Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2886514 C 2021/04/13

(11)(21) 2 886 514

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2013/09/27
(87) Date publication PCT/PCT Publication Date: 2014/04/03
(45) Date de délivrance/Issue Date: 2021/04/13
(85) Entrée phase nationale/National Entry: 2015/03/27
(86) N° demande PCT/PCT Application No.: GB 2013/052531
(87) N° publication PCT/PCT Publication No.: 2014/049373
(30) Priorités/Priorities: 2012/09/27 (US61/706,338); 2012/11/15 (GB1220554.8)

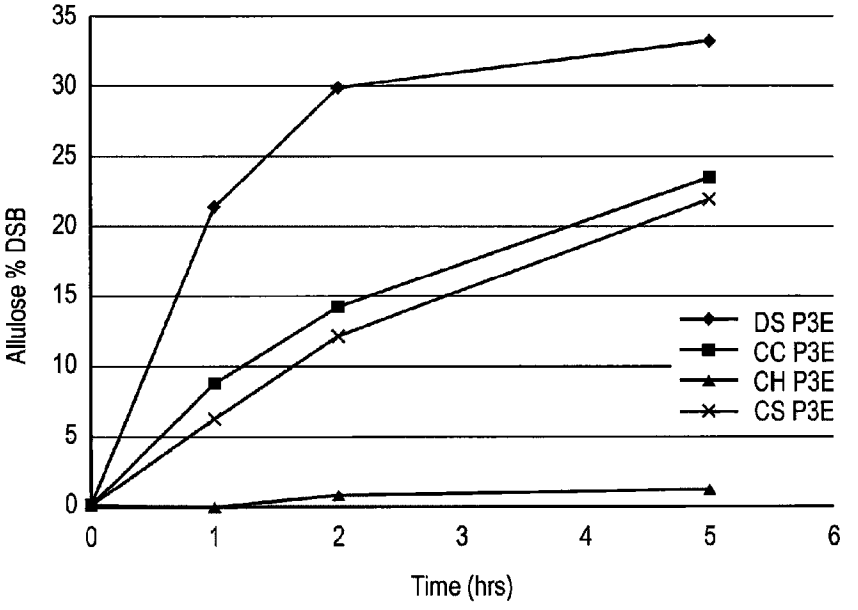
(51) Cl.Int./Int.Cl. *C12P 19/02* (2006.01), *C12N 9/90* (2006.01), *C12P 19/24* (2006.01)

(72) Inventeurs/Inventors:
WOODYER, RYAN DAVID, US;
ARMENTROUT, RICHARD W., US

(73) Propriétaire/Owner:
TATE & LYLE INGREDIENTS AMERICAS LLC, US

(74) Agent: MARKS & CLERK

(54) Titre : 3-EPIMERASE
(54) Title: 3-EPIMERASE



(57) Abrégé/Abstract:
A protein comprising a polypeptide sequence having at least 70% sequence identity to SEQ ID NO:6, SEQ ID NO:2or SEQ ID NO:4. The protein hasketose 3-epimerase activity.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2014 (03.04.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/049373 A1

(51) International Patent Classification:
C12P 19/02 (2006.01) *C12N 9/90* (2006.01)
C12P 19/24 (2006.01)

(21) International Application Number:
PCT/GB2013/052531

(22) International Filing Date:
27 September 2013 (27.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/706,338 27 September 2012 (27.09.2012) US
1220554.8 15 November 2012 (15.11.2012) GB

(71) Applicant: **TATE & LYLE INGREDIENTS AMERICAS LLC** [US/US]; 5450 Prairie Stone Parkway, Hoffman Estates, Illinois 60192 (US).

(71) Applicant *(for MG only)*: **TATE & LYLE TECHNOLOGY LIMITED** [GB/GB]; 1 Kingsway, London WC2B 6AT (GB).

(72) Inventors: **WOODYER, Ryan Daivd**; Carpentersville, Kane, Illinois (US). **ARMENTROUT, Richard W**; Carpentersville, Kane, Illinois (US).

(74) Agent: **ARENDS, William Gerrit**; Marks & Clerk LLP, 90 Long Acre, London, WC2E 9RA (GB).

(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, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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).

[Continued on next page]

(54) Title: 3-EPIMERASE

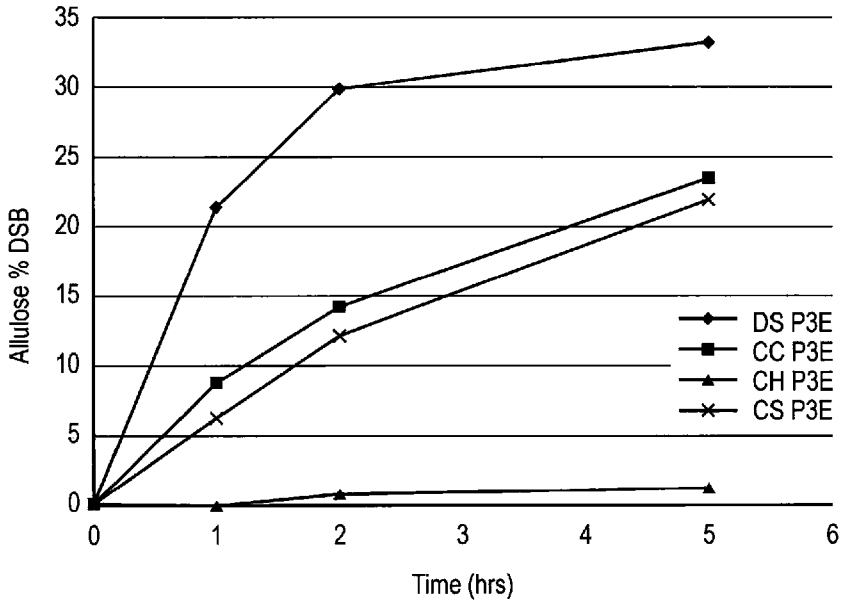


FIG. 6

(57) Abstract: A protein comprising a polypeptide sequence having at least 70% sequence identity to SEQ ID NO:6, SEQ ID NO:2or SEQ ID NO:4. The protein hasketose 3-epimerase activity.

WO 2014/049373 A1

WO 2014/049373 A1

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

3-EPIMERASE

Field of the Invention

The present invention relates to a protein having ketose 3-epimerase activity and a nucleic acid molecule encoding said protein. The invention also relates to a vector and a host cell comprising the nucleic acid molecule. The invention also relates to a method of synthesising allulose using the protein and to allulose produced in such a manner.

Background of the Invention

Allulose is a "zero-calorie" sweetener and has sweetness suggested to be similar to dextrose. It also has bulking and browning properties similar to those of other sugars. The primary target market for allulose is food and beverage manufacturers that currently use dextrose, fructose or HFCS in their products and that are looking to significantly reduce calories without significantly altering other properties imparted by the sugar component, for example, bulking, browning, texture and sweetness.

Allulose is not Generally Regarded As Safe (GRAS) in the United States but there is currently a GRAS notice pending (GRN400). Allulose is present in processed cane and beet molasses, steam treated coffee, wheat plant products and high fructose corn syrup. The typical total daily intake of allulose has been estimated to be greater than 0.2 grams per day. D-allulose is the C-3 epimer of D-fructose, and the structural difference between allulose and fructose results in allulose not being metabolized by the human body and thus having zero calories. Therefore, allulose is thought to be a promising candidate as a sweet bulking agent as it has no calories and is reported to be sweet while maintaining similar properties to typical monosaccharides.

Ketose-3-epimerases can interconvert fructose and allulose. US patent no. 8,030,035 and PCT publication no. WO2011/040708 disclose that D-psicose (an alternative name for allulose) can be produced by reacting a protein derived from *Agrobacterium tumefaciens,* and having psicose 3-epimerase activity, with D-fructose.

US patent publication no. 2011/0275138 discloses a ketose 3-epimerase derived from a microorganism of the *Rhizobium* genus. This protein shows a high specificity to D- or L-ketopentose and D- or L-ketohexose, and especially to D-fructose and D-psicose. This document also discloses a process for producing ketoses by using the protein.

Korean patent no. 100832339 discloses a Sinorhizobium YB-58 strain which is capable of converting fructose into psicose (i.e. allulose), and a method of producing psicose using a fungus body of the Sinorhizobium YB-58 strain.

Korean patent application no. 1020090098938 discloses a method of producing psicose using *E. coli* wherein the *E. coli* expresses a polynucleotide encoding a psicose 3-epimerase.

The present invention seeks to provide an improvement in the production of allulose over existing technology. The present invention seeks to provide a ketose-3-epimerase with higher rates of conversion and volumetric productivity in a whole cell system than previously reported.

Summary of the Invention

The present invention arises from the identification and characterisation of three ketose-3-epimerase enzymes, exemplary amino acid sequences of which are shown in SEQ. ID NOS. 2, 4 and 6. The ketose-3-epimerases may be used to convert fructose to allulose. These proteins had previously been identified as hypothetical proteins or as having tagatose epimerase activity. However, the present inventors have now surprisingly found that these enzymes have psicose-3-epimerase activity.

According to a first aspect of the present invention there is provided a protein comprising a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 2, SEQ ID NO: 4 or SEQ ID NO: 6, wherein the protein has ketose 3-epimerase activity.

Conveniently, the polypeptide sequence has at least 80%, 90%, 95% or 99% sequence identity, or has 100% sequence identity, to SEQ ID NO: 2, SEQ ID NO: 4 or SEQ ID NO: 6.

Advantageously, the polypeptide sequence comprises the sequence of SEQ ID NO: 13.

Preferably, the protein is immobilized on a solid substrate.

According to a second aspect of the present invention there is provided the use of a protein according to the first aspect of the invention for synthesizing allulose.

According to a third aspect of the present invention there is provided a nucleic acid molecule comprising a polynucleotide sequence encoding a protein according to the first aspect of the invention.

Advantageously, the nucleic acid molecule comprises a polynucleotide sequence which:

i) has at least 70%, 80%, 90%, 95% or 99% sequence identity, or has 100% sequence identity, to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3.

According to a fourth aspect of the present invention there is provided a vector comprising a nucleic acid molecule according to the third aspect of the invention.

According to a fifth aspect of the present invention there is provided a host cell comprising a recombinant nucleic acid molecule according to the third aspect of the invention.

Conveniently, the host cell is a yeast, bacterium or other microorganism, or is a mammalian, plant or other cell culture.

Preferably, the host cell is *E. coli*.

According to a sixth aspect of the present invention there is provided allulose produced by a protein according to the first aspect of the invention.

According to a seventh aspect of the present invention there is provided a method of producing allulose comprising:

i) providing a protein according to the first aspect of the invention; and

ii) contacting the protein with a fructose substrate under conditions such that the fructose substrate is converted into allulose. A method of producing allulose is also provided by the invention, the method comprising contacting a protein according to the first aspect of the invention with a fructose substrate under conditions such that the fructose substrate is converted to allulose.

Advantageously, the protein is present in a host cell.

Alternatively, the protein is in isolated form.

Conveniently, the conditions comprise maintaining the protein and the fructose substrate at a temperature between 25ºC and 75ºC, preferably between 50ºC and 60ºC, more preferably between 52ºC and 55ºC, more preferably 55ºC.

Preferably, the conditions comprise maintaining the protein and the fructose substrate between pH 4 and pH10.

Advantageously, the conditions comprise maintaining the fructose substrate concentration between 75% and 95% (W/V).

According to an eighth aspect of the present invention there is provided a nucleic acid molecule comprising a polynucleotide sequence which:

i) has at least 70% sequence identity to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3.

The nucleic acid molecule encodes a polypeptide having ketose 3-epimerase activity. The nucleic acid molecule may be in isolated form, according to one aspect of the invention.

According to a ninth aspect of the present invention there is provided a host cell comprising a recombinant nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide having ketose 3-epimerase activity, wherein the polynucleotide sequence:

i) has at least 70% sequence identity to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3.

According to a tenth aspect of the present invention there is provided a vector comprising a nucleic acid molecule comprising a polynucleotide encoding a polypeptide having ketose 3-epimerase activity, wherein the polynucleotide sequence:

i) has at least 70% sequence identity to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3.

According to an eleventh aspect of the present invention there is provided a method of producing allulose comprising the steps of:

i) providing a vector comprising a nucleic acid molecule having a polynucleotide sequence encoding a protein having ketose 3-epimerase activity wherein the polynucleotide sequence: a) has at least 70% sequence identity to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or b) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3;

ii) synthesising the protein having ketose 3-epimerase activity encoded by the polynucleotide sequence;

iii) contacting fructose with the protein having ketose 3-epimerase activity and maintaining the fructose and protein under conditions to permit the conversion of fructose to allulose; and

iv) at least partially purifying the allulose produced in step iii).

According to one aspect of the present invention there is provided a protein comprising a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 6, SEQ ID NO: 2 or SEQ ID NO: 4, wherein the protein has ketose 3-epimerase activity and the protein does not comprise a polypeptide sequence which is identical to SEQ ID NO: 6, SEQ ID NO: 2 or SEQ ID NO: 4.

According to a further aspect of the present invention there is provided use of a protein for synthesizing allulose, wherein the protein comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 6, SEQ ID NO: 2 or SEQ ID NO: 4, and has psicose 3-epimerase activity.

According to another aspect of the present invention there is provided a nucleic acid molecule comprising a polynucleotide sequence encoding a protein as described herein.

According to yet another aspect of the present invention there is provided a nucleic acid molecule as described herein, comprising a polynucleotide sequence which:

i) has at least 70%, 80%, 90%, 95% or 99% sequence identity, to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3,

wherein the nucleic acid molecule does not comprise a polynucleotide sequence which is identical to SEQ ID NO: 5, SEQ ID NO: 1 or SEQ ID NO: 3.

According to still another aspect of the present invention there is provided a host cell comprising a recombinant nucleic acid molecule as described herein.

According to a further aspect of the present invention there is provided allulose produced by a protein as described herein.

According to another aspect of the present invention a method of producing allulose comprising contacting a protein with a fructose substrate under conditions such that the fructose substrate is converted into allulose, wherein the protein comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 6, SEQ ID NO: 2 or SEQ ID NO: 4,and wherein the protein has psicose 3-epimerase activity.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also apply to amino acid polymers in which one or more amino acid residues is a modified residue, or a non-naturally occurring residue, such as an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The polypeptide may or may not be "isolated", that is to say removed from the components which exist around it when naturally occurring.

The term "amino acid" as used herein refers to naturally occurring and synthetic amino acids, as well as amino acid analogues and amino acid mimetics that have a function that is similar to naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those modified after translation in cells (e.g. hydroxyproline, gamma-carboxyglutamate, and O-phosphoserine). The phrase "amino acid analogue" refers to compounds that have the same basic chemical structure (an alpha carbon bound to a hydrogen, a carboxy group, an amino group, and an R group) as a naturally occurring amino acid but have a modified R group or modified backbones (e.g. homoserine, norleucine, methionine sulfoxide, methionine methyl sulphonium). The phrase "amino acid mimetic" refers to chemical compounds that have different structures from, but similar functions to, naturally occurring amino acids. It is to be appreciated that, owing to the degeneracy of the genetic code, nucleic acid molecules encoding a particular polypeptide may have a range of polynucleotide sequences. For example, the codons GCA, GCC, GCG and GCT all encode the amino acid alanine.

The percentage "identity" between two sequences may be determined using the BLASTP algorithm version 2.2.2 (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids

Res. 25:3389-3402) using default parameters. In particular, the BLAST algorithm can be accessed on the internet.

The term "ketose-3-epimerase activity" as used herein means that an enzyme is capable of catalyzing the inversion of stereochemistry of ketoses, in particular the conversion of fructose to allulose. For example, in one embodiment, "ketose-3-epimerase activity" is defined as being the capacity of an enzyme to increase the rate of interconversion of fructose to allulose by at least 10 micromol/min per mg of added enzyme (0.1 U/mg) over a reaction mixture under the same conditions in the absence of the enzyme. In alternative embodiments an increase in rate of interconversion of fructose to allulose of at least 0.05 U/mg or 0.2 U/mg is is considered to be "ketose-3-epimerase activity". A suitable assay for determining the activity of an enzyme in converting D-fructose into allulose is as follows. A reaction mixture comprising 1ml D-fructose (50g/L), Tris-HCL buffer (50mM, pH 8.0), and 0.5µM enzyme is incubated at 55°C for 2 minutes. The reaction is stopped after 10 minutes by boiling. The amount of D-allulose produced is determined by the HPLC method. One unit of enzyme activity is defined as the amount of enzyme catalysing the formation of 1µmol of D-allulose/min at pH 8.0 and 55°C (J. Agric. Food Chem. 2011, 59, 7785-7792).

The terms "gene", "polynucleotides", and "nucleic acid molecules" are used interchangeably herein to refer to a polymer of multiple nucleotides. The nucleic acid molecules may comprise naturally occurring nucleic acids (i.e. DNA or RNA) or may comprise artificial nucleic acids such as peptide nucleic acids, morpholin and locked nucleic acids as well as glycol nucleic acids and threose nucleic acids.

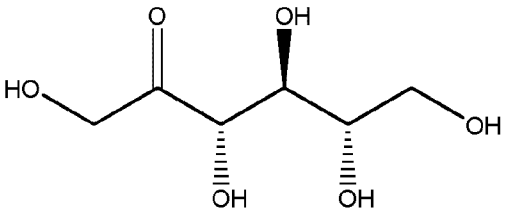
The term "nucleotide" as used herein refers to naturally occurring nucleotides and synthetic nucleotide analogues that are recognised by cellular enzymes.

The term "vector" as used herein refers to any natural or artificial construct containing a nucleic acid molecule in which the nucleic acid molecule can be subject to cellular transcription and/or translation enzymes. Exemplary vectors include: a plasmid, a virus (including bacteriophage), a cosmid, an artificial chromosome or a transposable element.

The term "host cell" as used herein refers to any biological cell which can be cultured in medium and used for the expression of a recombinant gene. Such host cells may be eukaryotic or prokaryotic and may be a microorganism such as a bacterial cell, or may be a cell from a cell line (such as an immortal mammalian cell line).

The term "highly stringent conditions" as used herein when referring to hybridization conditions means: at least about 6X SSC and 1% SDS at 65ºC, with a first wash for 10 minutes at about 42ºC with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65ºC. It is known in the art that hybridization techniques using a known nucleic acid as a probe under highly stringent conditions, such as those set forth in the specification, will identify structurally similar nucleic acids.

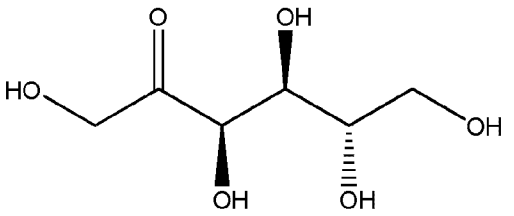
The term "allulose" as used herein refers to a monosaccharide sugar of the structure shown in Formula I. It is also known as "D-Psicose".



Formula (I)

The term "fructose" as used herein refers to a monosaccharide sugar having the structure shown in Formula II. Examples of fructose substrate include but are not limited to crystalline fructose and crystalline fructose greens. As used herein, "crystalline fructose greens" refers to a process stream created during fructose crystallization from the non-crystallizing portion of the crystallization mother liquor.

Formula (II)



The term "recombinant" as used herein refers to a nucleic acid molecule or a polypeptide which is located in a non-naturally occurring context and which has been produced by artificial intervention. For example, a first polypeptide isolated from other polypeptides or linked by a peptide bond to a second polypeptide sequence having a different amino acid sequence from any polypeptide with which the first polypeptide is associated in nature is a recombinant polypeptide.

Brief Description of the Figures

Figure 1 shows the amino acid sequence of the ketose-3-epimerase from *Clostridium scindens* in accordance with one embodiment of the present invention (SEQ ID NO: 2).

Figure 2 shows the amino acid sequence of the ketose-3-epimerase from *Clostridium hylemonae* in accordance with another embodiment of the present invention (SEQ ID NO: 4).

Figure 3 shows the amino acid sequence of the ketose-3-epimerase from *Desmospora sp.* in accordance with a further embodiment of the present invention (SEQ ID NO: 6).

Figure 4 shows the amino acid sequence of a previously known xylose isomerase from *Clostridium cellulolyticum* (SEQ ID NO: 8).

Figure 5 shows a sequence comparison between the three ketose-3-epimerases shown in Figures 1 to 3 and three previously known ketose-3-epimerases. Completely conserved residues are highlighted.

Figure 6 is a graph showing the rate of conversion of fructose to allulose by *E. coli* transformed to express an enzyme in accordance with an embodiment of the present invention and a control.

Figure 7 shows the optimised gene sequence (SEQ ID NO: 7) encoding the amino acid sequence shown in Figure 4 and a comparison of the optimised sequence with the original sequence.

Figure 8 shows the optimised gene sequence (SEQ ID NO: 5) encoding the amino acid sequence shown in Figure 3, and a comparison of the optimised sequence with the original sequence.

Figure 9 shows the optimised gene sequence (SEQ ID NO: 1) encoding the amino acid sequence shown in Figure 1, and a comparison of the optimised sequence with the original sequence.

Figure 10 shows the optimised gene sequence (SEQ ID NO: 3) encoding the amino acid sequence shown in Figure 2, and a comparison of the optimised sequence with the original sequence.

Figure 11 is a graph showing the preparation of conversion of fructose substrate to allulose by *E. coli* transformed to express an enzyme in accordance with one embodiment of the present invention (the ketose-3-epimerase from *Desmospora sp*) at 18L scale.

Figure 12 is a graph showing the rate of allulose conversion by enzymes according to embodiments of the invention (CH P3E, CS P3E and DS P3E) and a known ketose-3-epimerase (CC P3E).

Figure 13 is a graph showing the rate of allulose conversion by DS P3E in a 30ml fixed bed reactor packed with A568 resin.

Figure 14 is a graph showing the rate of allulose conversion by DS P3E in a 300ml fixed bed reactor packed with A568 resin.

Figure 15 shows the amino acid sequence of an artificial variant of the naturally occurring ketose-3-epimerase from *Desmospora sp.* (SEQ ID NO. 13).

Brief Description of the Sequence Listing

SEQ ID NO: 1 shows a gene sequence (optimised for expression in *E. coli*) encoding a ketose-3-epimerase from *Clostridium scindens*.

SEQ ID NO: 2 shows the amino acid sequence of the ketose-3-epimerase encoded by the gene sequence of SEQ ID NO. 1.

SEQ ID NO: 3 shows a gene sequence (optimised for expression in *E. coli*) encoding a ketose-3-epimerase from *Clostridium hylemonae*.

SEQ ID NO: 4 shows the amino acid sequence of the ketose-3-epimerase encoded by the gene sequence of SEQ ID NO. 3.

SEQ ID NO: 5 shows a gene sequence (optimised for expression in *E. coli*) encoding a ketose-3-epimerase from *Desmospora sp.* 8437.

SEQ ID NO: 6 shows the amino acid sequence of the ketose-3-epimerase encoded by the gene sequence of SEQ ID NO. 5.

SEQ ID NO: 7 shows a gene sequence (optimised for expression in *E. coli*) encoding a ketose-3-epimerase from *Clostridium cellulolyticum*.

SEQ ID NO: 8 shows the amino acid sequence of the ketose-3-epimerase encoded by the gene sequence of SEQ ID NO. 7.

SEQ ID NO: 9 shows the naturally occurring gene sequence encoding the ketose-3-epimerase from *Clostridium scindens*.

SEQ ID NO: 10 shows the naturally occurring gene sequence encoding the ketose-3-epimerase from *Clostridium hylemonae*.

SEQ ID NO: 11 shows the naturally occurring gene sequence encoding the ketose-3-epimerase from *Desmospora sp.* 8437.

SEQ ID NO: 12 shows the naturally occurring gene sequence encoding the ketose-3-epimerase from *Clostridium cellulolyticum.*

SEQ ID NO: 13 shows the amino acid sequence of an artificial variant of the ketose-3-epimerase of *Desmospora sp.* 8437.

Detailed Description

The present invention relates, in general terms, to a protein comprising a polypeptide having an amino acid sequence shown in one of SEQ. ID NO. 2, 4 or 6. The source organisms of the polypeptides of SEQ. ID NOS. 2, 4 and 6 are shown in Table 1.

Table 1

Source Organism	SEQ. ID NO.
Clostridium scindens ATCC 35704	2
Clostridium hylemonae DSM 15053	4
*Desmospora sp.*8437	6

However, in alternative embodiments, the polypeptide sequence is not identical to that shown in SEQ. ID NOs. 2, 4 or 6 but has at least 70% sequence identity thereto. It is

preferred that the polypeptide sequence has at least 80%, 90%, 95% or 99% sequence identity, or 100% sequence identity, to SEQ. ID NO. 2, 4 or 6.

For example, in one embodiment the polypeptide sequence comprises the sequence of SEQ. ID NO. 13 which has 89% sequence identity to SEQ. ID NO. 6. The polypeptide sequence has ketose 3-epimerase activity.

Thus in some embodiments, one or more amino acids of the peptides are omitted or are substituted for a different amino acid, preferably a similar amino acid. A similar amino acid is one which has a side chain moiety with related properties and the naturally occurring amino acids may be categorized into the following groups. The group having basic side chains: lysine, arginine, histidine. The group having acidic side chains: aspartic acid and glutamic acid. The group having uncharged polar side chains: aspargine, glutamine, serine, threonine and tyrosine. The group having non-polar side chains: glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan and cysteine. Therefore it is preferred to substitute amino acids within these groups.

It is generally preferred that the polypeptide conforms with the chemistry of naturally occurring polypeptides (although it may be synthesized *in vitro*) but in some alternative embodiments the polypeptide is a peptidomimetic, that is to say a modification of a polypeptide in a manner that will not naturally occur. Such peptidomimetics include the replacement of naturally occurring amino acids with synthetic amino acids and/or a modification of the polypeptide backbone. For example in some embodiments, the peptide bonds are replaced with a reverse peptide bond to generate a retro-inverso peptidomimetic (see Méziére *et al* J Immunol. 1997 Oct 1;159(7):3230-7) Alternatively, the amino acids are linked by a covalent bond other than a peptide bond but which maintains the spacing and orientation of the amino acid residues forming the polymer chain.

All such modified and unmodified polypeptides of the invention have ketose-3-epimerase polymerase activity. That is to say, the protein, when purified or expressed in a host cell, has the capacity to catalyze the conversion of fructose to allulose. Suitable conditions for testing the presence of ketose-3-epimerase activity are shown in Example 1.

The polypeptide of the invention may be contained within a whole cell or may be an isolated protein, a partially purified protein or an immobilized protein. Purification of the protein may be by standard methods such as cell disruption and filtration. Other standard methods are known to those skilled in the art.

In embodiments of the present invention, there is provided a nucleic acid molecule which comprises a polynucleotide sequence encoding a protein having an amino acid sequence with at least 70% sequence identity to SEQ. ID NO. 2, 4 or 6, where the protein has ketose-3-epimerase activity. For example, in one embodiment the nucleic acid molecule comprises a sequence encoding the polypeptide sequence of SEQ. IN NO. 13.

In addition to the sequence specifically encoding the protein of the invention, the nucleic acid molecule may contain other sequences such as primer sites, transcription factor binding sites, vector insertion sites and sequences which resist nucleolytic degradation (e.g. polyadenosine tails). The nucleic acid molecule may be DNA or RNA and may include synthetic nucleotides, provided that the polynucleotide is still capable of being translated in order to synthesize a protein of the invention.

As described above, the amino acid sequence of the protein of the present invention may differ from the specific sequences disclosed herein. In preferred embodiments, the nucleic acid molecule comprises a polynucleotide having the sequence of SEQ. ID NO. 1, 3 or 5, which has been optimised by expression in *E. coli* host cells. In alternative embodiments, the polynucleotide sequence has at least 70% sequence identity to any one of SEQ. ID NO. 1, 3 or 5 and encodes a protein which has ketose-3-epimerase activity. It is preferred that the polynucleotide sequence has at least 80%, 90%, 95% or 99% sequence identity, or 100% sequence identity, to one of SEQ. ID NO. 1, 3 or 5. In alternative embodiments the nucleic acid molecule comprises a polynucleotide sequence which hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 1, 3 or 5 and which encodes a protein which has ketose-3-epimerase activity. In some embodiments, there is provided a nucleic acid molecule comprising a polynucleotide having the sequence of SEQ. ID NO. 9, 10 or 11, which are the naturally occurring sequences of the enzymes.

In some embodiments, the nucleic acid molecule forms part of a vector such as a plasmid. In addition to the nucleic acid sequence described above, the plasmid comprises other elements such as a prokaryotic origin of replication (for example, the *E. coli* OR1 origin of replication) an autonomous replication sequence, a centromere sequence; a promoter sequence, upstream of the nucleic acid sequence, a terminator sequence located downstream of the nucleic acid sequence, an antibiotic resistance gene and/or a secretion signal sequence. A vector comprising an autonomous replication sequence is also a yeast artificial chromosome.

In some alternative embodiments, the vector is a virus, such as a bacteriophage and comprises, in addition to the nucleic acid sequence of the invention, nucleic acid sequences for replication of the bacteriophage, such as structural proteins, promoters, transcription activators and the like.

The nucleic acid molecule of the invention may be used to transfect or transform host cells in order to synthesize the protein of the invention. Suitable host cells include prokaryotic cells such as *E. coli* and eukaryotic cells such as yeast cells, or mammalian or plant cell lines. Host cells are transfected or transformed using techniques known in the art such as electroporation; calcium phosphate base methods; a biolistic technique or by use of a viral vector.

After transfection, the nucleic acid molecule of the invention is transcribed as necessary and translated. In some embodiments, the synthesized protein is allowed to remain in the host cell and cultures of the recombinant host cell are subsequently used. In other embodiments, the synthesized protein is extracted from the host cell, either by virtue of its being secreted from the cell due to, for example, the presence of secretion signal in the vector, or by lysis of the host cell and purification of the protein therefrom.

The protein of the present invention is used to catalyze the conversion of fructose to allulose. In some embodiments, the protein is present in host cells and is mixed, to form a conversion mixture, with a fructose substrate, such as borate buffered fructose substrate, at a concentration from 1 to 1000g/L under suitable conditions, such as incubation at a

temperature from 25°C to 75°C, pH from 4 to 10. The conversion mixture may also comprise a solvent and optionally additional co-solvents (in addition to water) for example ethanol, toluene and methanol. The fructose substrate may also contain other sugars such as glucose or sucrose. The protein catalyzes a conversion of the fructose substrate to allulose. In practice, not all fructose in the conversion mixture is converted to allulose so there is typically a subsequent step of extracting and purifying the allulose through evaporation and crystallisation. Residual fructose in the mixture may be removed by yeast fermentation.

In alternative embodiments, the protein of the present invention is provided in purified form and mixed with a fructose substrate together with suitable solvent for an entirely *in vitro* conversion. In one embodiment the conditions are pH 4-10, a temperature between 30°C and 70°C and a fructose concentration of 10-95% w/v, with water as the solvent. Alternative concentration ranges for fructose include but are not limited to 20-95%, 30-95%, 40-95%, 50-95%, 60-95%, 70-95%, 75-95%. It is particularly preferred that the fructose substrate is provided at a concentration between 70 and 95%. In other preferred embodiments, the fructose concentration is 75-95%.

The conversion reaction of ketoses is usually carried out using a substrate concentration of 1-60% (w/v), preferably about 5-50%. It is a particular advantage of the present invention that the protein can be used in ketose conversion reactions under usual operating conditions but with higher fructose concentrations than have previously been used. Thus, a greater volumetric productivity is possible with the protein of the present invention.

In some embodiments, the protein of the present invention is immobilised on a solid substrate. This provides the advantage that the enzyme has a longer usage life, can be packed in a smaller fixed bed reactor, and has greater tolerance to contaminants and to fluctuations in the conditions of the process. Exemplary solid substrates include ion exchange resins and polymer encapsulations. In some embodiments, the protein of the present invention is immobilised on Duolite® A568 resin. In some embodiments, the protein of the invention is immobilised on a substrate by weakly basic ion exchange (i.e. electrostatic interaction based on the charge of the protein and the charge of a substrate

such as a resin). In other embodiments the protein is immobilised by non-specific binding to porous regions of a substrate such as a resin.

In another embodiment the invention relates to a method of producing allulose. The method comprises the following steps.

1) Providing a vector comprising a nucleic acid molecule having at least 70% sequence identity to SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO: 5.

2) Transforming competent host cells with said vector.

3) Optionally, culturing the transformed host cells.

4) Mixing the transformed cells with a fructose substrate and maintaining under conditions to permit conversion of fructose to allulose.

5) Purifying the allulose produced using standard methods in the art, such as evaporation and crystallization.

In alternative embodiments, step 4) is omitted. Instead, the protein encoded by said nucleic acid molecule is isolated from the transformed host cell and optionally immobilized on a substrate. The protein is then mixed with the fructose substrate and maintained under conditions to permit conversion of fructose to allulose. Step 5) is then performed. In other embodiments, step 2) is omitted and the protein is instead synthesised by *in vitro* translation. Subsequently, the protein is isolated and mixed with the fructose substrate.

Allulose produced by the method of the present invention may be used in a product for human and/or animal consumption. In some embodiments, the product may be a food product, a beverage product, a pharmaceutical product, a nutritional product, a sports product, or a cosmetic product. For example, when the product is a food product, the food product can be selected from the group consisting of a confectionary product, a dessert product, a cereal product, baked goods, frozen dairy products, meats, dairy products, condiments, snack bars, soups, dressings, mixes, prepared foods, baby foods, diet preparations, syrups, food coatings, dried fruit, sauces, gravies, and jams/jellies. In some embodiments, the food product may comprise allulose produced by the method of the present invention as a coating or frosting formed on the surface of the product.

Alternatively, when the product is a beverage product, the beverage product can be selected from the group consisting of a carbonated beverage, a non-carbonated beverage, fruit-flavoured beverage, fruit-juice, tea, milk, coffee, and the like.

Examples

Example 1

In this example, host cells were transformed to express one of three putative ketose-3-epimerase enzymes. Transformed host cells were tested for ketose-3-epimerase activity by incubating them with a fructose substrate.

Materials

Borate Buffer 1M pH8:

i) 62g boric acid dissolved in 1l DI water
ii) adjust to pH8 with 10M NaOH
iii) store in 1L bottle in 4 ºC refrigerator

Borate Buffered Fructose Substrate:

i) 970g liquid fructose (77% DS) in 50ml Borate buffer pH8
ii) water up to 1L final volume
iii) adjust pH to 8 with 5M NaOH

Expression Medium LB – 4 x 2.8L baffled shake flasks:

i) 10g tryptone, 7g NaCl and10g yeast in 1L DI water
ii) autoclave

Methods

Three putative ketose-3-epimerase gene sequences and one control sequence were selected to be synthetically constructed by Genscript USA, Inc. The putative ketose-3-epimerase sequences encoded:

i) hypothetical protein CLOSCI_02526 from *Clostridium scindens* ATCC 35704 (accession ZP_02432281) (SEQ. ID NO. 2)

ii) hypothetical protein CLOHYLEM_05645 from *Clostridium hylemonae* DSM 15053 (accession ZP_03778576.1) (SEQ. ID NO. 4)

iii) D-tagatose 3-epimerase from *Desmospora sp.* 8437 (accession ZP_08466075) (SEQ. ID NO. 6).

The control sequence encoded a xylose isomerase protein from *Clostridium cellulolyticum* H10 (accession YP_002505284) (SEQ. ID NO. 8).

The genes were synthetically constructed with sequences optimized for expression in *E. coli* (see Figures 7 to 10) and each of the resulting four genes was cloned into an expression vector, pET15b. Other combinations of microorgansims and expression vectors known to one skilled in the art are expected to perform equally well.

Competent cells used for the transformation were prepared by inoculating 3ml Lysogeny Broth (LB) with *E. coli* BL21 (DE3) and allowing the bacteria to propagate overnight at 37ºC. 300ml LB was inoculated with this 3ml culture and the cells were grown at 37ºC with shaking to 0.7-1.0 OD (600). Optical densities (OD) were measured in a 1 cm cell at 600nm wavelength on a typical spectrophotometer. The cells were chilled on ice for 10 mins and then spun down at 7500xg at 4 ºC for 15 minutes. The media was poured off and the cells resuspended in 300 ml cold water. The spin was repeated and the cell resuspended in 150 ml cold water. The spin was repeated again and the cells were suspended in about 2ml cold sterile 10% glycerol. The cells were spun down as previously and were suspended in about 2ml cold sterile 10% glycerol. The suspension was divided into 100µl aliquots in sterile eppendorf tubes and stored at -80 ºC.

The expression vectors provided by Genscript were subsequently used to transform competent *E. coli* BL21 (DE3) by electroporation and positive transformants were selected

on ampicillin containing LB agar. 1L LB was poured into each of four 2.8L baffled flasks and was autoclaved. Once cool, 1ml of 100mg/l ampicillin was added to each flask aseptically and each flask was inoculated with 2-3ml of the overnight culture of competent cells prepared above (1 flask per expression strain). The cells were allowed to grow for about 3 hours at 37 ºC with 200rpm shaking in order to achieve an OD of 0.8-1.5. 1ml of a freshly prepared 1M isopropyl β-D-1-thiogalactopyranoside solution was added to each flask, the temperature was reduced to room temperature (i.e. 25-30 ºC) and induction was allowed to proceed for about 5 hours. The cells were spun down at about 5000xg for 30 minutes at 4 ºC and the supernatant decanted. The cell pellet was transferred to a weighed 50ml centrifuge tube and the cell mass was recorded. The cells were resuspended in a few ml of sterile glycerol (10%w/w) and were frozen at -80ºC.

The conversion activity of the cells was checked by mixing the whole cells into a borate buffered fructose substrate and analysing by HPLC using DP 1-4 method with a Ca^{2+} column. Four flasks containing 250ml of borate buffered fructose substrate were warmed to 55 ºC and the frozen cells were thawed at room temperature. The cells were pelleted at 6500xg and resuspended in DI water. 2g (wet weight) of cells were mixed in borate buffered fructose substrate and were incubated at 55 ºC with 90rpm mixing in a 1L baffled flask. Samples were taken at 0, 1, 2 and 5 hours, and were submitted for HPLC analysis.

HPLC analysis consisted of injection of 20 μL of a sample to be analysed at 0.1% (W/V) into a chromatographic system consisting of a water mobile phase with a flow rate between 0.1 and 1.5 mL/min and a stationary phase consisting of a resin of particle size between 1 and 10 μm in the Ca^{2+} form maintained at 80°C. Peaks were detected and quantitated by a refractive index detector and qualitatively assigned based on retention time of known standards.

Results and Discussion

Three protein sequences were identified to be tested as ketose-3-epimerase proteins. The sequences of these proteins are given in Figures 1 to 3 (SEQ. ID NOS 2, 4 and 6). The xylose isomerase from *Clostridium cellulolyticum* H10, used as a control, has previously

been suggested to produce allulose from fructose. Its amino acid sequence is shown in Figure 4 (SEQ. ID NO. 8).

The amino acid sequences of these proteins were aligned with those of other known ketose-3-epimerases and the aligned sequences are shown in Figure 5. Completely conserved residues are highlighted. There are very few conserved residues between these sequences, with fewer than 65% of the residues being conserved from one sequence to the next.

The degree of sequence identity between each of the sequences of SEQ. ID NOS. 2, 4, and 6 with each of Accession Nos NP_535228 and BAA24429 and SEQ. ID NO 8 was determined. The results are shown in Table 2. Between 40 and 63% sequence identity for the known ketose-3-epimerases was observed for each of the selected protein sequences. There was no overall strong homology based on the sequence alignment of all of the sequences. The selected proteins had genes optimized for expression by *E. coli* synthetically constructed and cloned into commercial expression vector pET15b by Genscript. Transformation of *E. coli* BL21 (DE3) was successful for each construct and frozen stocks of each strain were saved along with the expression vectors. Protein expression was carried out on the 1L scale and whole cells were harvested as the crude catalyst. Conversion activity was checked and Figure 6 shows the % DSB allulose produced during the experiment by the four different strains tested.

Table 2

		% Identical Protein Sequence		
Organism	SEQ ID NO.	*A. tumefaciens* (accession NP_535228)	*P. cichorii* (accession BAA24429)	*C. cellulolyticum* (accession NC_011898)
C. scindens	2	59	43	43
C. hylemonae	4	60	41	63
Desmospora sp.	6	50	43	51

Expression of the three putative ketose-3-epimerases was successfully carried out and all three could successfully convert fructose into allulose, confirming that each protein is, indeed, a ketose-3-epimerase.

The most active protein was DS P3E (D-tagatose 3-epimerase from *Desmospora sp.* 8437, SEQ. ID NO. 6) which was capable of converting 30% of a 750g/L fructose solution in just 2 hours utilizing 8g of wet cell weight per litre for a volumetric productivity of 112g/L/hour.

Example 2

In this example, the current best conditions for cell growth and conversion were carried out on an 18L scale to determine scaleability and to produce allulose for further sensory and clinical investigation. Following the conversion, an initial clean-up step to remove fructose was carried out. This example was for the purpose of identifying the scaleability of this process, any unforeseen problems with scale-up and the amount of allulose that could reasonably be produced in a laboratory.

Materials

Isopropylthiogalactopyranoside (IPTG)

Filter sterilized aqueous ampicillin solution 100mg/ml

Crystalline fructose greens

Liquid fructose (77% DS)

Growth media:

i) 25g NaCl, 25g Staleydex® 333, 6g glycerol, 50g tryptone (Difco), 60g yeast extract (Difco), 8g potassium phosphate dibasic and 8g potassium phosphate monobasic in 6l DI water

ii) adjust pH to ~7.8 with Tris base (solid)

iii) autoclave 1L per flask in 6x2.8L baffled flasks with foil on top

Tris buffer 1M pH8:

iv) 121g in 1L DI water

iv) adjust pH to 8 with HCl

iv) store in 1L bottle in 4 °C refrigerator

Method

To propagate the cells, six overnight cultures of 5ml LB medium supplemented with 100µg/ml ampicillin were started. The cultures were inoculated with the *E. coli* production strain (BL21-DE3 pET15b-DS-P3E expressing the protein of SEQ. ID NO. 6) and allowed to grow overnight (~16 hours) at 37°C. 6L of growth media was prepared and autoclaved as described above, 5ml of the overnight culture was added to each flask and this was shaken at 190 rpm at 37°C for 4 hours. 1mM IPTG was added to each flask by preparing a fresh 1M solution and adding 1ml per litre to the flasks. The temperature was reduced to 25 °C with continued shaking for 14-16 hours.

In order to harvest the cells, the cultures were centrifuged at 6000 rpm for 20 minutes using floor centrifuge in Mod 322 and 1L bottles (filled with not more than 800ml of media). The media was decanted into a kill bucket, to which 1% by volume bleach was added, and was allowed to sit for 30 minutes. The centrifuge tubes were weighed and 3ml DI water per gram of cells was added to the tubes. The cells were re-suspended using a spatula and vortex genie until a uniform cell slurry was obtained. The suspension was transferred to 40ml centrifuge tubes and re-pelleted at 6500xg. The wash was decanted into the kill bucket and the cells were re-suspended in the same volume of water.

The propagation and harvesting of cells was repeated with a second batch of cells.

The crystalline fructose greens conversion substrate was prepared by warming a 5 gallon (18.9L) bucket of crystalline fructose greens to room temperature and adding 16,506g crystalline fructose greens to a sanitised 5 gallon (18.9L) plastic bucket with an 18L calibration mark. 900ml of 1M Tris pH 8.0 prepared as above was added to the bucket, followed by water up to the 18L calibration mark, and was mixed using an overhead mixer until homogeneous. The mixture and the unused crystalline fructose greens were returned to a cold room for storage.

The liquid fructose conversion substrate was prepared by combining 17,460g liquid fructose (77% DS) and 500ml of 1M Tris pH 8.0 (prepared as above) in a sanitised 5

gallon (18.9L) plastic bucket with an 18L calibration mark. Water was added up to 18L calibration mark and an overhead mixer was used to mix until homogeneous.

For the whole cell conversion, 18L of prepared crystalline fructose greens conversion substrate was heated to 55ºC in a water bath and was gently mixed with an overhead mixer at about 150 rpm. The re-suspended cell paste obtained from the cell harvesting was added to a total of 100g wet weight of cells. After 5 hours a sample was removed and was submitted HPLC analysis. The reaction was stopped by refrigerating the entire bucket at 4 ºC. A sample was submitted for microbial analysis for *E. coli*, coliforms and TPC.

The whole cell conversion process was repeated with 18L of prepared liquid fructose conversion substrate. 120g of cells wet weight was used and samples were taken at 2 and 4 hours for HPLC Analysis.

Yeast fermentation was used to remove fructose from the crystalline fructose greens conversion substrate. The crystalline fructose greens conversion product was diluted with 2 volumes of water for a final concentration of ~250g/L of combined allulose and fructose in a total volume of 54L. The 54L of diluted mixture was split between four sanitised 5 gallon (18.9L) buckets, with approximately 13L per bucket. Two of the buckets were stored in a refrigerator. The remaining two buckets were set up with vigorous agitation from overhead mixers and aeration from 9L/min air pumps with diffusers for approximately 0.3VVM air flow. 120g dry active baker's yeast (Fleishman's brand) was added to each bucket, and these were mixed and aerated for 2 days (~36 hours) with occasional sampling for DP1-4 allulose analysis. The buckets were transferred to the cold room overnight to allow the yeast to settle. The supernatant was then transferred to two new clean sanitised buckets and the remaining yeast fraction was transferred to the two refrigerated crystalline fructose greens-containing buckets prepared above. The agitation and mixing process was repeated, followed by the removal of yeast. Following the yeast fermentation step, about 45L of supernatant was obtained and sterile filtered into 3 clean sanitsed buckets, which were stored at 4 ºC for further processing.

Results and Discussion

Approximately 220g of BL21 (DE3) pET-15b-DS P3E cells were obtained from 12L of culture and split into two 18L bioconversions as described above. Thus, the total whole cell biocatalyst concentration was 5.6g/L for the crystalline fructose greens conversion and 6.7g/L for the liquid fructose conversion.

Figure 11 shows that both conversions rapidly reached ~25%, calculated as allulose as a percentage of allulose+fructose. This is slightly lower but very near the conversion level previously achieved using this cell type and similar conversion media on the small scale. Conversion had already reached 22% after just 2 hours with the liquid fructose substrate.

In each 18L conversion, approximately 3.3kg of allulose was produced. There did not appear to be a significant difference between the two substrates.

The scale up from 250ml did not produce any unforeseen issues and proceeded as expected.

The microbial testing resulted in no live *E. coli* with a negative result and <3 coliforms per gram, and a total plate count of two. Therefore, a temperature of 55 ºC combined with a high percentage of DS syrup was sufficient to kill the whole cell biocatalyst.

The bioconversion of fructose to allulose using the newly identified enzyme DS P3E was successfully scaled up to 18L.

Example 3

An *E. coli* strain containing the newly identified DS P3E protein (SEQ ID NO: 6) was produced by two 10 L fermentations in a fermentation lab using a pH control feed batch culture method with glucose yeast extract media. The fermentations proceeded as expected.

During the fermentation batch growth and fed batch phases the cells grew exponentially with a doubling time of approximately 1 hr. Glucose concentration dropped from about 9 g/L to <1 g/L in about 5.5 hours (OD ~ 28). During the Induction phase for enzyme production, the OD continued to rise to about 130 and then was not observed to change significantly. Harvesting of the fermentation by centrifugation resulted in 4.5 kg (10 lbs) of wet cell paste or approximately 1.1 kg (2.5 lbs) dry cell weight.

Fructose substrate (836 kg DS (dry solids) basis) was diluted to 69 %DS (920 grams/L) with RO water and heated to 52°C and pH adjusted to 7.8. Low agitation (~50 rpms) was utilized to promote mixing throughout the reaction and the entire batch of 4.5 kg (wet paste) of expressed whole cells from above was added to the reaction and a time 0 sample was taken. This provided a 0.48 g/L biocatalyst load which is similar to the previously tested lab scale conversions, however, the substrate concentration was higher at 920 g/L. Samples were taken at 4 and 16 hrs and analyzed by HPLC.

No loss of DS was observed and no bioproducts were produced during the reaction. The reaction proceeded nearly to the equilibrium value of ~30% allulose at the end of the 16 hr reaction. At 4 hrs the reaction had already proceeded to 18% conversion. The volumetric conversion rate previously obtained using 0.5 g/L biocatalyst with 750 g/L substrate (Examples 1 and 2) was 46 g/L*hr or per unit biocatalyst 92 g/L*hr/gram biocatalyst. Here, using a higher substrate concentration and slightly lower temperature (52°C vs 55°C), the volumetric conversion rate was 41 g/L*hr or 85 g/L*hr/gram biocatalyst (calculated using 4 hr data point). This demonstrates the remarkable flexibility of the epimerase reaction. When the reaction was completed at 16 hrs, 230 kg of allulose were present in the 28:72 mixture of allulose:fructose.

Example 4

Conversion of fructose by the four different enzymes (SEQ ID No 2, 4, 6 and 8) was compared on Tris buffered fructose substrate at 750 g/L. Cells were induced at 16°C instead of 25-30°C and the rate of conversion was slower than in previous experiments. To 200 mL of the substrate, 2 g wet weight of resuspended cells was added in 500 mL baffled

flasks and incubated at 55°C with 90 rpm shaking. Samples were taken at 2 hrs and 3.5 hrs for HPLC analysis. The results are shown in Figure 12 in which CC P3E corresponds to SEQ ID NO. 8 and CH P3E, CS P3E and DS P3E correspond to SEQ ID NOs. 4, 2 and 6, respectively. In this experiment, all 4 strains expressing one of the proteins set forth in SEQ ID 2, 4, 6 or 8 appear to have approximately the same level of activity converting approximately 5% of the substrate to allulose in 3.5 hrs.

Example 5

In this example, the first trial of allulose production using immobilised enzymes was carried out. This example was for the purpose of improving enzyme utilisation.

Materials

Lyophilized enzyme powder as prepared from Codexis®, Lot D13007 or D13008

- *Desmospora sp.* Psicose 3-epimerase

Duolite® A568 (Dow)

Amberlite® XAD2 (Sigma)

Tris buffer 1M:

i) Prepared in water by dissolving at a concentration of 1M

ii) Adjusted pH to 8.0 with HCl

iii) Diluted to 100mM before use

Crystalline Fructose Greens, 80% dry solids with a composition of:

i) 90% DSB fructose

ii) 7% DSB dextrose

iii) 3% DP2+

iv) Other monosaccherides

$MnCl_2$ (Sigma)

Method

1) Small scale immobilisation

In order to test efficiency of immobilisation, a small fixed bed reaction was carried out in jacketed columns of approximately 30ml volume (11mm x 300mm column dimensions).

Both XAD2 and A568 resin was washed several times with water to remove fines (i.e. fine resin particles (i.e. broken/fractured particles) that are a by-product of resin manufacturing) and any residuals from manufacturing. 2g of lyophilised enzyme (i.e. epimerase) was dissolved in approximately 50 ml of water and split into two aliquots. The pH was measured and determined to be 6.5. Approximately 30ml of each resin was incubated with one aliquot of epimerase solution for approximately one hour at room temperature and light agitation. The resins were then packed into the jacketed column and peristaltic pumps were used to recycle the epimerase solution through the fixed bed for an additional two hours. The columns were then washed with 10 bed volumes of Tris buffer 100mM pH 8.0. The effluent at this point looked clear and free of protein as measured spectrophotometrically at A280. Crystalline fructose greens feed was prepared by diluting crystalline fructose greens down to 60% DS with RO water, followed by adjusting the pH to 8.0 and then addition of 28 ppm $MnCl_2$ and 10 mM Tris buffer pH 8.0.

Feed was then pumped through the 30ml fixed bed reactors at a rate of 8 bed volumes per hour (BV/h) with the jacketed columns heated with a recirculating water bath to a temperature of 57°C. The reactor effluent was collected and analysed by FT-IR to provide a relative concentration of allulose and fructose. This process was continued for a total of 5 days. The feed rate was adjusted from 8 bed volumes per hour down to 6 bed volumes per hour over the course of the test production run in order to maintain conversion rate.

2. Scale-up 300ml fixed bed reactor

In order to test scaled-up efficiency of immobilisation, a larger fixed bed reactor was created in jacketed columns of approximately 300ml volume (25mm x 600mm column dimensions). Both XAD2 and A568 resin was washed several times with water to remove fines and any residuals from manufacturing. 10g of lyophilised epimerase was dissolved in approximately 100 ml of water. Approximately 300ml of A568 resin was packed into the 300ml jacketed column and peristaltic pumps were used to recycle the epimerase solution through the fixed bed for approximately two hours at room temperature. The columns were then washed with 5 bed volumes of Tris buffer 100mM, pH 8.0, at room temperature. The effluent at this point looked clear and free of protein as measured spectrophotometrically at A280. Crystalline fructose greens feed was prepared by diluting crystalline fructose greens down to 60% DS with RO water, followed by adjusting the pH to 8.0 and then

addition of 28 ppm $MnCl_2$ and 10 mM Tris buffer pH 8.0. Feed was then pumped through the 30ml fixed bed reactors at a rate of 8 bed volumes per hour (BV/h) with the jacketed columns heated with a recirculating water bath to a temperature of 57°C. The reactor effluent was collected and analysed by FT-IR to provide a relative concentration of allulose and fructose. This process was continued for a total of 4 days. The feed rate was adjusted from eight bed volumes per hour down to two bed volumes per hour over the course of the test production run. Additionally, the column was allowed to sit at room temperature for two weeks and then restarted to determine the epimerase stability during column storage.

Results and Discussion

The 30ml column with XAD2 displayed no significant conversion for any of the samples checked and, therefore, no further analysis was performed. However, significant conversion was observed with Dowex® A568. Figure 13 shows the time course of reaction for the 30ml fixed bed reactor with A568 resin. More than 4kg of allulose was produced over the course of 120 hour fixed bed conversion containing just 1g of epimerase. The percentage conversion gradually decreased over the course of the 120 hours and the flow rate through the column was reduced at 72 hours to compensate. Near equilibrium concentrations of allulose were produced during the reaction.

Figure 14 displays the course of the 300ml reaction with A568 resin. Flow rate was decreased at 24 hours due to limitations in the amount of feed available (86l of feed were used in 72 hours). More than 20kg of allulose was produced in 72 hours from 10g of epimerase. The conversion rate was still high at the end of 72 hours, although some decline in performance was observed.

Epimerase stability in solution has previously been determined in flask reactions. More than 90% of activity is lost within 8 hours at 53°C. In this example, the conversion was carried out at 57°C. A higher temperature is advantageous in terms of reaction rate, equilibrium ratio, and microbiological stability. In this example, significant epimerase activity remained even after 120 hours.

In the large scale reaction, which was feed limited, 20kg of allulose was produced using 10g of epimerase, resulting in a net epimerase dosing rate of 0.05% (m/m). In the small

scale reaction, 4.8kg of allulose was produced from 1g of epimerase, resulting in a net epimerase dosing rate of 0.02% (m/m). In standard fructose production, immobilised glucoisomerase is used at a rate of 0.01-0.005% (m/), although this is over the course of 6-12 months of operation.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A protein comprising a polypeptide sequence having at least 89% sequence identity to SEQ ID NO: 6, wherein the protein has psicose 3-epimerase activity and the protein does not consist of a polypeptide sequence which is identical to SEQ ID NO: 6.

2. The protein according to claim 1, wherein the polypeptide sequence has at least 90%, 95% or 99% sequence identity, to SEQ ID NO: 6.

3. The protein according to claim 1 or 2, wherein the polypeptide sequence comprises the sequence of SEQ ID NO: 13.

4. The protein according to any one of claims 1 to 3, wherein the protein is immobilized on a solid substrate.

5. Use of a protein for synthesizing allulose, wherein the protein comprises a polypeptide sequence having at least 89% sequence identity to SEQ ID NO: 6, and has psicose 3-epimerase activity.

6. A nucleic acid molecule comprising a polynucleotide sequence encoding the protein according to any one of claims 1 to 4.

7. The nucleic acid molecule according to claim 6, comprising a polynucleotide sequence which:

i) has at least 90%, 95% or 99% sequence identity, to SEQ ID NO: 5; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, wherein hybridization under highly stringent conditions means at 6X SSC and 1% SDS at 65°C, with a first wash for 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65°C,

wherein the nucleic acid molecule does not comprise a polynucleotide sequence which is identical to SEQ ID NO: 5.

8. A vector comprising a nucleic acid molecule according to claim 6 or 7.

9. A host cell comprising a recombinant nucleic acid molecule according to claim 6 or 7.

10. The host cell according to claim 9, wherein the host cell is a yeast, bacterium or other microorganism, or is a mammalian or plant cell.

11. The host cell according to claim 10, wherein the host cell is *E. coli*.

12. A method of producing allulose comprising contacting a protein with a fructose substrate under conditions such that the fructose substrate is converted into allulose, wherein the protein comprises a polypeptide sequence having at least 89% sequence identity to SEQ ID NO: 6, wherein the protein has psicose 3-epimerase activity.

13. The method according to claim 12, wherein the protein is present in a host cell.

14. The method according to claim 12, wherein the protein is in isolated form.

15. The method according to any one of claims 12 to 14, wherein the conditions comprise maintaining the protein and the fructose substrate at a temperature between 25°C and 75°C.

16. The method according to any one of claims 12 to 15, wherein the conditions comprise maintaining the protein and the fructose substrate between pH 4 and pH 10.

17. The method according to any one of claims 12 to 16, wherein the conditions comprise maintaining the fructose substrate concentration between 75% and 95% (W/V).

18. A nucleic acid molecule comprising a polynucleotide sequence which:

i) has at least 89% sequence identity to SEQ ID NO: 5; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, wherein hybridization under highly stringent conditions means at 6X SSC and 1% SDS at 65°C, with a first wash for 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65°C.

19. A host cell comprising a recombinant nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide having psicose 3-epimerase activity, wherein the polynucleotide sequence:

i) has at least 89% sequence identity to SEQ ID NO: 5; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, wherein hybridization under highly stringent conditions means at 6X SSC and 1% SDS at 65°C, with a first wash for 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65°C.

20. A vector comprising a nucleic acid molecule comprising a polynucleotide encoding a polypeptide having psicose 3-epimerase activity, wherein the polynucleotide sequence:

i) has at least 89% sequence identity to SEQ ID NO: 5; or

ii) hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, wherein hybridization under highly stringent conditions means at 6X SSC and 1% SDS at 65°C, with a first wash for 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65°C.

21. A method of producing allulose comprising the steps of:

i) providing a vector comprising a nucleic acid molecule having a polynucleotide sequence encoding a protein having psicose 3-epimerase activity wherein the polynucleotide sequence: a) has at least 89% sequence identity to SEQ ID NO: 5; or b)

hybridizes under highly stringent conditions to a polynucleotide having a sequence complementary to the sequence set forth in SEQ ID NO: 5, wherein hybridization under highly stringent conditions means at 6X SSC and 1% SDS at 65°C, with a first wash for 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with a subsequent wash with 0.2 X SSC and 0.1% SDS at 65°C;

ii) synthesising the protein having psicose 3-epimerase activity encoded by the polynucleotide sequence;

iii) contacting fructose with the protein having psicose 3-epimerase activity and maintaining the fructose and protein under conditions to permit the conversion of fructose to allulose; and

iv) at least partially purifying the allulose produced in step iii).

Clostridium scindens Psicose- 3 -Epimerase (SEQ ID NO : 2)
MNRIGIFMNFWVKNWDADHVKYIKKVSGLGFDILEFQAQALLEMDKSRMDEVRQAAKDNGIELTYSLGLN
PKYDVASPDAKVREGGIEYLKRIVERIGYMEGKLLSGVNYAGWGSPDYIVDDKSEIVEHSIESVRQVIKT
AEDYDVTYCVEVVNRFEGIVMNTAKEAIEYVKQIDSDKIGILLDTYHMNIEEGSIGDAIRSVGGYLKNFH
TGENNRVVPGKGHLDWDEIFGALHDIDYQGRIVSEPFVQMGGEVARDIKVWRDLVEDPSEEVLDEEARFL
LNFEKDMIRKHYGIA

FIG. 1

Clostridium hylemonae Psicose - 3 - Epimerase (SEQ ID NO : 4)
MKHGIYYAYWEQEWAADYKRYVEKVAKLGFDILEIGAGPLPEYAEQDVKELKKCAQDNGITLTAGYGPTF
NHNIGSSDAGVREEALEWYKRLFEVLAELDIHLIGGALYSYWPVDFANADKTEDWKWSVEGMQRLAPAAA
KYDINGMEVLNRFESHILNTAEEGVKFVEEVGMDNVKVMLDTFHMNIEEQSIGGAIRRAGKLLGHFHTG
ECNRMVPGKGRIPWREIGDALRDIGYDGTAVMEPFVRMGGQVGADIKVWRDISRGADEAQLDDDARRALE
FQRYMLEWK

FIG. 2

Desmospora sp. Psicose- 3-Epimerase (SEQ ID NO : 6)
MKYGVYFAYWEDSWDVDFEKYVRKVKKLGFDILEVAALGLVNLPEEKLERLKQLAEQHDIILTAGIGLPK
EYDVSSTDKKVRRNGISFMKKVMDAMHQAGIHRIGGTVYSYWPVDYSCSFDKPAVRKHSIESVRELAEYA
RQYNITLLIETLNRFEQFLLNDAEEAVAYVKEVDEPNVKVMLDTFHMNIEEDHIADAIRYTGDHLGQLHI
GEANRKVPGKGSMPWTEIGQALKDIRYDGYVVMEPFIKTGGQVGRDIKLWRDLSGNATEEQLDRELAESL
EFVKAAFGE

FIG. 3

Clostridium cellulolyticum Psicose- 3-Epimerase (SEQ ID NO : 8)
MKHGIYYAYWEQEWEADYKYYIEKVAKLGFDILElAASPLPFYSDIQINELKACAHGNGITLTVGHGPSA
EQNLSSPDPDIRKNAKAFYTDLLKRLYKLDVHLIGGALYSYWPIDYTKTIDKKGDWERSVESVREVAKVA
EACGVDFCLEVLNRFENYLINTAQEGVDFVKQVDHNNVKVMLDTFHMNIEEDSIGGAIRTAGSYLGHLHT
GECNRKVPGRGRIPWVEIGEALADIGYNGSVVMEPFVRMGGTVGSNIKVWRDISNGADEKMLDREAQAAL
DFSRYVLECHKHS

FIG. 4

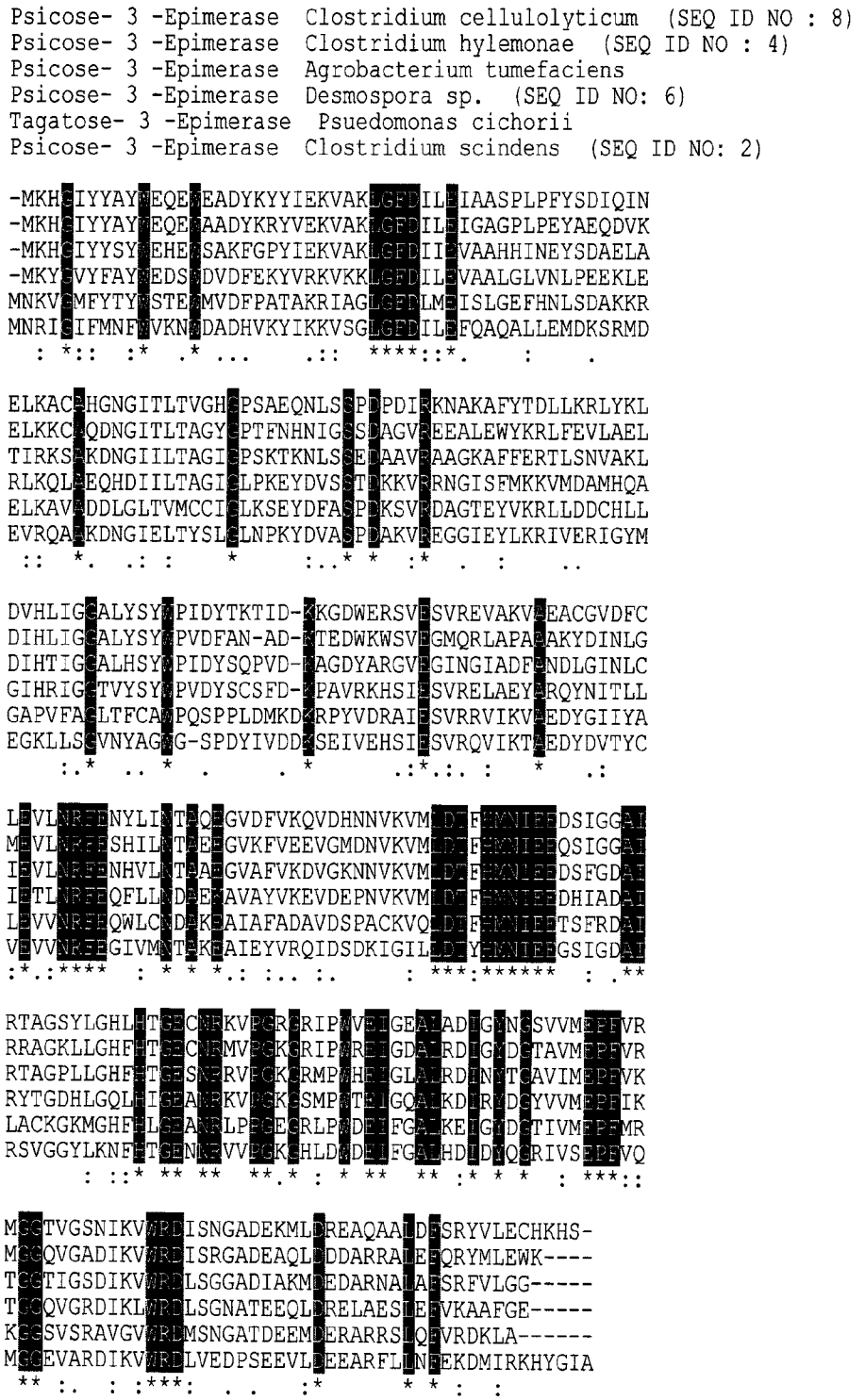

Psicose- 3 -Epimerase Clostridium cellulolyticum (SEQ ID NO : 8)
Psicose- 3 -Epimerase Clostridium hylemonae (SEQ ID NO : 4)
Psicose- 3 -Epimerase Agrobacterium tumefaciens
Psicose- 3 -Epimerase Desmospora sp. (SEQ ID NO: 6)
Tagatose- 3 -Epimerase Psuedomonas cichorii
Psicose- 3 -Epimerase Clostridium scindens (SEQ ID NO: 2)

FIG. 5

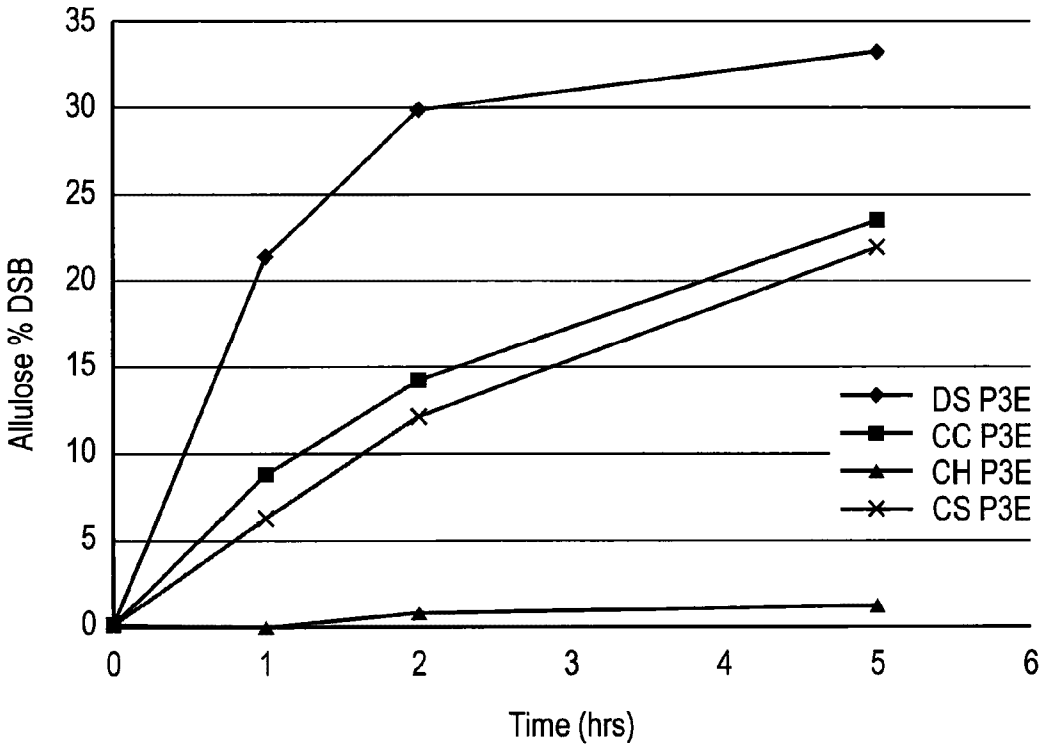

35
30
25
20
15
10
5
0
Allulose % DSB
0
1
2
3
4
5
6
Time (hrs)
DS P3E
CC P3E
CH P3E
CS P3E

FIG. 6

4/10

Clostridium cellulolyticum .
(Optimized Sequence Length:897, GC%:51.61) (SEQ ID NO: 7)
CATATG
AAGCACGGCATCTATTACGCCTATTGGGAACAAGAATGGGAAGCAGACTACAAGTATTACATCGAAAAGGTTGCG
AAGCTGGGTTTTGATATTCTGGAAATCGCGGCCTCACCGCTGCCGTTTTATTCGGACATTCAGATCAATGAACTG
AAAGCGTGCGCGCATGGCAACGGTATTACCCTGACGGTGGGCCACGGTCCGAGCGCGGAACAAAATCTGAGCAGC
CCGGACCCGGACATCCGTAAAAACGCAAAGGCTTTCTATACCGATCTGCTGAAACGCCTGTACAAGCTGGACGTT
CATCTGATTGGCGGTGCCCTGTATTCTTACTGGCCGATCGATTACACCAAGACGATCGATAAGAAGGGCGACTGG
GAACGTAGTGTTGAATCCGTCCGCGAAGTGGCCAAGGTTGCGGAAGCCTGCGGTGTCGATTTTTGTCTGGAAGTG
CTGAACCGTTTCGAAAATTACCTGATTAACACCGCACAGGAAGGCGTCGATTTTGTGAAACAAGTTGACCATAAC
AATGTCAAGGTGATGCTGGATACGTTCCACATGAATATCGAAGAAGACAGTATTGGCGGTGCGATCCGTACCGCC
GGCTCCTATCTGGGTCATCTGCACACGGGCGAATGCAATCGCAAAGTTCCGGGCCGTGGTCGCATTCCGTGGGTC
GAAATCGGTGAAGCACTGGCTGATATTGGCTACAACGGTTCAGTGGTTATGGAACCGTTTGTTCGTATGGGCGGC
ACCGTCGGCAGCAATATTAAAGTGTGGCGCGATATCTCTAACGGTGCAGATGAAAAGATGCTGGACCGTGAAGCT
CAGGCAGCTCTGGACTTCTCACGCTACGTGCTGGAATGTCATAAACACTCGTAA
AGATCTGGATCC

DNA Alignment (Optimized Region) (Upper: SEQ ID NO: 7; Lower: SEQ ID NO:12

```
Optimized 7    AAGCACGGCATCTATTACGCCTATTGGGAACAAGAATGGGAAGCAGACTACAAGTATTAC
Original 7     AAACATGGTATATACTACGCATATTGGGAACAAGAATGGGAAGCTGATTACAAATACTAT
Optimized 67   ATCGAAAAGGTTGCGAAGCTGGGTTTTGATATTCTGGAAATCGCGGCCTCACCGCTGCCG
Original 67    ATTGAGAAGGTTGCAAAGCTTGGTTTTGATATTCTAGAGATTGCAGCTTCACCGCTACCT
Optimized 127  TTTTATTCGGACATTCAGATCAATGAACTGAAAGCGTGCGCGCATGGCAACGGTATTACC
Original 127   TTTTACAGTGACATTCAGATTAATGAGCTCAAGGCATGTGCCCATGGCAATGGAATTACA
Optimized 187  CTGACGGTGGGCCACGGTCCGAGCGCGGAACAAAATCTGAGCAGCCCGGACCCGGACATC
Original 187   CTTACGGTAGGCCATGGGCCTAGTGCAGAACAAAACCTGTCTTCTCCCGACCCCGATATT
Optimized 247  CGTAAAAACGCAAAGGCTTTCTATACCGATCTGCTGAAACGCCTGTACAAGCTGGACGTT
Original 247   CGCAAAAATGCTAAAGCTTTTTATACCGATTTACTCAAACGACTTTACAAGCTGGATGTA
Optimized 307  CATCTGATTGGCGGTGCCCTGTATTCTTACTGGCCGATCGATTACACCAAGACGATCGAT
Original 307   CATTTGATAGGTGGGGCTTTATATTCTTATTGGCCGATAGATTACACAAAGACAATTGAT
Optimized 367  AAGAAGGGCGACTGGGAACGTAGTGTTGAATCCGTCCGCGAAGTGGCCAAGGTTGCGGAA
Original 367   AAAAAAGGCGATTGGGAACGCAGCGTTGAAAGTGTTCGAGAAGTTGCTAAGGTGGCCGAA
Optimized 427  GCCTGCGGTGTCGATTTTTGTCTGGAAGTGCTGAACCGTTTCGAAAATTACCTGATTAAC
Original 427   GCCTGTGGAGTGGATTTCTGCCTAGAGGTTCTTAATAGATTTGAGAATTATTTAATTAAC
Optimized 487  ACCGCACAGGAAGGCGTCGATTTTGTGAAACAAGTTGACCATAACAATGTCAAGGTGATG
Original 487   ACAGCACAAGAGGGTGTAGATTTTGTAAAACAGGTTGACCATAACAATGTAAAGGTAATG
Optimized 547  CTGGATACGTTCCACATGAATATCGAAGAAGACAGTATTGGCGGTGCGATCCGTACCGCC
Original 547   CTTGATACCTTCCATATGAATATTGAGGAAGATAGTATCGGAGGTGCAATCAGGACTGCG
Optimized 607  GGCTCCTATCTGGGTCATCTGCACACGGGCGAATGCAATCGCAAAGTTCCGGGCCGTGGT
Original 607   GGCTCTTACTTGGGACATTTACACACTGGCGAATGTAATCGTAAAGTTCCCGGCAGAGGA
Optimized 667  CGCATTCCGTGGGTCGAAATCGGTGAAGCACTGGCTGATATTGGCTACAACGGTTCAGTG
Original 667   AGAATTCCATGGGTAGAAATTGGTGAGGCTCTTGCTGACATAGGTTATAACGGTAGTGTT
Optimized 727  GTTATGGAACCGTTTGTTCGTATGGGCGGCACCGTCGGCAGCAATATTAAAGTGTGGCGC
Original 727   GTTATGGAACCTTTTGTTAGAATGGGCGGAACTGTCGGATCTAATATTAAGGTTTGGCGT
Optimized 787  GATATCTCTAACGGTGCAGATGAAAAGATGCTGGACCGTGAAGCTCAGGCAGCTCTGGAC
Original 787   GACATTAGTAACGGTGCAGATGAGAAAATGCTGGATAGAGAAGCACAGGCCGCACTTGAT
Optimized 847  TTCTCACGCTACGTGCTGGAATGTCATAAACACTCGTAA
Original 847   TTCTCCAGATATGTATTAGAATGTCATAAACACTCCTGA
```

FIG. 7

5/10

Desmospora sp. (SEQ ID NO: 5)
CATATG
AAATACGGTGTCTACTTTGCTTACTGGGAAGATTCGTGGGATGTTGACTTTGAAAAATACGTTCGCAAGGTGAAA
AAACTGGGCTTTGATATTCTGGAAGTTGCAGCACTGGGTCTGGTCAACCTGCCGGAAGAAAAACTGGAACGTCTG
AAGCAGCTGGCGGAACAACATGACATTATCCTGACCGCCGGCATTGGTCTGCCGAAAGAATATGATGTCAGCTCT
ACGGACAAAAAAGTGCGTCGCAATGGCATCTCCTTTATGAAAAAGGTTATGGATGCAATGCATCAGGCTGGTATT
CACCGTATTGGCGGCACCGTGTATAGCTACTGGCCGGTTGATTACAGTTGCTCCTTCGACAAACCGGCGGTTCGC
AAGCACTCAATTGAATCGGTCCGTGAACTGGCGGAATATGCCCGCCAGTACAACATTACCCTGCTGATCGAAACG
CTGAACCGCTTTGAACAATTCCTGCTGAATGATGCCGAAGAAGCGGTTGCCTATGTCAAAGAAGTGGATGAACCG
AACGTCAAGGTGATGCTGGACACCTTCCACATGAACATCGAAGAAGATCACATCGCAGACGCTATCCGTTACACG
GGCGATCATCTGGGTCAGCTGCACATCGGCGAAGCCAACCGCAAAGTGCCGGGCAAGGGTAGTATGCCGTGGACC
GAAATTGGCCAAGCACTGAAAGATATCCGTTATGACGGTTACGTGGTTATGGAACCGTTCATTAAAACCGGCGGT
CAGGTTGGCCGTGATATCAAACTGTGGCGCGACCTGAGCGGTAATGCAACGGAAGAACAACTGGATCGCGAACTG
GCTGAATCTCTGGAATTTGTGAAAGCAGCTTTCGGTGAATAA
AGATCTGGATCC

DNA Alignment (Optimized Region) (Upper: SEQ ID NO: 5; Lower: SEQ ID NO:11

Optimized 7 AAATACGGTGTCTACTTTGCTTACTGGGAAGATTCGTGGGATGTTGACTTTGAAAAATAC
Original 7 AAATACGGTGTCTATTTCGCTTACTGGGAAGACTCGTGGGATGTGGATTTCGAGAAGTAC
Optimized 67 GTTCGCAAGGTGAAAAAACTGGGCTTTGATATTCTGGAAGTTGCAGCACTGGGTCTGGTC
Original 67 GTGCGGAAAGTGAAAAAGTTGGGCTTCGACATCCTCGAAGTGGCGGCATTGGGTCTCGTC
Optimized 127 AACCTGCCGGAAGAAAAACTGGAACGTCTGAAGCAGCTGGCGGAACAACATGACATTATC
Original 127 AACCTTCCGGAGGAGAAACTGGAGCGGCTGAAACAACTCGCCGAACAGCACGATATCATC
Optimized 187 CTGACCGCCGGCATTGGTCTGCCGAAAGAATATGATGTCAGCTCTACGGACAAAAAAGTG
Original 187 CTGACGGCCGGGATCGGCCTGCCAAAGGAATACGATGTCTCGTCAACTGACAAAAAGGTG
Optimized 247 CGTCGCAATGGCATCTCCTTTATGAAAAAGGTTATGGATGCAATGCATCAGGCTGGTATT
Original 247 CGCCGGAACGGCATCTCCTTCATGAAGAAAGTGATGGACGCGATGCATCAGGCCGGCATC
Optimized 307 CACCGTATTGGCGGCACCGTGTATAGCTACTGGCCGGTTGATTACAGTTGCTCCTTCGAC
Original 307 CACCGGATCGGCGGCACGGTCTACTCGTATTGGCCGGTTGACTACAGTTGCTCCTTCGAC
Optimized 367 AAACCGGCGGTTCGCAAGCACTCAATTGAATCGGTCCGTGAACTGGCGGAATATGCCCGC
Original 367 AAGCCGGCCGTAAGGAAGCACAGCATCGAAAGCGTCAGAGAGCTGGCGGAGTACGCACGG
Optimized 427 CAGTACAACATTACCCTGCTGATCGAAACGCTGAACCGCTTTGAACAATTCCTGCTGAAT
Original 427 CAGTACAACATCACACTCCTCATCGAAACGCTCAACCGGTTTGAGCAGTTTCTCCTGAAC
Optimized 487 GATGCCGAAGAAGCGGTTGCCTATGTCAAAGAAGTGGATGAACCGAACGTCAAGGTGATG
Original 487 GACGCGGAGGAAGCAGTCGCCTATGTGAAGGAAGTGGACGAGCCGAATGTGAAAGTCATG
Optimized 547 CTGGACACCTTCCACATGAACATCGAAGAAGATCACATCGCAGACGCTATCCGTTACACG
Original 547 CTCGACACATTCCACATGAACATCGAGGAAGACCACATTGCCGATGCCATCCGCTACACC
Optimized 607 GGCGATCATCTGGGTCAGCTGCACATCGGCGAAGCCAACCGCAAAGTGCCGGGCAAGGGT
Original 607 GGTGACCACCTCGGCCAACTGCACATCGGCGAAGCGAATCGGAAAGTCCCGGGCAAGGGT
Optimized 667 AGTATGCCGTGGACCGAAATTGGCCAAGCACTGAAAGATATCCGTTATGACGGTTACGTG
Original 667 TCGATGCCTTGGACAGAAATCGGACAGGCGCTGAAAGACATTCGCTACGATGGCTACGTT
Optimized 727 GTTATGGAACCGTTCATTAAAACCGGCGGTCAGGTTGGCCGTGATATCAAACTGTGGCGC
Original 727 GTCATGGAACCCTTCATCAAAACCGGCGGACAGGTCGGCCGGGACATCAAGCTCTGGCGC
Optimized 787 GACCTGAGCGGTAATGCAACGGAAGAACAACTGGATCGCGAACTGGCTGAATCTCTGGAA
Original 787 GATCTGTCGGGAAATGCGACGGAGGAACAGTTGGACCGGGAGCTGGCAGAGTCGCTGGAA
Optimized 847 TTTGTGAAAGCAGCTTTCGGTGAATAA
Original 847 TTTGTGAAAGCGGCGTTCGGGGAGTAA

FIG. 8

6/10

Clostridium scindens
(Optimized Sequence Length:903, GC%:49.13) (SEQ ID NO: 1)
CATATG
AATCGTATTGGCATTTTTATGAATTTTTGGGTGAAGAACTGGGACGCTGACCACGTTAAGTACATCAAGAAGGTG
TCGGGCCTGGGCTTTGATATTCTGGAATTTCAGGCACAAGCTCTGCTGGAAATGGATAAATCTCGTATGGACGAA
GTGCGCCAGGCGGCCAAGGATAACGGCATTGAACTGACCTATTCTCTGGGTCTGAATCCGAAATACGATGTGGCA
AGTCCGGACGCTAAGGTTCGTGAAGGCGGTATCGAATATCTGAAACGTATTGTGGAACGCATCGGCTACATGGAA
GGCAAGCTGCTGTCAGGCGTTAACTATGCGGGCTGGGGTTCGCCGGATTACATTGTCGATGACAAAAGCGAAATT
GTGGAACATAGCATCGAAAGCGTGCGTCAGGTCATCAAAACCGCCGAAGATTATGACGTGACGTACTGCGTTGAA
GTGGTTAACCGCTTTGAAGGCATTGTTATGAATACCGCGAAAGAAGCCATTGAATATGTCAAACAAATCGATAGC
GACAAGATTGGTATCCTGCTGGATACGTACCACATGAACATCGAAGAAGGCAGTATTGGTGATGCGATCCGTTCC
GTTGGCGGTTATCTGAAAAATTTCCACACGGGCGAAAACAATCGCGTCGTGCCGGGCAAGGGTCATCTGGATTGG
GACGAAATTTTTGGCGCACTGCACGATATTGACTACCAGGGTCGCATCGTCTCCGAACCGTTCGTGCAAATGGGC
GGTGAAGTGGCTCGTGATATCAAAGTTTGGCGCGATCTGGTCGAAGACCCGAGCGAAGAAGTTCTGGATGAAGAA
GCGCGTTTTCTGCTGAATTTCGAAAAAGACATGATTCGCAAGCACTATGGTATCGCCTAA
AGATCTGGATCC

DNA Alignment (Optimized Region) (Upper: SEQ ID NO: 1; Lower: SEQ ID NO: 9)

```
Optimized 7    AATCGTATTGGCATTTTTATGAATTTTTGGGTGAAGAACTGGGACGCTGACCACGTTAAG
Original 7     AACAGAATAGGAATATTTATGAATTTCTGGGTTAAGAACTGGGATGCAGATCATGTCAAG
Optimized 67   TACATCAAGAAGGTGTCGGGCCTGGGCTTTGATATTCTGGAATTTCAGGCACAAGCTCTG
Original 67    TATATTAAAAAGGTATCCGGCCTTGGATTTGATATTCTGGAATTCCAGGCCCAGGCGCTT
Optimized 127  CTGGAAATGGATAAATCTCGTATGGACGAAGTGCGCCAGGCGGCCAAGGATAACGGCATT
Original 127   CTGGAGATGGATAAGAGCAGGATGGATGAGGTCAGGCAGGCGGCAAAGGACAATGGAATC
Optimized 187  GAACTGACCTATTCTCTGGGTCTGAATCCGAAATACGATGTGGCAAGTCCGGACGCTAAG
Original 187   GAACTGACCTACAGCCTTGGGCTGAATCCTAAGTACGATGTCGCAAGCCCGGATGCAAAA
Optimized 247  GTTCGTGAAGGCGGTATCGAATATCTGAAACGTATTGTGGAACGCATCGGCTACATGGAA
Original 247   GTCAGGGAAGGCGGAATCGAATATCTGAAGCGGATCGTGGAGCGGATTGGATACATGGAA
Optimized 307  GGCAAGCTGCTGTCAGGCGTTAACTATGCGGGCTGGGGTTCGCCGGATTACATTGTCGAT
Original 307   GGAAAACTGCTTTCCGGAGTCAACTATGCCGGCTGGGGAAGCCCGGACTATATCGTGGAT
Optimized 367  GACAAAAGCGAAATTGTGGAACATAGCATCGAAAGCGTGCGTCAGGTCATCAAAACCGCC
Original 367   GACAAAAGCGAGATCGTGGAGCACAGCATCGAAAGCGTCCGCCAGGTCATTAAGACGGCA
Optimized 427  GAAGATTATGACGTGACGTACTGCGTTGAAGTGGTTAACCGCTTTGAAGGCATTGTTATG
Original 427   GAAGATTATGACGTGACTTACTGCGTGGAGGTCGTGAACCGGTTTGAGGGCATCGTGATG
Optimized 487  AATACCGCGAAAGAAGCCATTGAATATGTCAAACAAATCGATAGCGACAAGATTGGTATC
Original 487   AATACGGCAAAGGAAGCCATCGAGTACGTGAAGCAGATTGACAGTGATAAGATCGGAATC
Optimized 547  CTGCTGGATACGTACCACATGAACATCGAAGAAGGCAGTATTGGTGATGCGATCCGTTCC
Original 547   CTGCTGGATACCTATCATATGAACATCGAGGAAGGCTCTATAGGAGACGCCATCCGATCT
Optimized 607  GTTGGCGGTTATCTGAAAAATTTCCACACGGGCGAAAACAATCGCGTCGTGCCGGGCAAG
Original 607   GTAGGCGGATATCTGAAGAACTTCCACACTGGAGAGAACAACCGGGTCGTTCCGGGGAAG
Optimized 667  GGTCATCTGGATTGGGACGAAATTTTTGGCGCACTGCACGATATTGACTACCAGGGTCGC
Original 667   GGGCACCTCGACTGGGATGAAATATTTGGAGCGCTCCATGATATCGATTATCAGGGAAGG
Optimized 727  ATCGTCTCCGAACCGTTCGTGCAAATGGGCGGTGAAGTGGCTCGTGATATCAAAGTTTGG
Original 727   ATCGTGTCAGAGCCGTTCGTCCAGATGGGCGGGGAAGTCGCAAGAGACATCAAGGTATGG
Optimized 787  CGCGATCTGGTCGAAGACCCGAGCGAAGAAGTTCTGGATGAAGAAGCGCGTTTTCTGCTG
Original 787   AGAGATCTGGTGGAAGATCCTTCAGAAGAAGTGCTGGATGAGGAGGCGCGCTTCCTTCTG
Optimized 847  AATTTCGAAAAAGACATGATTCGCAAGCACTATGGTATCGCCTAA
Original 847   AATTTTGAAAAGGATATGATCCGGAAGCACTATGGCATAGCGTAA
```

FIG. 9

7/10

Clostridium hylemonae
(Optimized Sequence Length:885, GC%:52.20) (SEQ ID NO: 3)
CATATG
AAACACGGTATCTATTACGCCTATTGGGAACAAGAATGGGCAGCAGACTACAAACGCTATGTGGAAAAAGTGGCA
AAACTGGGCTTCGATATTCTGGAAATCGGCGCCGGTCCGCTGCCGGAATATGCAGAACAGGACGTTAAAGAACTG
AAAAAGTGCGCTCAAGATAACGGCATTACCCTGACGGCGGGCTACGGTCCGACCTTTAACCATAATATCGGCAGC
TCTGATGCTGGTGTGCGTGAAGAAGCGCTGGAATGGTATAAACGCCTGTTCGAAGTTCTGGCCGAACTGGACATT
CACCTGATCGGCGGTGCACTGTATAGTTACTGGCCGGTCGATTTTGCTAACGCGGACAAAACGGAAGATTGGAAG
TGGTCCGTGGAGGGTATGCAGCGTCTGGCCCCGGCGGCGGCAAAATACGATATTAACCTGGGTATGGAAGTTCTG
AATCGCTTTGAATCACATATCCTGAATACCGCCGAAGAAGGCGTCAAATTCGTGGAAGAAGTTGGTATGGACAAC
GTGAAGGTTATGCTGGATACGTTCCACATGAATATTGAAGAACAATCGATTGGCGGTGCCATCCGTCGCGCAGGC
AAACTGCTGGGTCATTTTCACACCGGCGAATGTAATCGTATGGTGCCGGGCAAGGGTCGTATTCCGTGGCGCGAA
ATCGGTGACGCTCTGCGTGATATCGGCTACGACGGTACGGCAGTCATGGAACCGTTCGTGCGTATGGGTGGTCAG
GTTGGTGCAGATATTAAAGTCTGGCGTGACATCTCTCGCGGTGCCGATGAAGCACAGCTGGATGACGATGCTCGT
CGCGCGCTGGAATTTCAACGCTATATGCTGGAATGGAAGTAA
AGATCTGGATCC

DNA Alignment (Optimized Region) (Upper: SEQ ID NO: 3; Lower: SEQ ID NO:10

```
Optimized 7    AAACACGGTATCTATTACGCCTATTGGGAACAAGAATGGGCAGCAGACTACAAACGCTAT
Original 7     AAACATGGTATCTATTATGCATACTGGGAACAAGAATGGGCGGCCGACTACAAGCGCTAT
Optimized 67   GTGGAAAAAGTGGCAAAACTGGGCTTCGATATTCTGGAAATCGGCGCCGGTCCGCTGCCG
Original 67    GTTGAAAAGGTGGCAAAGCTTGGGTTTGACATTCTGGAGATCGGCGCTGGGCCGCTGCCG
Optimized 127  GAATATGCAGAACAGGACGTTAAAGAACTGAAAAAGTGCGCTCAAGATAACGGCATTACC
Original 127   GAATACGCAGAGCAGGATGTGAAGGAACTGAAGAAATGTGCGCAGGACAATGGGATCACG
Optimized 187  CTGACGGCGGGCTACGGTCCGACCTTTAACCATAATATCGGCAGCTCTGATGCTGGTGTG
Original 187   CTGACGGCCGGATATGGTCCGACGTTCAACCACAATATCGGTTCTTCAGACGCCGGGGTA
Optimized 247  CGTGAAGAAGCGCTGGAATGGTATAAACGCCTGTTCGAAGTTCTGGCCGAACTGGACATT
Original 247   AGGGAAGAGGCGCTGGAATGGTATAAGAGGTTATTTGAAGTGCTGGCAGAGCTTGATATC
Optimized 307  CACCTGATCGGCGGTGCACTGTATAGTTACTGGCCGGTCGATTTTGCTAACGCGGACAAA
Original 307   CACCTGATCGGAGGGGCGCTCTATTCTTACTGGCCTGTCGATTTTGCAAACGCCGATAAA
Optimized 367  ACGGAAGATTGGAAGTGGTCCGTGGAGGGTATGCAGCGTCTGGCCCCGGCGGCGGCAAAA
Original 367   ACGGAAGACTGGAAGTGGAGTGTAGAGGGCATGCAGAGGCTGGCGCCGGCCGCGGCCAAA
Optimized 427  TACGATATTAACCTGGGTATGGAAGTTCTGAATCGCTTTGAATCACATATCCTGAATACC
Original 427   TATGACATCAACCTGGGCATGGAAGTTCTGAACCGGTTTGAGAGCCATATCCTGAATACA
Optimized 487  GCCGAAGAAGGCGTCAAATTCGTGGAAGAAGTTGGTATGGACAACGTGAAGGTTATGCTG
Original 487   GCCGAGGAAGGTGTGAAGTTTGTAGAGGAAGTCGGCATGGACAACGTAAAGGTCATGCTG
Optimized 547  GATACGTTCCACATGAATATTGAAGAACAATCGATTGGCGGTGCCATCCGTCGCGCAGGC
Original 547   GATACATTCCATATGAATATAGAAGAGCAAAGCATAGGCGGCGCGATCCGCCGGGCAGGA
Optimized 607  AAACTGCTGGGTCATTTTCACACCGGCGAATGTAATCGTATGGTGCCGGGCAAGGGTCGT
Original 607   AAACTGCTCGGGCATTTCCACACCGGAGAATGCAACCGCATGGTGCCCGGGAAGGGACGT
Optimized 667  ATTCCGTGGCGCGAAATCGGTGACGCTCTGCGTGATATCGGCTACGACGGTACGGCAGTC
Original 667   ATTCCATGGCGTGAGATAGGGGATGCTCTCCGTGATATCGGATATGACGGAACTGCTGTA
Optimized 727  ATGGAACCGTTCGTGCGTATGGGTGGTCAGGTTGGTGCAGATATTAAAGTCTGGCGTGAC
Original 727   ATGGAGCCGTTCGTTCGCATGGGAGGACAGGTCGGCGCTGATATCAAGGTGTGGAGAGAC
Optimized 787  ATCTCTCGCGGTGCCGATGAAGCACAGCTGGATGACGATGCTCGTCGCGCGCTGGAATTT
Original 787   ATAAGCCGTGGAGCAGACGAGGCACAGCTTGACGATGACGCGCGCCGTGCGCTGGAGTTC
Optimized 847  CAACGCTATATGCTGGAATGGAAGTAA
Original 847   CAGAGATATATGCTGGAGTGGAAGTAA
```

FIG. 10

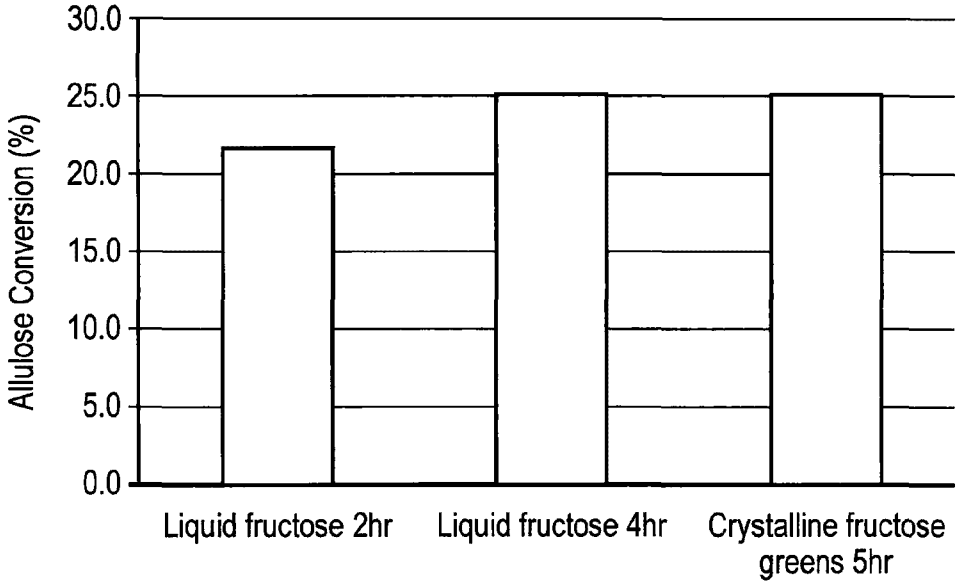

30.0
25.0
20.0
15.0
10.0
5.0
0.0
Allulose Conversion (%)
Liquid fructose 2hr
Liquid fructose 4hr
Crystalline fructose greens 5hr

FIG. 11

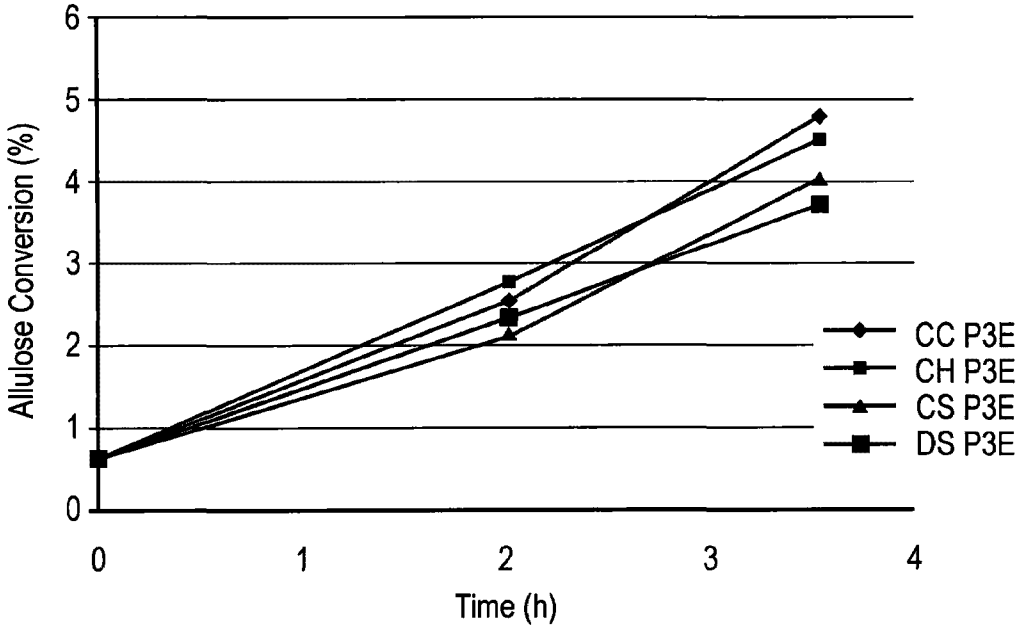

6
5
4
3
2
1
0
Allulose Conversion (%)
0
1
2
3
4
Time (h)
CC P3E
CH P3E
CS P3E
DS P3E

FIG. 12

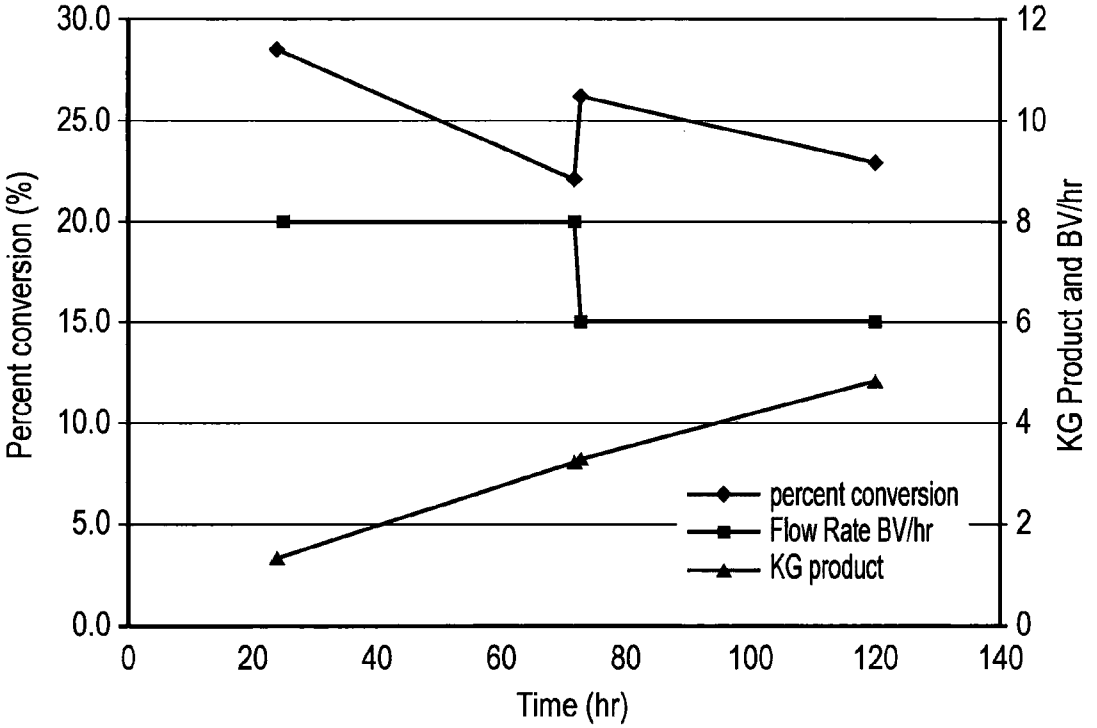

30.0
25.0
20.0
15.0
10.0
5.0
0.0
Percent conversion (%)
12
10
8
6
4
2
0
KG Product and BV/hr
0
20
40
60
80
100
120
140
Time (hr)
percent conversion
Flow Rate BV/hr
KG product

FIG. 13

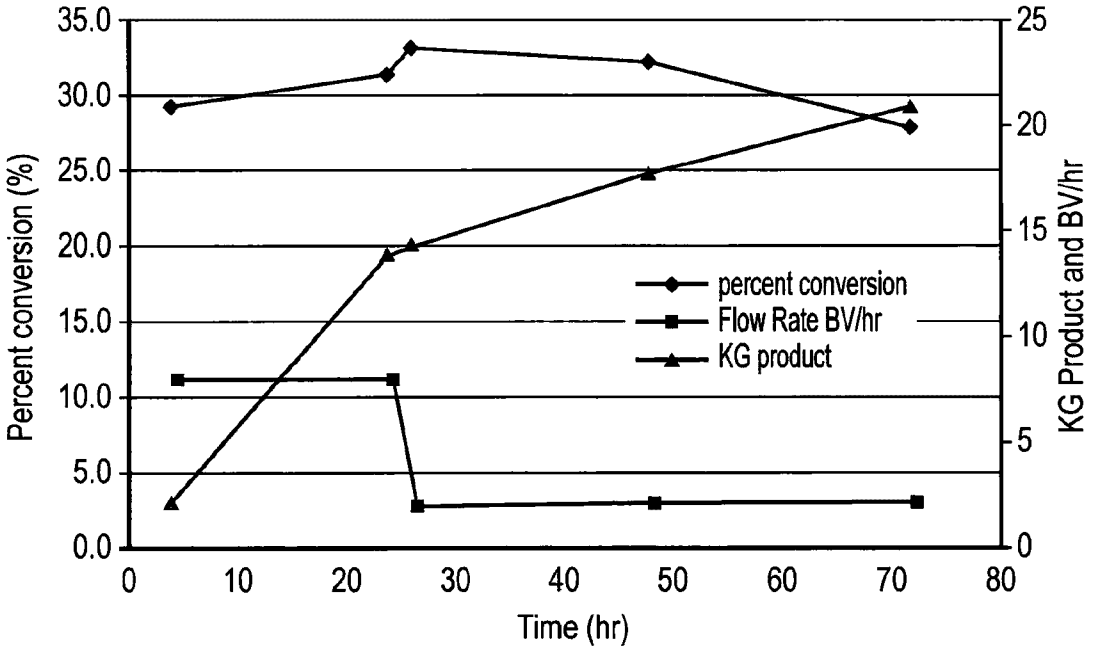
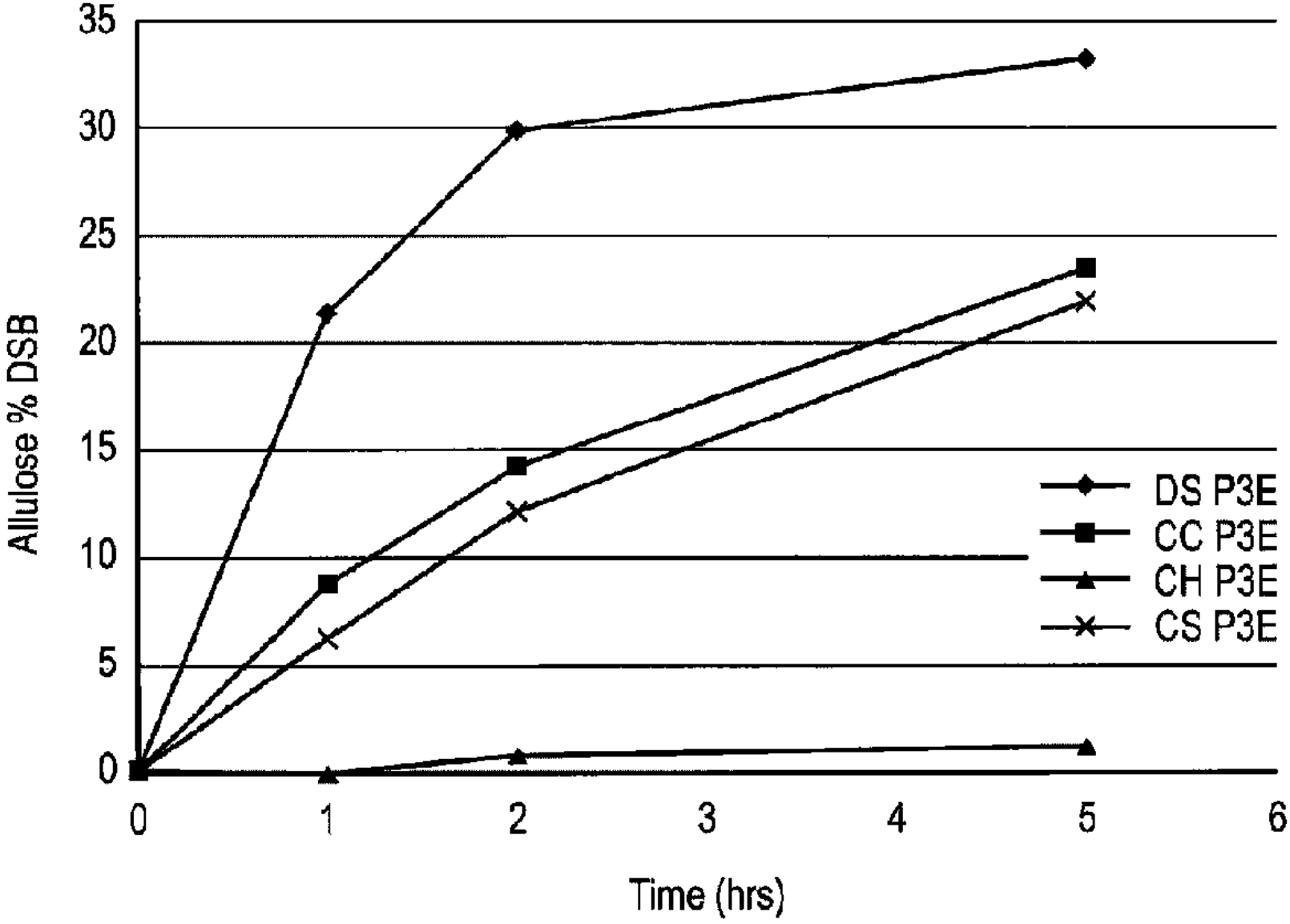

35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Percent conversion (%)
25
20
15
10
5
0
KG Product and BV/hr
0
10
20
30
40
50
60
70
80
Time (hr)
percent conversion
Flow Rate BV/hr
KG product

FIG. 14

MRYGIYYAYWEDSWDADFEKYVKKVKKLGFDIIEVAALGFVNLPEEKLETLRQLA
EQHDIILTAGYGLPKEYNVSSPDKKVRRNGISFMKKVLDAMHQLGIHRIGGTVFS
YWPVDYSCSFDKPAVRKHAIESVREVAEYARQYNITLAIEVLNRFEQFVLNDAEE
AIAYVKEVGEPNVKVMLDTFHMNIEEDHFADAIRYAGDLLGQLHIGEANRKVPGK
GSLPWTEIGQALKDIRYDGYVIMEPFVKTGGTVGRDVKLWRDMSGNATEEQLDRE
LAESLEFVRAAFGE (SEQ ID NO. 13)

FIG. 15


35
30
25
20
15
10
5
0
Allulose % DSB
0
1
2
3
4
5
6
Time (hrs)
DS P3E
CC P3E
CH P3E
CS P3E