial organism. However, *S. cerevisiae* lacks an endogenous xylose catabolic pathway and thus is unable to natively utilize the second most abundant sugar in lignocellulosic biomass, xylose. Decades of research have been focused on improving xylose
- 20
25
30

catabolic pathways in recombinant *S. cerevisiae*, but little effort has been focused on the first committed step of the process—xylose transport, an outstanding limitation in the efficient conversion of lignocellulosic sugars. There is a need in the art for efficient transport systems for xylose in yeast. Provided herein are solutions to these and other problems in the art.

5 BRIEF SUMMARY OF THE INVENTION

[0005] Accordingly, provided herein, *inter alia*, are compositions and methods useful for transporting xylose, arabinose, galactose and other monosaccharides and polysaccharides into a yeast cell.

[0006] In a first aspect is a recombinant xylose transporter protein including a transporter
10 motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia* *GXS1* protein. The transporter motif sequence is -G-G/F-X¹-X²-X³-G-. X¹ is D, C, G, H, I, L, or F. X² is A, D, C, E, G, H, or I. X³ is N, C, Q, F, G, L, M, S, T, or P. The transporter motif sequence is not -G-G-L-I-F-G- or -G-G-F-I-F-G-.

[0007] In another aspect is a recombinant galactose-arabinose transporter protein including a
15 transporter motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia* *GXS1* protein. The transporter motif sequence is -G-G/F-X⁴-X⁵-X⁶-G-. X⁴ is D, C, F, G, H, L, R, T, or P. X⁵ is A, C, E, F, H, K, S, P, or V. X⁶ is R, D, E, F, H, I, M, T, or Y. The sequence is not -G-G-L-V-Y-G-, or -G-G-F-V-F-G-.

[0008] Also provided herein are yeast cells that include a recombinant hexose or pentose
20 transporter protein described herein. In one aspect the yeast cell includes a recombinant xylose transporter protein described herein. In another aspect the yeast cell includes a recombinant galactose-arabinose transporter described herein.

[0009] Provided herein are nucleic acid sequences that encode a recombinant hexose or
25 pentose transporter protein described herein. In one aspect the nucleic acid encodes a recombinant xylose transporter protein described herein. In another aspect the nucleic acid encodes a recombinant galactose-arabinose transporter protein described herein.

[0010] Further provided herein are methods of transporting a hexose or pentose into a yeast
30 cell using the recombinant transporter proteins described herein. In one aspect is a method of transporting xylose into a yeast cell by contacting a yeast cell having a recombinant xylose transporter protein described herein with a xylose compound described herein. The xylose transporter protein is allowed to transport the xylose compound into the yeast cell. In another

aspect is a method of transporting galactose or arabinose into a yeast cell by contacting a yeast cell having a recombinant galactose-arabinose transporter protein described herein with a galactose compound or an arabinose compound described herein. The recombinant galactose-arabinose transporter protein is allowed to transport the galactose compound or the arabinose compound into the yeast cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Figure 1 – Sequence categorization and phenotypic classification of native and heterologous transporters. **A)** The distribution of phenotypic classes for 46 cloned wild type major facilitator superfamily transporters. **B)** The distribution of each sequence category present in each phenotypic class. Transporters containing the conserved motif are enriched in the phenotypic classes that confer growth on xylose. **C)** Weblogos of the phenotypic classes illustrate enrichment of the G-G/F-XXXG motif in TMS1. *Abbreviations:* $\mu_{\text{all}} = 0$: no growth the five carbon sources tested. $\mu_X = 0$: growth on hexoses but not xylose. $\mu_X < \mu_G$: growth on xylose is less than that on glucose. $\mu_X > \mu_G$: growth on xylose is greater than that on glucose.

[0012] Figure 2 – Classification tree of fractional change in carbon source growth profile. This figure depicts hypothetical fractional change data in order to demonstrate how these phenotypes were classified. Little fractional change across all sugars indicates that the substitution does not control efficiency or selectivity in this background. Amplification or attenuation of growth rates across all carbon sources indicates an efficiency substitution. Amplification of growth on one sugar, ideally xylose, and attenuation of all others indicates a selectivity substitution.

[0013] Figure 3 – Fractional change of saturation mutagenesis libraries of *C. intermedia* *GXS1*. **A)** Fractional change in growth by substitutions at position 38. **B)** Fractional change in growth by substitutions at position 39. **C)** Fractional change in growth by substitutions at position 40. The solid line is the confidence line for no growth based on the negative control sample.

[0014] Figure 4 – Growth characterization of *C. intermedia* *gxs1* triple mutants. **A)** Fractional change from wild type for the two triple mutants and an empty vector control. **B)** Average growth curves on xylose based on optical density at 600 nm. **C)** Average growth curves on glucose based on optical density at 600 nm.

[0015] Figure 5 – Further characterization of *C. intermedia* *gxs1* Phe³⁸ Ile³⁹ Met⁴⁰ triple mutant. **A)** Glucose uptake at high cell density for *S. cerevisiae* *EX.12* expressing wild type,

Phe³⁸ Ile³⁹ Met⁴⁰, and empty vector. **B)** Xylose uptake at high cell density for *S. cerevisiae* EX.12 expressing wild type, Phe³⁸ Ile³⁹ Met⁴⁰, and empty vector. **C)** Inhibition of growth rate on xylose with increasing glucose concentration. **D)** $V_{\max}$ of both the wild type and the mutant. **E)** K_M of both the wild type and triple mutant. Error is based on standard deviation of biological replicates.

5 **[0016] Figure 6 – Growth characterization of *S. stipitis* RGT2 and mutants. A) Fractional change from wild type for the two single mutants and an empty vector control. B) Average growth curves on xylose based on optical density at 600 nm. C) Average growth curves on glucose based on optical density at 600 nm.**

10 **[0017] Figure 7 – Growth characterization of *S. cerevisiae* HXT7 and mutants. A) Fractional change from wild type for the mutants and an empty vector control. B) Average growth curves on xylose based on optical density at 600 nm. C) Average growth curves on glucose based on optical density at 600 nm.**

15 **[0018] Figure 8 - Maximum exponential growth rates for all cloned native and heterologous transporters. Bar chart of growth rate (μ) calculated from growth curves of *S. cerevisiae* EX.12 measured on a Bioscreen C. Carbon source profiling on five different sugars allows better functional classification than measuring only glucose and xylose. Error is standard deviation of biological triplicates. A) Transporters cloned in the initial study measured for the first time in *S. cerevisiae* EX.12. B) Novel transporters identified and characterized. Abbreviations: Empty – empty vector control strain. A.t. – *Arabidopsis thaliana*. C.i. – *Candida intermedia*. C.n. – *Cryptococcus neoformans*. D.h. – *Debaryomyces hansenii*. S.c. – *Saccharomyces cerevisiae*. S.s. – *Scheffersomyces stipitis*. Y.l. – *Yarrowia lipolytica*.**

25 **[0019] Figure 9 - High cell density cofermentation in *S. cerevisiae* EX.12. Cells were inoculated at OD 20 in a mixture of 10 g/L glucose and 10 g/L xylose. Optical density, glucose, xylose, and ethanol concentration was measured over the length of the fermentation. Note that the triple mutant does not consume either xylose or glucose, nor is an appreciable amount of ethanol produced in this multiple knockout strain. A) Optical density over time. B) Glucose concentration in the media over time. C) Xylose concentration in the media over time. D) Ethanol concentration in the media over time.**

30 **[0020] Figure 10 - High cell density cofermentation in *S. cerevisiae* YSX3. Cells were inoculated at OD 20 in a mixture of 10 g/L glucose and 10 g/L xylose. Optical density, glucose, xylose, and ethanol concentration was measured over the length of the fermentation. Note that**

the triple mutant does not appreciably alter the fermentation dynamics in a strain that is expressing the full suite of transporters. **A)** Optical density over time **B)** Glucose concentration in the media over time. **C)** Xylose concentration in the media over time. **D)** Ethanol concentration in the media over time.

5 **[0021] Figure 11** - Growth curves of transporters of interest. Optical density measurements from the Bioscreen C were plotted over time. Each line represents the growth curve for *S. cerevisiae* *EX.12* expressing a transporter on a particular carbon source. **A)** *D.h. 2D01474*. **B)** *S.s. RGT2*. **C)** *D.h. 2E01166*. **D)** *D.h. 2B05060*. **E)** *S.c. STL1*. **F)** *S.s. AUT1*.

[0022] Figure 12 - Phylogenetic tree and growth rate. Phylogram constructed in TreeView of a
10 ClustalW multiple sequence alignment with the full amino acid sequences of all transporters. To the right of the phylogram is plotted the exponential growth rate of *S. cerevisiae* *EX.12* conferred by transporter expression. A blue line and a green line are placed across the chart to mark the upper limit of no growth for glucose and xylose, respectively. Note the most robust glucose growth phenotypes are clustered in the HXT family and related transporters. Some of the more
15 desirable growth phenotypes for xylose growth are clustered in the transporters related to *C.i. GXS1* and *S.s. XUT3*.

[0023] Figure 13 - Relatedness based on G-G/F-XXXG motif and growth rate data. Phylogram constructed in TreeView of a ClustalW multiple sequence alignment of the G-G/F-
20 XXG motif of each transporter. To the right of the phylogram is plotted the exponential growth rate of *S. cerevisiae* *EX.12* conferred by transporter expression. Two lines are placed across the chart to mark the upper limit of no growth for glucose and xylose. Arranging the transporters in this fashion remarkably clusters conferred phenotype better than basing the alignment on the whole amino acid sequence. This is further evidence of the influence the G-G/F-XXG motif has over monosaccharide uptake.

25 **[0024] Figure 14**— Carbon source profile comparison. **A)** *C.i. GXS1* and mutants. **B)** *S.s. RGT2* and mutants. **C)** *S.c. HXT7* and mutants. Note that these values are maximum exponential growth rates, and therefore may produce different comparisons than the late-stage linear exponential portions of the growth curves.

[0025] Figure 15 – Growth characterization of *C. intermedia gxs1* rationally designed triple
30 mutants. Fractional change from wild type is calculated on a variety of carbon sources for five mutants with differing transporter motif sequences (e.g. FLS, FIS, FIM, RPT, TPT, *VP which

contains a stop codon in the motif) compared to the negative control with no transporter motif sequence. The T³⁸P³⁹T⁴⁰ mutant shows a distinct preference toward galactose and away from the other sugars tested.

[0026] **Figure 16** – Growth curves of rational *gxs1* mutants by mutation. Growth curves are presented for the mutants described in Figure 15 on glucose, xylose, galactose, fructose, and mannose. Data is presented in graphs separated by mutant.

[0027] **Figure 17** – Growth curves of rational *gxs1* mutants by mutation. Growth curves are presented for the mutants described in Figure 15 on glucose, xylose, galactose, fructose, and mannose. Data is presented in graphs separated by carbon source.

[0028] **Figure 18** – Rewiring *xut3* transporter proteins through the equivalent of the 297 residue from *C. intermedia GXS1*. **A)** Identification of previously identified mutations in the *xut3* mutant transporter. **B)** Saturation mutagenesis was performed on the equivalent of the 297 residue from *C. intermedia GXS1*. Fractional change on growth of various carbon sources was measured and the results illustrated that this residue can control sugar transporter preference.

DETAILED DESCRIPTION OF THE INVENTION

[0029] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell culture, molecular genetics, organic chemistry, and nucleic acid chemistry and hybridization described below are those well-known and commonly employed in the art. Standard techniques are well known and commonly used in the art for nucleic acid and peptide synthesis. The techniques and procedures are generally performed according to conventional methods in the art and various general references (*see generally*, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference), which are provided throughout this document.

[0030] "Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, and complements thereof. The term "polynucleotide" refers to a linear sequence of nucleotides. The term "nucleotide" typically refers to a single unit of a polynucleotide, *i.e.*, a monomer. Nucleotides can be ribonucleotides, deoxyribonucleotides, or modified versions thereof. Examples of polynucleotides contemplated herein include single

and double stranded DNA, single and double stranded RNA (including siRNA), and hybrid molecules having mixtures of single and double stranded DNA and RNA. Nucleic acid as used herein also refers nucleic acids that have the same basic chemical structure as a naturally occurring nucleic acids. Such analogues have modified sugars and/or modified ring substituents, but retain the same basic chemical structure as the naturally occurring nucleic acid. A nucleic acid mimetic refers to chemical compounds that have a structure that is different the general chemical structure of a nucleic acid, but that functions in a manner similar to a naturally occurring nucleic acid. Examples of such analogues include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

[0031] "Synthetic mRNA" as used herein refers to any mRNA derived through non-natural means such as standard oligonucleotide synthesis techniques or cloning techniques. Such mRNA may also include non-proteinogenic derivatives of naturally occurring nucleotides. Additionally, "synthetic mRNA" herein also includes mRNA that has been expressed through recombinant techniques or exogenously, using any expression vehicle, including but not limited to prokaryotic cells, eukaryotic cell lines, and viral methods. "Synthetic mRNA" includes such mRNA that has been purified or otherwise obtained from an expression vehicle or system.

[0032] The words "complementary" or "complementarity" refer to the ability of a nucleic acid in a polynucleotide to form a base pair with another nucleic acid in a second polynucleotide. For example, the sequence A-G-T is complementary to the sequence T-C-A. Complementarity may be partial, in which only some of the nucleic acids match according to base pairing, or complete, where all the nucleic acids match according to base pairing.

[0033] Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are near each other.

[0034] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one

or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

[0035] The term "recombinant" when used with reference to, for example, a cell, nucleic acid, or protein, indicates that the cell, nucleic acid, or protein, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express genes otherwise modified from those found in the native form of a cell (e.g. genes encoding a mutation in a native or non-native transporter protein, such as a transporter motif sequence described herein). For example, a recombinant protein may be a protein that is expressed by a cell or organism that has been modified by the introduction of a heterologous nucleic acid (e.g. encoding the recombinant protein).

[0036] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0037] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0038] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants

refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are “silent variations,” which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual probe sequences.

[0039] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0040] The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (*see, e.g., Creighton, Proteins* (1984)).

[0041] A “yeast cell” as used herein, refers to a eukaryotic unicellular microorganism carrying out metabolic or other function sufficient to preserve or replicate its genomic DNA. Yeast cells may carry out fermentation of sugars described herein. Fermentation may convert the sugar to a biofuel or biochemical as set forth herein. Yeast cells referenced herein include, for example, those species listed in Figure 19 or Figure 20. Yeast cells referenced herein include, for example,

the following species: *Kluyveromyces lactis*, *Torulaspora delbrueckii*, *Zygosaccharomyces rouxii*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Candida intermedia*, *Cryptococcus neoformans*, *Debaryomyces hansenii*, *Phaffia rhodozyma*, or *Scheffersomyces stipitis*.

[0042] The term “biofuel” as used herein refers to a convenient energy containing substance produced from living organisms (e.g. biomass conversion to a fuel). Thus, biofuels may be produced through, for example, fermentation of carbohydrates (e.g. sugars) found in biomass (e.g. lignocellulosic biomass). Biofuels may be solid, liquid, or gas forms. Biofuels include, for example, ethanol, biodiesel, vegetable oil, ether (oxygenated fuels), or gas (e.g. methane).

[0043] The term “biochemical” as used herein refers to chemicals produced by living organisms. Biochemicals herein include alcohols (e.g. butanol, isobutanol, 2,3-butanediol, propanol); sugars (e.g. erythritol, mannitol, riboflavin); carotenoids (e.g. β -carotene, lycopene, astaxanthin); fatty acids (e.g. ricinoleic acid, linolenic acid, tetracetyl phytosphingosine); amino acids (e.g. valine, lysine, threonine); aromatics (e.g. indigo, vanillin, styrene, p-hydroxystyrene); flavonoids (e.g. naringenin, genistein, kaempferol, quercetin, chrysin, apigenin, luteolin,); stillbenoids (e.g. resveratrol); terpenoids (e.g. β -amyrin, taxadiene, miltiradiene, paclitaxel, artemisinin, bisabolane); polyketides (e.g. aureothin, spectinabilin, lovastatin, geodin); or organic acids (e.g. citric acid, succinic acid, malic acid, lactic acid, polylactic acid, adipic acid, glucaric acid) produced by living organisms (e.g. a yeast cell). See e.g. Curran K.A., Alper H.S., Metabolic Engineering 14:289-297 (2012).

[0044] A “transporter motif sequence” as used herein refers to an amino acid sequence that, when present in a protein (e.g. a sugar transporter protein such as a MFS transporter protein), increases the ability of the protein to transport a sugar or sugar-containing compound into a yeast cell. The transporter motif sequence may impart a hexose sugar transport preference or pentose sugar transport preference to the protein. Thus, for example, the transporter motif sequence may impart preference to hexose sugars to a transporter protein, thereby allowing the transporter protein to preferentially transport hexoses into a yeast cell. The transporter motif sequence may impart preference to a single hexose (e.g. galactose). The transporter motif sequence may impart preference to more than one hexose sugar (galactose and mannose). The transporter motif sequence may impart preference to pentose sugars to a transporter protein, thereby allowing the transporter protein to preferentially transport pentose into a yeast cell. The transporter motif sequence may impart preference to a single pentose (e.g. xylose). The transporter motif sequence may impart preference to more than one pentose sugar (e.g. xylose and arabinose). The

transporter motif sequence may impart preference for at least two sugars (e.g. galactose and arabinose).

[0045] The transporter motif sequence described herein corresponds to residues corresponding to positions 36-41 of the *Candida intermedia GXS1* protein (“GXS1 motif sequence”). One skilled in the art will immediately recognize the identity and location of residues corresponding to positions 36-41 of the *Candida intermedia GXS1* protein in other transporter proteins with different numbering systems. For example, by performing a simple sequence alignment with *Candida intermedia GXS1* protein the identity and location of residues corresponding to positions 36-41 of the *Candida intermedia GXS1* protein are identified in other yeast transport proteins as illustrated in Figures 19 and 20. Insertion (e.g. substitution) of a transporter motif sequence into a yeast transport protein may thereby be performed resulting in a functional yeast transporter protein with an altered sugar transport preference (e.g. changing a preference for hexoses to a preference for pentoses). For example, amino acid residue positions 75-81 of *S. cerevisiae HXT7* protein correspond to amino acid residue positions 36-41 of the *Candida intermedia GXS1* protein. See e.g. Example 2 and SEQ ID NO:1.

SEQ ID NO:1

1	MGLEDNRMVKRFVNVGEKKAGSTAMAIIVGLFAASGGVLFGYDTGTISGVMTMDYVLARY	60
61	PSNKHSTADESSLIVSILSVGTFFGALCAPFLNDTLGRRWCLILSALIVFNIGAILQVI	120
121	STAIPLLCAGRVIAGFGVGLISATIPLYQSETAPKWIRGAIVSCYQWAITIGLFLASCVN	180
181	KGTEHMTNSGSYRIPLAIQCLWGLILGIGMIFLPETPRFWISKGNQEKAESLARLRKLP	240
241	IDHPDSLEELRDITAAYEFETVYGKSSWSQVFSHKNHQLKRLFTGVAIQAFQQLTGVNFI	300
301	FYYGTTFFKRAGVNGFTISLATNIVNVGSTIPGILLMEVLGRRNMLMGATGMSLSQLIV	360
361	AIVGVATSENNKSSQSVLVAFCIFIAFFAATWGPCA WVVVGELFPLRTRAKSVSLCTAS	420
421	NWLWNWGIAYATPYMVDEDKGNLGSNVFFIWGGFNLACVFFAWYFIYETKGLSLEQVDEL	480
481	YEHVSKAWKSKGFVPSKHSFREQVDQQMDSKTEAIMSEEASV	522

[0046] A “transporter protein” as used herein refers to a transmembrane protein which transports sugars (e.g. hexoses and pentoses) into a yeast cell. The transporter protein may be a yeast transporter protein. The transporter protein may be a transporter protein belonging to the major facilitator superfamily (“MFS”) transporter proteins. A transporter protein may transport a hexose (e.g. galactose) into a yeast cell. A transporter protein may transport a pentose (e.g. xylose or arabinose) into a yeast cell. A transporter protein may be engineered, using the transporter motif sequences described herein, to alter its sugar preference (e.g. a transporter protein having a preference to transport a hexose compound may be converted to a transporter protein having a preference to transport a pentose compound). A transporter protein may be

characterized as a transporter protein derived from a particular organism. Where a transporter protein is derived from a particular organism, the endogenous sequence of the transporter protein may be maintained and residues corresponding to positions 36-41 of the *Candida intermedia* *GXS1* protein may be replaced with a transporter motif sequence. For example, a *C. intermedia* *gxs1* transporter protein is a *gxs1* transporter protein, a homolog thereof, or a functional fragment thereof, found in *C. intermedia* SEQ ID NO:1. Amino acids 75-81 of *S. cerevisiae hxt7* transporter protein may be replaced with a transporter motif sequence thereby forming a transporter protein with desired sugar transport characteristics described herein. The transporter protein may be a protein, functional fragment, or homolog thereof, identified by the following NCBI gene ID numbers: 836043, 831564, AJ937350.1, AJ875406.1, 2901237, 2913528, 8998057, 8999011, 50419288, 948529, 4839826, 4852047, 4851844, 4840896, 4840252, 4841106, 4851701, 2907283, 2906708, 2908504, 2909312, 2909701, 4935064, 851943, 856640, 856640, 851946, 856494, 8998297, 2902950, 2902912, 853207, 852149, 855023, 853216, 853236, 850536, 855398, 4836720, 4836632, 4840859, 2913215, 2902914, 2910370, 4838168, 2901237.

[0047] A “xylose compound” is xylose or a xylose-containing compound including at least one xylose moiety. Thus as used herein, the term xylose compound represents a single xylose, a chain including one or more xylose moieties, or a xylose moiety covalently or non-covalently bound to another chemical moiety (e.g. another sugar forming a xylose containing

polysaccharide or xylose bound to lignin). An “arabinose compound” is arabinose or an arabinose-containing compound including at least one arabinose moiety. Thus as used herein, the term arabinose compound represents a single arabinose, a chain including one or more arabinose moieties, or an arabinose moiety covalently or non-covalently bound to another chemical moiety (e.g. another sugar forming a arabinose containing polysaccharide or arabinose bound to lignin).

A “galactose compound” is galactose or a galactose-containing compound including at least one galactose moiety. Thus as used herein, the term galactose compound represents a single galactose, a chain including one or more galactose moieties, or a galactose moiety covalently or non-covalently bound to another chemical moiety (e.g. another sugar forming a galactose containing polysaccharide or bound to lignin).

[0048] Polysaccharides herein include hexose-only polysaccharides, pentose-only polysaccharides, and hexose-pentose mixture polysaccharides. The xylose compound, the arabinose compound, or the galactose compound may be derived from or form part of a

lignocellulosic biomass (e.g. plant dry matter that may used in as a source for pentose compounds or hexose compounds and for production of biofuels or biochemicals), hemicellulose, or other natural or synthetic sources for xylose, arabinose, or galactose. "Derived from" refers to extraction, removal, purification, or otherwise freeing a xylose compound, arabinose compound, or galactose compound from a source (e.g. lignocellulosic biomass) by either chemical processes (e.g. acid hydrolysis, ammonium explosion, or ionic liquids extraction) or through natural biological processes by organisms capable of using such sources for energy.

[0049] A "pentose compound" or "pentose" is a monosaccharide-containing compound having 5 carbon atoms. Pentose compounds include aldopentoses (e.g. pentose compounds having an aldehyde moiety at carbon 1) and ketopentoses (e.g. pentose compounds having a ketone moiety at carbon 2 or carbon 3). Pentose compounds include, for example, D/L-arabinose, D/L-lyxose, D/L-ribose, D/L-xylose, D/L-ribulose, and D/L-xylulose. The term "monosaccharide-containing" refers to a compound that includes at least one monosaccharide.

[0050] A "hexose compound" "or "hexose" is a monosaccharide-containing compound having 6 carbon atoms. Hexose compounds include aldohexoses (e.g. hexose compounds having an aldehyde moiety at carbon 1) and ketohexoses (e.g. hexose compounds having a ketone moiety at carbon 2). Hexose compounds include, for example, D/L-allose, D/L-altrose, D/L-glucose, D/L-mannose, D/L-gluose, D/L-idose, D/L-galactose, and D/L-talose.

[0051] The word "expression" or "expressed" as used herein in reference to a DNA nucleic acid sequence (e.g. a gene) means the transcriptional and/or translational product of that sequence. The level of expression of a DNA molecule in a cell may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of protein encoded by that DNA produced by the cell (Sambrook *et al.*, 1989 *Molecular Cloning: A Laboratory Manual*, 18.1-18.88). The level of expression of a DNA molecule may also be determined by the activity of the protein.

[0052] The term "gene" means the segment of DNA involved in producing a protein; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons). The leader, the trailer as well as the introns include regulatory elements that are necessary during the transcription and the translation of a gene. A "protein gene product" is a protein expressed from a particular gene.

[0053] “Contacting” is used in accordance with its plain ordinary meaning and refers to the process of allowing at least two distinct species (e.g. chemical compounds including biomolecules or cells) to become sufficiently proximal to react, interact or physically touch. It should be appreciated; however, the resulting reaction product or interaction can be produced directly between the added reagents or from an intermediate from one or more of the added reagents which can be produced in the reaction mixture.

[0054] The term “contacting” may include allowing two species to react, interact, or physically touch, wherein the two species may be a compound described herein (e.g. xylose compound, arabinose compound, or galactose compound) and a protein or enzyme described herein.

Contacting may include allowing the compound described herein to interact with a protein or enzyme that is involved in transporting hexose compounds or pentose compounds into a yeast cell.

[0055] Provided herein are recombinant hexose and pentose transporter proteins. In one aspect is a recombinant xylose transporter protein. The recombinant xylose transporter protein includes a transporter motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia* *GXS1* protein. The transporter motif sequence has the sequence -G-G/F-X¹-X²-X³-G-. X¹ is D, C, G, H, I, L, or F. X² is A, D, C, E, G, H, or I. X³ is N, C, Q, F, G, L, M, S, T, or P. In embodiments, the transporter motif sequence is not -G-G-L-I-F-G- or -G-G-F-I-F-G-.

[0056] X¹ may be D, C, G, I, L, or F. X¹ may be D, C, G, H, or F. X¹ may be D. X¹ may be C. X¹ may be G. X¹ may be I. X¹ may be L. X¹ may be H. X¹ may be F. X² may be D, C, E, G, H, or I. X² may be E, G, H, or I. X² may be H or I. X² may be H. X² may be I. X³ may be N, Q, F, M, S, T, or P. X³ may be F, M, S, or T. X³ may be S, T, or M. X³ may be T. X³ may be S. X³ may be M. When X¹ is F, X² may be I and X³ may be M or S.

[0057] The transporter motif sequence may be -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, -G-F-F-I-S-G-, -G-G-F-I-T-G-, -G-F-F-I-T-G-, -G-G-F-L-M-G-, -G-F-F-L-M-G-, -G-G-F-L-S-G-, -G-F-F-L-S-G-, -G-G-F-L-T-G-, -G-F-F-L-T-G-, -G-G-F-H-M-G-, -G-F-F-H-M-G-, -G-G-F-H-S-G-, -G-F-F-H-S-G-, -G-G-F-H-T-G- or -G-F-F-H-T-G-. The transporter motif sequence may be -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, -G-F-F-I-S-G-, -G-G-F-I-T-G-, or -G-F-F-I-T-G-. The transporter motif sequence may be -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, or -G-F-F-I-S-G-. The transporter motif sequence may be -G-G-F-I-M-G-, or -G-F-F-I-M-

G-. The transporter motif sequence may be -G-G-F-I-M-G-. The transporter motif sequence may be -G-F-F-I-M-G-. The transporter motif sequence may be -G-G-F-I-S-G-. The transporter motif sequence may be -G-F-F-I-S-G-. The transporter motif sequence may be -G-G-F-I-T-G-. The transporter motif sequence may be -G-F-F-I-T-G-. The transporter motif sequence may be -G-G-F-L-M-G-. The transporter motif sequence may be -G-F-F-L-M-G-. The transporter motif sequence may be -G-G-F-L-S-G-. The transporter motif sequence may be -G-F-F-L-S-G-. The transporter motif sequence may be -G-G-F-L-T-G-. The transporter motif sequence may be -G-F-F-L-T-G-. The transporter motif sequence may be -G-G-F-H-M-G-. The transporter motif sequence may be -G-F-F-H-M-G-. The transporter motif sequence may be -G-G-F-H-S-G-. The transporter motif sequence may be -G-F-F-H-S-G-. The transporter motif sequence may be -G-G-F-H-T-G-. The transporter motif sequence may be -G-F-F-H-T-G-.

[0058] The recombinant xylose transporter protein described herein may further include a mutation of an amino acid at the residue position corresponding to 297 of *Candida intermedia GXSI* protein. The amino acid at the residue position corresponding to 297 of *Candida intermedia GXSI* protein may be substituted with a Met, Ala, Ser, or Asn residue. The amino acid may be substituted with Met. The amino acid may be substituted with Ala. The amino acid may be substituted with Ser. The amino acid may be substituted with Asn. The recombinant xylose transporter protein may include a -G-G-F-I-M-G- transporter motif sequence and a Met substitution at the position corresponding to 297 of *Candida intermedia GXSI* protein. The mutations of the amino acid at the residue position corresponding to 297 of *Candida intermedia GXSI* protein may prevent transport of hexoses by the recombinant xylose transporter. The mutations of the amino acid at the residue position corresponding to 297 of *Candida intermedia GXSI* protein, in combination with the transporter motif sequences described herein, may prevent transport of hexoses by the recombinant xylose transporter.

[0059] The recombinant xylose transporter protein may be derived from a sugar transporter protein (e.g. a transporter protein (e.g. a MFS transporter protein), a homolog thereof, or a functional fragment thereof, found in in a cell). The xylose transporter protein may be derived from a yeast cell transporter protein (e.g. a transporter protein, a homolog thereof, or a functional fragment thereof, found in in a yeast cell). The yeast cell transporter protein may be a MFS transporter protein. The yeast cell may be a species set forth in Figure 19 or Figure 20. The recombinant xylose transporter protein may be derived from a *C. intermedia gxs1* transporter protein (e.g. a *gxs1* transporter protein, a homolog thereof, or a functional fragment thereof,

found in *C. intermedia* SEQ ID NO:1), a *S. stipitis* *rgt2* transporter protein (e.g. a *rgt2* transporter protein, a homolog thereof, or a functional fragment thereof, found in *S. stipitis*), or a *S. cerevisiae* *hxt7* transporter protein (e.g. a *hxt7* transporter protein, a homolog thereof, or a functional fragment thereof, found in *S. cerevisiae*). The recombinant xylose transporter protein
 5 may be derived from a *C. intermedia* *gxs1* transporter protein. The recombinant xylose transporter protein may be derived from a *S. stipitis* *rgt2* transporter protein. The recombinant xylose transporter protein may be derived from a *S. cerevisiae* *hxt7* transporter protein.

[0060] In another aspect is a recombinant galactose-arabinose transporter protein. The recombinant galactose-arabinose transporter protein includes a transporter motif sequence
 10 corresponding to residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia* *GXS1* protein. The transporter motif sequence has the sequence -G-G/F-X⁴-X⁵-X⁶-G-. X⁴ is D, C, F, G, H, L, R, T, or P. X⁵ is A, C, E, F, H, K, S, P, or V. X⁶ is R, D, E, F, H, I, M, T, or Y. The sequence is not -G-G-L-V-Y-G-, or -G-G-F-V-F-G.

[0061] X⁴ may be D, F, G, L, R, or T. X⁴ may be R, T, H, or F. X⁴ may be R. X⁴ may be T. X⁴
 15 may be H. X⁴ may be F. X⁵ may be A, E, F, P, H, or V. X⁵ may be P, H, or V. X⁵ may be P. X⁵ may be H. X⁵ may be V. X⁶ may be T, H, F, M, or Y. X⁶ may be F or Y. X⁶ may be T or M. X⁶ may be T. X⁶ may be H. X⁶ may be F. X⁶ may be M. X⁶ may be Y. When X⁴ is F or T, X⁵ may be P or I, and X⁶ may be M or T.

[0062] The transporter motif sequence may be -G-G-F-H-M-G-, -G-F-F-H-M-G-, -G-G-R-P-
 20 T-G-, -G-F-R-P-T-G-, -G-G-T-P-T-G-, or -G-F-T-P-T-G-. The transporter motif sequence may be -G-G-F-H-M-G-, -G-F-F-H-M-G-. The transporter motif sequence may be -G-G-R-P-T-G-, -G-F-R-P-T-G-. The transporter motif sequence may be -G-G-T-P-T-G-, or -G-F-T-P-T-G-. The transporter motif sequence may be -G-G-F-H-M-G-. The transporter motif sequence may be -G-F-F-H-M-G-. The transporter motif sequence may be -G-G-R-P-T-G-. The transporter motif
 25 sequence may be -G-F-R-P-T-G-. The transporter motif sequence may be -G-G-T-P-T-G-. The transporter motif sequence may be -G-F-T-P-T-G-.

[0063] The recombinant galactose-arabinose transporter protein described herein may include a mutation of an amino acid at the residue position corresponding to 297 of *Candida intermedia* *GXS1* protein. The amino acid at the residue position corresponding to 297 of *Candida*
 30 *intermedia* *GXS1* protein may be substituted with a Met, Thr, Ala, or Ile residue. The amino acid may be substituted with Met. The amino acid may be substituted with Thr. The amino acid may

be substituted with Ala. The amino acid may be substituted with Ile. The recombinant galactose-arabinose transporter protein may include a -G-G-T-P-T-G- transporter motif sequence and a Met substitution at the position corresponding to 297 of *Candida intermedia* *GXS1* protein. The mutations of the amino acid at the residue position corresponding to 297 of *Candida intermedia* *GXS1* protein may prevent transport of hexoses, other than galactose, by the recombinant galactose-arabinose transporter. The mutations of the amino acid at the residue position corresponding to 297 of *Candida intermedia* *GXS1* protein, in combination with the transporter motif sequences described herein, may prevent transport of hexoses, other than galactose, by the recombinant galactose-arabinose transporter.

[0064] The recombinant galactose-arabinose transporter protein may be derived from a sugar transporter protein (e.g. a transporter protein (e.g. a MFS transporter protein), a homolog thereof, or a functional fragment thereof, found in in a cell). The recombinant galactose-arabinose transporter protein may be derived from a yeast cell transporter protein (e.g. a transporter protein, a homolog thereof, or a functional fragment thereof, found in in a yeast cell). The transporter protein may be a MFS transporter protein. The yeast cell may be a species set forth in Figure 19 or Figure 20. The recombinant galactose-arabinose transporter protein may be derived from a *C. intermedia* *gxs1* transporter protein (e.g. a *gxs1* transporter protein, a homolog thereof, or a functional fragment thereof, found in *C. intermedia* SEQ ID NO:1), a *S. stipitis* *rgt2* transporter protein (e.g. a *rgt2* transporter protein, a homolog thereof, or a functional fragment thereof, found in *S. stipitis*), a *S. cerevisiae* *hxt7* transporter protein (e.g. a *hxt7* transporter protein, a homolog thereof, or a functional fragment thereof, found in *S. cerevisiae*), or a *S. cerevisiae* *GAL2* transporter protein (e.g. a *GAL2* transporter protein, a homolog thereof, or a functional fragment thereof, found in *S. cerevisiae*). The recombinant galactose-arabinose transporter protein may be derived from a *C. intermedia* *gxs1* transporter protein. The recombinant galactose-arabinose transporter protein may be derived from a *S. stipitis* *rgt2* transporter protein. The recombinant galactose-arabinose transporter protein may be derived from a *S. cerevisiae* *hxt7* transporter protein. The recombinant galactose-arabinose transporter protein may be derived from a *S. cerevisiae* *GAL2* transporter protein.

[0065] Further provided herein are nucleic acid sequences encoding the hexose or pentose transporter proteins described herein. In one aspect is a nucleic acid encoding a recombinant xylose transporter protein described herein. In another aspect is a nucleic acid encoding a recombinant galactose-arabinose transporter protein described herein. The nucleic acids may be

RNA or DNA. The nucleic acids may be single- or double-stranded RNA or single- or double-stranded DNA. The nucleic acids may be located on a plasmid or other vector (e.g. a yeast artificial chromosome (YAC)). The nucleic acids may be introduced and expressed by a yeast cell using conventional techniques known to those in the art.

- 5 [0066] Provided herein are yeast cells that include a hexose or pentose transporter protein described herein. In one aspect is a yeast cell that includes a recombinant xylose transporter protein described herein. The yeast cell including a recombinant xylose transporter protein described herein may be a species as set forth in Figure 19 or Figure 20. The yeast cell including a recombinant xylose transporter protein described herein may be a *S. stipitis* yeast cell, a *C.*
 10 *intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell. The yeast cell including a recombinant xylose transporter protein described herein may be capable of growth when placed in the presence of pentoses. The yeast cell including a recombinant xylose transporter protein described herein may be capable of growth, or have significantly increased growth compared to a yeast cell lacking the recombinant xylose
 15 transporter protein when placed in the presence of a xylose compound. The xylose compound is described herein. The xylose compound may be derived from lignocellulosic biomass.

- [0067] The xylose compound may be present at a concentration of about 0.05 g/L to about 20 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 15 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 10 g/L. The
 20 xylose compound may be present at a concentration of about 0.05 g/L to about 5 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 4 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 3 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 2 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 1 g/L. The xylose
 25 compound may be present at a concentration of about 0.05 g/L to about 0.5 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 0.1 g/L. The xylose compound may be present at a concentration of about 0.05 g/L. The xylose compound may be present at a concentration of about 0.1 g/L. The xylose compound may be present at a concentration of about 0.5 g/L. The xylose compound may be present at a concentration of about
 30 0.1 g/L. The xylose compound may be present at a concentration of about 0.5 g/L. The xylose compound may be present at a concentration of about 1 g/L. The xylose compound may be present at a concentration of about 2 g/L. The xylose compound may be present at a

concentration of about 3 g/L. The xylose compound may be present at a concentration of about 4 g/L. The xylose compound may be present at a concentration of about 5 g/L. The xylose compound may be present at a concentration of about 10 g/L. The xylose compound may be present at a concentration of about 15 g/L. The xylose compound may be present at a concentration of about 20 g/L.

[0068] The xylose compound may be present at a concentration of about 0.05 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 250 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 200 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 150 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 100 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 50 g/L. The xylose compound may be present at a concentration of about 0.05 g/L to about 25 g/L. The xylose compound may be present at a concentration of about 1 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 20 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 30 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 40 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 50 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 75 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 100 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 125 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 150 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 175 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 200 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 225 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 250 g/L to about 300 g/L. The xylose compound may be present at a concentration of about 275 g/L to about 300 g/L.

[0069] The xylose compound may be present at a concentration of about 10 g/L to about 275 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 250 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 225 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 200 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 175 g/L. The xylose

compound may be present at a concentration of about 10 g/L to about 150 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 125 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 100 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 75 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 50 g/L. The xylose compound may be present at a concentration of about 10 g/L to about 25 g/L.

[0070] The xylose compound may be present at a concentration of about 25 g/L. The xylose compound may be present at a concentration of about 50 g/L. The xylose compound may be present at a concentration of about 75 g/L. The xylose compound may be present at a

concentration of about 100 g/L. The xylose compound may be present at a concentration of about 125 g/L. The xylose compound may be present at a concentration of about 150 g/L. The xylose compound may be present at a concentration of about 175 g/L. The xylose compound may be present at a concentration of about 200 g/L. The xylose compound may be present at a concentration of about 225 g/L. The xylose compound may be present at a concentration of about 250 g/L. The xylose compound may be present at a concentration of about 275 g/L. The xylose compound may be present at a concentration of about 300 g/L.

[0071] The yeast cell including a recombinant xylose transporter protein described herein may be incapable of growth, or have significantly impaired growth compared to a yeast cell lacking the recombinant xylose transporter protein when placed in the presence of only hexoses. The

hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 20 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 15 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 10 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 5 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 4 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 3 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 2 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 1 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 0.5 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L to about 0.1 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.05 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.1 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.5 g/L. The hexose (e.g. glucose) may be present at a

concentration of about 0.1 g/L. The hexose (e.g. glucose) may be present at a concentration of about 0.5 g/L. The hexose (e.g. glucose) may be present at a concentration of about 1 g/L. The hexose (e.g. glucose) may be present at a concentration of about 2 g/L. The hexose (e.g. glucose) may be present at a concentration of about 3 g/L. The hexose (e.g. glucose) may be present at a concentration of about 4 g/L. The hexose (e.g. glucose) may be present at a concentration of about 5 g/L. The hexose (e.g. glucose) may be present at a concentration of about 10 g/L. The hexose (e.g. glucose) may be present at a concentration of about 15 g/L. The hexose (e.g. glucose) may be present at a concentration of about 20 g/L.

[0072] The recombinant xylose transporter protein of the yeast cell may include a transporter motif sequence as set forth herein. The yeast cell may metabolize the xylose compound. The yeast cell may convert xylose compound to a biofuel (e.g. ethanol) or a biochemical described herein. The yeast cell may convert xylose compound to a biofuel (e.g. ethanol). The yeast cell may convert xylose compound to a biochemical described herein.

[0073] In another aspect is a yeast cell that includes a recombinant galactose-arabinose transporter protein described herein. The yeast cell including a recombinant galactose-arabinose transporter protein described herein may be a species as set forth in Figure 19 or Figure 20. The yeast cell including the recombinant galactose-arabinose transporter protein may be a *S. stipitis* yeast cell, a *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell. The yeast cell including the recombinant galactose-arabinose transporter protein may be capable of growth, or have significantly increased growth compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein when placed in the presence of pentoses (e.g. arabinose). The yeast cell including the recombinant galactose-arabinose transporter protein may be capable of growth, or have significantly increased growth compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein when placed in the presence of an arabinose compound. The arabinose compound is described herein. The arabinose compound may be derived from lignocellulosic biomass.

[0074] The arabinose compound may be present at a concentration of about 0.05 g/L to about 20 g/L. The arabinose compound may be present at a concentration of about 0.05 g/L to about 15 g/L. The arabinose compound may be present at a concentration of about 0.05 g/L to about 10 g/L. The arabinose compound may be present at a concentration of about 0.05 g/L to about 5 g/L. The arabinose compound may be present at a concentration of about 0.05 g/L to about 4 g/L. The arabinose compound may be present at a concentration of about 0.05 g/L to about 3 g/L. The

The arabinose compound may be present at a concentration of about 200 g/L to about 300 g/L.

The arabinose compound may be present at a concentration of about 225 g/L to about 300 g/L.

The arabinose compound may be present at a concentration of about 250 g/L to about 300 g/L.

The arabinose compound may be present at a concentration of about 275 g/L to about 300 g/L.

- 5 **[0076]** The arabinose compound may be present at a concentration of about 10 g/L to about 275 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 250 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 225 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 200 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 175 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 150 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 125 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 100 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 75 g/L. The arabinose compound may be present at a concentration of about 10 g/L to about 50 g/L.
- 10 The arabinose compound may be present at a concentration of about 10 g/L to about 25 g/L.

- [0077]** The arabinose compound may be present at a concentration of about 25 g/L. The arabinose compound may be present at a concentration of about 50 g/L. The arabinose compound may be present at a concentration of about 75 g/L. The arabinose compound may be present at a concentration of about 100 g/L. The arabinose compound may be present at a concentration of about 125 g/L. The arabinose compound may be present at a concentration of about 150 g/L. The arabinose compound may be present at a concentration of about 175 g/L. The arabinose compound may be present at a concentration of about 200 g/L. The arabinose compound may be present at a concentration of about 225 g/L. The arabinose compound may be present at a concentration of about 250 g/L. The arabinose compound may be present at a concentration of about 275 g/L. The arabinose compound may be present at a concentration of about 300 g/L.
- 20
- 25

- [0078]** The yeast cell including the recombinant galactose-arabinose transporter protein may be incapable of growth, or have significantly impaired growth compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein when placed in the presence of hexoses such as glucose or mannose (i.e. the recombinant galactose-arabinose transporter protein does not transport glucose or mannose). The hexose (e.g. glucose) may be present in a concentration as set forth herein. The yeast cell including the recombinant galactose-arabinose transporter protein
- 30

may be capable of growth, or have significantly increased growth compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein when placed in the presence of a galactose compound. The galactose compound is described herein. The galactose compound may be derived from lignocellulosic biomass.

5 **[0079]** The galactose compound may be present at a concentration of about 0.05 g/L to about 20 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 15 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 10 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 5 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 4 g/L. The
10 galactose compound may be present at a concentration of about 0.05 g/L to about 3 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 2 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 1 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 0.5 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 0.1 g/L. The
15 galactose compound may be present at a concentration of about 0.05 g/L. The galactose compound may be present at a concentration of about 0.1 g/L. The galactose compound may be present at a concentration of about 0.5 g/L. The galactose compound may be present at a concentration of about 0.1 g/L. The galactose compound may be present at a concentration of about 0.5 g/L. The galactose compound may be present at a concentration of about 1 g/L. The
20 galactose compound may be present at a concentration of about 2 g/L. The galactose compound may be present at a concentration of about 3 g/L. The galactose compound may be present at a concentration of about 4 g/L. The galactose compound may be present at a concentration of about 5 g/L. The galactose compound may be present at a concentration of about 10 g/L. The galactose compound may be present at a concentration of about 15 g/L. The galactose compound
25 may be present at a concentration of about 20 g/L.

[0080] The galactose compound may be present at a concentration of about 0.05 g/L to about 300 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 250 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 200 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about
30 150 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 100 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 50 g/L. The galactose compound may be present at a concentration of about 0.05 g/L to about 25

g/L. The galactose compound may be present at a concentration of about 1 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 10 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 20 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 30 g/L to about 300 g/L.

5 The galactose compound may be present at a concentration of about 40 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 50 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 75 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 100 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 125 g/L to about 300 g/L.

10 The galactose compound may be present at a concentration of about 150 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 175 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 200 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 225 g/L to about 300 g/L.

The galactose compound may be present at a concentration of about 250 g/L to about 300 g/L.

15 The galactose compound may be present at a concentration of about 275 g/L to about 300 g/L.

[0081] The galactose compound may be present at a concentration of about 10 g/L to about 275 g/L. The galactose compound may be present at a concentration of about 10 g/L to about 250

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 225

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 200

20 g/L. The galactose compound may be present at a concentration of about 10 g/L to about 175

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 150

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 125

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 100

g/L. The galactose compound may be present at a concentration of about 10 g/L to about 75 g/L.

25 The galactose compound may be present at a concentration of about 10 g/L to about 50 g/L. The galactose compound may be present at a concentration of about 10 g/L to about 25 g/L.

[0082] The galactose compound may be present at a concentration of about 25 g/L. The

galactose compound may be present at a concentration of about 50 g/L. The galactose compound

may be present at a concentration of about 75 g/L. The galactose compound may be present at a

30 concentration of about 100 g/L. The galactose compound may be present at a concentration of

about 125 g/L. The galactose compound may be present at a concentration of about 150 g/L. The

galactose compound may be present at a concentration of about 175 g/L. The galactose

compound may be present at a concentration of about 200 g/L. The galactose compound may be present at a concentration of about 225 g/L. The galactose compound may be present at a concentration of about 250 g/L. The galactose compound may be present at a concentration of about 275 g/L. The galactose compound may be present at a concentration of about 300 g/L.

5 [0083] The yeast cell including the recombinant galactose-arabinose transporter protein may be capable of growth, or have significantly increased growth when compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein when placed in the presence of an arabinose compound and a galactose compound. The arabinose compound is described herein and may be present in a concentration described herein. The galactose compound is described
10 herein and may be present in a concentration described herein. The arabinose compound may be derived from lignocellulosic biomass. The galactose compound may be derived from lignocellulosic biomass.

[0084] The recombinant galactose-arabinose transporter protein of the yeast cell may include a transporter motif sequence as set forth herein. The yeast cell may metabolize the arabinose
15 compound. The yeast cell may metabolize the galactose compound. The yeast cell may convert the arabinose compound to a biofuel (e.g. ethanol) or a biochemical described herein. The yeast cell may convert the galactose compound to a biofuel (e.g. ethanol) or a biochemical described herein. The yeast cell may convert the arabinose compound to a biofuel (e.g. ethanol). The yeast cell may convert the arabinose compound to a biochemical described herein. The yeast cell may
20 convert the galactose compound to a biofuel (e.g. ethanol). The yeast cell may convert the galactose compound to a biochemical described herein.

[0085] Also provided herein are methods of transporting hexose or pentose moieties into a yeast cell. In one aspect is a method for transporting xylose into a yeast cell. The method includes contacting a yeast cell having a recombinant xylose transport protein described herein
25 with a xylose compound described herein. The recombinant xylose transport protein is allowed to transport the xylose compound into the cell. The yeast cell may be a yeast cell described herein. The yeast cell may be a *S. stipitis* yeast cell, a *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell.

[0086] The xylose compound may be derived from lignocellulosic biomass, hemicellulose, or
30 xylan. The xylose compound may be derived from lignocellulosic biomass. The xylose compound may be derived from hemicellulose. The xylose compound may be derived from

xylan. The yeast cell may metabolize the xylose compound. The yeast cell may preferentially grow in the presence of a xylose compound and may not grow using only another sugar source (e.g. glucose) when compared to a yeast cell lacking the recombinant xylose transporter protein.

The xylose compound may be present in a concentration described herein. The yeast cell may
5 convert the xylose compound to a biofuel (e.g. ethanol) or to a biochemical described herein. The yeast cell may convert the xylose compound to a biofuel (e.g. ethanol). The yeast cell may convert the xylose compound to a biochemical described herein.

[0087] The recombinant xylose transport protein may have a binding affinity of about 1 mM to about 0.02 mM for a xylose compound. The recombinant xylose transport protein may have a
10 binding affinity of about 0.8 mM to about 0.02 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.05 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.1 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.2 mM for a xylose compound. The recombinant xylose
15 transport protein may have a binding affinity of 0.8 mM to about 0.3 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.4 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.5 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.6 mM for a xylose
20 compound. The recombinant xylose transport protein may have a binding affinity of 0.8 mM to about 0.7 mM for a xylose compound.

[0088] The recombinant xylose transport protein may have a binding affinity of at least 0.02 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.05 mM for a xylose compound. The recombinant xylose transport protein
25 may have a binding affinity of at least 0.1 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.2 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.3 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.4 mM for a xylose compound. The recombinant xylose transport protein may have a binding
30 affinity of at least 0.5 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.6 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.7 mM for a xylose compound. The

recombinant xylose transport protein may have a binding affinity of at least 0.8 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 0.9 mM for a xylose compound. The recombinant xylose transport protein may have a binding affinity of at least 1 mM for a xylose compound.

- 5 [0089] The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 7 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 8 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 9 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 10 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 11 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 12 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 13 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 14 nmol min⁻¹ gDCW⁻¹ to about 15 nmol min⁻¹ gDCW⁻¹.
- 20 [0090] The recombinant xylose transport protein may have a rate of at least 7 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 8 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 9 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 10 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 11 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 12 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 13 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 14 nmol min⁻¹ gDCW⁻¹ of transporting a
- 30

xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least 15 nmol min⁻¹ gDCW⁻¹ of transporting a xylose compound into a yeast cell.

[0091] The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 10 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹.

The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 20 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 30 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 40 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 50 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 60 nmol min⁻¹ gDCW⁻¹ to about 70 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 80 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 90 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 100 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 110 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 120 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 130 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 140 nmol min⁻¹ gDCW⁻¹ to about 150 nmol min⁻¹ gDCW⁻¹.

[0092] The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 10 nmol min⁻¹ gDCW⁻¹ to about 140 nmol min⁻¹ gDCW⁻¹.

30 The recombinant xylose transport protein may have a rate of transporting a xylose compound into a yeast cell of about 10 nmol min⁻¹ gDCW⁻¹ to about 130 nmol min⁻¹ gDCW⁻¹. The recombinant xylose transport protein may have a rate of transporting a xylose compound into a

have a rate of at least $100 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least $110 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least $120 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least $130 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least $140 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell. The recombinant xylose transport protein may have a rate of at least $150 \text{ nmol min}^{-1} \text{ gDCW}^{-1}$ of transporting a xylose compound into a yeast cell.

[0094] In another aspect is a method of transporting galactose or arabinose into a yeast cell. The method includes contacting a yeast cell including a recombinant galactose-arabinose transport protein described herein, with a galactose compound or an arabinose compound described herein. The recombinant galactose-arabinose transport protein is allowed to transport the galactose compound or the arabinose compound into the yeast cell. The yeast cell may be a yeast cell described herein. The yeast cell may be a *S. stipitis* yeast cell, a *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell.

[0095] In the presence of arabinose, the recombinant galactose-arabinose transport protein may transport arabinose into a yeast cell. The arabinose compound may be present at a concentration as set forth herein. The arabinose compound may be derived from lignocellulosic biomass, hemicellulose, or arabinoxylan. The arabinose compound may be derived from lignocellulosic biomass. The arabinose compound may be derived from hemicellulose. The arabinose compound may be derived from arabinoxylan. The yeast cell may metabolize the arabinose compound. The yeast cell may preferentially grow in the presence of an arabinose compound and may not grow using only another sugar source (e.g. glucose) as compared to a yeast cell lacking the recombinant galactose-arabinose transporter protein. The yeast cell may convert the arabinose compound to a biofuel (e.g. ethanol) or to a biochemical (e.g. an organic acid.) The yeast cell may convert the arabinose compound to a biofuel (e.g. ethanol). The yeast cell may convert the arabinose compound to a biochemical described herein.

[0096] In the presence of galactose, the recombinant galactose-arabinose transport protein transports galactose into a yeast cell. The galactose compound may be at a concentration as set forth herein. The galactose compound may be derived from lignocellulosic biomass, hemicellulose, or galactan. The galactose compound may be derived from lignocellulosic

biomass. The galactose compound may be derived from hemicellulose. The galactose compound may be derived from galactan. The yeast cell may metabolize the galactose compound. The yeast cell may preferentially grow in the presence of a galactose compound and may not grow using only another sugar source (e.g. glucose). The yeast cell may convert the galactose compound to a biofuel (e.g. ethanol) or to a biochemical described herein. The yeast cell may convert the galactose compound to a biofuel (e.g. ethanol). The yeast cell may convert the galactose compound to a biochemical described herein.

I. Embodiments

[0097] Embodiment 1 A recombinant xylose transporter protein comprising a transporter motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia* *GXS1* protein, wherein said transporter motif sequence is -G-G/F-X¹-X²-X³-G-; wherein, X¹ is D, C, G, H, I, L, or F; X² is A, D, C, E, G, H, or I; X³ is N, C, Q, F, G, L, M, S, T, or P; and wherein, said transporter motif sequence is not -G-G-L-I-F-G- or -G-G-F-I-F-G-.

[0098] Embodiment 2 The recombinant xylose transporter protein of embodiment 1, wherein, X¹ is D, C, G, H, or F; X² is H or I; and X³ is S, T, or M.

[0099] Embodiment 3 The recombinant xylose transporter protein of embodiment 1 or 2, wherein X¹ is F, X² is I, and X³ is M or S.

[0100] Embodiment 4 The recombinant xylose transporter protein of any one of embodiments 1 to 3, wherein said transporter motif sequence is -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, -G-F-F-I-S-G-, -G-G-F-I-T-G-, -G-F-F-I-T-G-, -G-G-F-L-M-G-, -G-F-F-L-M-G-, -G-G-F-L-S-G-, -G-F-F-L-S-G-, -G-G-F-L-T-G-, -G-F-F-L-T-G-, -G-G-F-H-M-G-, -G-F-F-H-M-G-, -G-G-F-H-S-G-, -G-F-F-H-S-G-, -G-G-F-H-T-G- or -G-F-F-H-T-G-.

[0101] Embodiment 5 The recombinant xylose transporter protein of any one of embodiments 1 to 4, wherein said transporter motif sequence is -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, or -G-F-F-I-S-G-.

[0102] Embodiment 6 The recombinant xylose transporter protein of any one of embodiments 1 to 5 further comprising a mutation of an amino acid at the residue position corresponding to 297 of *Candida intermedia* *GXS1* protein.

[0103] Embodiment 7 The recombinant xylose transporter protein of any one of embodiments 1 to 6, wherein said amino acid at the residue position corresponding to 297 of *Candida intermedia GXS1* protein is substituted with a Met, Ala, Ser, or Asn residue.

[0104] Embodiment 8 The recombinant xylose transporter protein of any one of embodiments 1 to 7, wherein said recombinant xylose transporter protein is derived from a *C. intermedia gxs1* transporter protein, a *S. stipitis rgt2* transporter protein, or a *S. cerevisiae hxt7* transporter protein.

[0105] Embodiment 9 A recombinant galactose-arabinose transporter protein comprising a transporter motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida intermedia GXS1* protein, wherein said transporter motif sequence is -G-G/F-X⁴-X⁵-X⁶-G-; wherein, X⁴ is D, C, F, G, H, L, R, T, or P; X⁵ is A, C, E, F, H, K, S, P, or V; X⁶ is R, D, E, F, H, I, M, T, or Y; and wherein said sequence is not -G-G-L-V-Y-G-, or -G-G-F-V-F-G-.

[0106] Embodiment 10 The recombinant galactose-arabinose transporter protein of embodiment 9, wherein, X⁴ is R, T, H, or F; X⁵ is P, H, or V; and X⁶ is T, H, F, M, or Y.

[0107] Embodiment 11 The recombinant galactose-arabinose transporter protein of embodiment 9, wherein X⁴ is F or T, X⁵ is P or I, and X⁶ is M or T.

[0108] Embodiment 12 The recombinant galactose-arabinose transporter protein of embodiment 10 or 11, wherein said transporter motif sequence is -G-G-F-H-M-G-, -G-F-F-H-M-G-, -G-G-R-P-T-G-, -G-F-R-P-T-G-, -G-G-T-P-T-G-, or -G-F-T-P-T-G-.

[0109] Embodiment 13 The recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 12, wherein said galactose-arabinose transporter protein further comprises a mutation of an amino acid at the residue position corresponding to 297 of *Candida intermedia GXS1* protein.

[0110] Embodiment 14 The recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 13, wherein said amino acid at the residue position corresponding to 297 of *Candida intermedia GXS1* protein is substituted with a Met, Thr, Ala, or Ile residue.

[0111] Embodiment 15 The recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 14, wherein said recombinant galactose-arabinose transporter protein is derived from a *C. intermedia gxs1* transporter protein, a *S. stipitis rgt2* transporter protein, a *S. cerevisiae hxt7* transporter protein, or a *S. cerevisiae GAL2* protein.

[0112] Embodiment 16 A yeast cell comprising the recombinant xylose transporter protein of any one of embodiments 1 to 8.

[0113] Embodiment 17 A yeast cell comprising the recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 15.

5 [0114] Embodiment 18 A nucleic acid encoding the recombinant xylose transporter protein of any one of embodiments 1 to 8.

[0115] Embodiment 19 A nucleic acid encoding the recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 15.

10 [0116] Embodiment 20 A method of transporting xylose into a yeast cell, said method comprising: contacting a yeast cell comprising the recombinant xylose transporter protein of any one of embodiments 1 to 8 with a xylose compound; and allowing said recombinant xylose transporter protein to transport said xylose compound into said yeast cell.

[0117] Embodiment 21 The method of embodiment 20, wherein said xylose compound forms part of lignocellulosic biomass, hemicellulose, or xylan.

15 [0118] Embodiment 22 The method of embodiment 20 or 21, wherein said yeast cell metabolizes said xylose compound.

[0119] Embodiment 23 The method of any one of embodiments 20 to 22, wherein said yeast cell converts said xylose compound to a biofuel.

20 [0120] Embodiment 24 The method of any one of embodiments 20 to 23, wherein said recombinant xylose transporter protein has a binding affinity of at least 0.7 mM for said xylose compound.

[0121] Embodiment 25 The method of any one of embodiments 20 to 24, wherein said recombinant xylose transporter protein has a rate of at least 15 nmol min⁻¹ gDCW⁻¹ of transporting said xylose compound into said yeast cell.

25 [0122] Embodiment 26 A method of transporting galactose or arabinose into a yeast cell, said method comprising: contacting a yeast cell comprising the recombinant galactose-arabinose transporter protein of any one of embodiments 9 to 15 with a galactose compound or an arabinose compound; and allowing said recombinant galactose-arabinose transporter protein to transport said galactose compound or said arabinose compound into said yeast cell.

[0123] Embodiment 27 The method of embodiment 26, wherein said recombinant galactose-arabinose transporter protein is contacted with an arabinose compound.

[0124] Embodiment 28 The method of any one of embodiments 26 to 27, wherein said arabinose compound forms part of lignocellulosic biomass, hemicellulose or arabinoxylan.

5 [0125] Embodiment 29 The method of any one of embodiments 26 to 28, wherein said yeast cell metabolizes said arabinose compound.

[0126] Embodiment 30 The method of any one of embodiments 26 to 29, wherein said yeast cell converts said arabinose compound to a biofuel.

10 [0127] Embodiment 31 The method of any one of embodiments 26 to 30, wherein said recombinant galactose-arabinose transporter protein is contacted with a galactose compound.

[0128] Embodiment 32 The method of any one of embodiments 26 to 31, wherein said galactose compound forms a part of lignocellulosic biomass, hemicellulose, or galactan.

[0129] Embodiment 33 The method of any one of embodiments 26 to 32, wherein said yeast cell metabolizes said galactose compound.

15 [0130] Embodiment 34 The method of any one of embodiments 26 to 33, wherein said yeast cell converts said galactose compound to a biofuel.

[0131] Embodiment 35 The method of embodiment 20, wherein said yeast cell is a *S. stipitis* yeast cell, a *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell.

20 [0132] Embodiment 36 The method of embodiment 26, wherein said yeast cell is a *S. stipitis* yeast cell, a *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast cell.

II. Examples

25 [0133] Example 1: Identification of the G-G/F-XXX-G motif that controls sugar transport preference.

[0134] A multiple sequence alignment of 26 previously cloned transporters (36) indicates that Phe⁴⁰ was part of a highly conserved glycine-rich motif of the form G-G/F-XXX-G, where X represents a variable, but usually nonpolar amino acid residue. In *C. intermedia* GXS1, the wild

type motif is G³⁶G³⁷V³⁸L³⁹F⁴⁰G⁴¹. The high conservation of this motif suggested it could be responsible for xylose uptake, transporter efficiency, and monosaccharide selectivity. To further corroborate this hypothesis, an additional 20 putative transporters were identified using a BLAST search seeded with transporters functionally characterized in *S. cerevisiae* EX.12, a recombinant strain lacking endogenous monosaccharide transporters (Figure 8 and Table 1) (26, 38). The vast majority of these transporters were functional and all possessed a similar motif. Among these transporters, *D. hansenii* 2D01474 confers much faster growth on xylose than on glucose and *S. stipitis* RGT2 confers the fastest growth on xylose of all the *S. stipitis* derived transporters in this study.

[0135] Following the functional characterization, motif sequence was correlated with transporter carbon source growth profile. Four major phenotypic classifications were made: (a) transporters that failed to function heterologously ($\mu_{\text{all}} = 0$), (b) transporters that conferred growth on a hexose but not xylose ($\mu_X = 0$), (c) transporters that conferred growth on xylose but not as fast as glucose ($\mu_X < \mu_G$), and (d) transporters that conferred a higher growth rate on xylose than on glucose ($\mu_X > \mu_G$). Figure 1A displays the relative proportions of each of these classifications in the group of 46 transporters studied. To characterize the sequence, four major motif classifications were made: (a) a full G-G/F-XXXG motif, (b) a related S-G-XXXG motif, (c) a motif unrelated to the glycine rich motif, and (d) the lack of homology to other transporters at both the motif and surrounding residues. Figure 1B depicts the distribution of the four sequence motif classifications within the four phenotypic classifications. Strikingly, there is a clear enrichment of the G-G/F-XXXG motif among the functional transporters that enable high xylose transport rates. In fact, this motif is exclusively seen in phenotype class (d) where $\mu_X > \mu_G$. The enrichment and convergence of the variable residues within the motif is displayed in Figure 1C. It should be noted that the consensus sequence from this analysis appears to be G-G/F-XX-F-G. Yet, variations at the consensus F residue led to the discovery of the motif, therefore this position was considered variable. Figure 1C highlights the strong correlation between sequence motif and xylose transport function and suggests an important role of TMS1 on sugar recognition.

[0136] *Identification of potentiating variable residues within the G-G/F-X-X-X-G motif.*

[0137] To examine the role of the variable region, complete saturation mutagenesis was performed for each of the three residues (Val³⁸, Leu³⁹, and Phe⁴⁰) in *C. intermedia* GXSI and evaluated the impact on carbon source growth profile as measured by growth rate. Previous studies demonstrate that growth rate in this test strain is a good surrogate for transporter kinetics

(36, 38). Specifically, the fractional change in growth rate of *S. cerevisiae* EX.12 on glucose, xylose, galactose, fructose, or mannose as the sole carbon source was evaluated compared to the wild-type transporter. The impact of each residue can be classified as having no change, altered efficiency, altered selectivity, or a combination of the three (Figure 2). For creating xylose
5 specific transporters, the goal is to identify mutations that attenuate hexose growth while either amplifying or maintaining xylose growth.

[0138] Members of the *C. intermedia* gxs1 Val³⁸ saturation library (Figure 3A) display differential exponential growth rates with the most significant one being the Phe³⁸ substitution.

This mutant confers a selectivity phenotype that almost completely attenuates glucose
10 exponential growth rate while amplifying exponential xylose growth rate by 50%. Other substitutions that confer desirable selectivity phenotypes are Asp³⁸, Cys³⁸, Gly³⁸, and His³⁸. All of these affect the growth profile in different patterns, but none as significantly as Phe³⁸. Three substitutions, Ile³⁸, Leu³⁸, and Met³⁸, differentially amplify growth on multiple sugars while glucose growth remains unchanged. The Leu³⁸ substitution in particular increases exponential
15 xylose growth rate by 73% without altering glucose exponential growth rate significantly. Ala³⁸ attenuates growth on glucose only. Nearly all of the remaining substitutions attenuate growth, yet many preferentially attenuate growth on hexoses. In this subset, Lys³⁸ attenuates growth on glucose, fructose, and mannose without affecting growth rate on xylose. The frequency of selectivity and differentially attenuating phenotypes arising at this residue indicates that position
20 38 predominately influences monosaccharide selectivity.

[0139] Table 1: Exponential growth rate values for each cloned transporter

	GeneID	GenBank	Glucose		Xylose		Galactose		Fructose		Mannose	
			μ (1/hr)	st. dev	μ (1/hr)	st. dev	μ (1/hr)	st. dev	μ (1/hr)	st. dev	μ (1/hr)	st. dev
Empty	-	-	0.028	0.002	0.033	0.004	0.054	0.004	0.041	0.008	0.038	0.002
At5g59250	836043		0.031	0.004	0.045	0.003	0.057	0.007	0.046	0.006	0.037	0.001
At5g17010	831564		0.024	0.002	0.032	0.002	0.046	0.004	0.032	0.002	0.036	0.005
Ci GXF1		AJ937350.1	0.206	0.002	0.101	0.003	0.072	0.000	0.278	0.003	0.268	0.004
Ci GXS1		AJ875406.1	0.065	0.008	0.062	0.002	0.074	0.004	0.082	0.003	0.038	0.000
Dh DD02167	2901237		0.064	0.005	0.065	0.008	0.111	0.007	0.093	0.007	0.123	0.010
Dh 2B14276	2913528		0.056	0.005	0.026	0.011	0.094	0.009	0.081	0.017	0.114	0.009
Dh 2A14300	8998057		0.036	0.004	0.040	0.013	0.051	0.004	0.035	0.007	0.032	0.001
Dh 2F19140	8999011		0.026	0.002	0.036	0.005	0.046	0.002	0.034	0.009	0.035	0.004
Dh XylHP	50419288		0.116	0.003	0.096	0.010	0.100	0.014	0.123	0.006	0.126	0.004
Ec XylE	548529		0.026	0.002	0.039	0.005	0.050	0.005	0.024	0.002	0.030	0.009
Ss XUT1	4639826		0.085	0.007	0.060	0.007	0.109	0.006	0.060	0.003	0.096	0.003
Ss XUT2	4852047		0.042	0.005	0.047	0.004	0.067	0.005	0.041	0.002	0.035	0.011
Ss XUT3	4851844		0.064	0.008	0.078	0.005	0.099	0.007	0.090	0.010	0.109	0.009
Ss XUT4	4840896		0.027	0.005	0.024	0.001	0.050	0.009	0.029	0.003	0.043	0.006
Ss XUT5	4840252		0.029	0.001	0.029	0.008	0.020	0.003	0.031	0.002	0.027	0.002
Ss XUT6	4841106		0.044	0.004	0.021	0.003	0.053	0.007	0.032	0.003	0.029	0.002
Ss XUT7	4851701		0.028	0.008	0.032	0.005	0.048	0.001	0.030	0.006	0.033	0.007
Yi DB06391	2907283		0.030	0.006	0.063	0.002	0.054	0.006	0.051	0.007	0.058	0.009
Yi DB01342	2906708		0.031	0.003	0.036	0.003	0.062	0.001	0.038	0.010	0.035	0.005
Yi F06776	2908504		0.032	0.005	0.039	0.004	0.049	0.003	0.033	0.008	0.039	0.008
Yi DC06424	2909312		0.063	0.013	0.069	0.001	0.075	0.010	0.048	0.005	0.069	0.010
Yi OC08943	2909701		0.031	0.005	0.039	0.003	0.046	0.004	0.029	0.007	0.027	0.003
Cn BC3990	4935064		0.031	0.008	0.034	0.010	0.045	0.001	0.032	0.005	0.030	0.004
Sc HXT7	851943		0.191	0.013	0.099	0.015	0.130	0.008	0.254	0.008	0.278	0.005
Sc HXT13	856640		0.072	0.003	0.045	0.003	0.055	0.011	0.059	0.003	0.132	0.008
Sc GAL2	856640		0.105	0.013	0.107	0.003	0.078	0.009	0.154	0.003	0.237	0.008
Sc HXT3	851946		0.280	0.003	0.083	0.006	0.053	0.002	0.314	0.006	0.258	0.007
Sc HXT11	856494		0.297	0.008	0.088	0.006	0.118	0.003	0.303	0.003	0.229	0.004
Dh 2C02530	8998297		0.132	0.007	0.099	0.002	0.055	0.006	0.075	0.002	0.138	0.005
Dh 2E01166	2902960		0.071	0.010	0.061	0.010	0.174	0.002	0.054	0.007	0.081	0.002
Dh 2E01296	2902912		0.074	0.011	0.069	0.006	0.056	0.005	0.041	0.002	0.040	0.002
Sc MAL11	853207		0.148	0.004	0.085	0.003	0.117	0.001	0.036	0.006	0.034	0.002
Sc STL1	852149		0.034	0.001	0.069	0.003	0.089	0.004	0.034	0.006	0.029	0.004
Sc HXT2	855023		0.278	0.009	0.101	0.003	0.079	0.009	0.287	0.002	0.256	0.002
Sc HXT8	853216		0.120	0.005	0.060	0.000	0.064	0.001	0.111	0.003	0.172	0.007
Sc HXT9	853236		0.205	0.026	0.098	0.005	0.065	0.005	0.249	0.022	0.207	0.011
Sc HXT10	850536		0.079	0.009	0.078	0.008	0.124	0.008	0.125	0.003	0.137	0.009
Sc HXT14	855396		0.042	0.009	0.032	0.006	0.061	0.005	0.051	0.003	0.051	0.012
Ss AUT1	4835720		0.046	0.005	0.038	0.004	0.104	0.009	0.036	0.004	0.046	0.019
Ss HGT2	4836632		0.113	0.007	0.024	0.010	0.120	0.007	0.100	0.002	0.179	0.002
Ss RGT2	4840659		0.126	0.013	0.119	0.001	0.068	0.008	0.036	0.007	0.053	0.006
Dh 2B05060	2913215		0.033	0.003	0.050	0.002	0.112	0.001	0.028	0.001	0.029	0.001
Dh 2E01386	2902914		0.032	0.004	0.035	0.001	0.069	0.002	0.137	0.030	0.045	0.015
Yi OD00132	2910370		0.026	0.006	0.034	0.002	0.039	0.005	0.035	0.007	0.035	0.005
Ss STL1	4838168		0.037	0.006	0.055	0.006	0.060	0.005	0.061	0.004	0.045	0.008
Dh 2D01474	2901237		0.072	0.000	0.104	0.004	0.082	0.005	0.085	0.006	0.051	0.004

[0140] Nearly all members of the Leu³⁹ saturation library (Figure 3B) display uniform attenuation patterns across sugars. Thus, residue 39 appears to greatly control transporter efficiency. Nevertheless, several of these substitutions differentially attenuate growth. Specifically, Asp³⁹, Cys³⁹, Gly³⁹, His³⁹, Ile³⁹, and Phe³⁹ reduce exponential growth on hexoses without drastically altering xylose growth rate. Of these, His³⁹ and Ile³⁹ establish the greatest difference between the hexose and pentose growth rates.

[0141] Members of the Phe⁴⁰ library (Figure 3C) display differential carbon source selectivity similar to Val³⁸ and have the greatest frequency of selectivity substitutions. Specifically, amino acid substitutions that confer a selectivity phenotype for xylose over glucose are Asn⁴⁰, Cys⁴⁰, Gly⁴⁰, Leu⁴⁰, Met⁴⁰, Ser⁴⁰, and Thr⁴⁰. Of these, Ser⁴⁰ and Met⁴⁰ appear as the most significant.

5 There are several attenuating substitutions that can be seen at residue 40 including Arg⁴⁰, Asp⁴⁰, Glu⁴⁰, Ile⁴⁰, Lys⁴⁰, Pro⁴⁰, and Tyr⁴⁰. Of these, Pro⁴⁰ appears as the only one that does not attenuate growth on xylose. Finally, Ala⁴⁰, His⁴⁰, and Trp⁴⁰ confer increased growth on most of the monosaccharides tested. In summary, residues 38 and 40 appear to play a role in transporter selectivity while residue 39 appears to play a role for controlling net transporter efficiency. In
10 general, hydrophobic residues of moderate to large size were beneficial for xylose growth, while charged residues were not (also seen with the evaluated transporters in Figure 1C). These motif design guidelines may be used to reprogram transporter function.

[0142] *Rewiring C. intermedia GXS1 into a xylose specific transporter.*

[0143] Using the design guidelines discovered above, triple mutants were constructed to
15 investigate the synergy between xylose favoring substitutions (in particular, Phe³⁸, Ile³⁹, and Ser⁴⁰ / Met⁴⁰). Both Phe³⁸ Ile³⁹ Ser⁴⁰ and Phe³⁸ Ile³⁹ Met⁴⁰ attenuate glucose exponential growth while maintaining or slightly increasing xylose exponential growth (Figure 4A), with the Phe³⁸ Ile³⁹ Met⁴⁰ triple mutant attenuating glucose growth to the same level as the negative control. Average growth curves on xylose and glucose (Figure 4B-C) highlight that both triple mutants
20 maintain wild-type xylose growth profile while severely attenuate glucose growth. Further characterization of the best mutant, *gxs1* Phe³⁸ Ile³⁹ Met⁴⁰ was performed. First, to assay transport capacity, high cell density fermentations with xylose and glucose were performed (Figure 5A-B). The Phe³⁸ Ile³⁹ Met⁴⁰ triple mutant displayed no appreciable glucose uptake whereas xylose uptake has become more efficient compared to the wild-type *GXS1*. These results
25 display a rewiring of the sugar uptake ratio. However, despite minimizing glucose transport capacity, glucose at levels of 5 g/L still appear to inhibit xylose growth (Figure 5C). This finding is corroborated by high cell density cofermentations (Figure 9 and Figure 10).

[0144] Radiolabelled xylose uptake experiments were performed to quantify the improvement of transport kinetics in the Phe³⁸ Ile³⁹ Met⁴⁰ triple mutant. The improvements in xylose utilization
30 observed at high cell density culturing were mainly due to a doubling in $V_{\max}$ (Figure 5D). An increased K_M was observed as well (Figure 5E), a phenotype observed in previous efforts to engineer this transporter (38). Nevertheless, the binding affinity is still quite high for practical

culturing at a value corresponding to around 0.1 g/L (Table 2). These kinetics experiments were also performed in the presence of glucose and no radiolabelled xylose uptake was detected indicating that while glucose cannot pass through the transporter, it can still bind and inhibit xylose uptake. Hence, binding appears to occur at a different residue.

5 [0145] **Table 2:** kinetics values calculated from radiolabeled xylose uptake

Gene in p414-TEF	K_M (mM)	V_{MAX} (nmol min ⁻¹ gDCW ⁻¹)
C.I. GXS1	0.0256 ± 0.0659	7.23 ± 0.6
C.I. GXS1 F³⁸I³⁹M⁴⁰	0.721 ± 0.116	15.01 ± 2.38

[0146] *The G-G/F-XXXG motif can be used to rewire other transporters*

[0147] To test how broad these design guidelines are for transporters, the conserved G-G/F-XXXG motif was utilized to reengineer the sugar preference of other predominately hexose transporters. Specifically, two transporters, *S. stipitis* RGT2 and *S. cerevisiae* HXT7, were selected based on evolutionary distance from GXS1. *S. stipitis* RGT2 is closely related to *C. intermedia* GXS1, while the native HXT transporters are more distant (Figures 11 and 12). First, the impact of rewiring the closely related transporter, *S. stipitis* RGT2 was investigated. This transporter contains a G³⁶G³⁷I³⁸L³⁹F⁴⁰G⁴¹ motif and two separate point mutations were characterized, Phe³⁸ and Met⁴⁰. In both cases, glucose growth has been completely attenuated (Figure 6). Most striking is the Met⁴⁰ mutation, which eliminates growth on all carbon sources but xylose and galactose. By modifying the motif in RGT2, two additional mutant proteins were generated that transport xylose, but not glucose.

[0148] Second, the potential to rewire *S. cerevisiae* HXT7, a more distantly related protein yet is able to efficiently transport hexoses and xylose in yeast, was evaluated (32, 42). Given the proficiency of hexose transport by this protein, rewiring to attenuate growth on hexoses presents a greater challenge. The native motif within *S. cerevisiae* HXT7 is G³⁶G³⁷F³⁸V³⁹F⁴⁰G⁴¹. Two double mutations to this motif - Ile³⁹Met⁴⁰ and His³⁹Met⁴⁰ were initially evaluated. Figure 7 demonstrates that the Ile³⁹Met⁴⁰ double mutant amplified xylose exponential growth and attenuated growth on all hexoses save glucose whereas the His³⁹Met⁴⁰ double mutant attenuated glucose growth yet also severely attenuated xylose exponential growth. Previous studies have indicated that mutations at Asp³⁴⁰ can eliminate glucose transport (39) in HXT7 and transport of nearly all monosaccharides is severely attenuated with this mutation was verify herein (Figure 7).

Coupling the Met³⁴⁰ mutation with the Ile³⁹Met⁴⁰ double mutant resulted in robust growth on xylose while maintaining the inability to transport glucose. With this triple mutant, a robust hexose transporter was converted to a xylose transporter unable to support growth on glucose.

[0149] Thus, a short, six residue motif of the form G-G/F-XXXG in TMS1 was identified that exerts control over selectivity and efficiency of monosaccharide transport of MFS family transporters. This motif is conserved among functional transporters and highly enriched in transporters that confer growth on xylose. Altering the composition of the variable region changes the sugar uptake profiles of these transporters and can thus be used to rewire transporter function. Altering the residues in this domain can eliminate glucose transport while retaining xylose transport, a major step forward for molecular transporter engineering. As a result, several transporter mutants were created that support the transport of xylose and not glucose.

[0150] Hydrophobic, nonpolar, and moderate to large size residues often attenuated glucose compared to xylose. Amino acids such as Phe, Ile, Ser, and Met were among the most effective substitutions that differentially amplified xylose growth rate. While many of these residues are found naturally in wild type motif sequences (Figure 2), the combinations found herein (particularly Phe³⁸ Ile³⁹ Met⁴⁰) are not found naturally. Hypotheses concerning transporter substrate recognition and transport mechanism may be formed based on these results. Without being bound to any particular theory, the advantage of large and nonpolar residues suggests that glucose growth attenuation is due to steric exclusion. The larger side chains may physically restrict the size of the pore, allowing the smaller xylose molecule to bind and traverse more efficiently than larger hexoses. A similar hypothesis has been proposed to explain an observed correlation between amino acid size and transporter function for glucose (43). This hypothesis is supported by the crystal structure of a related MFS transporter, *E. coli xylE* (41). Based on the structure, *E. coli xylE* Phe²⁴, an analogous residue to *C. intermedia gxs1* Phe⁴⁰, appears to interact with sugars as they pass through the pore. *E. coli xylE* is too dissimilar from yeast MFS transporters to enable structure prediction, yet this evidence suggests that this residue appears to play a role in all MFS sugar transporters.

[0151] Transporters from *Neurospora crassa* and *S. stipitis* were found to be exclusive for xylose in uptake assays (35), but are unable to support robust growth of recombinant *S. cerevisiae* on xylose. The *Escherichia coli xylE* transporter is xylose specific when expressed in its native host (44), but is inhibited by glucose and remains non-functional in *S. cerevisiae* despite attempts at directed evolution. Prior to this work, no evidence has demonstrated a defined

transporter engineering approach that is able to effectively eliminate glucose transport while amplifying xylose transport and supporting robust xylose growth. The mutants generated in this study demonstrate this desirable phenotype and provide evidence that the G-G/F-XXXG motif controls transport phenotype in a large number of MFS transport proteins.

5 [0152] It is also important to note that altering this motif in *C. intermedia* GXS1 not only had an impact on glucose uptake, but also had an impact on the kinetics of xylose uptake.

Specifically, the K_m for xylose was significantly increased compared to wild type, indicating that exclusion of glucose was obtained at the expense of reduced affinity for xylose. Nevertheless, the affinity for xylose remains sufficiently high for nearly all fermentation conditions ($K_M = 0.721 \pm$
10 0.116 mM, or approximately 0.1 g/L), and was partially compensated by a doubling in V_{max} (Figure 5). This result suggests a complex set of interactions between the transporter and sugar substrate, and is similar to other mutants of *C. intermedia* GXS1 (38).

[0153] In the course of identifying and validating this motif, several novel native and heterologous transporters were identified and shown to possess previously unreported

15 phenotypes (Figure 13). The transporter *D. hansenii* 2D01474 can natively support growth on xylose compared with glucose. The transporter *S. stipitis* RGT2 confers the fastest growth rate on xylose over any ORF cloned from *S. stipitis*. Both of these transporters are closely related to *C. intermedia* GXS1 (Figure 12) and may present a new class of related transporters that make excellent starting scaffolds for engineering exclusive xylose uptake. Of the remaining novel
20 ORFs studied here, one group (*D. hansenii* 2E01166, *D. hansenii* 2B05060, *S. cerevisiae* STL1, and *S. stipitis* AUT1) confer higher exponential growth rates on galactose than any other sugar tested. This hexose transport profile is indicative of the potential for L-arabinose transport, since the galactose transporter *S. cerevisiae* GAL2 is one of the few transporters able to facilitate L-arabinose (45). This correlation is likely due to the similar stereochemistry between L-arabinose
25 and galactose.

[0154] As discovered herein, substitution at the -XXX- positions of the transporter motif sequence uncovered several interesting phenotypes. Indeed, substitution with Thr and Pro (e.g. a transporter motif sequence of -G-G-T³⁸P³⁹T⁴⁰-G-) results in selective galactose uptake in the modified transporter protein. Such exclusive uptake, as discussed herein, is also indicative of L-
30 arabinose uptake ability (Figures 15 & 16). Thus, the work described herein shows transporter proteins can successfully be engineered into galactose and arabinose transporters.

[0155] This work describes a conserved G-G/F-XXXG motif and an engineering approach to modify this motif. This motif allowed for the rewiring of several transporters and yielded the mutant transporters *C. intermedia* gxs1 Phe³⁸ Ile³⁹ Met⁴⁰, *S. stipitis* rgt2 Phe³⁸ and Met⁴⁰, and *S. cerevisiae* hxt7 Ile³⁹ Met⁴⁰ Met³⁴⁰ that do not transport glucose yet support *S. cerevisiae* EX.12 growth on xylose. This motif also yielded *C. intermedia* Thr²⁸Pro³⁹Thr⁴⁰ that supports *S. cerevisiae* EX.12 growth on galactose, and no other sugar tested. These major facilitator superfamily transporters are channels and thus a substrate molecule interacts with many residues during transport. Yet, no other residues discovered to date display the degree to which glucose transport can be attenuated and xylose transport amplified than the residues in the G-G/F-XXXG motif. Thus, this study provides further insight into the residues responsible for monosaccharide transport in MFS proteins while establishing a platform for engineering a specific, efficient xylose transporter.

[0156] Materials and Methods

[0157] *Strains, media, and plasmids* – Molecular cloning and standard culturing techniques with *E. coli* DH10B were performed according to Sambrook (46). *S. cerevisiae* EX.12 was used for all yeast experiments and was constructed as previously described (38). All transporters were cloned into p414-TEF, a standard yeast shuttle vector created by Mumberg (47). Yeast synthetic complete media was used for culture and experimental growth media. CSM-Trp was used when *S. cerevisiae* EX.12 was carrying a transporter. Carbon sources were provided at 20 g/L.

[0158] *Transporter Cloning* – Potential xylose transporters were identified from literature and BLAST search. To obtain this list of 46, we combined 26 transporters from our previous survey of transporters (36) along with 20 additional transporters identified through homology search using *C. intermedia* GXS1 and *S. cerevisiae* STL1 as a template. Details on cloning and transporter libraries are described herein. Primers are listed in Table 4 (cloning), Table 5 (saturation mutagenesis), and Table 6 (point mutations).

[0159] Table 4: primers used for cloning putative transporters.

Name	Target ORF	Orientation	T _m	R. Enz.	Sequence	SEQ ID NO:
EY194	ScHXT3	F	55	XmaI	TATTCCTCCCGGGatgaattcaact ccagattttaatatctcc	2
EY208	ScHXT3	R	64	ClaI	CGGTATCCATCGATttattttcttg ccgaacatttttctt	3
EY265	HXT1	F	64	XmaI	TATTCCTCCCGGGatgaattcaact cccgatctaataatc	4
EY266	HXT1	R	64	ClaI	CGGTATCCATCGATttattttcttg ctaaacaaactcttg	5
EY554	SsHGT2	F	60	SpeI	GGACTAGTatgagctacgaagata aactcg	6
EY555	SsHGT2	R	60	SalI	TATTCCTCGACTtaaggcttttctc ctcagaactt	7
EY558	Dh2C02530	F	60	SpeI	GGACTAGTatgggttacgaagata aattagt	8
EY559	Dh2C02530	R	60	SalI	TATTCCTCGACTtaagtcattgtg agaagtatcgc	9
EY560	Dh2E01166	F	61	SpeI	GGACTAGTatgggatatgaagaaa agttgg	10
EY561	Dh2E01166	R	59	SalI	TATTCCTCGACTcaagcaatgtg atctgc	11
EY562	Dh2E01298	F	61	SpeI	GGACTAGTatgggatacgaagata aattactagg	12
EY563	Dh2E01298	R	60	SalI	TATTCCTCGACTtaagcaatatg gacagcactag	13
EY564	ScMAL11	F	60	SpeI	GGACTAGTatgaaaaatatcattt cattggtaag	14
EY565	ScMAL11	R	60	SalI	TATTCCTCGACTtaacattttatc agctgcattttaat	15
EY566	ScSTL1	F	60	SpeI	GGACTAGTatgaaggattttaaatt tatcgaattt	16
EY567	ScSTL1	R	60	SalI	TATTCCTCGACTcaaccctcaaa atttgct	17
EY568	SsRGT2	F	60	SpeI	GGACTAGTatgggtttagaagaca gtgct	18
EY569	SsRGT2	R	61	SalI	TATTCCTCGACTtatacagaagc ttcttcaacttcag	19
EY572	SsAUT1	F	61	XmaI	TATTCCTCCCGGGatgagtgctgac gaaaaagtc	20
EY573	SsAUT1	R	61	XhoI	TATTCCTCGAGctactcgacata agagacttctgg	21
EY574	HXT8	F	61	XmaI	TATTCCTCCCGGGatgactgatcgt aaaaccaactt	22
EY575	HXT8	R	61	XhoI	TATTCCTCGAGctaaaaacattct tttgtagaagggtt	23
EY576	HXT2	F	62	XmaI	TATTCCTCCCGGGatgtctgaattc gctactagcc	24
EY577	HXT2	R	63	XhoI	TATTCCTCGAGttatttcctcgga aactcttttttct	25
EY578	HXT9	F	62	XmaI	TATTCCTCCCGGGatgtccggtggt	26

Name	Target ORF	Orientation	T _m	R. Enz.	Sequence	SEQ ID NO:
					aataatacatcc	
EY579	HXT9	R	62	XhoI	TATTCCCTCGAGttagctggaaaa gaacctcttg	27
EY580	HXT10	F	60	XmaI	TATTCCTCCGGGatgggttagttca agtgtttcca	28
EY581	HXT10	R	60	XhoI	TATTCCCTCGAGttatttactatc aacaataactaatgggtgtac	29
EY582	HXT14	F	61	XmaI	TATTCCTCCGGGatgactgctcag attccgtat	30
EY583	HXT14	R	61	XhoI	TATTCCCTCGAGctactccgggttc aaatatatttattg	31
EY644	SsSTL1	F	61	SpeI	GGACTAGTatggcatatcttgatt ggttaac	32
EY645	SsSTL1	R	62	XmaI	TATTCCTCCGGGctaggctgcttt agggttttctg	33
EY646	DhE01386	F	61	SpeI	GGACTAGTatgtataaaatattggt caaaaactaacact	34
EY647	DhE01386	R	61	XmaI	TATTCCTCCGGGttaaaacttccgc aggcttaa	35
EY648	DhB05060	F	63	SpeI	GGACTAGTatggcttttaaaatct tttctagaacc	36
EY649	DhB05060	R	63	XmaI	TATTCCTCCGGGttaagcattagg agttaagataccttctg	37
EY650	Y10D00132	F	63	SpeI	GGACTAGTatgggttttggacgag aaaaag	38
EY651	Y10D00132	R	63	XmaI	TATTCCTCCGGGttaaacgaactc ggcagtgt	39
EY700	DhA2D01474	F	55	XmaI	TATTCCTCCGGGatgggttttagaa gataatgc	40
EY701	DhA2D01474	R	56	XhoI	TATTCCCTCGAGttagactgaagt gggtttcaat	41

[0160] Table 5: primers used for saturation mutagenesis of *C. intermedia* GXS1.

Name	Target A.A.	Orientation	T _m	Kit Used	Sequence	SEQ ID NO.:
EY630	GXS1Sat40	F	79	Sat Multi	tttgctgcttctggtggtgtcc ttNNKggatacgatactggtac ta	42
EY675	GXS1 39LSat	F	79	Sat Multi	tttgctgcttctggtggtgtcn nkttcggatacgatactggtac t	43
EY676	GXS1 38VSat	R	79	Sat Multi	gtctttttgctgcttctggtgg tnnkcttttcggatacgatact ggtac	44
EY711	GXS1 38Sat	F	79	Sat Multi	gtctttttgctgcttctggtgg tyrkcttttcggatacgatact ggtac	45
EY712	GXS1 39LSat	F	79	Sat Multi	tttgctgcttctggtggtgtcv vgttcggatacgatactggtac t	46
EY723	GXS139DDK	F	78	Sat Multi	tttgctgcttctggtggtgtcD DKttcggatacgatactggtac t	47
EY727	GXS1 38Y	F	78	Quik	gctgcttctggtggttatcttt tcggatacgatac	48
EY728	GXS1 38Y	R	78	Quik	gtatcgtatccgaaaagataac caccagaagcagc	49
EY729	GXS1 38C	F	78	Quik	gctgcttctggtggttgctcttt tcggatacgatac	50
EY730	GXS1 38C	R	78	Quik	gtatcgtatccgaaaagacaac caccagaagcagc	51
EY731	GXS1 38H	F	78	Quik	gctgcttctggtggtcatcttt tcggatacgatac	52
EY732	GXS1 38H	R	78	Quik	gctgcttctggtggtcatcttt tcggatacgatac	53
EY733	GXS1 39H	F	78	Quik	tgcttctggtggtgtccatttc ggatacgatactg	54
EY734	GXS1 39H	R	78	Quik	cagtatcgtatccgaaatggac accaccagaagca	55
EY735	GXS1 39M	F	78	Quik	gctgcttctggtggtgtcatgt tcggatacgatactggt	56
EY736	GXS1 39M	R	78	Quik	accagtatcgtatccgaacatg acaccaccagaagcagc	57

[0161] Table 6: Primers used for point mutations.

Name	Target A.A.	Orientation	T _m	Sequence	SEQ ID NO.:
EY702	SsRGT2-F40M	F	78	cagccttcggtggtatccttatgggttatga cactggt	58
EY703	SsRGT2-F40M	R	78	accagtgtcataaaccataaggataaccaccg aaggctg	59
EY737	GXS1 FLS	F	78	ttttgctgcttctggtggtttcctttctgg atacg	60
EY738	GXS1 FLS	R	78	cgtatccagaaaggaaaccaccagaagcagc aaaaa	61
EY739	GXS1 FIS	F	78	ttttgctgcttctggtggtttcatttctgga taccgatactgg	62
EY740	GXS1 FIS	R	78	ccagtatcgtatccagaaatgaaaccaccag aagcagcaaaa	63
EY741	GXS1 FIM	F	78	ttttgctgcttctggtggtttcattatggga taccgatactgg	64
EY742	GXS1 FIM	R	78	ccagtatcgtatcccataatgaaaccaccag aagcagcaaaa	65
EY747	ScHXT7M340	F	78	ctatgattcaatctctacaacaattgacagg tatgaactatttcttctactatggtactact atattt	66
EY748	ScHXT7M340	R	78	aaaatagtagtagtagtagtagtagtagtagt tcataacctgtcaattggtgtagagattgaat catag	67
EY749	ScHXT7 FHM	F	78	catgatcgcccttgggtggtttccatatgggt tggtgatactggtaccattt	68
EY750	ScHXT7 FHM	R	78	aaatgggtaccagtatcccaaccatatggaa accaccaaaggcgatcatg	69
EY753	ScHXT10 63M	F	78	ctgatgattgccttgggtggttccattatgg gttgggatacagg	70
EY754	ScHXT10 63M	R	78	cctgtatcccaaccataatgaatccaccaa aggcaatcatcag	71
EY760	SsRGT2-38F	F	78	gttcgcagccttcggtggtttccttttcggt ta	72
EY761	SsRGT2-38F	R	78	taaccgaaaaggaaaccaccgaaggctgcga ac	73
EY766	ScHXT7-FIM	F	78	catgatcgcccttgggtggtttcattatgggt tggtgatactggtaccattt	74
EY767	ScHXT7-FIM	R	78	aaatgggtaccagtatcccaaccataatgaa accaccaaaggcgatcatg	75

[0162] *Growth rate measurements* – All exponential growth rates were measured and calculated according to the method previously described using a Bioscreen C (Growth Curves
5 USA, Piscataway, New Jersey) and a MATLAB script (36, 38).

[0163] *Fractional change* – Fractional change in growth rate from wild type was calculated by taking the difference between the growth rates of the mutant and wild type over the growth rate

of the wild type for each individual carbon source. Error was propagated using the least squares method based on the standard deviation in exponential growth rates of the mutant and the wild type.

[0164] *High cell density fermentation* – High cell density experiments were conducted as

previously described (38). Yeast cultures were suspended at OD in 20 g/L glucose, 10 g/L glucose and 10 g/L xylose, or 20 g/L xylose. Supernatant concentration of xylose and/or glucose was measured using a YSI Life Sciences Bioanalyzer 7100MBS.

[0165] *Radiolabeled xylose uptake* – Uptake of ^{14}C labeled xylose was used to determine the Michaelis-Menten parameters for *C. intermedia* *GXS1* and the Phe³⁸ Ile³⁹ Met⁴⁰ triple mutant.

The method was performed as previously described (38).

[0166] *Growth rate measurements* – All exponential growth rates were measured and

calculated according to the method previously described using a Bioscreen C (Growth Curves USA, Piscataway, New Jersey) and a MATLAB script. The Bioscreen C measures online optical density for easy and accurate measurement of the growth curves of up to 200 strains at one time.

Error was calculated based on biological triplicate in all cases. In all cases, the Bioscreen C was set to maintain a temperature of 30 °C, employ high continuous shaking, and to measure optical density every 10 minutes. A single carbon source per well was used in all experiments save one. Growth on xylose in the presence of increasing concentrations of glucose was measured for *C. intermedia* *gxs1* Phe38 Ile39 Met40.

[0167] It is important to note that the environment of the Bioscreen C does not support cultures reaching high optical density and observed values are below OD₆₀₀ of 2. This does not reflect the optical densities reached in flasks, which typically approach OD₆₀₀ of 10.

[0168] *Transporter Cloning* - Each of these transporters was functionally analyzed for conferred growth rate on xylose and glucose in *S. c. EX.12*. Genomic DNA and PCR were

performed as previously described (36). Using this approach, open reading frames from *Scheffersomyces stipitis*, *Debaryomyces hansenii*, *Yarrowia lipolytica*, and *Saccharomyces cerevisiae* were cloned using primers listed in Table 4. Mutant transporters and saturation library construction is described below and Primers are listed in Table 5 (saturation) and Table 6 (point).

[0169] *Saturation mutagenesis and point mutation* – The Strategene Multi mutagenesis kit was used to generate saturation mutagenesis libraries at positions 38, 39, and 40 in *C.i.* *GXS1*. Each codon was replaced with the degenerate NNK sequence recommended for use when creating

saturation mutagenesis libraries. It is important to note that the wild type codon was represented in the NNK library for both Val38 and Leu39 thus alternative 3 primers that did not contain the wild type sequence were designed. This subsequently necessitated the design of specific point mutation primers to access certain residues and the use of the Stratagene Quikchange kit. Some
5 single point mutation primers were ordered to complete the saturation libraries. The Stratagene Quikchange mutagenesis kit was used to generate all rational single, double, and triple mutants. Primers are listed in Table 5 (saturation) and Table 6 (point).

[0170] **Example 2:** Sequence alignment of 54 sequences from major facilitator superfamily sugar transporter proteins. The transporter motif sequence is shown as **bolded** residues and corresponds as described herein to residue positions 36-41 of *C. intermedia* *GXS1* protein.

5	Dh2C02530p	KFRNPLDKTPNIYNVFIASISCI SGLMFG IDISSMSLPIGDDKYIKYFHK----- 63
	Dh2E01166p	KLRLFLDKRLPNINIIYVIATISCI SGLMFG IDISSMSAPLSNDAYLKVFGT----- 63
	Dh2E01298p	KFRNPLDKFPNIHNVIYVVGISCI SGLMFG IDISSMSLPIGDDKYLDYFNS----- 63
	SsHGT2	KFRFLDKRLPNINIIYVIASISCI SGLMFG IDISSMSAPIGEDDYKNPFNN----- 63
	Dh2A14300p	SLNKELDKFHTTYNIYVIAMITTI SGLMFG PDVSSISAPISEPSYRRFFNY----- 61
	Y10B06391p	QVGALQHRFPKLNHPYLTAAVAT MGGLLFG PDISSVSAPVDTKPYKEYFGY----- 59
10	Y10B01342p	-----MYKVHNPYLTAAVAT MGGMLFG PDISSVSAPVGEDNYMNYFGH----- 43
	BmHGT2	-----MGRITNPYVLTALACT GGLLFG PDISSMSAIISPPNYLTYFGPKDLTVECPD 52
	At5g59250	LASDAPESFSWSSVILPFIFF-AL GGLLFG YDIGATSGATLSLQSP---ALSGTTWFNF 139
	At5g17010	---HVPENYSVVAAILPFLFP-AL GGLLFG YEGIGATSCATISLQEPMTLLSYAVPFSAV 89
	SsAUT1	LNAEATNKWHIPRLIGVIALGSM AAAVQGM DESINGANLFPYKAPGVD---TMHNSD 161
15	Y10D00132	LNREITNKWDHMPKVYYLVVCCSI AAAVQGM DETIVINGANIIFPAQPGIKEDSGVVSRRS 180
	BmSTL1	-----FLGMRGIKLNWAIGFAAS AGLLFG YDQGVLSLYTLPSWNAQFPEINTAAVGD 73
	SsXUT6	AKTNSYLGLRGHKLNFVAVSCFAGV GGLLFG YDQGVMSGLLTLPSFENTFPAMK----- 75
	Dh2E01386p	--KTNTMGLRGKPLRVAITICCTI GSLFG YDQGLMSGIITGKQPNEEFPPTHTGT----- 59
	Dh2B05060p	--RTNTMGLRGKRLRMVMTVVATL GSLFG YDQGLMSGIITGEOQFNAEFPPTAGK----- 60
20	SsSTL1	--RTNTFGLRGKRLRAFITVVAVT GSLFG YDQGLMSGIITADQFNSEFPATR----- 60
	ScSTL1	--RTSHWGLTGKRLRYFITIASMT GSLFG YDQGLMASLITGKQPNYEFPPATKENG---D 70
	BmHXT10	----IDVGLRGNWLLTVITASCA AGLLVG YDNGVMGCVVGLGEFNKTFNNPD----- 66
	SsXUT2	-----GKQVSYAVTFTCELA FLIFG IEQGIIGNLINNQDPLNTFGNPTG----- 53
	CnB3990p	--HKTQRRLVGHNLISVSVFSLI GVWLF GYDQGVMSGIITGPYPKAYFNQPTS----- 62
25	Y10F06776p	-----MFSLTGKPLLYFTSVFVSL GVVLF GYDQGVMSGIITGFYPKYFHEPTR----- 49
	BmXUT3	VGATGAKGLIKNARTFAIAVFAS MGLIY GYNQGMFGQILSMHSPQEASGVKGIT----- 78
	SsXUT1	AGKSGVAGLVANSRSFFIAVFAS LGLLV GYNQGMFGQISGMYSFSAIGVEIKQD---- 77
	SsXUT3	AHGNVVTIMMKDPVVFVLVILFAS LGLLV GYDQGVISGIVTMESF--GAKFPRIFM----- 63
	SsXUT3-A	AHGNVVTIMMKDPVVFVLVILFAS LGLLV GYDQGVISGIVTMESF--GAKFPRIFM----- 63
30	SsXUT3-B	AHGNVVTIMMKDPVVFVLVILFAS LGLLV GYDQGVISGIVTMESF--GAKFPRIFM----- 63
	DhXylHP	SKGNIITVMSKDLPLVFCIIAFAS LGLLV GYDQGVISGIVTMESF--AAKFPRIFS----- 64
	ScGAL2	PIETPKKPMSEYVTVSLLCLCVA FGFMFG WDGTGTISGFVQTDPLRRFEG-MKHKDGTT-- 113
	ScHXT8	EVVVPEKPASAYATVSIIMCLMA FGFMFG WDGTGTISGFVNQTDPLRRFEGNYSHSKNT-- 109
	ScHXT1	AVAPPNTGKGVYVTVSICCVMA FGGFIFG WDGTGTISGFVAQTDPLRRFEG-MKHHDGS-- 107
35	ScHXT3	VLTNPTGKGAIVTVSICCVMA FGGFIFG WDGTGTISGFVAQTDPLRRFEG-MKHHDGS-- 104
	ScHXT7	VVEIPKRPASAYVTVSICMIMIA FGGFIFG WDGTGTISGFVNQTDPLRRFEG-MKHKDGTT-- 107
	ScHXT9	PIDLQKPLSAYTTVAILCLMIA FGGFIFG WDGTGTISGFVNLSDPIRRFEG-QKNDKGT-- 103
	ScHXT2	NAELPAKPIAAWYTVICLCLMIA FGGFIFG WDGTGTISGFVNQTDPKRRFEG-QMKSDGT-- 98
	ScHXT10	SLDIPYKPIIAWYTVMGLCLMIA FGGFIFG WDGTGTISGFVNQTDPKRRFEG-ELQRDGS-- 91
40	CiGXF1	QVDAPOKGFKDYIVISIFCFMVA FGGFIFG WDGTGTISGFVNMSDFPKDRFG-QHHADGT-- 86
	ScHXT13	NVEPPKRGLIGYLVIIYLLCYPIS FGFLPG WDGSGITAGFINMDNPKMNFSGSYKHSTGE-- 100
	BmGXF1	-MVFQVRGTPIGALTFLFIAMLAS MGGFLFG WDGTGTISGLTQMADFRORFATVDNPDIAIG- 58
	ScHXT14	GQAAKISHNASLHIPVLLCLVIS LGGFIFG WDIGTIGGMTNMVSPQEKFGTTNIIHDDDET 105
	BmGXS1	GPVARPASVKQSLPAIILVAAASAF GGVLF GYDTGTISGLIVMPNQETFGKPVPGSTTGA 74
45	BmRGT2	GPVARPASVKQSLPAIILVAAASAF GGVLF GYDTGTISGLIVMPNQETFGKPVPGSTTGA 74
	CiGXS1	FVNVEKKKAGSTAMAIIVGLFAAS GGVLF GYDTGTISGVMTMDYVLARY-----PSNK- 64
	CiGXS1-A	FVNVEKKKAGSTAMAIIVGLFAAS GGVLF GYDTGTISGVMTMDYVLARY-----PSNK- 64
	CiGXS1-B	FVNVEKKKAGSTAMAIIVGLFAAF GGVLF GYDTGTISGVMTMDYVLARY-----PSNK- 64
	Dh2D01474	YVNVEKKRAGSASMGIFVGAFAAF GGVLF GYDTGTISGIMAMNYVKGEF-----PANK- 64
50	Dh2D02167p	YVNVEKKRAGSASMGIFVGAFAAF GGVLF GYDTGTISGIMAMNYVKGEF-----PANK- 64
	SsRGT2	YINFGKKKAGSTTMGICVGLFAAF GGILF GYDTGTISGIMAMDYVTARF-----PSNH- 64
	Y10C06424p	IINRGEKPEGSFAFMAAFVAVFA GGILF GYDTGTISGVMAMPFVKKTF-----TDDG- 58
	Y10C08943p	-----MAIIVAVFA GGLLY GYDTGTIAGIMTMGYVKEHF-----TDFGK 41
	Dh2B14278p	YKKMQQKS--SSSAITVGLVAA AGGLY GYDTGLINDIMEMTYVKDNF-----PANG- 69
55	EcXylE	-----MNTQYNSSYIFSTILVAT LGLLFG YDTAVISGTVESLHTVVFAPQNLSESAAN- 54
	SsXUT5	RSIGPLIPRNKHLFYGSVLLMSIV HPTIM GYDSMMVGSILNLDAYVNYFH----- 53
	ScMAL11	KSMTLKQALLKYPKAALWSILVST TLVME GYDTALLSALYALPVFQPKFGTINGEGS--- 148

[0171] **Example 3:** Sequence alignment of 57 sequences from major facilitator superfamily sugar transporter proteins. **Bolded** residues correspond to the alignment of conserved residue corresponding to 297 of *C. intermedia* *GXS1* protein.

5	Dh2C02530p	WAQAWQQLTGMNTLMYYIVYVFMAGYEG-DANLVASSIQYCLNTGMTIPALYFMDKLGR	340
	Dh2E01166p	FAQIWQQLTGMNTLMYYIVYVFMAGYHG-DANLVASSIQYCINFAMTIPALYLMCKVGR	340
	Dh2E01298p	FAQIWQQLTGMNTLMYYIVYVFMAGYQG-DANLIASSIQYVLFVMTAPSLYLMCKLGR	340
	SsHGT2	FAQIWQQLTGMNVMMYYIVYIFNMAGYSN-NANLVASSIQYVLNTAATVPALFLMDYIGR	340
	Dh2A14300p	SAQIWNQLTGMNVMMYYIVYIFEMVGYTG-NTVLVSSSIQYVINFGVTLIALPLSDYVGR	336
	Y10B06391p	WAQIWQQLTGMNVMMYYIVLIFTMAGYTG-NANLVASSIQYVINMIMTIPALLFIDRVGR	336
10	Y10B01342p	WAQIWQQLTGMNMIMYYVVIIFKMAGYSGKSAVIVSGSIQYIINVMTIPALLFIDKIGR	320
	BmHGT2	FTQIWSQLTGMNVMMYYLSYVFMAGITG-NIALISNGIQYVINVMTVPALLYVDRWGR	347
	At5g59250	GLVLFQQITGQPSVLYYAGSILQTAGFSAADATRVSVIIGVFKLLMTWVAVKVDDLGR	424
	At5g17010	GLVLFQQ-----LIMTGVAVVVIDRLGR	334
	SsAUT1	FIVMFQQFCGINVIAYYSSSIFVQSGFSQTSALIASWGFGMLNFTPAIPAFFTIDRFGR	441
15	Y10D00132	FIVMFQQFCGINVIAYYSSNIFMESGFGAIQALLASFGFGAINFVPALPAVYTIDTFGR	459
	BmSTL1	MSQMFQQISGINLITYYIKTLQEQLGFSIDINSRILAAANGTEYPIASWAAVFFIEKMGR	353
	SsXUT6	WSQIMQQITGNIITYYAGTIFESYIGMSPFMSRILAAALNGTEYFLVSLIAFYTVRVLGR	363
	Dh2E01386p	STQFFQQFTGCNAAIYYSTVLPENSIGLTGKPLILGGVFATYIALSTIPSFFLIDRLGR	344
	Dh2B05060p	SGQFFQQFTGCNAAIYYSTVLPEDTIHLERRLALILGGVFATYIALSTIPSFFLVDTLGR	345
20	SsSTL1	STQFFQQFTGCNAAIYYSTVLPQDTIGLERMALIIGGVFATVYAIPTIPSFFLVDTLGR	342
	SsSTL1	STQFFQQFTGCNAAIYYSTVLPNKTIKLDYRLSMIIGGVFATYIALSTIPSGFFLIEKLGR	356
	BmHXT10	FIQAAQQLSGINNALIYYSGTLFSSQISGLDSKKSALFAGGLNMCLILGSTISIFLIDRVGR	346
	SsXUT2	MSMFAQQLSGYNVNNYYITFVLINSVGIEDNLALILGGVAVICFTVGSVLPVTFADRMGR	330
	CnBC3990p	SSQLFAQLNGINVISYAPLVFEQAG-WIGRDAILMTGINALFYVASSLPWYLMDRAGR	334
25	Y10F06776p	SSQMFQQLNGINVISYAPLVFEEAG-WVGRSAILMTGINGIVYVCSTIPPWYLVLDKWR	322
	Dh2F19140p	FSQMFQQLNGINMVSYAPMIPELAG-WVGRQAILMTGINSIVYVLTSTIPPWYLVLDGWR	293
	SsXUT4	FSQMFQQLNGINMVSYAPMIPESAG-WVGRQAILMTGINSIYYIFSTIPPWYLVDSWGR	293
	SsXUT7	SALGFAQFNGINIISYAPMVFEAG-FNNSKALLMTGINSIVYWFSTIPPWYLVLDHWR	274
	BmXUT3	LIMLFQQWGTGNFILIYAPPIFKQJGLSGNTISLLASGVVGIPLATIPAVLYIDSWGR	382
30	SsXUT1	LIMTFQQWGTGNFILIYAPPIFSSGLSGNTISLLASGVVGIPLATIPAVLVVDRLGR	381
	SsXUT3	AVMFFQQFPGCNAIYYAPTIFTQLGMNSTTSLTGLTGLYGINCLSTLPAVFLIDRCGR	381
	SsXUT3-A	AVMFFQQFPGCNAIYYAPTIFTQLGMNSTTSLTGLTGLYGINCLSTLPAVFLIDRCGR	381
	SsXUT3-B	AVMFFQQFPGCNAIYYAPTILTRLGMNSTTSLTGLTGLYGINCLSTLPAVFLIDRCGR	381
	DhXyl1HP	AVMFFQQFPGCNAIYYAPTIFSQLGMDSENTALLGTGVYGINCLSTIPALFAIDRFGR	382
35	SsGAL2	FVQMFQQLTGNIFYFFYYGTIVFKSVGLDD---SFETSIVIGVGNFASFSSFLWTVENLGH	392
	SsHXT8	MINSLQQLTGDNYFFYYGTTFKSVGMND---SFETSIVLGIVNPFASFSSFLYSVDKLR	388
	SsHXT1	MIQSLQQLTGDNYFFYYGTTFVQAVGLSD---SFETSIVFGVGNPFSTCCSLYTVDRFGR	386
	SsHXT3	MIQSLQQLTGDNYFFYYGTTFVNAVGMSD---SFETSIVFGVGNPFSTCCSLYTVDRFGR	383
	SsHXT7	MIQSLQQLTGDNYFFYYGTTFKAVGLSD---SFETSIVLGIVNPFASFVGVYVVERYGR	386
40	SsHXT9	MIQSLQQLTGDNYFFYYGTTFKSVGLKD---SFQTSIIIGVGNPFSSFIAYVTIERFGR	382
	SsHXT2	MIQSLQQLTGNIFYFFYYGTTFNAVGMKD---SFQTSIVLGIVNPFASFVALYTVDKFGR	377
	SsHXT10	VIQSLQQLTGCNIFYFFYYGTTFNAVGMQD---SFETSIVLGAVNPFASFVALYIVDKFGR	370
	CiGXF1	MLQSLQQLTGDNYFFYYGTTFQAVGLKD---SFQTSIILGIVNPFASFVGVYVIERLGR	365
	SsHXT13	LVQTFQLTGCNIFYFFYYGTTFKSVGLTD---GFETSIVLGTVNPFSTIIAVMVVDKIGR	379
45	BmGXF1	TLQAGQQFTGANIFYFFYPGTAFITSVGLSD---SFVTQIILGAVNPFASFGLYILERFGR	340
	BmHXT14	MIMAFQQLSGINIFYFFYGTSTVFKGVIKID---PYITSIISSVNPLSTILGIYYVEKWGH	403
	BmGXS1	FIQAFQQLTGNIFYFFYGTTFKFSALPGTN---PFIFSVISNVNVVTVPGMYMMERLGR	354
	BmRGT2	FIQAFQQLTGNIFYFFYGTTFKFSALPGTN---PFIFSVISNVNVVTVPGMYMMERLGR	730
	CiGXS1	AIQAFQQLTGVNIFYFFYGTTFKFRAGVN---GFTISLATNIVNVGSTIPGILLMEVLGR	342
50	CiGXS1-A	AIQAFQQLTGVNIFYFFYGTTFKFRAGVN---GFTISLATNIVNVGSTIPGILLMEVLGR	342
	CiGXS1-B	AIQAFQQLTGVNIFYFFYGTTFKFRAGVN---GFTISLATNIVNVGSTIPGILLMEVLGR	342
	Dh2D01474	ALQAFQQLTGVNIFYFPGTSFFKSAGIEN---EFLISLATSIVNVGMTVPGIFLIELVGR	343
	Dh0D02167p	ALQAFQQLTGVNIFYFPGTSFFKSAGIEN---EFLISLATSIVNVGMTVPGIFLIELVGR	343
	SsRGT2	GIQALQQLTGNIFYFFYGTTFKFGSGIKN---EFLIQMATNIVNVGSTVPGILLVEIIGR	343
55	Y10C06424p	AIQALQQLTGNIFYFFYGTTFEKKSNISN---PFLIQMITNIVNVMTIPGIMFVDRVGR	336
	Y10C08943p	SIQALQQLTGNIFYFFYGTTFEKKTAGIKD---PFVVSMTSNAVNFATPLPGILFVDKVR	319
	Dh2B14278p	GVQAFQSSGNIFYFFYGVNFFASSGIKN---YYLMSFVTVAVNTLFTIPGIIIEVIGR	351
	EcXylE	MLSIFQQFVGINVVLYAPEVFKTLGAST-DIALLOTIIVGVINLTFTVLAIMTVDKFGR	341
	SsXUT5	TQAIIVTEMAGSEVGSYYFSIIILTQAGVKDSNDRRLRVNIVMSSWSLVIALSGCLMFDRIGR	331
60	SsMAL11	CLTWVAQNSSGAVLLGYSTYFFERAGMAT-DKAFTFSLIQYCLGLAGTLCSSWVISGRVGR	431

[0172] References

1. Reijenga KA, *et al.* (2001). *Biophysical Journal* 80(2):626-634.
2. Gardonyi M, Jeppsson M, Liden G, Gorwa-Grausland MF, & Hahn-Hagerdal B (2003). *Biotechnology and Bioengineering* 82(7):818-824.
- 5 3. Elbing K, *et al.* (2004). *Applied and Environmental Microbiology* 70(9):5323-5330.
4. Wahlbom CF, Otero RRC, van Zyl WH, Hahn-Hagerdal B, & Jonsson LJ (2003). *Applied and Environmental Microbiology* 69(2):740-746.
5. Bengtsson O, *et al.* (2008). *Yeast* 25(11):835-847.
6. Jeffries TW & Jin YS (2004). *Applied Microbiology and Biotechnology* 63(5):495-509.
- 10 7. Hahn-Hagerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, & Gorwa-Grauslund MF (2007). *Applied Microbiology and Biotechnology* 74(5):937-953.
8. Martin CH, Nielsen DR, Solomon KV, & Prather KLJ (2009). *Chemistry & Biology* 16(3):277-286.
9. Tyo KEJ, Kocharin K, & Nielsen J (2010). *Current Opinion in Microbiology*
15 13(3):255-262.
10. Curran KA & Alper HS (2012). *Metabolic Engineering* 14(4):289-297.
11. Hahn-Hagerdal B, Galbe M, Gorwa-Grauslund MF, Liden G, & Zacchi G (2006). *Trends in Biotechnology* 24(12):549-556.
12. Almeida JR, *et al.* (2007). *Journal of Chemical Technology & Biotechnology*
20 82(4):340-349.
13. Van Vleet JH & Jeffries TW (2009). *Curr Opin Biotechnol* 20(3):300-306.
14. Zhang F, Rodriguez S, & Keasling JD (2011). *Current Opinion in Biotechnology* 22(6):775-783.
15. Liu L, Redden H, & Alper HS (2013). *Current Opinion in Biotechnology*. DOI:
25 10.1016/j.copbio.2013.03.005.
16. Hong KK & Nielsen J (2012). *Cellular and molecular life sciences : CMLS* 69(16):2671-2690.

17. Bae JY, Laplaza J, & Jeffries TW (2008). *Applied Biochemistry and Biotechnology* 145(1-3):69-78.
18. Karhumaa K, Pahlman AK, Hahn-Hagerdal B, Levander F, & Gorwa-Grauslund MF (2009). *Yeast* 26(7):371-382.
- 5 19. Runquist D, Hahn-Hagerdal B, & Bettiga M (2010). *Applied and Environmental Microbiology* 76(23):7796-7802.
20. Krahulec S, Klimacek M, & Nidetzky B (2012). *Journal of Biotechnology* 158(4):192-202.
21. Lee SM, Jellison T, & Alper HS (2012). *Applied and Environmental Microbiology*
10 78(16):5708-5716.
22. Scalcinati G, *et al.* (2012). *FEMS Yeast Res* 12(5):582-597.
23. Jojima T, Omumasaba CA, Inui M, & Yukawa H (2010). *Applied Microbiology and Biotechnology* 85(3):471-480.
24. Young E, Lee SM, & Alper H (2010). *Biotechnology for Biofuels* 3(24):24.
- 15 25. Boles E & Hollenberg CP (1997). *FEMS Microbiology Reviews* 21(1):85-111.
26. Wieczorke R, *et al.* (1999). *FEBS Letters* 464(3):123-128.
27. Pao SS, Paulsen IT, & Saier MH, Jr. (1998). *Microbiology and Molecular Biology Reviews* 62(1):1-34.
28. Ozcan S & Johnston M (1999). *Microbiology and Molecular Biology Reviews*
20 63(3):554.
29. Sedlak M & Ho NWY (2004) Char. *Yeast* 21(8):671-684.
30. Subtil T & Boles E (2012). *Biotechnology for Biofuels* 5:14.
31. Leandro MJ, Goncalves P, & Spencer-Martins I (2006). *Biochemical Journal* 395:543-549.
- 25 32. Saloheimo A, *et al.* (2007). *Applied Microbiology and Biotechnology* 74(5):1041-1052.
33. Hector RE, Qureshi N, Hughes SR, & Cotta MA (2008). *Applied Microbiology and Biotechnology* 80(4):675-684.
34. Katahira S, *et al.* (2008). *Enzyme and Microbial Technology* 43(2):115-119.

35. Du J, Li SJ, & Zhao HM (2010). *Molecular Biosystems* 6(11):2150-2156.
36. Young E, Poucher A, Comer A, Bailey A, & Alper H (2011). *Applied and Environmental Microbiology* 77(10):3311-3319.
37. Leandro MJ, Fonseca C, & Goncalves P (2009). *FEMS Yeast Research* 9(4):511-525.
- 5 38. Young EM, Comer AD, Huang HS, & Alper HS (2012) A. *Metabolic Engineering* 14(4):401-411.
39. Kasahara T & Kasahara M (2010). *Journal of Biological Chemistry* 285(34):26263-26268.
40. Ha SJ, *et al.* (2013). *Applied and Environmental Microbiology* 79(5):1500-1507.
- 10 41. Sun L, *et al.* (2012) Cry. *Nature* 490(7420):361-366.
42. Hamacher T, Becker J, Gardonyi M, Hahn-Hagerdal B, & Boles E (2002). *Microbiology* 148(Pt 9):2783-2788.
43. Kasahara T, Shimogawara K, & Kasahara M (2011). *Biochemistry* 50(40):8674-8681.
44. Davis EO & Henderson PJF (1987). *Journal of Biological Chemistry* 262(29):13928-13932.
- 15 45. Subtil T & Boles E (2011). *Biotechnology for Biofuels* 4:38.
46. Sambrook J (2000) *Molecular Cloning: A Laboratory Manual*. ed Russell DW (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
47. Mumberg D, Muller R, & Funk M (1995). *Gene* 156(1):119-122.
- 20

WHAT IS CLAIMED IS:

1 1. A recombinant xylose transporter protein comprising a transporter motif
 2 sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and 41 of *Candida*
 3 *intermedia* *GXS1* protein, wherein said transporter motif sequence is -G-G/F-X¹-X²-X³-G-;
 4 wherein,
 5 X¹ is D, C, G, H, I, L, or F;
 6 X² is A, D, C, E, G, H, or I;
 7 X³ is N, C, Q, F, G, L, M, S, T, or P; and
 8 wherein, said transporter motif sequence is not -G-G-L-I-F-G- or -G-G-F-I-F-G-.

1 2. The recombinant xylose transporter protein of claim 1, wherein,
 2 X¹ is D, C, G, H, or F;
 3 X² is H or I; and
 4 X³ is S, T, or M.

1 3. The recombinant xylose transporter protein of claim 1, wherein X¹ is F, X²
 2 is I, and X³ is M or S.

1 4. The recombinant xylose transporter protein of claim 1, wherein said
 2 transporter motif sequence is -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, -G-F-F-I-S-G-,
 3 -G-G-F-I-T-G-, -G-F-F-I-T-G-, -G-G-F-L-M-G-, -G-F-F-L-M-G-, -G-G-F-L-S-G-,
 4 -G-F-F-L-S-G-, -G-G-F-L-T-G-, -G-F-F-L-T-G-, -G-G-F-H-M-G-, -G-F-F-H-M-G-,
 5 -G-G-F-H-S-G-, -G-F-F-H-S-G-, -G-G-F-H-T-G- or -G-F-F-H-T-G-.

1 5. The recombinant xylose transporter protein of claim 4, wherein said
 2 transporter motif sequence is -G-G-F-I-M-G-, -G-F-F-I-M-G-, -G-G-F-I-S-G-, or -G-F-F-I-S-G-.

1 6. The recombinant xylose transporter protein of claim 5 further comprising
 2 a mutation of an amino acid at the residue position corresponding to 297 of *Candida intermedia*
 3 *GXS1* protein.

1 7. The recombinant xylose transporter protein of claim 6, wherein said amino
 2 acid at the residue position corresponding to 297 of *Candida intermedia* *GXS1* protein is
 3 substituted with a Met, Ala, Ser, or Asn residue.

1 8. The recombinant xylose transporter protein of claim 7, wherein said
2 recombinant xylose transporter protein is derived from a *C. intermedia gxs1* transporter protein,
3 a *S. stipitis rgt2* transporter protein, or a *S. cerevisiae hxt7* transporter protein.

1 9. A recombinant galactose-arabinose transporter protein comprising a
2 transporter motif sequence corresponding to amino acid residue positions 36, 37, 38, 39, 40, and
3 41 of *Candida intermedia GXS1* protein, wherein said transporter motif sequence is -G-G/F-X⁴-
4 X⁵-X⁶-G-;

5 wherein,

6 X⁴ is D, C, F, G, H, L, R, T, or P;

7 X⁵ is A, C, E, F, H, K, S, P, or V;

8 X⁶ is R, D, E, F, H, I, M, T, or Y; and

9 wherein said sequence is not -G-G-L-V-Y-G-, or -G-G-F-V-F-G-.

1 10. The recombinant galactose-arabinose transporter protein of claim 9,
2 wherein,

3 X⁴ is R, T, H, or F;

4 X⁵ is P, H, or V; and

5 X⁶ is T, H, F, M, or Y.

1 11. The recombinant galactose-arabinose transporter protein of claim 9,
2 wherein

3 X⁴ is F or T;

4 X⁵ is P or I; and

5 X⁶ is M or T.

1 12. The recombinant galactose-arabinose transporter protein of claim 11,
2 wherein said transporter motif sequence is -G-G-F-H-M-G-, -G-F-F-H-M-G-,
3 -G-G-R-P-T-G-, -G-F-R-P-T-G-, -G-G-T-P-T-G-, or -G-F-T-P-T-G-.

1 13. The recombinant galactose-arabinose transporter protein of claim 12,
2 wherein said galactose-arabinose transporter protein further comprises a mutation of an amino
3 acid at the residue position corresponding to 297 of *Candida intermedia GXS1* protein.

1 14. The recombinant galactose-arabinose transporter protein of claim 13,
2 wherein said amino acid at the residue position corresponding to 297 of *Candida intermedia*
3 *GXS1* protein is substituted with a Met, Thr, Ala, or Ile residue.

1 15. The recombinant galactose-arabinose transporter protein of claim 14,
2 wherein said recombinant galactose-arabinose transporter protein is derived from a *C. intermedia*
3 *gxs1* transporter protein, a *S. stipitis rgt2* transporter protein, a *S. cerevisiae hxt7* transporter
4 protein, or a *S. cerevisiae GAL2* protein.

1 16. A yeast cell comprising the recombinant xylose transporter protein of
2 claim 1.

1 17. A yeast cell comprising the recombinant galactose-arabinose transporter
2 protein of claim 9.

1 18. A nucleic acid encoding the recombinant xylose transporter protein of
2 claim 1.

1 19. A nucleic acid encoding the recombinant galactose-arabinose transporter
2 protein of claim 9.

1 20. A method of transporting xylose into a yeast cell, said method comprising:
2 (i) contacting a yeast cell comprising the recombinant xylose transporter protein
3 of claim 1 with a xylose compound; and
4 (ii) allowing said recombinant xylose transporter protein to transport said xylose
5 compound into said yeast cell.

1 21. The method of claim 20, wherein said xylose compound forms part of
2 lignocellulosic biomass, hemicellulose, or xylan.

1 22. The method of claim 20, wherein said yeast cell metabolizes said xylose
2 compound.

1 23. The method of claim 20, wherein said yeast cell converts said xylose
2 compound to a biofuel.

1 24. The method of claim 20, wherein said recombinant xylose transporter
2 protein has a binding affinity of at least 0.7 mM for said xylose compound.

1 25. The method of claim 20, wherein said recombinant xylose transporter
2 protein has a rate of at least 15 nmol min⁻¹ gDCW⁻¹ of transporting said xylose compound into
3 said yeast cell.

1 26. A method of transporting galactose or arabinose into a yeast cell, said
2 method comprising:

3 (i) contacting a yeast cell comprising the recombinant galactose-arabinose
4 transporter protein of claim 9 with a galactose compound or an arabinose compound; and

5 (ii) allowing said recombinant galactose-arabinose transporter protein to transport
6 said galactose compound or said arabinose compound into said yeast cell.

1 27. The method of claim 26, wherein said recombinant galactose-arabinose
2 transporter protein is contacted with an arabinose compound.

1 28. The method of claim 27, wherein said arabinose compound forms part of
2 lignocellulosic biomass, hemicellulose or arabinoxylan.

1 29. The method of claim 26, wherein said yeast cell metabolizes said
2 arabinose compound.

1 30. The method of claim 26, wherein said yeast cell converts said arabinose
2 compound to a biofuel.

1 31. The method of claim 26, wherein said recombinant galactose-arabinose
2 transporter protein is contacted with a galactose compound.

1 32. The method of claim 31, wherein said galactose compound forms a part of
2 lignocellulosic biomass, hemicellulose, or galactan.

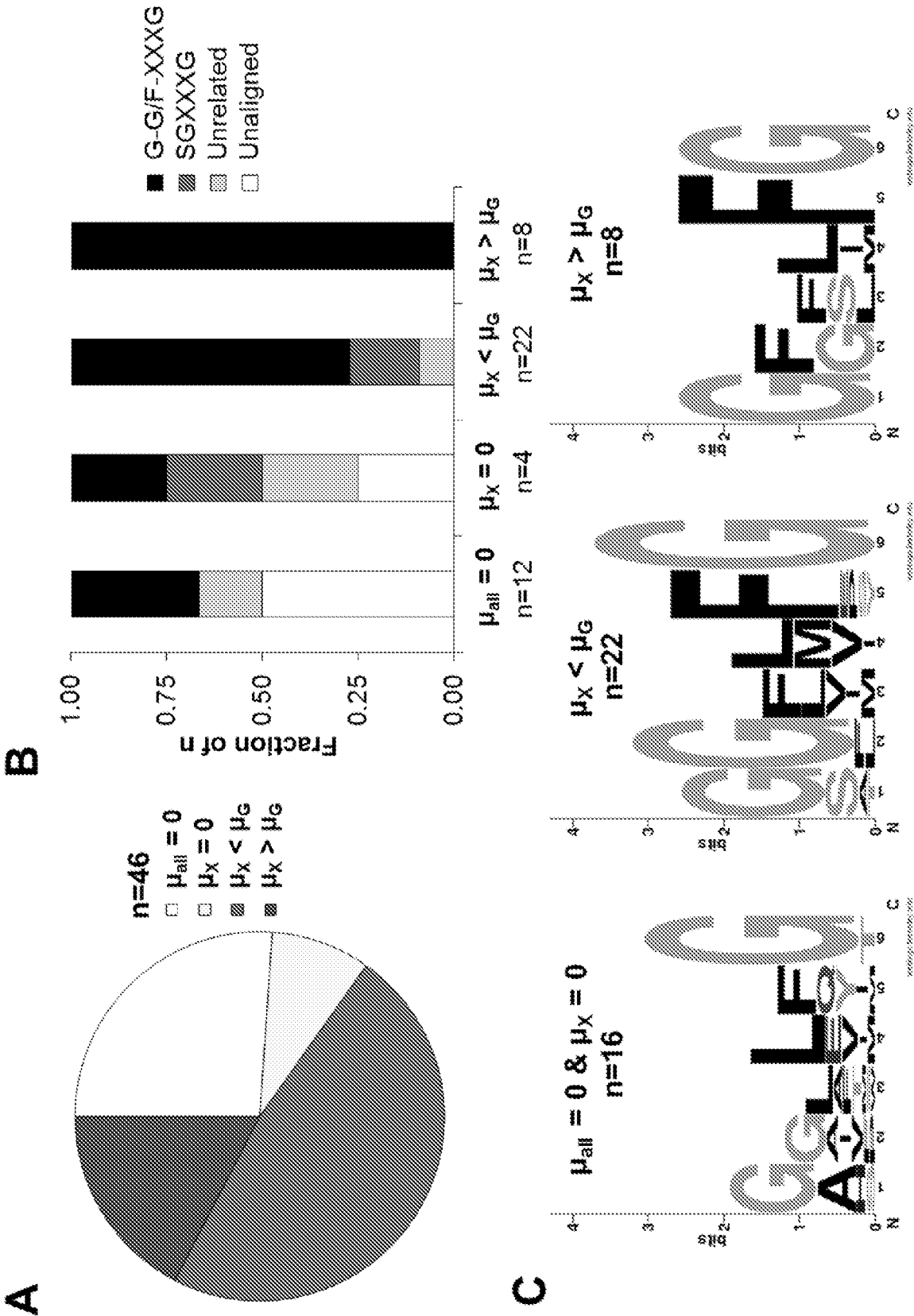
1 33. The method of claim 31, wherein said yeast cell metabolizes said
2 galactose compound.

1 34. The method of claim 31, wherein said yeast cell converts said galactose
2 compound to a biofuel.

1 35. The method of claim 20, wherein said yeast cell is a *S. stipitis* yeast cell, a
2 *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast
3 cell.

1 36. The method of claim 26, wherein said yeast cell is a *S. stipitis* yeast cell, a
2 *C. intermedia* yeast cell, a *S. cerevisiae* yeast cell, a *D. hansenii* yeast cell, or a *Y. lipolytica* yeast
3 cell.

Figure 1



2/22

Figure 2

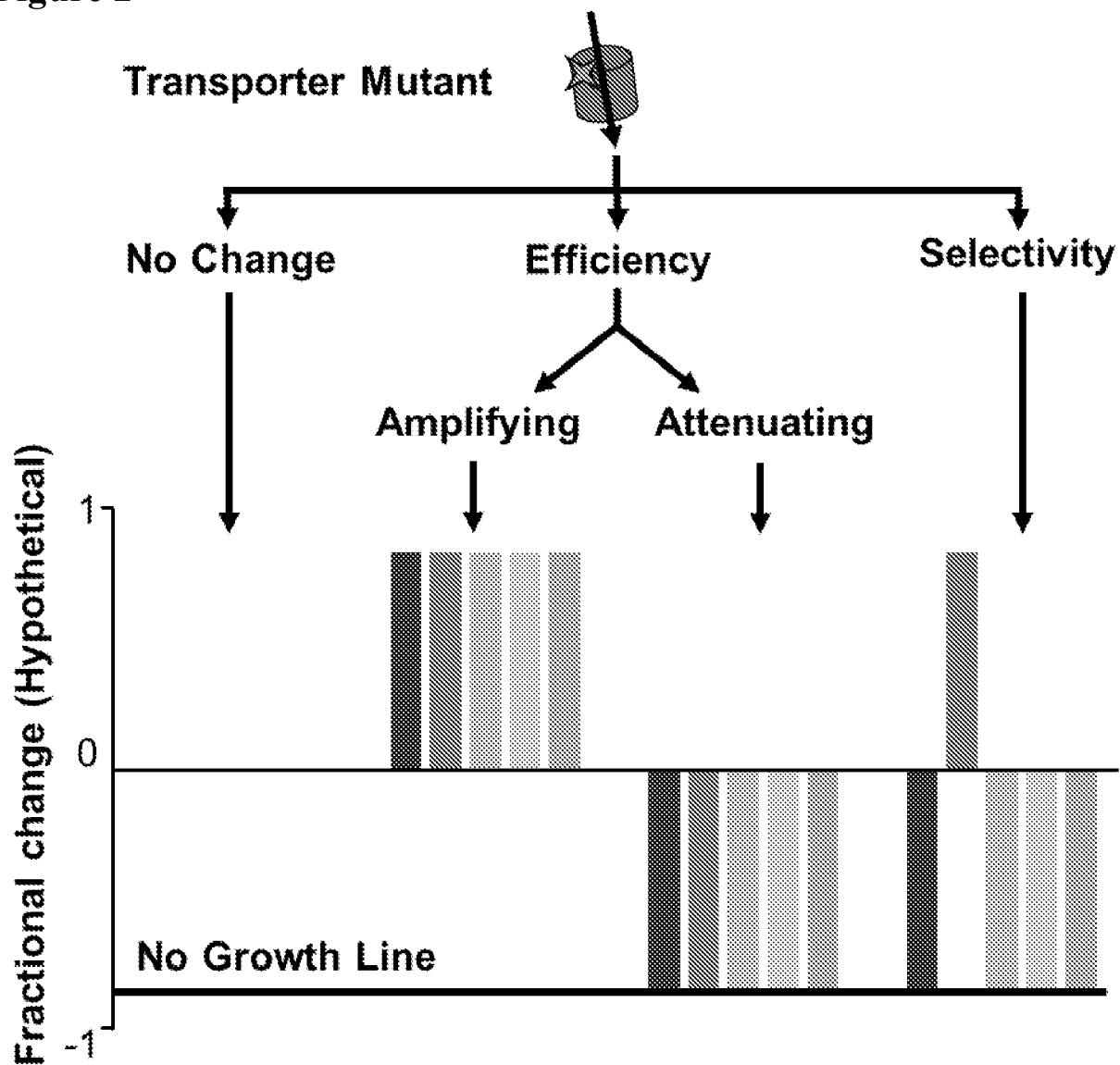
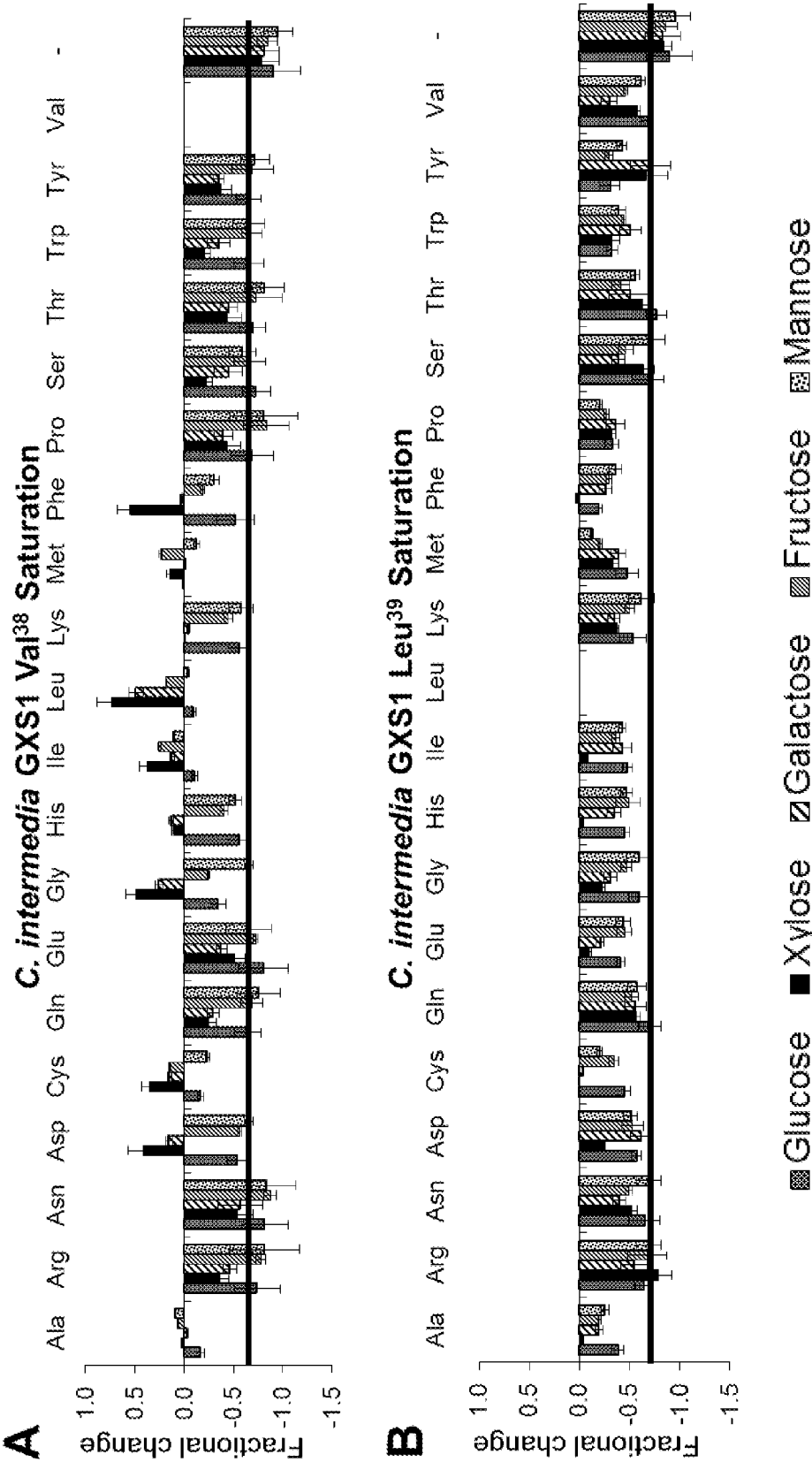
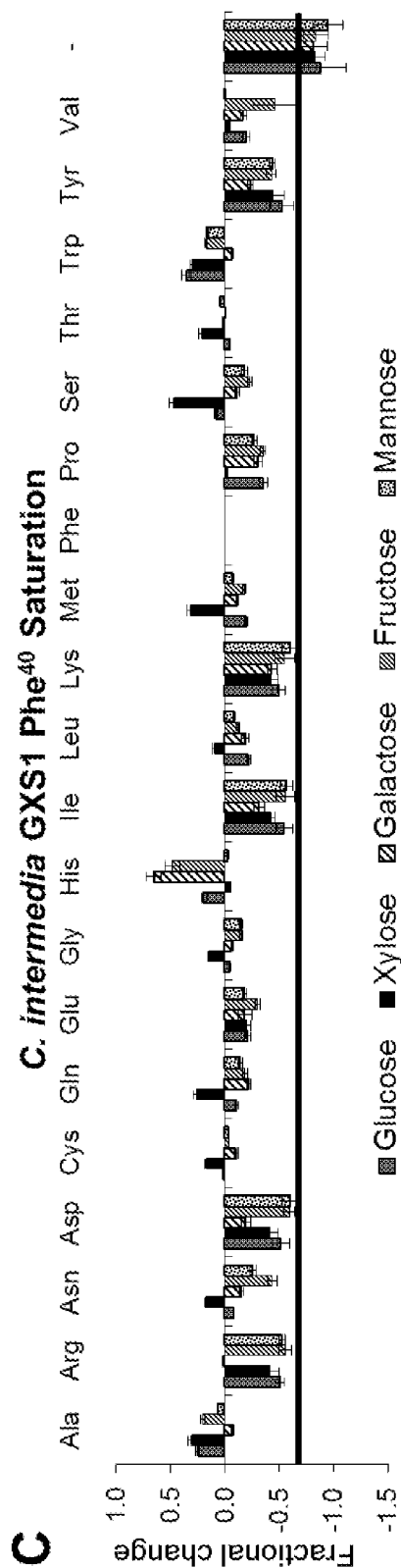


Figure 3



4/22

Figure 3 (cont.)



5/22

Figure 4

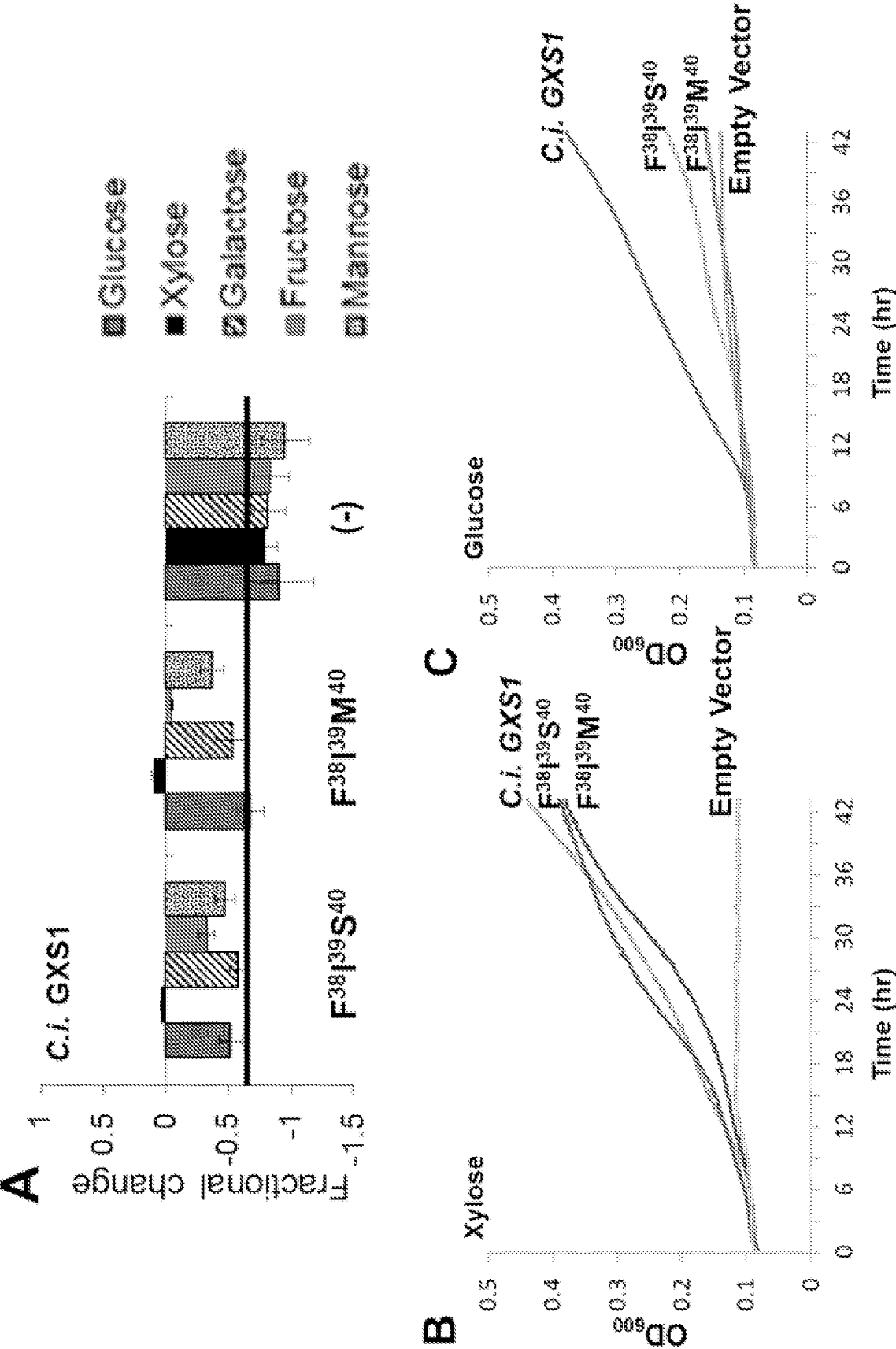


Figure 5

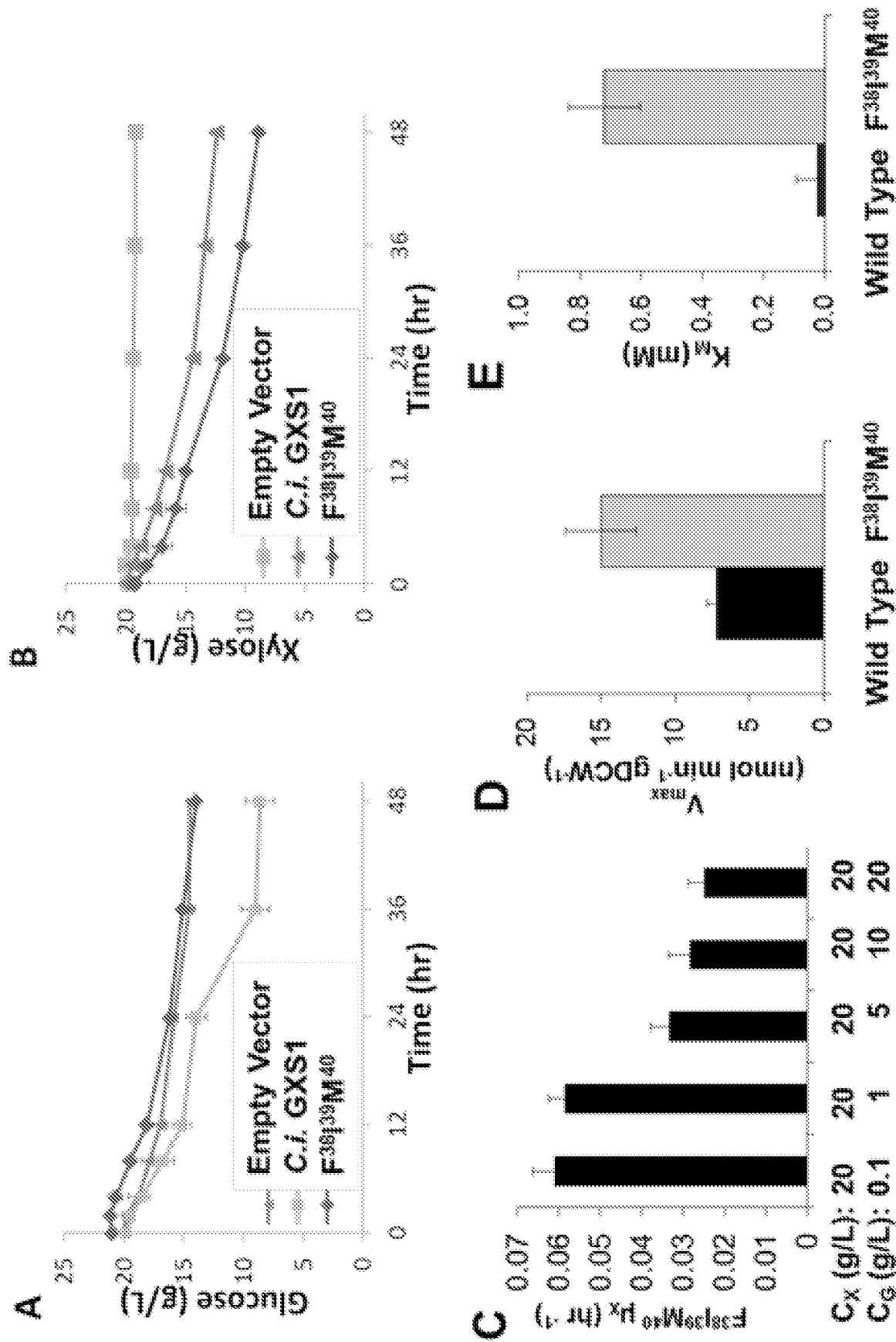
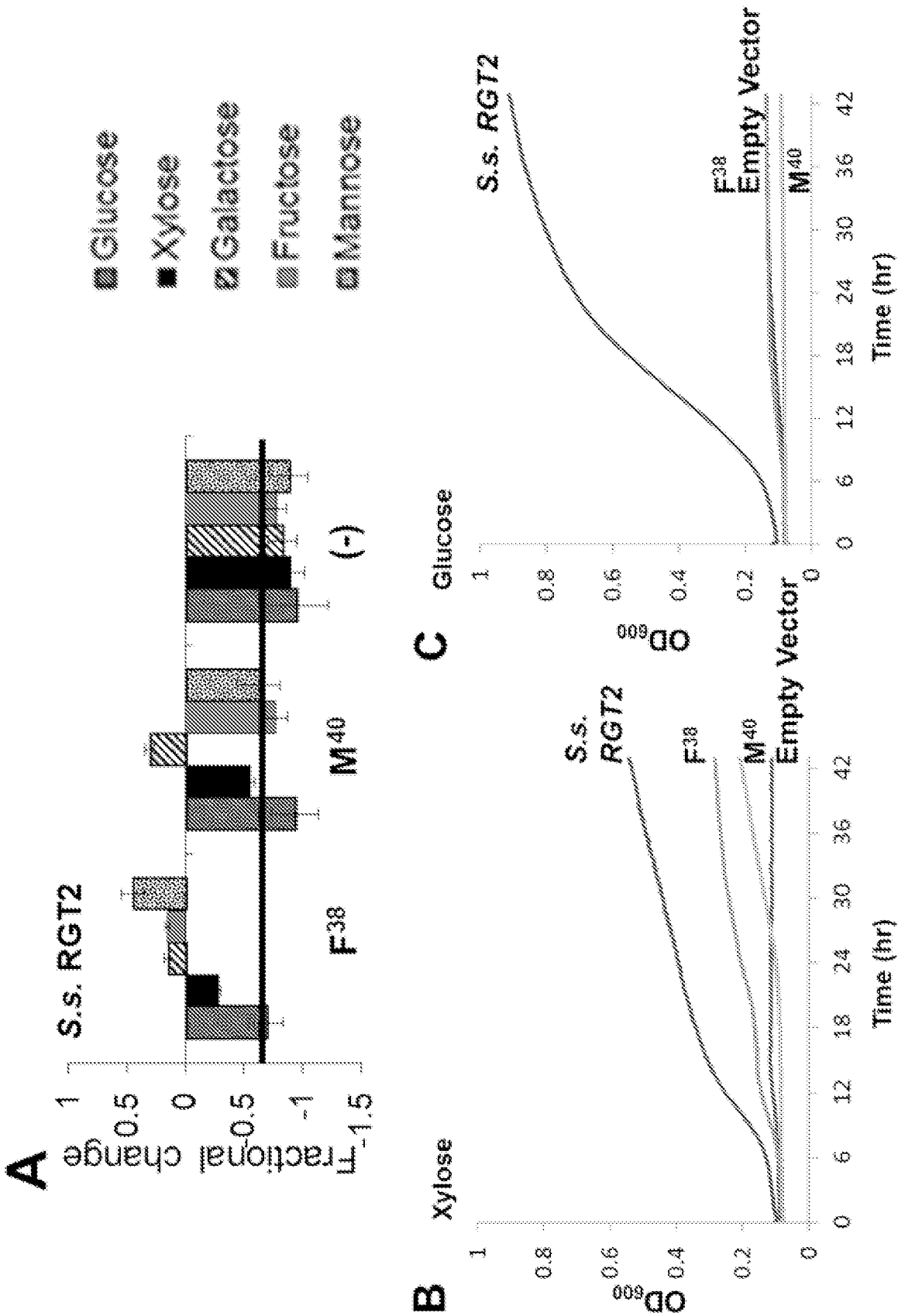
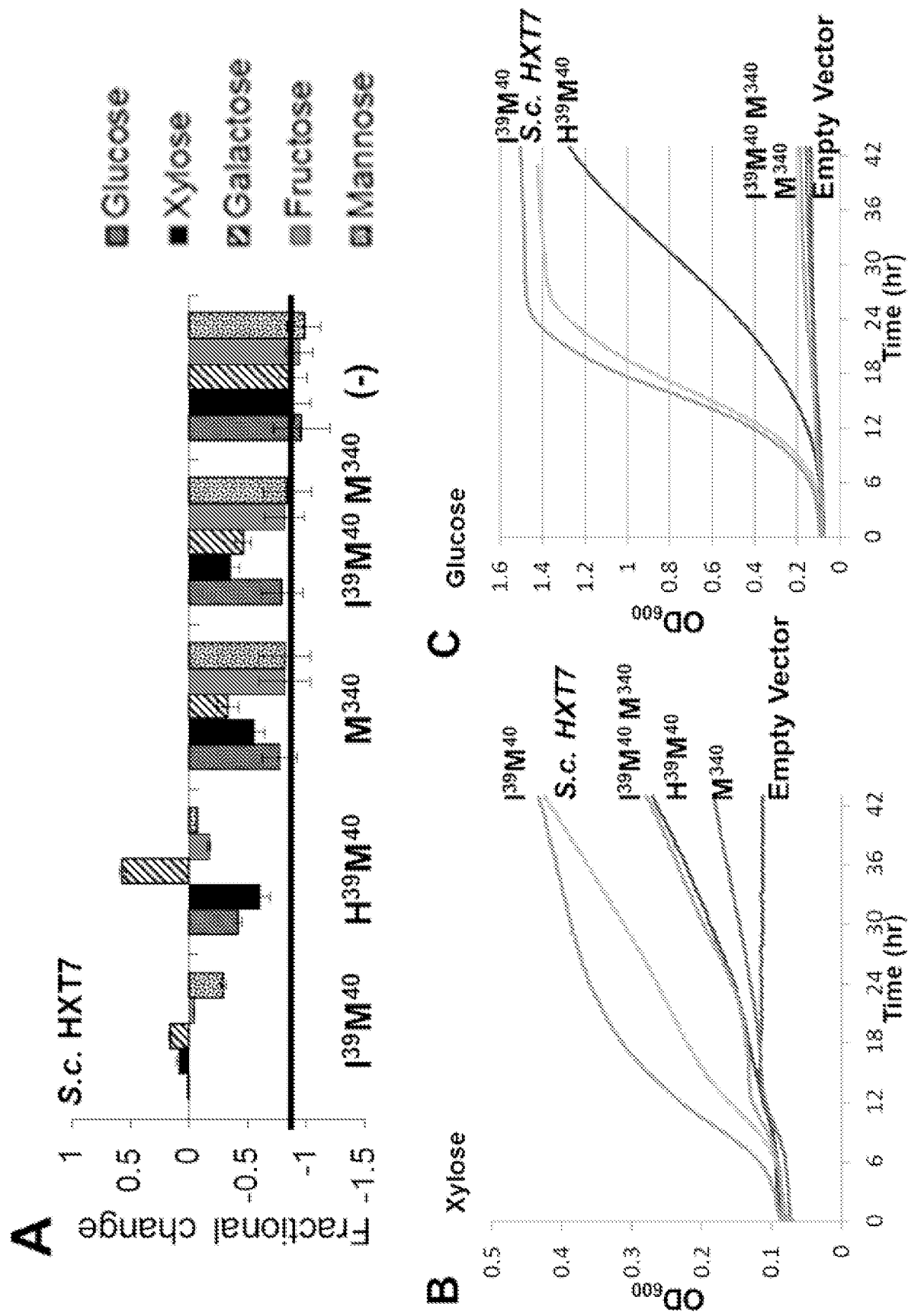


Figure 6



8/22

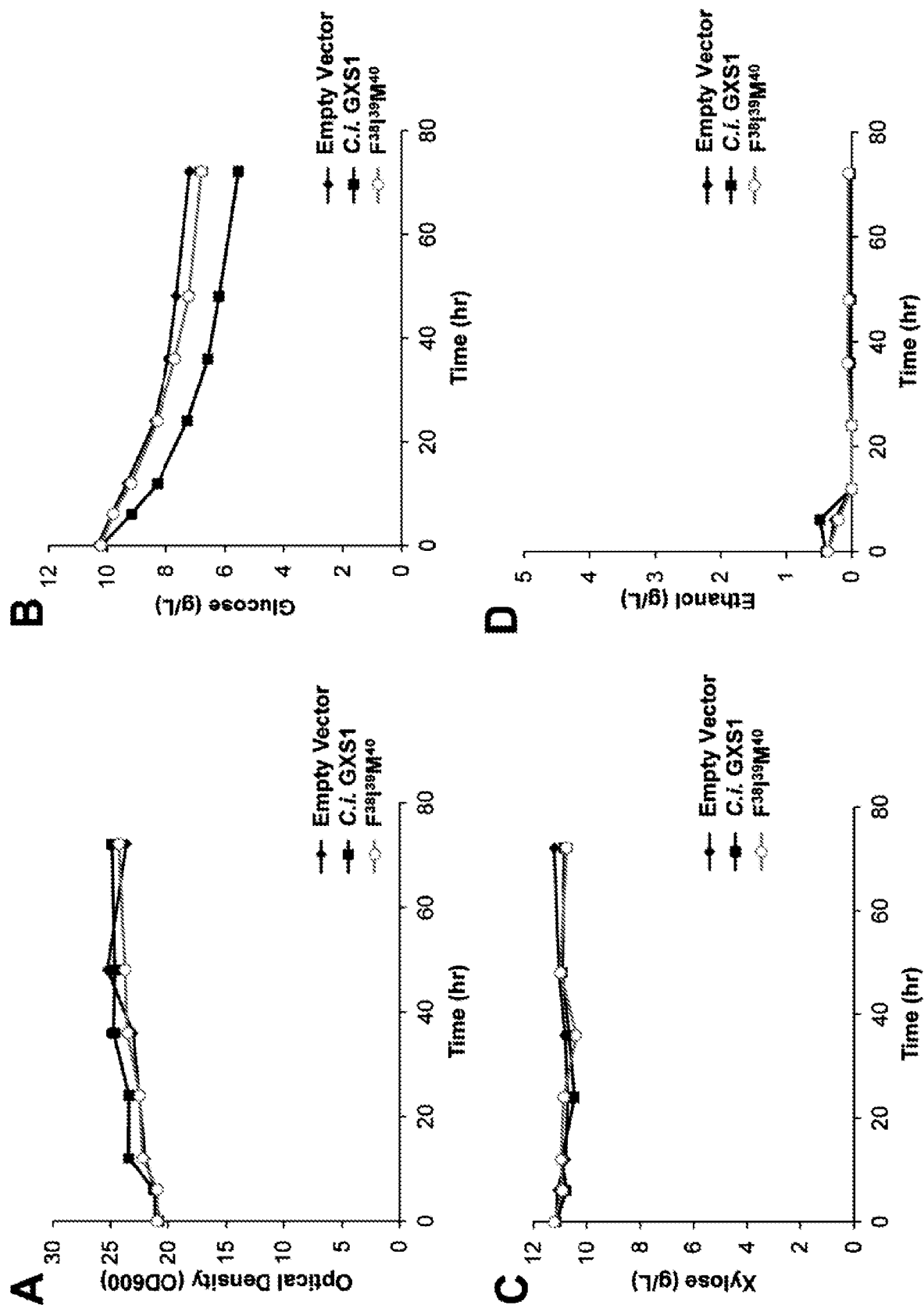
Figure 7





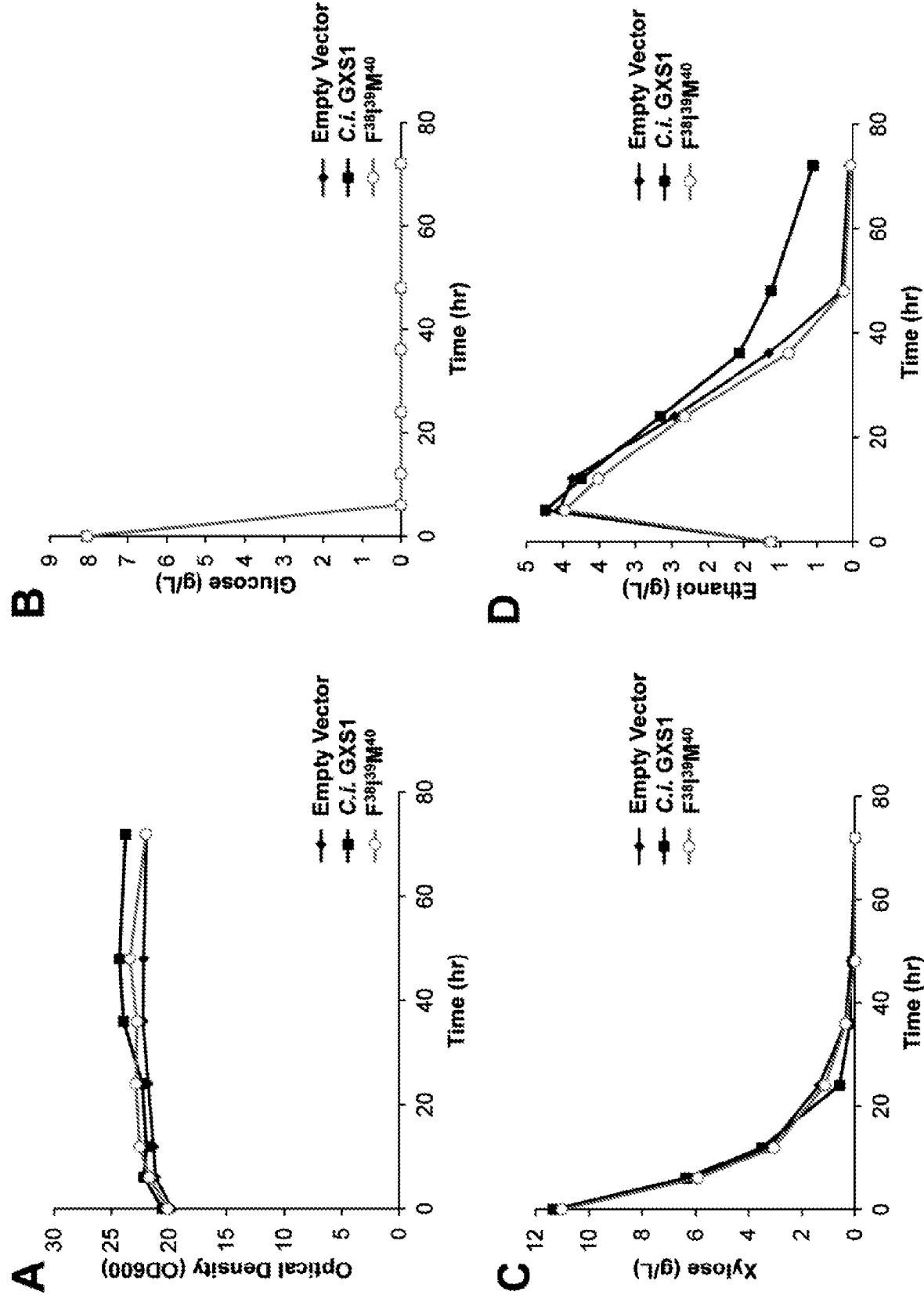
10/22

Figure 9



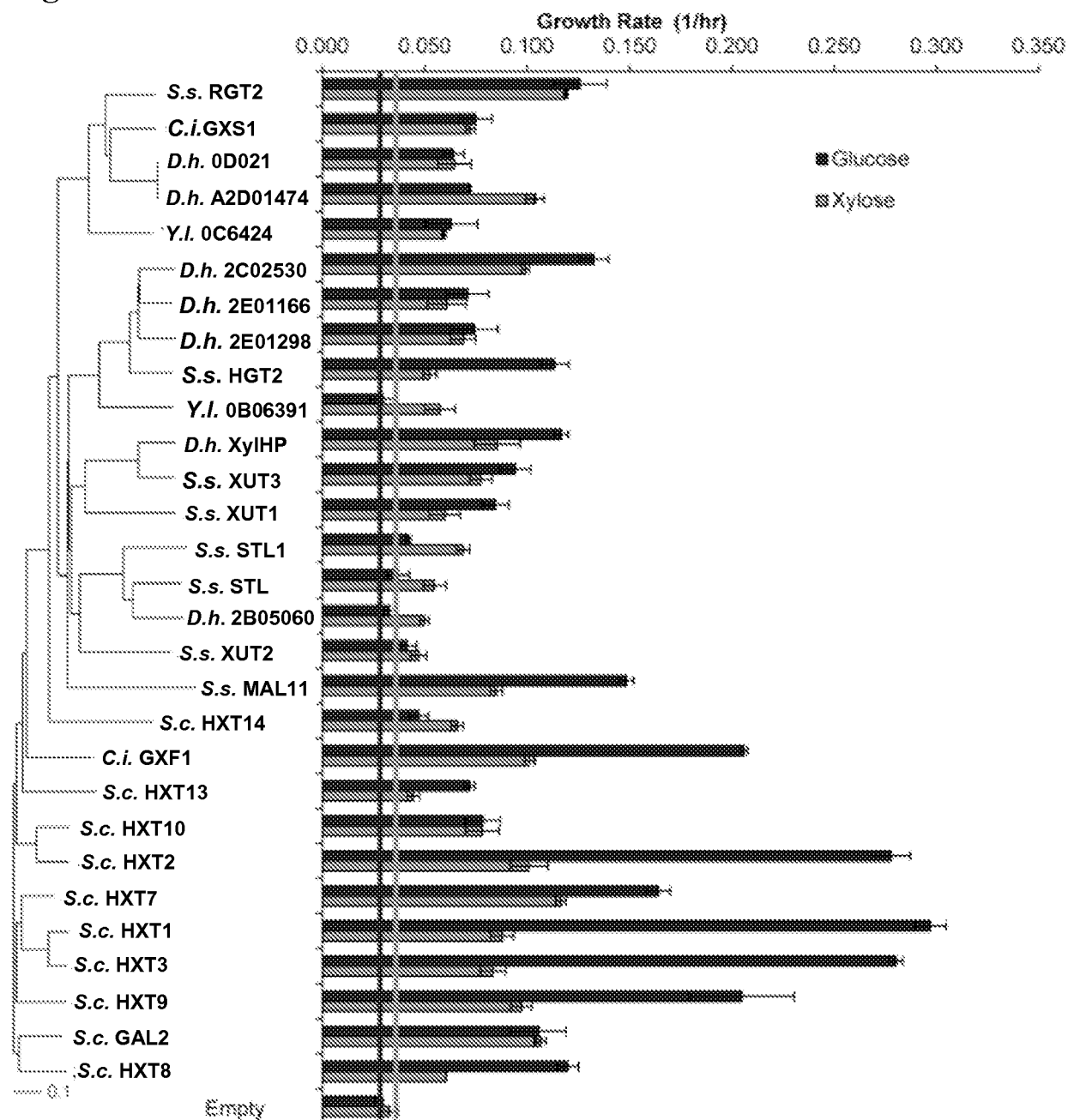
11/22

Figure 10



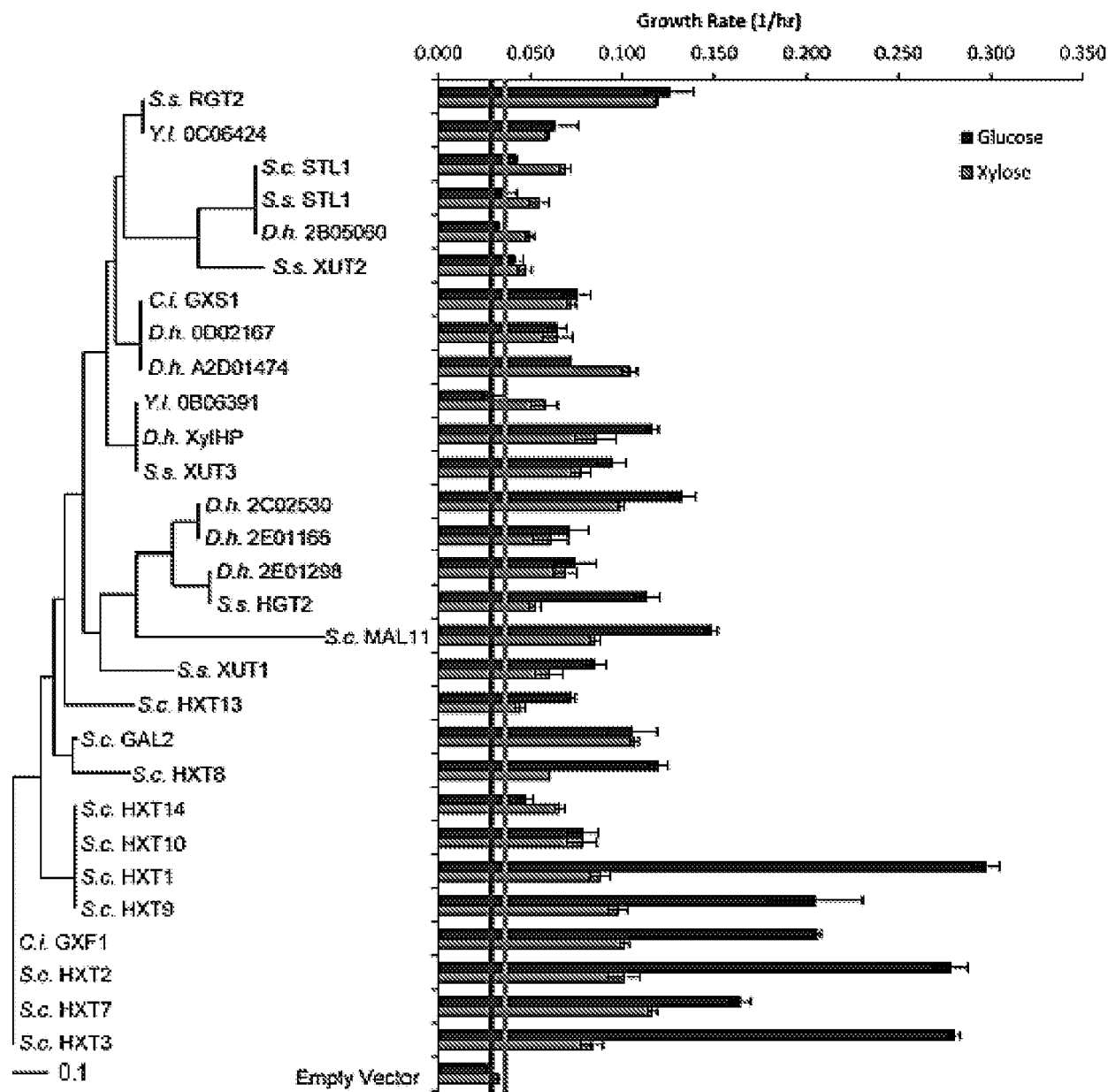
12/22

Figure 11



13/22

Figure 12



14/22

Figure 13A

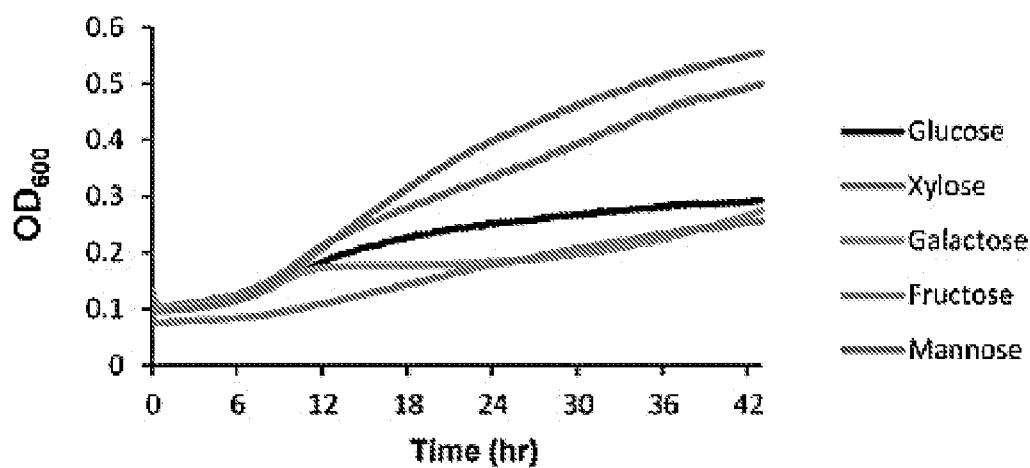
D.h. 2D01474

Figure 13B

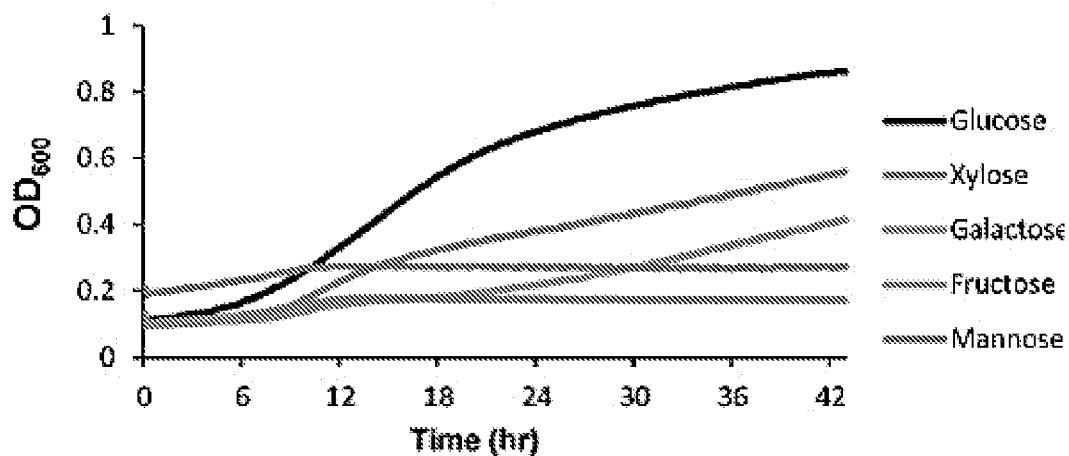
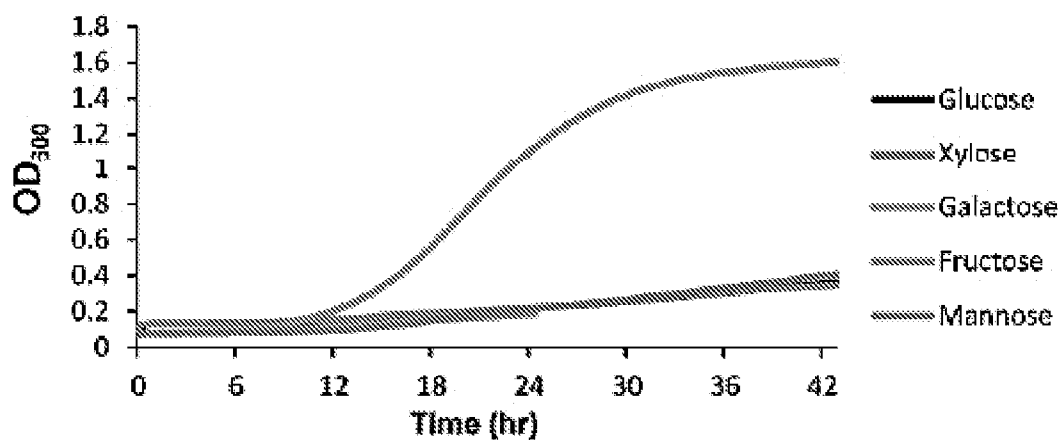
S.s. RGT2

Figure 13C

D.h. 2E01166

15/22

Figure 13D

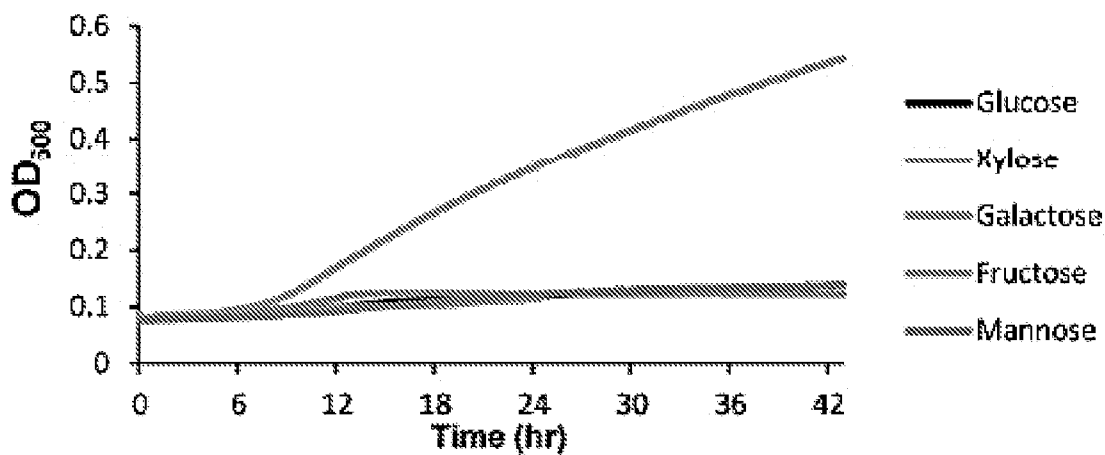
***D.h.* 2B05060**

Figure 13E

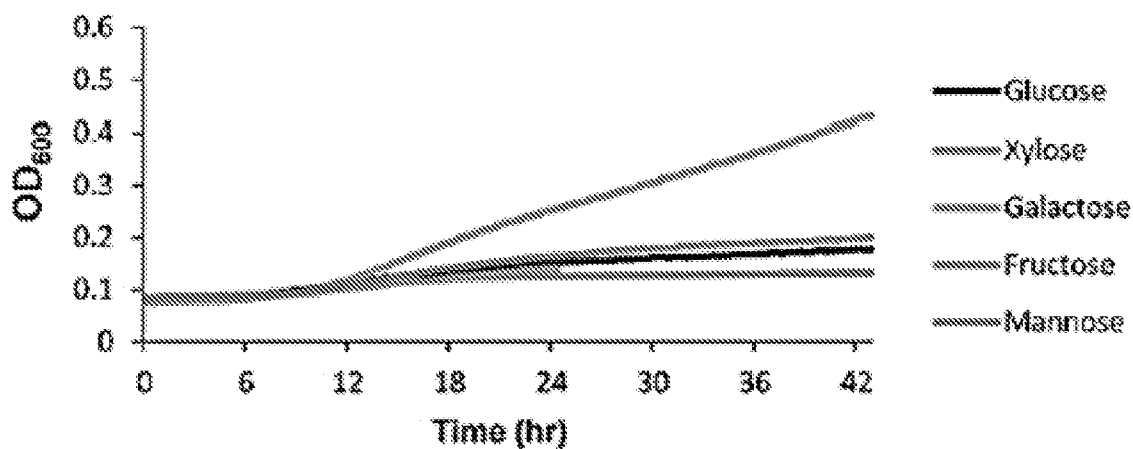
***S.c.* STL1**

Figure 13F

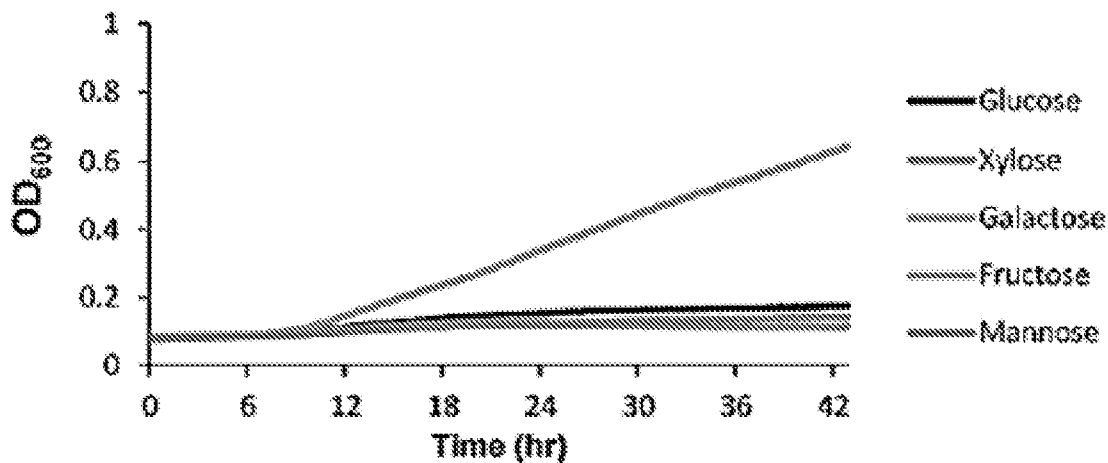
***S.s.* AUT1**

Figure 14

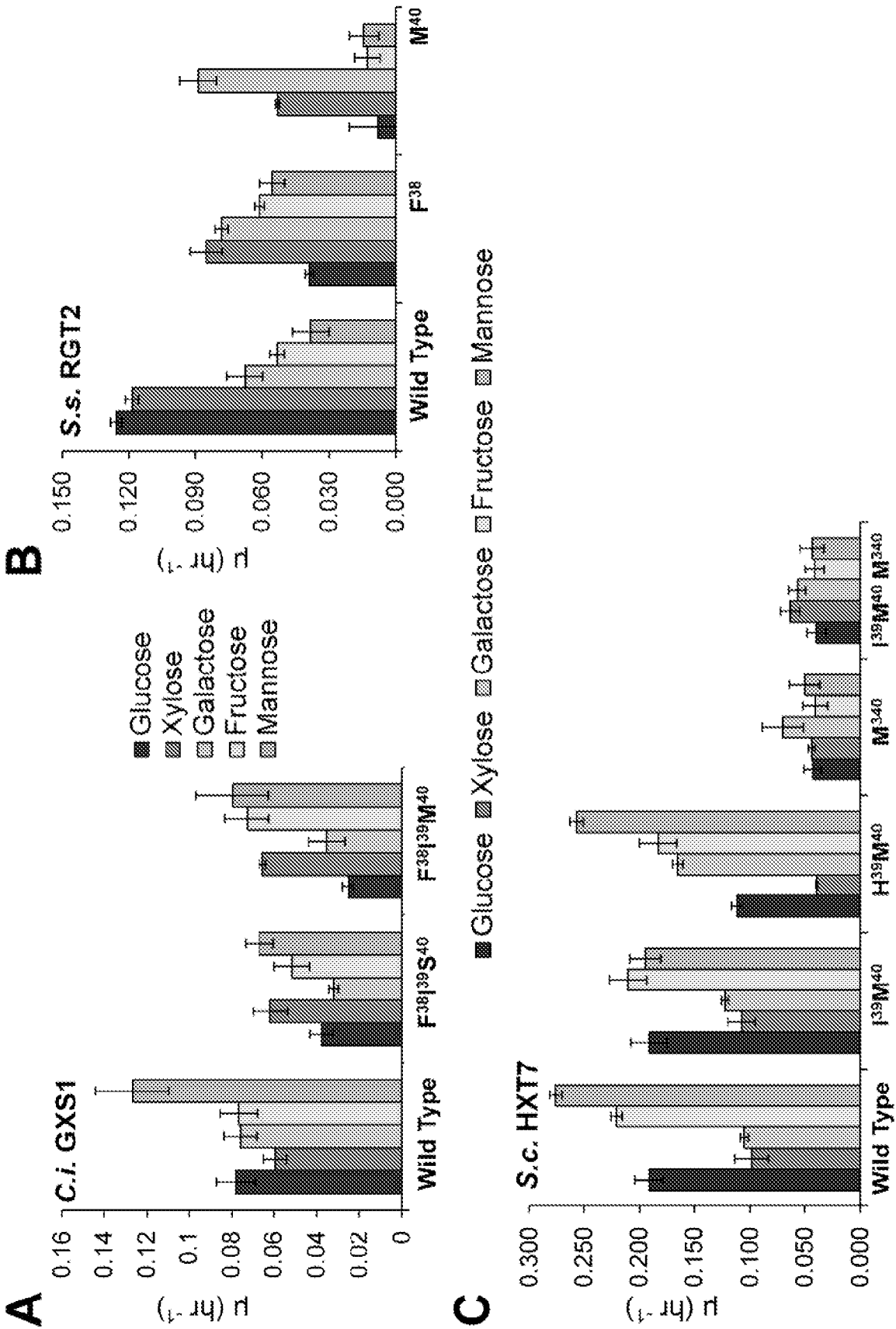
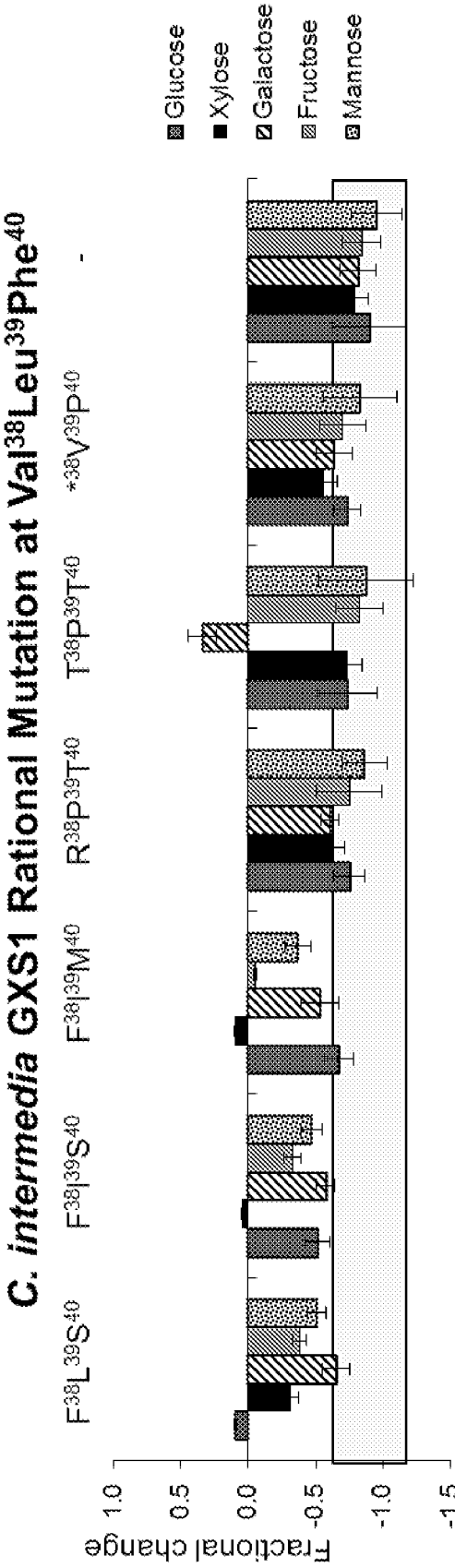
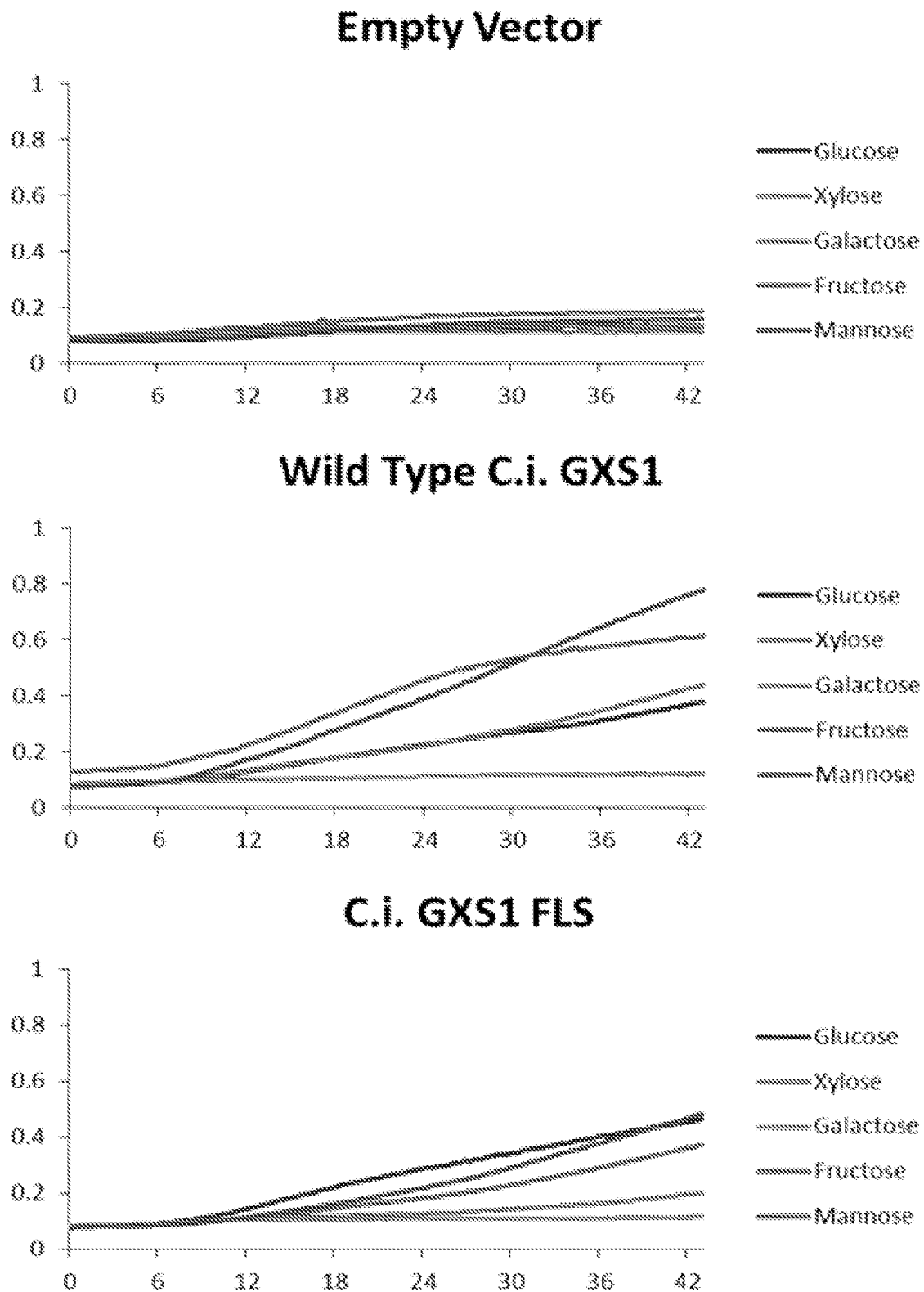


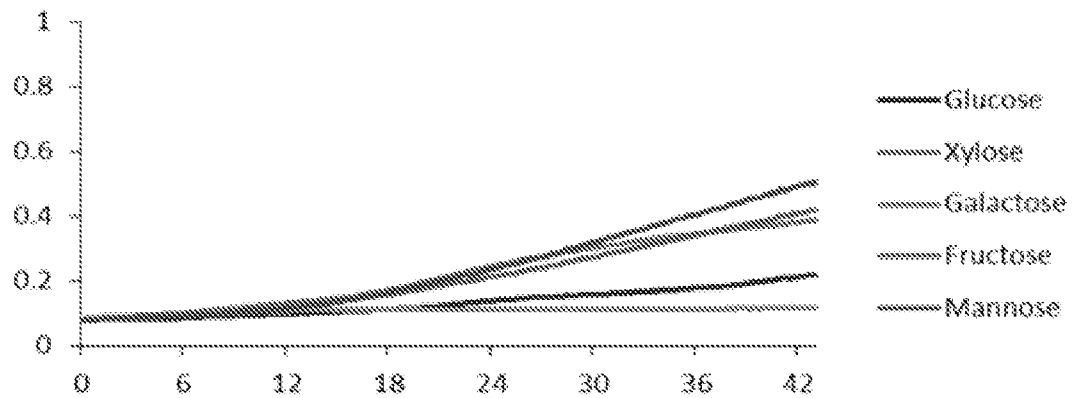
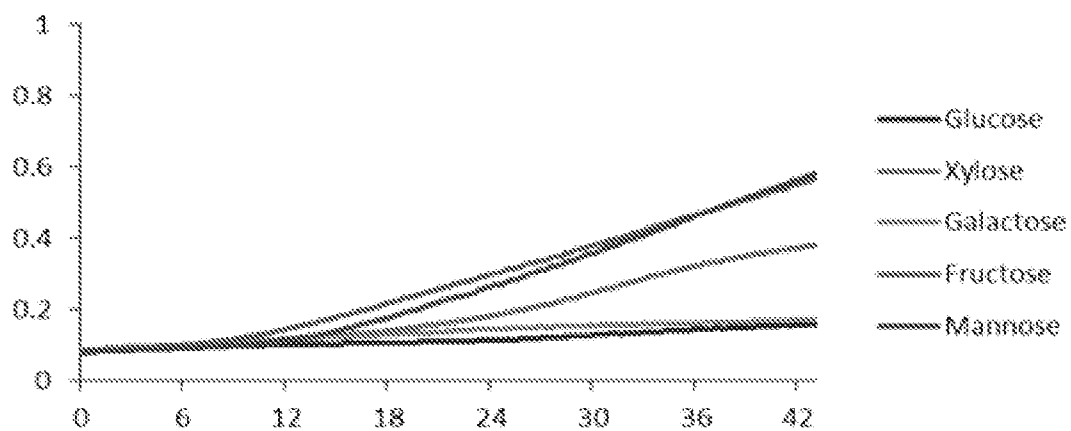
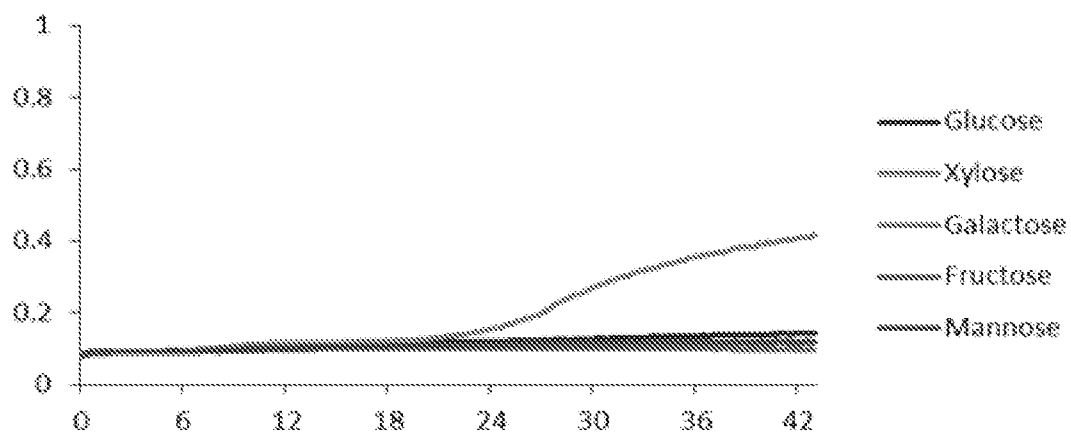
Figure 15



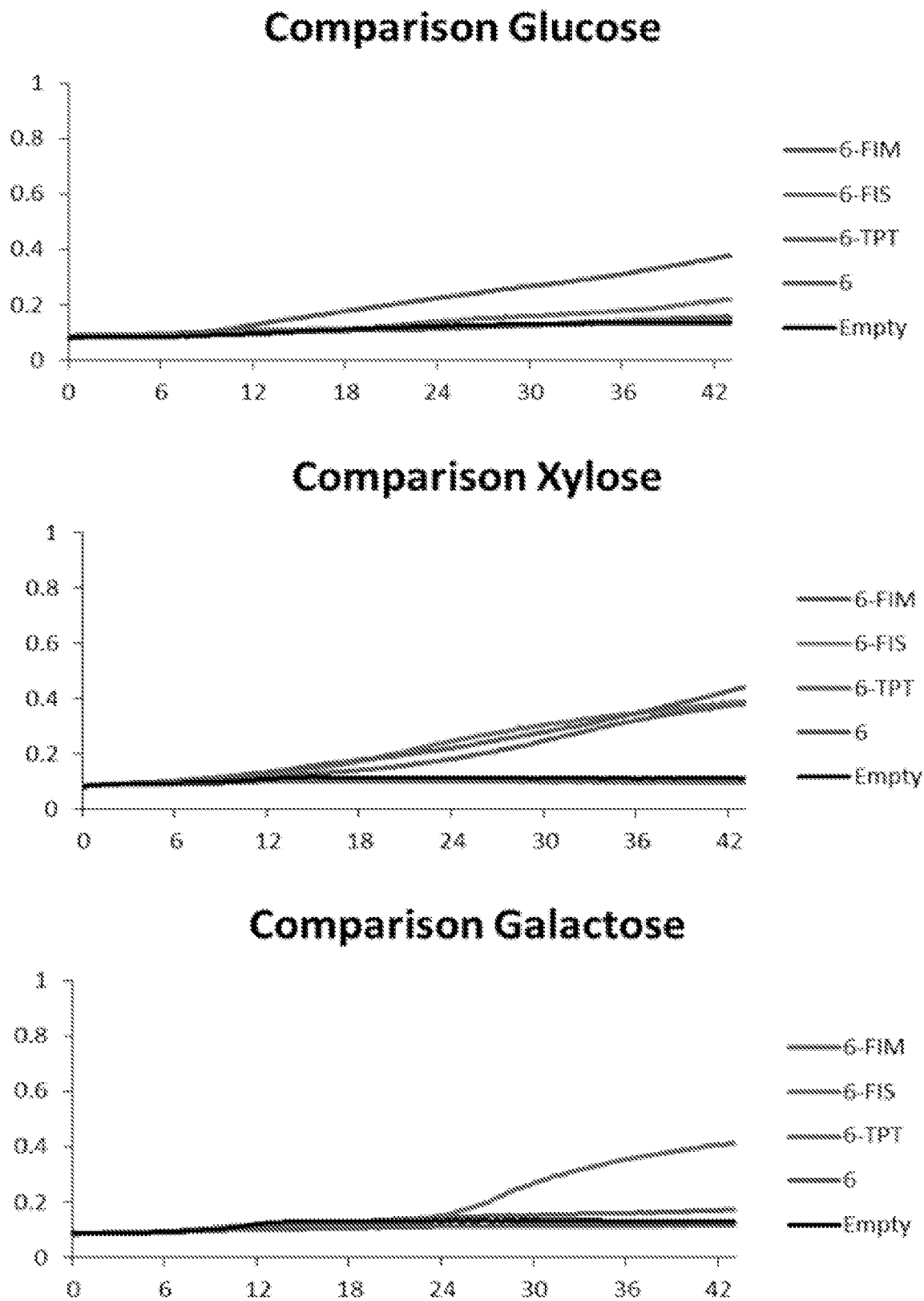
18/22

Figure 16

19/22

Figure 16 (cont.)**C.i. GXS1 FIS****C.i. GXS1 FIM****C.i. GXS1 TPT**

20/22

Figure 17

21/22

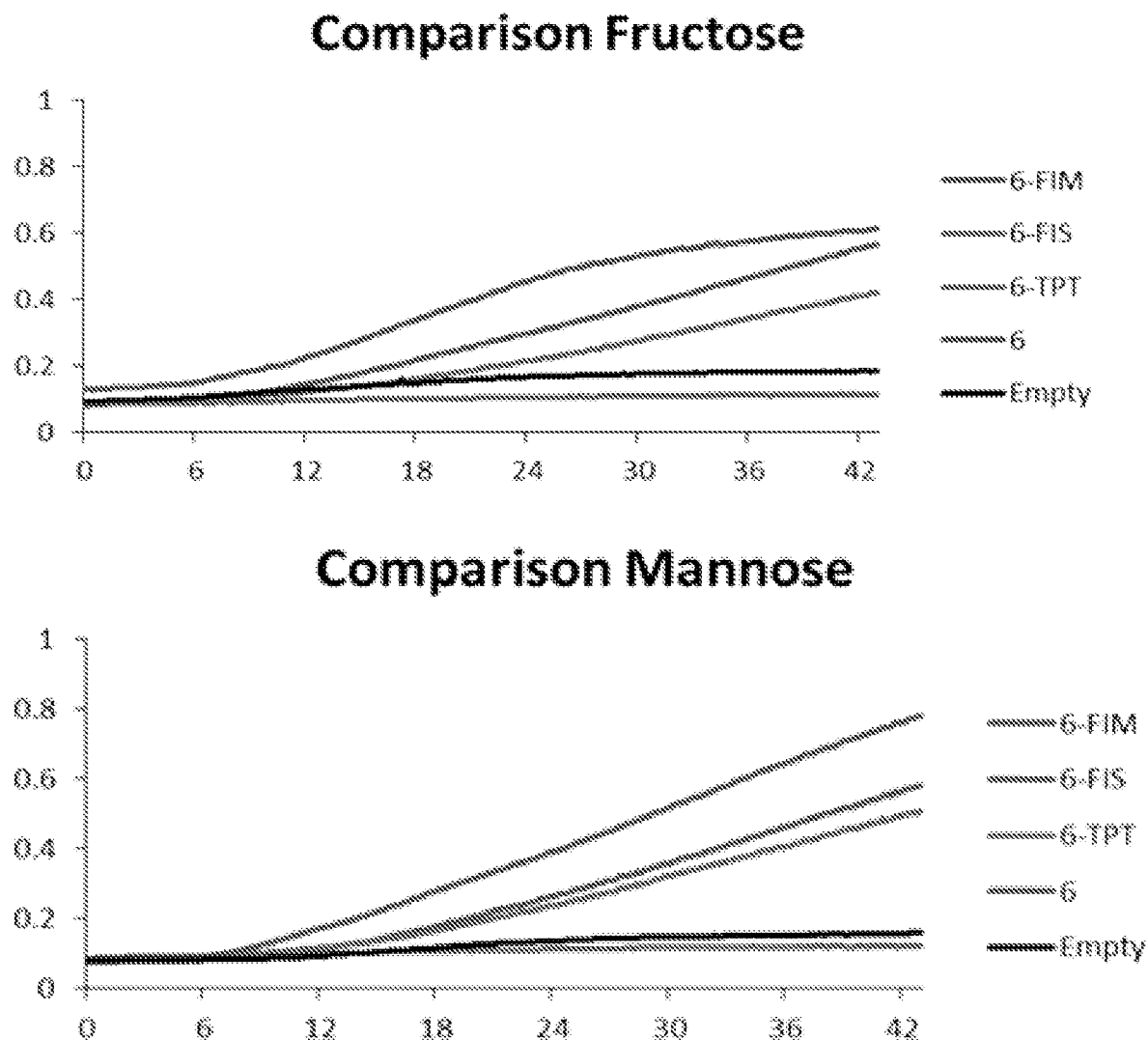
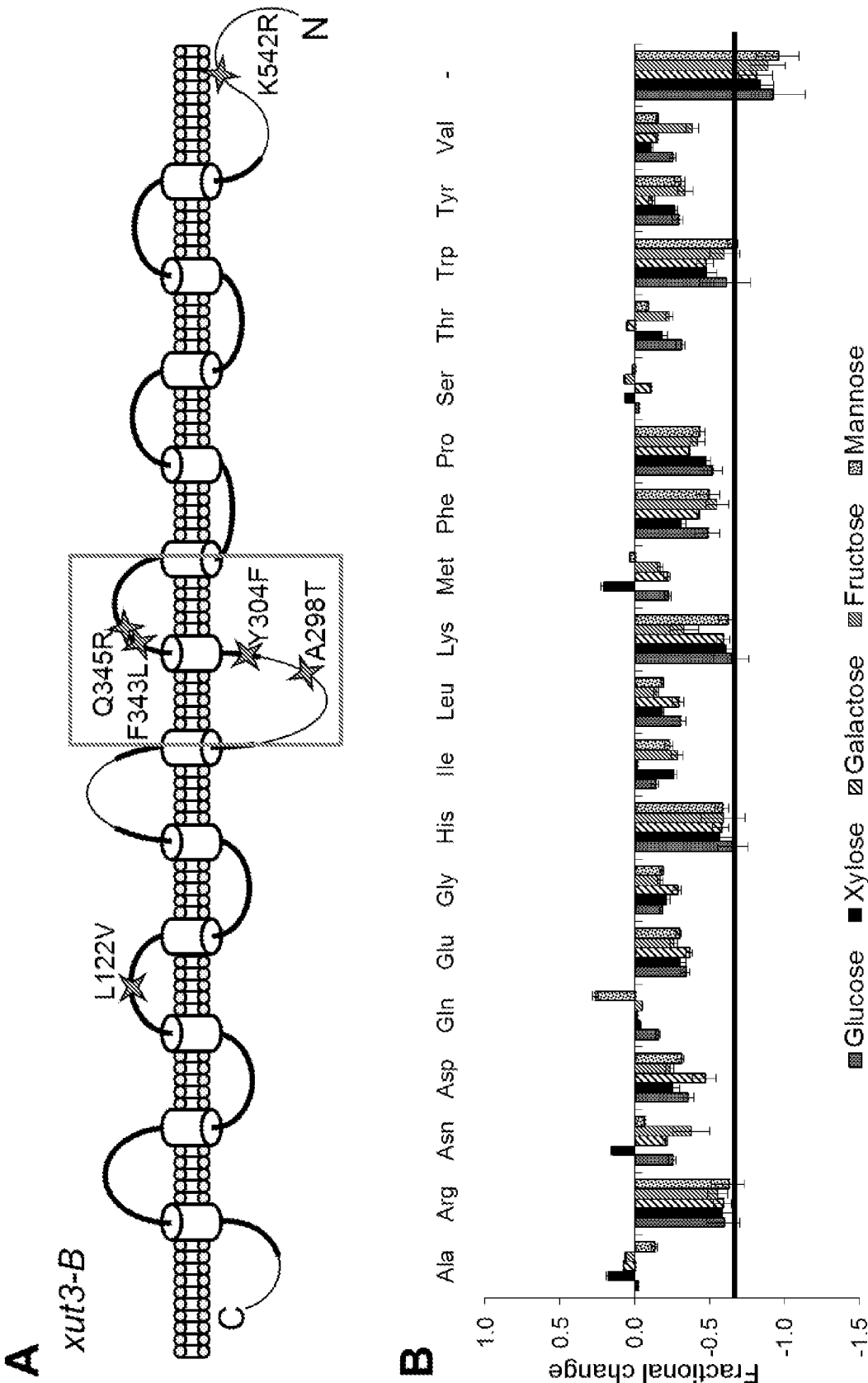
Figure 17 (cont.)

Figure 18



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 14/64168

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C07K 14/40, C12N 1/19, C12N 15/09, C12P 7/10 (2015.01)

CPC - C07K 14/40, Y02E 50/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C07K 14/40, C12N 1/19, C12N 15/09, C12P 7/10 (2015.01)

CPC - C07K 14/40, Y02E 50/16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - C07K 14/39, C12N 15/81, C12N 15/09

(keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, PubWEST (USPT, PGPB, EPAB, JPAB), Google Web, Google Scholar, NCBI Protein

Search terms: GX51, glucose-xylose symporter 1, glucose xylose symporter 1, glucose/xylose symporter 1, Candida intermedia, sugar, xylose, galactose, arabinose, monosaccharide, transport, transporter, motif, recombinant, engineered, substitution, mutation, lignocell

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YOUNG et al., A molecular transporter engineering approach to improving xylose catabolism in <i>Saccharomyces cerevisiae</i> . Metab Eng, July 2012, Vol 14, No 4, pp 401-411. Especially abstract; p 404, col 2, para 3.	1-36
A	CAI44932, GenBank Accession No. CAI44932, glucose/xylose symporter 1 [<i>Candida intermedia</i>], 24 October 2006 [online]. [Retrieved on 21 January 2015]. Retrieved from the internet <URL: http://www.ncbi.nlm.nih.gov/protein/CAI44932.1 > Entire document	1-36
A	US 2011/0020910 A1 (GLASS et al.) 27 January 2011 (27.01.2011) para [0053], [0225]	1-36
A	WO 2012097091 A2 (SALMON et al.) 19 July 2012 (19.07.2012) p 56, ln 28 – p 57, ln 4	1-36
X, P	YOUNG et al., Rewiring yeast sugar transporter preference through modifying a conserved protein motif. Proc Natl Acad Sci U S A, 7 January 2014, Vol 111, No 1, pp 131-136. Especially abstract; p 132, col 2, para 3 – p 133, col 2, para 3; p 133, Fig 3A, 3B, and 3C	1-36
A	US 2011/0195448 A1 (LIPPMEIER et al.) 11 August 2011 (11.08.2011) entire document	1-36
A	WO 2013155481 A1 (DOUDNA et al.) 17 October 2013 (17.10.2013) entire document	1-36

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

21 January 2015 (21.01.2015)

Date of mailing of the international search report

04 FEB 2015

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774