Improved process for producing a fermentation product from a lignocellulose-containing material

Jendresen, Christian Bille; Nielsen, Alex Toftgaard

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 August 2017 (31.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/144670 A1

(51) **International Patent Classification**:
C12P 5/00 (2006.01) ***C12N 9/10*** (2006.01)
C12P 11/00 (2006.01)

(21) **International Application Number**:
PCT/EP2017/054344

(22) **International Filing Date**:
24 February 2017 (24.02.2017)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
16157232.6 24 February 2016 (24.02.2016) EP

(71) **Applicant**: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) **Inventors**: **JENDRESEN, Christian Bille**; Rovsingsgade 15, 3. th., 2100 Copenhagen Ø (DK). **NIELSEN, Alex Toftgaard**; Pennehave 3E, 2960 Rungsted Kyst (DK).

(74) **Agent**: **ZACCO DENMARK A/S**; Arne Jacobsens Allé 15, 2300 Copenhagen S (DK).

(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, DJ, 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, KH, KN, KP, KR, KW, 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))*

WO 2017/144670 A1

(54) **Title**: IMPROVED PROCESS FOR PRODUCING A FERMENTATION PRODUCT FROM A LIGNOCELLULOSE-CONTAINING MATERIAL

(57) **Abstract**: The present invention relates to the production of hydrolyzates from a lignocellulose-containing material, and to fermentation of the hydrolyzates. More specifically, the present invention relates to the detoxification of phenolic inhibitors and toxins formed during the processing of lignocellulose-containing material by sulfating the phenolic inhibitors and toxins using aryl sulfotransferase (EC 2.8.2.1) and sulfate transporter.

Improved process for producing a fermentation product from a lignocellulose-containing material

Technical field of the invention

The present invention relates to the production of hydrolyzates from a lignocellulose-containing material, and to fermentation of the hydrolyzates. More specifically, the present invention relates to the detoxification of phenolic inhibitors and toxins formed during the processing of lignocellulose-containing material by sulfating the phenolic inhibitors and toxins.

Background of the invention

It is desired to change energy management increasingly from fossil fuels to renewable energies. The European Union, for instance, has established the serious ambition to increase the proportion of renewable energies to at least 10% in the traffic sector until the year 2020 (Directive 2009/28/EC of the European Parliament and of the Council). One renewable energy source of increasing importance is biofuel. In contrast to electric energy obtained from renewable energy sources, biofuel is well storable and universally usable. Biogas and biofuel are obtained from the conversion of biomass.

At present, the majority of biomass used for the production of biofuel predominantly originates from edible and forage crops. The production of biofuel by using edible and forage crops is designated as biofuel production of the first generation. In temperate climate zones such as, e.g., in Europe and in the U.S.A., mostly corn, wheat, rye and sugar beets are used. In the tropical climate zone such as, e.g., in Brazil, mostly sugar cane is used. Therefore, the production of biofuel by production methods of the first generation directly competes with the production of foodstuff. In the past years, this led to perceptible price rises of foodstuff with severe consequences for the nutrition of the population of the Third World. Exemplarily, from 2010 to 2011, the wheat price increased by 44% and the corn price even increased by 66%. Further, the production of biofuel by using forage crops is comparably ineffective and large cultivated farming areas are typically used. Further, extensive fertilization and manuring of large areas is required. For these reasons, several methods have been developed to generate biofuel by using lignocellulosic biomass. Here,

not only crops, but also inedible plants and trees are used. Furthermore, the waste residues of the crops and trees are used, such as straw, leaves and bark. The production of biofuel by using lignocellulosic biomass is also designated as biofuel production of the second generation. The biofuel production of the second generation has the advantage that it is not in direct competition with food production and fertilizers are often abdicable. A large spectrum of biomass resources can be used, as overall biomass averagely comprises approximately 70% lignocellulose. Therefore, in principle, the production of biofuel by using lignocellulosic biomass is a promising approach to overcome many of the above-referenced problems.

However, the efficient production of fermentation products such as biofuel by using lignocellulosic biomass is still hampered by the poor conversion of lignocellulose into sugars, and by the generation of inhibitory phenolic compounds during the pre-processing and hydrolysis of the biomass. These compounds are inhibitory to the microorganisms that are used for producing biofuels and biochemicals from the biomass hydrolyzate.

Summary of the invention

The object of the present invention is to decrease the concentration of inhibitory phenolic compounds that are present in the biomass hydrolysate. The invention also focuses on improving the tolerance of the production organisms towards inhibitory phenolic compounds. This is done by modifying inhibitory compounds such as phenolic lignin derivatives formed during the processing of the biomass such to make them less toxic for the production organisms.

The present invention thus provides in a first aspect a process for producing a hydrolyzed product from a lignocellulose-containing material, comprising the steps of:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate; and

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

More particularly, the present invention provides a process for producing a hydrolyzed product from a lignocellulose-containing material, comprising the steps of:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b);

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The present invention provides in a further aspect a process for the production of a fermentation product, from a lignocellulose-containing material, comprising the steps:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

(d) fermenting the hydrolyzate obtained in step (b) using a fermenting organism, thereby obtaining a fermentation product.

More particularly, the present invention provides a process for the production of a fermentation product, from a lignocellulose-containing material, comprising the steps:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

(d) fermenting the hydrolyzate obtained in step (b) using a fermenting organism, thereby obtaining a fermentation product;

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has

been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The invention can be further summarized by the following items:

1. A process for producing a hydrolyzed product from a lignocellulose-containing material, comprising the steps of:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b);

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

2. The process according to item 1, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

3. The process according to item 1, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

4. The process according to item 1, wherein the heterologous polypeptide is selected from the group consisting of:

i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

5. The process according to item 3 or 4, wherein the recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide.

6. The process according to item 5, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

7. The process according to item 5 or 6, wherein the exogenous nucleic acid molecule is a vector.

8. The process according to item 5 or 6, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said first recombinant host cell.

9. The process according to any one of items 1-8, wherein said recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

10. The process according to item 9, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter.

11. The process according to item 10, wherein the increase in the number of copies of the gene or genes is achieved by introducing into said recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

12. The process according to any one of items 1-11, wherein said recombinant host cell comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a sulfate transporter.

13. The process according to item 11, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

14. The process according to any one of items 11-13, wherein the exogenous nucleic acid molecule is a vector.

15. The process according to any one of items 11-13, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said first recombinant host cell.

16. The process according to any one of items 9-15, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

17. The process according to any one of item 9-16, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

18. The process according to any one of items 9-17, wherein the sulfate transporter is a bacterial sulfate transporter.

19. The process according to any one of items 9-18, wherein the sulfate transporter is selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

20. The process according to any one of items 9-19, wherein the sulfate transporter is a member of the CysZ family.

21. The process according to any one of items 9-20, wherein the sulfate transporter is a CysZ protein.

22. The process according to any one of items 9-21, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

23. The process according to any one of items 9-21, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

24. The process according to any one of items 9-19, wherein the sulfate transporter is a sulfate-transporting ATPase.

25. The process according to any one of items 9-19, wherein the sulfate transporter is a member of the SulT (cysPTWA) family.

26. The process according to any one of items 9-19, wherein the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

27. The process according to any one of items 24 to 26, where the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

28. The process according to item 26 or 27, wherein the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

29. The process according to any one of items 26 to 28, wherein the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

30. The process according to any one of items 26 to 29, wherein the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

31. The process according to any one of items 26 to 30, wherein the periplasmic binding protein is CysP.

32. The process according to any one of items 26 to 30, wherein the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

33. The process according to any one of items 26 to 30, wherein the periplasmic binding protein is Sbp.

34. The process according to any one of items 26 to 30, wherein the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

35. The process according to any one of items 9-19, wherein the sulfate transporter is a member of the SulP family.

36. The process according to any one of items 9-19, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

37. The process according to any one of items 9-19, wherein the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

38. The process according to any one of items 9-19, wherein the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

39. The process according to any one of items 1-38, wherein the recombinant host cell has been further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

40. The process according to item 39, wherein the ATP sulfurylase is encoded by the genes cysD and cysN.

41. The process according to item 40 or 41, wherein the ATP sulfurylase comprises a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29.

42. The process according to any one of items 1-41, wherein said recombinant host cell has been further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

43. The process according to item 42, wherein the said APS kinase is encoded by the gene cysC.

44. The process according to item 42 or 43, wherein the APS kinase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32.

45. The process according to any one of items 1-44, wherein said recombinant host cell has been further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

46. The process according to item 45, wherein said PAP phosphatase is encoded by the gene cycQ.

47. The process according to item 46 or 47, wherein the PAP phosphatase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37.

48. The process according to any one of items 39-47, wherein the increase in protein expression is achieved by increasing the number of copies of the encoding gene or genes.

49. The process according to item 48, wherein the increase in the number of copies of the gene or genes is achieved by introducing into said first recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

50. The process according to any one of items 39-49, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

51. The process according to any one of items 39-50, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

52. The process according to any one of items 1-51, wherein the recombinant host cell is selected from the group consisting of bacteria, yeasts, fungi, and algae.

53. The process according to any one of items 1-52, wherein the recombinant host cell is a bacterium.

54. The process according to item 53, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium*, *Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus,* or *Yersinia*.

55. The process according to item 53 or 54, wherein the bacterium is a bacterium of the genus *Bacillus*.

56. The process according to item 23, wherein the bacterium is *Bacillus subtilis*.

57. The process according to item 53 or 54, wherein the bacterium is a bacterium of the genus *Lactococcus*.

58. The process according to item 57, wherein the bacterium is *Lactococcus lactis*.

59. The process according to item 53 or 54, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

60. The process according to item 59, wherein the bacterium is *Pseudomonas putida*.

61. The process according to item 53 or 54, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

62. The process according to item 61, wherein the bacterium is *Corynebacterium glutamicum*.

63. The process according to item 53 or 54, wherein the bacterium is a bacterium of the genus *Escherichia*.

64. The process according to item 63, wherein the bacterium is *Escherichia coli*.

65. The process according to any one of item 1-52, wherein the recombinant host cell is a yeast.

66. The process according to item 65, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

67. The process according to any one of items 1-52, wherein the recombinant host cell is a fungus.

68. The process according to item 67, wherein the fungus is a fungus of the genus *Aspergillus*.

69. The process according to any one of items 1-52, wherein the recombinant host cell is an algae cell.

70. The process according to item 69, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

71. The process according to any one of items 1-70, wherein in step (a) the lignocellulose-containing material is pre-treated chemically, mechanically and/or biologically.

72. The process according to any one of items 1-71 wherein in step (a) the lignocellulose-containing material is pre-treated chemically.

73. The process according to any one of items 1-72, wherein in step (a) the lignocellulose-containing material is pre-treated mechanically.

74. The process according to any one of items 1-73, wherein in step (a) the lignocellulose-containing material is pre-treated biologically.

75. The process according to any one of items 1-74, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed chemically.

76. The process according to item 75, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed by acid treatment.

77. The process according to item 76, wherein the acid treatment is dilute acid treatment.

78. The process according to any one of items 1-74, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed enzymatically.

79. The process according to item 78, wherein in step (b) hydrolysis is carried out in the presence of at least one cellulolytic enzyme.

80. The process according to item 79, wherein the at least one cellulolytic enzyme is selected from the group consisting of cellobiohydrolases, endoglucanases and beta-glucosidases and combinations thereof.

81. The process according to any one of items 78-80, wherein in step (b) hydrolysis is carried out in the presence of at least one hemicellulolytic enzyme.

82. The process according to item 81, wherein the at least one hemicellulolytic enzyme is a hemicellulose.

83. The process according to item 81 or 82, wherein the at least one hemicellulolytic enzyme is selected from the group consisting of xylanases, arabinofuranosidases, acetyl xylan esterase, feruloyl esterase, glucuronidases, endo-galactanase, mannases, endo or exo arabinases, exo-galactanses and combinations thereof.

84. The process according to any one of items 1-83, wherein step (c) is carried out during or after the pre-treatment step (a) and/or during or after the hydrolysis step (b).

85. The process according to any one of items 1-84, wherein step (c) is carried out after step (a).

86. The process according to any one of items 1-85, wherein step (c) is carried out after step (b).

87. The process according to any one of items 1-84, wherein step (a) and step (c) are carried out simultaneously.

88. The process according to any one of items 1-84, wherein step (b) and step (c) are carried out simultaneously.

89. A process for the production of a fermentation product from a lignocellulose-containing material, comprising the steps:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

(d) fermenting the hydrolyzate obtained in step (b) using a fermenting organism, thereby obtaining a fermentation product;

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

90. The process according to item 89, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

91. The process according to item 89, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

92. The process according to claim 89, wherein the heterologous polypeptide is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

93. The process according to item 89 to 92, wherein the recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide.

94. The process according to item 93, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

95. The process according to item 93 or 94, wherein the exogenous nucleic acid molecule is a vector.

96. The process according to item 93 or 94, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said first recombinant host cell.

97. The process according to any one of items 89-96, wherein said recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

98. The process according to item 97, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter.

99. The process according to item 98, wherein the increase in the number of copies of the gene or genes is achieved by introducing into said recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

100. The process according to any one of items 89-99, wherein said recombinant host cell comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a sulfate transporter.

101. The process according to item 100, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

102. The process according to any one of items 99-101, wherein the exogenous nucleic acid molecule is a vector.

103. The process according to any one of items 99-101, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said first recombinant host cell.

104. The process according to any one of items 97-103, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

105. The process according to any one of item 97-104, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

106. The process according to any one of items 97-105, wherein the sulfate transporter is a bacterial sulfate transporter.

107. The process according to any one of items 97-106, wherein the sulfate transporter is selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

108. The process according to any one of items 97-107, wherein the sulfate transporter is a member of the CysZ family.

109. The process according to any one of items 97-107, wherein the sulfate transporter is a CysZ protein.

110. The process according to any one of items 97-109, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

111. The process according to any one of items 97-109, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

112. The process according to any one of items 97-107, wherein the sulfate transporter is a sulfate-transporting ATPase.

113. The process according to any one of items 97-107, wherein the sulfate transporter is a member of the SulT (cysPTWA) family.

114. The process according to any one of items 97-107, wherein the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

115. The process according to any one of items 112 to 114, where the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

116. The process according to item 114 or 115, wherein the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

117. The process according to any one of items 114 to 116, wherein the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

118. The process according to any one of items 114 to 117, wherein the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

119. The process according to any one of items 114 to 118, wherein the periplasmic binding protein is CysP.

120. The process according to any one of items 114 to 118, wherein the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

121. The process according to any one of items 114 to 118, wherein the periplasmic binding protein is Sbp.

122. The process according to any one of items 114 to 118, wherein the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

123. The process according to any one of items 97-107, wherein the sulfate transporter is a member of the SulP family.

124. The process according to any one of items 97-107, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

125. The process according to any one of items 97-107, wherein the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

126. The process according to any one of items 97-107, wherein the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

126. The process according to any one of items 89-125, wherein the recombinant host cell has been further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

127. The process according to item 126, wherein the ATP sulfurylase is encoded by the genes cysD and cysN.

128. The process according to item 126 or 127, wherein the ATP sulfurylase comprises a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29.

129. The process according to any one of items 89-128, wherein said recombinant host cell has been further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

130. The process according to item 129, wherein the said APS kinase is encoded by the gene cysC.

131. The process according to item 129 or 130, wherein the APS kinase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32.

132. The process according to any one of items 89-131, wherein said recombinant host cell has been further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

133. The process according to item 132, wherein said PAP phosphatase is encoded by the gene cycQ.

134. The process according to item 132 or 133, wherein the PAP phosphatase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37.

135. The process according to any one of items 126-134, wherein the increase in protein expression is achieved by increasing the number of copies of the encoding gene or genes.

136. The process according to item 135, wherein the increase in the number of copies of the gene or genes is achieved by using one or more vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

137. The process according to any one of items 126-136, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

138. The process according to any one of items 126-137, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

139. The process according to any one of items 89 to 138, wherein the recombinant host cell is independently selected from the group consisting of bacteria, yeasts, fungi, and algae.

140. The process according to any one of items 89 to 139 wherein the recombinant host cell is a bacterium.

141. The process according to item 140, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium*, *Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus,* or *Yersinia*.

142. The process according to item 141, wherein the bacterium is a bacterium of the genus *Bacillus*.

143. The process according to item 142, wherein the bacterium is *Bacillus subtilis*.

144. The process according to item 141, wherein the bacterium is a bacterium of the genus *Lactococcus*.

145. The process according to item 144, wherein the bacterium is *Lactococcus lactis*.

146. The process according to item 141, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

147. The process according to item 146, wherein the bacterium is *Pseudomonas putida*.

148. The process according to item 147, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

149. The process according to item 148, wherein the bacterium is *Corynebacterium glutamicum*.

150. The process according to item 147, wherein the bacterium is a bacterium of the genus *Escherichia*.

151. The process according to item 150, wherein the bacterium is *Escherichia coli*.

152. The process according to any one of item 89-139, wherein the recombinant host cell is a yeast.

153. The process according to item 152, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

154. The process according to any one of items 89-139, wherein the recombinant host cell is a fungus.

155. The process according to item 154, wherein the fungus is a fungus of the genus *Aspergillus*.

156. The process according to any one of items 89-139, wherein the recombinant host cell is an algae cell.

157. The process according to item 156, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

158. The process according to any one of items 89-157, wherein in step (a) the lignocellulose-containing material is pre-treated chemically, mechanically and/or biologically.

159. The process according to any one of items 89-158, wherein in step (a) the lignocellulose-containing material is pre-treated chemically.

160. The process according to any one of items 89-159, wherein in step (a) the lignocellulose-containing material is pre-treated mechanically.

161. The process according to any one of items 89-160, wherein in step (a) the lignocellulose-containing material is pre-treated biologically.

162. The process according to any one of items 89-161, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed chemically.

163. The process according to item 162, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed by acid treatment.

164. The process according to item 163, wherein the acid treatment is dilute acid treatment.

165. The process according to any one of items 89-161, wherein in step (b) the pre-treated lignocellulose-containing material is hydrolyzed enzymatically.

166. The process according to item 165, wherein in step (b) hydrolysis is carried out in the presence of at least one cellulolytic enzyme.

167. The process according to item 166, wherein the at least one cellulolytic enzyme is selected from the group consisting of cellobiohydrolases, endoglucanases and beta-glucosidases and combinations thereof.

168. The process according to any one of items 165-167, wherein in step (b) hydrolysis is carried out in the presence of at least one hemicellulolytic enzyme.

169. The process according to item 168, wherein the at least one hemicellulolytic enzyme is a hemicellulose.

170. The process according to item 168 or 169, wherein the at least one hemicellulolytic enzyme is selected from the group consisting of xylanases, arabinofuranosidases, acetyl xylan esterase, feruloyl esterase, glucuronidases, endo-galactanase, mannases, endo or exo arabinases, exo-galactanses and combinations thereof.

171. The process according to any one of items 89-170, wherein the fermenting organism in step (d) is the recombinant host cell used in step (c).

172. The process according to any one of items 89-171, wherein step (c) is carried out during or after the pre-treatment step (a) and/or during or after the hydrolysis step (b).

173. The process according to any one of items 89-171, wherein step (c) is carried out after step (a).

174. The process according to any one of items 89-173, wherein step (c) is carried out after step (b).

175. The process according to any one of items 89-171, wherein step (a) and step (c) are carried out simultaneously.

176. The process according to any one of items 89-171, wherein step (b) and step (c) are carried out simultaneously.

177. The process according to any one of items 89-176, wherein step (d) is carried out after step (d).

178. The process according to any one of items 89-176, wherein step (d) and step (b) are carried out simultaneously.

179. The process according to any one of items 89-176, wherein step (c) is carried out after step (b) and step (d) is carried out after step (c).

180. The process according to any one of items 89-76, wherein steps (b) to (d) are carried out simultaneously.

181. The process according to any one of items 89-180, wherein the fermentation product is a biofuel or biochemical.

182. The process according to any one of items 89-181, wherein the fermentation product is an alcohol.

183. The process according to any one of items 89-181, wherein the fermentation product is ethanol.

184. The process according to any one of items 89-181, wherein the fermentation product is an organic acid.

185. The process according to any one of items 89-181, wherein the fermentation product is a ketone.

186. The process according to any one of items 89-181, wherein the fermentation product is a diamine.

187. The process according to any one of items 89-181, wherein the fermentation product is a diol.

188. The process according to any one of items 89-181, wherein the fermentation product is a triol.

189. The process according to any one of items 89-181, wherein the fermentation product is a diene.

190. The process according to any one of items 89-181, wherein the fermentation product is an isoprenoid.

Brief description of the figures

Figure 1: Map of plasmid for expression of SULT1A1 from *Rattus norvegicus* in *Escherichia coli*

Figure 2: Map of plasmid for over-expression of cysDNC in *E. coli.*

Figure 3: Map of plasmid for over-expression of cysDNCQ in *E. coli.*

Figure 4: Toxicity of unsulfated or sulfated products

Figure 5: Growth curves of cultures without sulfotransferase (solid lines) or with SULT1A1 (dotted lines) growing in M9 without supplements (black), with 8% biomass hydrolysate (BH8%, dark grey), or with 12% biomass hydrolysate (BH12%, light grey).

Figure 6: Map of plasmid for over-expression of cysZ in *E. coli*

Figure 7: Map of plasmid for over-expression of cysPTWA in *E. coli*

Figure 8: Concentrations of zosteric acid in culture media with *E. coli* over-expressing SULT1A1 from *Rattus norvegicus*, either alone or in combination with over-expressing cysDNCQ and cysZ or cysPTWA

Figure 9: Map of plasmid for the expression of CysZ and CysDNCQ

Detailed description of the invention

Lignocellulose is a mass of insoluble organic material. Typically, lignocellulose is composed of approximately 40-50% cellulose, approximately 20-30% hemicellulose and 20-30% lignin. Cellulose is a linear polymer composed of glucose polymerized by an alpha-1,4 glycosidic linkage. It typically forms microcrystalline structures which can only be poorly dissolved and hydrolyzed. Hemicellulose is a heteropolysaccharide that is composed of different hexoses,

pentoses and glucuronic acid. The hemicellulose xylane is often found in grass and wood. Lignin is an insoluble polymer of aromatic alcohols known as monolignols, such as coniferyl alcohol, sinapyl alcohol and paracoumaryl alcohol. The cellulose microfibrils are conjugated with another by hemicellulose and/or lignin by covalent and non-covalent bonds. These bonds, in particular the covalent bonds, are highly stable and nearly inert against chemical and biological hydrolysis.

Lignocellulose is poorly accessible for most of the cellulose degrading enzymes. Therefore, a pretreatment of lignocellulose is an important step to obtain higher yields of sugars that can be further converted into downstream products.

However, during the pretreatment, undesirable fermentation inhibiting agents are generated. Further, lignin cannot be recovered by most of the methods employed in the art. The pretreatment of lignocellulose is typically the most expensive and laborious step of the production of bioethanol and the costs for said pretreatment step sums up to approximately 20-30% of the total costs of bioethanol. However, in comparison with bioethanol production with no pretreatment, the pretreatment of lignocellulose-containing biomass can still reduce the costs per liter bioethanol approximately 6-fold due to higher yields.

In general, the pretreatment can be enabled by physical, chemical and physicochemical means. Physical pretreatment may be, e.g., grinding, crushing, irradiation (e.g., gamma irradiation, cathode ray or microwave irradiation) and/or explosion (e.g., steam explosion, $C0_2$ explosion or $S0_2$ explosion). Chemical pretreatment may be, e.g., treatment with bases (basic hydrolysis) (e.g., sodium hydroxide solution and/or ammonia solution) or treatment with diluted acids (acidic hydrolysis) (e.g., sulfuric acid, hydrochloric acid, phosphoric acid and/or nitric acid). Physicochemical pretreatment may be, e.g., gas treatment (e.g., treatment with chlorine dioxide and/or sulfur dioxide), oxidation (e.g., hydrogen peroxide, active or oxygen or ozone treatment) and/or extraction of lignin (e.g., by a butyl alcohol solution and/or by an ethanol solution). Often, two or more of the aforementioned methods are also combined with another. Biomass subjected to grinding, crushing or the extraction of lignin is typically used for the production of biofuel, in particular ethanol. Biomass subjected to irradiation, explosion, any chemical treatment, gas treatment or oxidation is typically used for the production of biofuel. However, the pre-treatment

methods described above, result in the generation of inhibitors that interfere with the further fermentation steps. These inhibitors typically have to be removed by costly and laborious means. Especially acidic and heat-based pretreatment often leads to the production of inhibitors of a following fermentation steps, which may severely hamper the production of biofuel or biochemical such as amino acids. These inhibitors are often weak acids (e.g., acetic acid, formic acid, ferulic acid), furan derivatives (e.g., furfural and 5-hydroxymethylfurfural) and/or lignin derivatives (phenolic compounds/phenol derivatives such as vanillin and 4-hydroxybenzaldehyde). It has been shown that many types of yeast are inhibited by phenol derivatives which may occur upon pretreatment with acids or steam. Therefore, it is an object of the present invention to convert the inhibitors to non-toxic compounds, or at least convert them to less toxic/inhibiting compounds.

Thus, the present invention provides in a first aspect a process for producing a hydrolyzed product from a lignocellulose-containing material, comprising the steps of:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate; and

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b);

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The term "lignocellulose-containing material" used herein refers to material that comprises, or primarily consists of, cellulose, hemicellulose, and lignin. The term is synonymous with "lignocellulosic material". Such material is often referred to as "biomass".

The structure of lignocellulose is not directly accessible to hydrolysis, and in particular to enzymatic hydrolysis. Therefore, the lignocellulose-containing material has to be pre-treated, e.g., by acid hydrolysis under adequate conditions of pressure and temperature, in order to break the lignin seal and disrupt the crystalline structure of cellulose. This causes solubilization of the hemicellulose and cellulose fractions. The lignocellulose-containing

material may be pre-treated in any suitable way. Pre-treatment may be carried out before and/or during hydrolysis and/or fermentation. According to certain embodiments, the pre-treated material is hydrolyzed, preferably enzymatically, before and/or during fermentation. The goal of pre-treatment is to separate and/or release cellulose; hemicellulose and/or lignin and this way improve the rate of hydrolysis. Pre-treatment methods such as wet-oxidation and alkaline pre-treatment targets lignin, while dilute acid and auto- hydrolysis targets hemicellulose. Steam explosion is an example of a pre-treatment that targets cellulose.

When lignocellulose-containing material is pre-treated, degradation products that may inhibit enzymes and/or may be toxic to fermenting organisms are produced. Many of the phenolic compounds are released from the degrading lignin. Examples of such phenolic compounds are 4-OH benzyl alcohol, 4-OH benzaldehyde, 4-OH benzoic acid, coumaric acid, ferulic acid, phenol, guaiacol, pyrogallol, pyrogallol mono methyl ether, vanillyl alcohol, vanillin, isovanillin, vanillic acid, isovanillic acid, homovanillic acid, gallic acid, 2-O-methyl gallic acid, syringyl alcohol, syringylaldehyde, syringic acid, homocatechol, ethyl vanillin, creosol, coniferyl aldehyde, isoeugenol, hydroquinone, eugenol or combinations thereof. Other inhibitory compounds can be found in, e.g., Luo et al., 2002, Biomass and Bioenergy 22: 125-138, which reference is hereby incorporated by reference.

According to the invention the pre-treatment applied in step (a) may be a conventional pre-treatment step using techniques well known in the art. Examples of suitable pre-treatments are disclosed above. In a preferred embodiment, pre-treatment takes place in aqueous slurry.

Any lignocellulose-containing material is contemplated according to the present invention. The lignocellulose-containing material may be any material containing lignocellulose. In a preferred embodiment the lignocellulose-containing material contains at least 30 wt. %, preferably at least 50 wt. %, more preferably at least 70 wt. %, even more preferably at least 90 wt. % lignocellulose. It is to be understood that the lignocellulose-containing material may also comprise other constituents such as proteinaceous material, starch, sugars, such as fermentable sugars and/or un-fermentable sugars.

Lignocellulose-containing material is generally found, for example, in the stems, leaves, hulls, husks, and cobs of plants or leaves, branches, and wood of trees. Lignocellulose-containing material can also be, but is not limited to, herbaceous material, agricultural residues, forestry residues, municipal solid wastes, waste paper, and pulp and paper mill residues. It is understood herein that lignocellulose-containing material may be in the form of plant cell wall material containing lignin, cellulose, and hemi-cellulose in a mixed matrix.

According to certain embodiments, the lignocellulose-containing material comprises one or more of corn stover, corn fiber, rice straw, pine wood, wood chips, poplar, bagasse, paper and pulp processing waste.

Other examples of lignocellulose-containing material include hardwood, such as poplar and birch, softwood, cereal straw, such as wheat straw, switchgrass, municipal solid waste, industrial organic waste, office paper, or mixtures thereof.

The lignocellulose-containing material may according to the invention be chemically, mechanically and/or biologically pre-treated before hydrolysis and/or fermentation. Mechanical treatment (often referred to as physical treatment) may be used alone or in combination with subsequent or simultaneous hydrolysis, especially enzymatic hydrolysis.

Preferably, chemical, mechanical and/or biological pre-treatment is carried out prior to the hydrolysis and/or fermentation. Alternatively, the chemical, mechanical and/or biological pre-treatment may be carried out simultaneously with hydrolysis, such as simultaneously with addition of one or more cellulolytic enzymes, or other enzyme activities mentioned below, to release, e.g., fermentable sugars, such as glucose and/or maltose.

According to certain embodiments, the lignocellulose-containing material is pre-treated chemically. As used herein, the term “pre-treated chemically” or "chemical treatment" refers to any chemical pre-treatment which promotes the separation and/or release of cellulose, hemicellulose and/or lignin. Examples of suitable chemical pre-treatments include treatment with; for example, dilute acid, lime, alkaline, organic solvent, ammonia, sulfur dioxide, carbon dioxide. Further, wet oxidation and pH-controlled hydrothermolysis are also considered chemical pre-treatment. In a preferred embodiment the chemical pre-treatment is acid treatment, more preferably, a continuous dilute and/or mild acid treatment, such as, treatment with sulfuric acid, or another organic acid, such as acetic

acid, citric acid, tartaric acid, succinic acid, hydrogen chloride or mixtures thereof. Other acids may also be used. Mild acid treatment means that the treatment pH lies in the range from about 1 to about 5, such as from about 1 to about 3. In a specific embodiment the acid concentration is in the range from about 0.1 to about 2.0 wt. % acid, preferably sulphuric acid. The acid may be contacted with the lignocellulose-containing material and the mixture may be held at a temperature in the range from about 160 to about 220°C, such as from about 165 to about 195°C, for periods ranging from minutes to seconds, e.g., from about 1 to about 60 minutes, such as from about 2 to about 30 minutes or about 3 to about 12 minutes. Addition of strong acids, such as sulphuric acid, may be applied to remove hemicellulose. This enhances the digestibility of cellulose.

Other techniques are also contemplated. Cellulose solvent treatment has been shown to convert about 90% of cellulose to glucose. It has also been shown that enzymatic hydrolysis could be greatly enhanced when the lignocellulose structure is disrupted. Alkaline H_2O_2, ozone, glycerol, dioxane, phenol, or ethylene glycol are among solvents known to disrupt cellulose structure and promote hydrolysis.

Alkaline chemical pre-treatment with base, e.g., NaOH, Na_2CO_3 and/or ammonia or the like, is also contemplated according to the invention. Pre-treatment methods using ammonia are described in, e.g., WO 2006/110891, WO 2006/110899, WO 2006/110900, and WO 2006/110901 (which are hereby incorporated by reference).

Wet oxidation techniques involve use of oxidizing agents, such as: sulphite based oxidizing agents or the like. Examples of solvent pre-treatments include treatment with DMSO (Dimethyl Sulfoxide) or the like. Chemical pre-treatment is generally carried out for 1 to 60 minutes, such as from 5 to 30 minutes, but may be carried out for shorter or longer periods of time dependent on the material to be pre-treated.

Other examples of suitable pre-treatment methods are described by Schell et al., 2003, Appl. Biochem and Biotechn. Vol. 105-108: 69-85, and Mosier et al., 2005, Bioresource Technology 96: 673-686, and U.S. Publication No. 2002/0164730, which references are hereby all incorporated by reference.

According to certain embodiments, the lignocellulose-containing material is pre-treated mechanically. As used herein, the term “pre-treated mechanically” or "mechanical pre-

treatment" refers to any mechanical (or physical) treatment which promotes the separation and/or release of cellulose, hemicellulose and/or lignin from lignocellulose-containing material. For example, mechanical pre-treatment includes various types of milling, irradiation, steaming/steam explosion, and hydrothermolysis.

Mechanical pre-treatment includes comminution (mechanical reduction of the size). Comminution includes dry milling, wet milling and vibratory ball milling. Mechanical pre-treatment may involve high pressure and/or high temperature (steam explosion). I n an embodiment of the invention high pressure means pressure in the range from about 300 to about 600 psi, such as from about 400 to about 500 psi, such as at about 450 psi. In an embodiment of the invention high temperature means temperatures in the range from about 100 to about 300°C, such as from about 140 to about 235°C. In a preferred embodiment, mechanical pre-treatment is a batch-process, steam gun hydrolyzer system which uses high pressure and high temperature as defined above. A Sunds Hydrolyzer (available from Sunds Defibrator AB (Sweden) may be used for this.

According to particular embodiments, both chemical and mechanical pre-treatments are carried out. For instance, the pre-treatment step may involve dilute or mild acid treatment and high temperature and/or pressure treatment. The chemical and mechanical pre-treatment may be carried out sequentially or simultaneously, as desired.

Accordingly, in a particular embodiment, the lignocellulose-containing material is subjected to both chemical and mechanical pre-treatment to promote the separation and/or release of cellulose, hemicellulose and/or lignin.

In one embodiment the pre-treatment is carried out as a dilute and/or mild acid steam explosion step. In another preferred embodiment pre-treatment is carried out as an ammonia fiber explosion step (or AFEX pre-treatment step).

According to certain embodiments, the lignocellulose-containing material is pre-treated biologically. As used herein, the term “pre-treated biologically” or "biological pre-treatment" refers to any biological pre-treatment which promotes the separation and/or release of cellulose, hemicellulose, and/or lignin from the lignocellulose-containing material. Biological pre-treatment techniques can involve applying lignin-solubilizing microorganisms (see, for example, Hsu, 1996, Pretreatment of biomass, in Handbook on

Bioethanol: Production and Utilization, Wyman, ed., Taylor & Francis, Washington, DC, 179-212; Ghosh and Singh, 1993, Physicochemical and biological treatments for enzymatic/microbial conversion of lignocellulosic biomass, Adv. Appl. Microbiol. 39: 295-333; McMillan, 1994, Pretreating lignocellulosic biomass: a review, in Enzymatic Conversion of Biomass for Fuels Production, Himmel, Baker, and Overend, eds., ACS Symposium Series 566, American Chemical Society, Washington, DC, chapter 15; Gong, Cao, Du, and Tsao, 1999, Ethanol production from renewable resources, in Advances in Biochemical Engineering/Biotechnology, Scheper, ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241 ; Olsson and Hahn-Hagerdal, 1996, Fermentation of lignocellulosic hydrolysates for ethanol production, Enz. Microb. Tech. 18: 312-331; and Vallander and Eriksson, 1990, Production of ethanol from lignocellulosic materials: State of the art, Adv. Biochem. EngJBiotechnol. 42: 63-95).

The pre-treated lignocellulose-containing material is then hydrolyzed to break down cellulose and/or hemicellulose. This process converts the carbohydrate polymers into fermentable sugars which, by using a fermenting organism, e.g. a bacterium or yeast, may be fermented into a desired fermentation product, such as ethanol.

According to certain embodiments, in step (b) the pre-treated lignocellulose-containing material is hydrolyzed chemically, e.g., by acid treatment, such as dilute acid treatment. Suitable conditions for chemical hydrolysis, and more particular acid hydrolysis, of lignocellulose-containing material are well known to one skilled in the art.

According to certain other embodiments, in step (b) the pre-treated lignocellulose-containing material is hydrolyzed enzymatically, e.g., by one or more cellulolytic enzymes, to form a hydrolyzate.

The enzyme(s) used for hydrolysis is (are) capable of directly or indirectly converting carbohydrate polymers into fermentable sugars which can be fermented into a desired fermentation product, such as ethanol.

Hydrolysis in step (b) may also be carried out in the presence of one or more cellulolytic enzymes and/or one or more hemicellulolytic enzymes. According to certain embodiments, hydrolysis in step (b) is carried out in the presence of at least one cellulolytic enzyme. According to certain other embodiments, hydrolysis in step (b) is carried out in the

presence of at least one hemicellulolytic enzyme. According to certain embodiments, hydrolysis in step (b) is carried out in the presence of a combination of at least one cellulolytic enzyme and at least one hemicellulolytic enzyme.

The term "cellulolytic enzymes" as used herein are understood as comprising cellobiohydrolases (EC 3.2.1.91), e.g., cellobiohydrolase I and cellobiohydrolase II, as well as the endoglucanases (EC 3.2.1.4) and beta-glucosidases (EC 3.2.1.21).

1,4-beta-D-glucan cellobiohydrolases (E.C.3.2.1.91), also referred herein as "cellobiohydrolases", catalyze the hydrolysis of 1 ,4-beta-D-glucosidic linkages in cellulose, cello-oligosaccharides, or any beta-1 ,4-linked glucose containing polymer, releasing cellobiose from the reducing or non-reducing ends of the chain.

Endoglucanases (E.C.3.2.1.4) catalyze the endo hydrolysis of 1 ,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (such as carboxy methyl cellulose and hydroxy ethyl cellulose), lichenin, beta-1 ,4 bonds in mixed beta-1 ,3 glucans such as cereal beta-D-glucans or xyloglucans and other plant material containing cellulosic parts. The authorized name is endo-1,4-beta-D-glucan 4-glucanohydrolase, but the abbreviated term endoglucanase is used in the present specification. Endoglucanase activity may be determined using carboxy methyl cellulose (CMC) hydrolysis according to the procedure of Ghose, 1987, Pure andAppl. Chem. 59: 257-268.

"Beta-glucosidases"(E.C.3.2.1.21), also referred herein as “beta-D-glucoside glucohydrolases”, catalyze the hydrolysis of terminal non-reducing beta-D-glucose residues with the release of beta-D-glucose. For purposes of the present invention, beta-glucosidase activity is determined according to the basic procedure described by Venturi et al., 2002, J. Basic Microbiol. 42: 55-66. One unit of beta-glucosidase activity is defined as 1.0 µmole of p-nitrophenol produced per minute at 50°C, pH 5 from 4 mM p-nitrophenyl-beta-D-glucopyranoside as substrate in 100 mM sodium citrate, 0.01% TWEEN(R) 20.

In order to be efficient, the digestion of cellulose may require several types of enzymes acting cooperatively. At least three categories of enzymes are often employed in converting cellulose into glucose: endoglucanases (EC 3.2.1.4) that cut the cellulose chains at random; cellobiohydrolases (EC 3.2.1.91) which cleave cellobiosyl units from the cellulose chain ends and beta-glucosidases (EC 3.2.1.21) that convert cellobiose and soluble cellodextrins into

glucose. Among these three categories of enzymes involved in the biodegradation of cellulose, cellobiohydrolases are the key enzymes for the degradation of native crystalline cellulose.

The term "cellobiohydrolase I" is defined herein as a cellulose 1 ,4-beta-cellobiosidase (also referred to as Exo-glucanase, Exo-cellobiohydrolase or 1 ,4- beta-cellobiohydrolase) activity, as defined in the enzyme class EC 3.2.1.91, which catalyzes the hydrolysis of 1 ,4-beta-D-glucosidic linkages in cellulose and cellotetraose, by the release of cellobiose from the non-reducing ends of the chains. The definition of the term "cellobiohydrolase II activity" is identical, except that cellobiohydrolase II attacks from the reducing ends of the chains.

Hence, according to certain embodiments, at least one, such as at least two, at least three, at least four, at least five, at least six or at least seven, cellulolytic enzymes selected from the group consisting of endoglucanases, cellobiohydrolases, beta-glucosidases and combinations thereof is (are) employed in step (b).

The cellulolytic enzyme may comprise a carbohydrate-binding module (CBM) which enhances the binding of the enzyme to a lignocellulose-containing fiber and increases the efficacy of the catalytic active part of the enzyme. A CBM is defined as contiguous amino acid sequence within a carbohydrate-active enzyme with a discreet fold having carbohydrate- binding activity.

According to particular embodiments, the cellulolytic enzyme is a cellulolytic enzyme preparation, such as a preparation described in U.S. application no. 60/941,251, which is hereby incorporated by reference. According to more particular embodiments, the cellulolytic enzyme preparation comprises a polypeptide having cellulolytic enhancing activity (GH61A), such as the GH61A enzyme from *Thermoascus aurantiacus* as disclosed, e.g., in WO2005/074656 (hereby incorporated by reference). The cellulolytic enzyme preparation may further comprise a beta-glucosidase, such as a beta-glucosidase derived from a strain of the genus *Humicola*, *Trichoderma*, *Aspergillus or Penicillium*, including the *Humicola insolens* CEL45A endoglucanase core/ *Aspergillus oryzae* beta-glucosidase fusion protein disclosed in, e.g., U.S. application no. US 11/781,151 or PCT/US2007/074038 (Novozymes). According to certain embodiment, the cellulolytic enzyme preparation may also comprise a CBH II, such as *Thielavia terrestris* cellobiohydrolase II (CEL6A). According to

other certain embodiments, the cellulolytic enzyme preparation also comprises a cellulase enzyme preparation, such as a cellulase enzyme preparation derived from *Trichoderma reesei*.

According to certain embodiments, the cellulolytic enzyme preparation comprises a polypeptide having cellulolytic enhancing activity (GH61A); a cellobiohydrolase, such as Thielavia terrestris cellobiohydrolase II (CEL6A), a beta-glucosidase (e.g., the fusion protein disclosed in U.S. application no. 60/832,511) and cellulolytic enzymes, e.g., derived from *Trichoderma reesei*.

The term "cellulolytic enhancing activity" is defined herein as a biological activity that enhances the hydrolysis of a lignocellulose derived material by proteins having cellulolytic activity. For purposes of the present invention, cellulolytic enhancing activity is determined by measuring the increase in reducing sugars or in the increase of the total of cellobiose and glucose from the hydrolysis of a lignocellulose derived material, e.g., pre-treated lignocellulose-containing material by cellulolytic protein under the following conditions: 1-50 mg of total protein/g of cellulose in PCS (pre-treated corn stover), wherein total protein is comprised of 80-99.5% w/w cellulolytic protein/g of cellulose in PCS and 0.5-20% w/w protein of cellulolytic enhancing activity for 1-7 day at 50°C compared to a control hydrolysis with equal total protein loading without cellulolytic enhancing activity (1 -50 mg of cellulolytic protein/g of cellulose in PCS).

According to certain embodiments, the hydrolysis in step (b) is carried out in the presence of at least one cellulolytic enzyme in combination with a polypeptide having enhancing activity. According to particular embodiments, the polypeptide having enhancing activity is a family GH61A polypeptide. For example, WO2005/074647 discloses isolated polypeptides having cellulolytic enhancing activity and polynucleotides thereof from *Thielavia terrestris*, WO2005/074656 discloses an isolated polypeptide having cellulolytic enhancing activity and a polynucleotide thereof from *Thermoascus aurantiacus*, and U.S. Application Publication No. 2007/0077630 discloses an isolated polypeptide having cellulolytic enhancing activity and a polynucleotide thereof from *Trichoderma reesei*.

According to certain embodiments, the cellulolytic enzyme is the commercially available product CELLUCLAST(R) 1.5L or CELLUZYME(TM) available from Novozymes A/S, Denmark, or ACCELERASE(TM) 1000 available from Genencor Inc., USA.

The cellulolytic enzyme may, in accordance to particular embodiments, be derived from a fungal source, such as from a strain of the genus *Trichoderma*, such as from a strain of *Trichoderma reesei*; or from a strain of the genus *Humicola*, such as from a strain of *Humicola insolens*; or from a strain of the genus *Chrysosporium*, such as from a strain of *Chrysosporium lucknowense*.

According to certain embodiments, at least one endoglucanase is employed in step (b).

According to particular embodiments, the endoglucanase is derived from a strain of the genus *Trichoderma*, such as from a strain of *Trichoderma reesei*; from a strain of the genus *Humicola*, such as from a strain of *Humicola insolens*; or from a strain of the genus *Chrysosporium*, such as from a strain of *Chrysosporium lucknowense*.

According to certain embodiments, at least one beta-glucosidase is employed in step (b).

According to certain embodiments, the beta-glucosidase is of fungal origin, such as derived from a strain of the genus *Trichoderma*, *Aspergillus* or *Penicillium*. According to particular embodiments, the beta-glucosidase is a derived from *Trichoderma reesei*, such as the beta-glucosidase encoded by the *bgl1* gene (see Fig. 1 of EP 562003). According to other particular embodiments, the beta-glucosidase is derived from *Aspergillus oryzae*, *Aspergillus fumigatus* (e.g., recombinantly produced in *Aspergillus oryzae* according to Example 22 of WO02/095014) or *Aspergillus niger*.

According to certain embodiments, at least one cellobiohydrolase is employed in step (b).

The one or more cellulolytic enzymes may be dosed in the range from 0.1-100 FPU per gram dry solids (DS), preferably 0.5-50 FPU per gram dry solids, especially 1-20 FPU per gram dry solids. The cellulolytic enzyme may be dosed in the range from 1-1000 EGU per gram dry solids, preferably 10-500 EGU per gram dry solids, especially 50 to 200 EGU per gram dry solids.

According to certain embodiments, at least 1 mg cellulolytic enzyme per gram dry solids, such as at least 2 mg or at least 3 mg cellulolytic enzyme per gram dry solids, such as between 5 and 10 mg cellulolytic enzyme(s) is(are) used for hydrolysis.

Hemicellulose polymers can be broken down by hemicellulases and/or acid hydrolysis to release its five and six carbon sugar components. The six carbon sugars (hexoses), such as glucose, galactose, arabinose, and mannose, can readily be fermented to, e.g., ethanol, acetone, butanol, glycerol, citric acid, fumaric acid, etc. by suitable fermenting organisms including yeast. Preferred for ethanol fermentation is yeast of the species Saccharomyces cerevisiae, preferably strains which are resistant towards high levels of ethanol, i.e., up to, e.g., about 10, about 12 or about 15 vol. % ethanol or more, such as about 20 vol. % ethanol.

Accordingly, in step (b) the pre-treated lignocellulose-containing material may be subjected to at least one, such as at least two or at least three, hemicellulolytic enzyme, such as a hemicellulase.

Any hemicellulase suitable for use in hydrolyzing hemicellulose, such as into xylose, may be used. Suitable hemicellulases include xylanases, arabinofuranosidases, acetyl xylan esterase, feruloyl esterase, glucuronidases, endo-galactanase, mannases, endo or exo arabinases, exo-galactanses, and mixtures of two or more thereof. The hemicellulase for use in the present invention may be an exo-acting hemicellulase, and more particularly an exo-acting hemicellulase which has the ability to hydrolyze hemicellulose under acidic conditions of below pH 7, preferably pH 3-7. An example of hemicellulase suitable for use in the present invention includes VISCOZYME(TM) (available from Novozymes A/S, Denmark).

According to certain embodiments, the hemicellulase is a xylanase. According to particular embodiments, the xylanase is of microbial origin, such as of fungal origin (e.g., derived from a strain of the genus *Trichoderma*, *Meripilus*, *Humicola*, *Aspergillus*, or *Fusarium*) or of bacterial origin (e.g., derived from a strain of the genus *Bacillus*). According to particular embodiments, the xylanase is derived from a filamentous fungus, such as from a strain of the genus *Aspergillus*, such as from an *Aspergillus aculeatus* strain; or from a strain of the genus *Humicola*, such as from an *Humicola lanuginose* strain. The xylanase may also be an endo-1 ,4-beta-xylanase, such as an endo-1 ,4-beta-xylanase of GH10 or GH11. Examples of

commercial xylanases include SHEARZYME(TM) and BIOFEED WHEAT(TM) from Novozymes A/S, Denmark.

The hemicellulase may be added in an amount effective to hydrolyze hemicellulose, such as, in amounts from about 0.001 to 0.5 wt.-% of dry solids, more preferably from about 0.05 to 0.5 wt.-% of dry solids.

Xylanases may be added in amounts of 0.001-1.0 g/kg dry solids, preferably in the amounts of 0.005-0.5 g/kg dry solids, and most preferably from 0.05-0.10 g/kg dry solids.

Other hydrolytic enzymes may also be present during hydrolysis. Contemplated enzymes include alpha-amylases; glucoamylases or another carbohydrate-source generating enzymes, such as beta-amylases, maltogenic amylases and/or alpha-glucosidases; proteases; or mixtures of two of more thereof.

Hydrolysis may according to certain embodiment be carried out as a fed batch process where the pre-treated lignocellulose-containing material is fed gradually to an, e.g., enzyme containing hydrolysis solution.

Enzymatic hydrolysis may be carried out in a suitable aqueous environment under conditions which can readily be determined by one skilled in the art. According to particular embodiments, hydrolysis is carried out at suitable, preferably optimal conditions for the enzyme(s) in question.

Suitable process time, temperature and pH conditions can readily be determined by one skilled in the art. For example, hydrolysis in step (b) may be carried out at a temperature in the range from about 20°C to about 80°C, such as from about 20°C to about 70°C, from about 20°C to about 60°C, from about 20°C to about 50°C, from about 20°C to about 40°C, from about 20°C to about 37°C from about 25°C to about 60°C, from about 25°C to about 50°C, from about 25°C to about 40°C or from about 25°C to about 37°C. According to certain embodiments, hydrolysis in step (b) is carried out at a temperature in the range from about 25°C to about 40°C.

Hydrolysis in step (b) may, for example, be carried out at a pH ranging from about pH 1 to about pH 9, but will normally range from about pH 5.0 to about pH 9.0, such as from about pH 5.5 to about pH 8.0, such as from about pH 6 to about 7.5, such as from about pH 6.5 to

about pH 7, such as at about pH 6 or pH 7. According to certain embodiments, hydrolysis in step (b) is carried out at a pH ranging from about pH 5.5 to about pH 8.0. According to certain other embodiments, hydrolysis in step (b) is carried out at a pH ranging from about pH 6 to about pH 8, such as from about pH 6.5 to about pH 7.5.

Hydrolysis in step (b) may, for example, be carried out for at least about 4 hours, such as for at least about 8 hours, at least about 12 hours, at least about 18 hours, at least about 24 hours, at least about 48 hours, at least about 72 hours, at least about 96 hours, at least about one week, at least about two weeks, at least about one month or at least about 3 months. According to certain embodiments, hydrolysis in step (b) is carried out for at least about 8 hours.

Since a range of toxic/inhibiting phenolic compounds are produced (e.g. released) during the pre-treatment step (a) and/or the hydrolysis step (b) it is desirable to make them non-toxic, or at least less toxic, especially if the hydrolyzed product is to be further fermented by a fermenting organism. This is done by enzymatically sulfating the phenolic compounds produced (e.g. released) during the pre-treatment step (a) and/or the hydrolysis step (b).

In the present context, it should be understood that the phenolic compounds includes those compounds in which a hydroxyl group is directly attached to a benzenoid carbon atom, and which compounds may or may not contain other substituent groups.

A special group of enzymes that are suitable for sulfating phenolic compounds are aryl sulfotransferases (EC 2.8.2.1). The conversion of inhibitory or toxic phenolic compounds to their sulfated derivative has been shown by the inventors to result in detoxification. As an example, it has been shown that presence of an aryl sulfotransferase increases the tolerance of *E. coli* towards inhibitory or toxic phenolic compounds such as ferulic acid (Example 4).

The sulfating step (c) is performed by addition of a recombinant host cell.

A recombinant host cell utilized in accordance with the present invention is a recombinant host cell comprising (e.g., expressing) a polypeptide having aryl sulfotransferase activity. Generally, the polypeptide having aryl sulfotransferase activity will be heterologous to the

host cells, which means that the polypeptide is normally not found in or made (i.e. expressed) by the host cells, but derived from a different species.

According to certain embodiments, in step (c) a recombinant host cell is employed comprising a heterologous polypeptide having an aryl sulfotransferase activity.

The polypeptide having aryl sulfotransferase activity may be a sulfotransferase 1A1 enzyme, a sulfotransferase 1A2 enzyme, a sulfotransferase 1A3 enzyme, a sulfotransferase 1B1 enzyme, a sulfotransferase 1C1 enzyme, a sulfotransferase 1C2 enzyme, a sulfotransferase 1C4 enzyme, or a sulfotransferase 1E1 enzyme.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme. According to certain other embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A2 enzyme. According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1B1 enzyme. According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C1 enzyme. According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C2 enzyme. According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C4 enzyme. According to other certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1E1 enzyme (estrogen sulfotransferase), such as the sulfotransferase 1E1 from *Gallus gallus domesticus*.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a mammalian aryl sulfotransferase, such as a mammalian sulfotransferase 1A1 enzyme.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is an aryl sulfotransferase from *Rattus norvegicus* or a variant thereof. Such variant may have at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence of the aryl sulfotransferase from *Rattus norvegicus*. Such variant may also have an amino acid sequence of the sulfotransferase from *Rattus norvegicus*, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference aryl sulfotransferase. The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity may be a polypeptide selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 77, 79, 81 or 83 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 77, 79, 81 or 83 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 77, 79, 81 or 83 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity may be a polypeptide selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, a polypeptide having aryl sulfotransferase activity is a polypeptide according to 1a). Accordingly, a polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 1. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 2. According to yet other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 3. According to yet other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 4. According to yet other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 5. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 6. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 7. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 8. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 9. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 10. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 11. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 12. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 13.

According to other certain embodiments, a polypeptide having aryl sulfotransferase activity is a polypeptide according to 1b). Accordingly, a polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence which has at least about

70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to ii) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has

at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1.

Preferably, a polypeptide according to 1b) has aryl sulfotransferase activity. More preferably, a polypeptide according to 1b) has a aryl sulfotransferase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other embodiments, a polypeptide according to 1b) has aryl

sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 77. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 79. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 81. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 83.

According to other certain embodiments, a polypeptide having aryl sulfotransferase activity is a polypeptide according to 1c). Accordingly, a polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 or more, such as 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or more, 110 or more, 120 or more, 130 or more, 140 or more, or 150 or more, amino acid residues are

substituted, deleted, and/or inserted. According to particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 50,

about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (e.g., SEQ ID NO: 1). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 1c) has aryl sulfotransferase activity. More preferably, a polypeptide according to 1c) has a aryl sulfotransferase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid

sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 77. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 79. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 81. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 83.

With “similar” aryl sulfotransferase activity it is meant that the polypeptide according to 1b) or 1c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least

about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, at of the aryl sulfotransferase activity of the reference polypeptide (e.g., SEQ ID NO: 1).

The aryl sulfotransferase activity may for instance be determined in accordance to the following method: Aryl sulfotransferase activity may be determined by the reaction of radioactively sulfur labelled PAPS, [^{35}S]PAPS, with the substrate in presence of the polypeptide of interest. This is described previously, for example by Hattori *et al* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). The reaction takes place in a buffer such as 250 µL 50 mM sodium phosphate pH 6.8 with 1 µM [^{35}S]PAPS (3.7kBq) with 100 µM accepting compound for a period of 30 min at 30°C. The reaction is stopped by addition of 100 µL of a 1:1 mixture of 0.1 M barium acetate and barium hydroxide. 50 µL of 0.1 M zinc sulfate is added, followed by centrifugation at 1,200 × g for 5 min. 300 µL of the supernatant is then transferred to a new container and 50 µL of an equal volume of 0.1 M barium hydroxide and 0.1 M zinc sulfate is added. The mixture is then centrifuged at 13,000 × g for 5 min, and 300-µL aliquots of the supernatant are mixed with 2.5 mL of Cleasol I (Nacalai Tesque, Kyoto, Japan). The radioactivity is then measured by scintillation.

Alternatively, the activity of a sulfotransferase may be detected by direct measurement of the product by analytical methods such as high performance liquid chromatography (HPLC) or liquid chromatography in combination with mass spectrometry (LC-MS).

Sulfate supply may be a limiting factor in an enzymatic sulfation reaction, and hence in the conversion of inhibitory or toxic phenolic compounds to their sulfated derivative. Here, the present inventors have demonstrated that the conversion of inhibitory or toxic phenolic compounds to their sulfated derivatives can be significantly increased if the sulfate uptake by the recombinant host cell is increased.

Therefore, a recombinant host cell employed according to the invention may be one which has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

Sulfate uptake by a given cell may be determined by a S^{35}-sulfate based method as described, e.g., by Mansilla and Mendoza (Microbiology, 2000, **146**, 815-821). Generally,

cells are first grown in a defined minimal medium, such as M9 minimal medium, supplemented with glutathione as sulphur source to exponential phase. Cells are collected, washed and then resuspended in minimal medium. The measurement of sulfate uptake is performed by incubating for 5 min at 30°C the cell suspension containing 10^8 cells ml^{-1}, 0,01 mM sodium sulfate and approximately 10^6 cpm $^{35}SO_4^{2-}$ ml^{-1} (1050 Ci $mmol^{-1}$). The incubation period is terminated by filtering the cell suspension through a 0,45 µm Millipore filter, followed by washing the filters with 5 ml minimal medium containing 2 mM magnesium sulfate and 2 mM sodium thiosulfate. Filters are transferred to polyethylene vials containing 2 ml Optiphase 'HiSafe 3' scintillation fluid (Wallac) and the radioactivity counted in an LKB Primo liquid scintillation counter. Uptake rates are expressed in nmol sulfate min^{-1} (g cellular protein)$^{-1}$.

More particularly, a recombinant host cell employed according to the present invention may be modified to have an increased protein expression of sulfate transporter compared to the identical host cell that does not carry said modification. By "increased protein expression" it is meant that the amount of the sulfate transporter protein produced by the thus modified host cell is increased compared an identical host cell that does not carry said modification. More particularly, by "increased expression" it is meant that the amount of the sulfate transporter protein produced by the thus modified host cell is increased by at least 10%, such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared an identical host cell that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry or Western Blotting.

An increase in protein expression may be achieved by any suitable means well-known to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of the gene or genes encoding the sulfate transporter in the host cell, such as by introducing into the host cell a exogenous nucleic acid, such as a

vector, comprising the gene or genes encoding the sulfate transporter operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule. An increase in protein expression may also be achieved by integration of at least a second copy of the gene or genes encoding the sulfate transporter into the genome of the host cell. An increase in protein expression may also be achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding the sulfate transporter. An increase in protein expression may also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the sulfate transporter. By modifying the sequence of the ribosome binding site the translation initiation rate may be increased, thus increasing the translation efficiency.

According to certain embodiments, the increase in the number of copies of the gene or genes is achieved by introducing into the recombinant host cell one or more (such as two or three) exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

According to certain embodiments, a recombinant host cell is provided which comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more (such as two, three or four) nucleotide sequences encoding a sulfate transporter. Suitably, the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

According to certain embodiments, the exogenous nucleic acid molecule is stably integrated into the genome of the recombinant host cell.

The sulfate transporter may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the sulfate transporter is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the sulfate transporter is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to certain embodiments, the sulfate transporter is a bacterial sulfate transporter. With "bacterial sulfate transporter" it is meant that the sulfate transporter is naturally derived from a bacterium, such as *Escherichia coli*.

The sulfate transporter employed in accordance of the invention may be any suitable sulfate transporter which is functional in the respective host cell.

According to certain embodiments, the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

According to certain embodiments, the sulfate transporter is a bacterial sulfate transporter selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

According to particular embodiments, the sulfate transporter is a CysZ protein.

Members of the CysZ family (TCDB 2.A.121) are high affinity, high specificity proton-dependent sulfate transporters which mediates sulfate uptake. Non-limiting examples of CysZ proteins are those found in bacteria, such as E. coli (NCBI: NP_416908.1) S. typhimurium (NCBI: NP_456966.1), K. pneumoniae (NCBI: CDO15722.1), P. fluorescens (NCBI: AEV64873.1), S. sonnei (NCBI: AAZ89133.1), V. anguillarum (NCBI: AEH33702.1), B. japonicum (NCBI: KOY11972.1) and C. glutamicum (NCBI: CAF20834.1) to only name a few.

Accordingly, a sulfate transporter for use according to the invention may for instance be the CysZ protein from Escherichia coli (SEQ ID NO: 14). Further information regarding CysZ of *Escherichia coli* is available at EcoCyc (www.biocyc.org) under Accession number EG10003. See also NCBI Reference Sequence Database under NCBI Reference Sequence: NP_416908.1.

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 2a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 2b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 2c). According to particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 14). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 2b) or 2c) has sulfate transporter activity. More preferably, a polypeptide according to 2b) or 2c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14. With "similar" sulfate transporter activity it is meant that the polypeptide according to 2b) or 2c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 14).

Another suitable sulfate transporter for use according to the invention may for instance be the CysZ protein from *Corynebacterium glutamicum* (SEQ ID NO: 15). Further information regarding CysZ of *C. glutamicum* is available at NCBI under accession number CAF20834.1.

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 3a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 3b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO:

15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 3c). According to particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 15). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 3b) or 3c) has sulfate transporter activity. More preferably, a polypeptide according to 3b) or 3c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15. With "similar" sulfate transporter activity it is meant that the polypeptide according to 3b) or 3c)

has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 15).

According to certain embodiments, the sulfate transporter is a sulfate-transporting ATPase, such as a member of the SulT (cysPTWA) family.

According to particular embodiments, the sulfate transporter is bacterial sulfate transporter of the SulT (cysPTWA) family.

Sulfate transporters of the SulT (cysPTWA) family from proteobacteria (TCDB 3.A.1.6.1), such as *E. coli*, are generally constituted by: (i) one of two periplasmic proteins, Sbp, the sulfate binding protein, or CysP, the thiosulfate-binding protein; (ii) membrane proteins CysT (synonym: CysU) and CysW; and (iii) the ATP-binding protein CysA. The SulT subunits are encoded by the cysPTWA operon and by the sbp gene, located either in another chromosomal region or instead of cysP in the same operon. Non-limiting examples of sulfate transporters of the SulT (cysPTWA) family are those found in *Escherichia coli, Salmonella typhimurium* and *Rhodobacter capsulatus*.

CysT, CysW, CysA, CysP and Sbp of *Escherichia coli* are set forth in SEQ ID NO: 16 to 20, respectively. Further information regarding CysT, CysW, CysA, CysP and Sbp of *Escherichia coli* is available at EcoCyc (www.biocyc.org) under Accession numbers EG10197, EG10198, EG10183, EG10195 and EG10929, respectively.

CysT, CysW, CysA, CysP and Sbp of *Salmonella typhimurium* are set forth in SEQ ID NO: 21 to 25, respectively.

According to certain embodiments, the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

According to certain embodiments, the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide

sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

According to certain embodiments, the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16 or 21;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16 or 21; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16 or 21, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the first membrane subunit is a polypeptide according to 4a). According to particular embodiments, the first membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, the first membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 21.

According to other certain embodiments, the first membrane subunit is a polypeptide according to 4b). Accordingly, a first membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has

at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16.

Alternatively, a first membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21.

According to other certain embodiments, the first membrane subunit is a polypeptide according to 4c). According to particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 21, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 21, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 21, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 16 or 21). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The

alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 4b) or 4c) assembles with the proteins CysW, CysA and CysP/Sbp to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17 or 22;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17 or 22; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17 or 22, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the second membrane subunit is a polypeptide according to 5a). According to particular embodiments, the second membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, the second membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 22.

According to other certain embodiments, the second membrane subunit is a polypeptide according to 5b). Accordingly, a second membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17.

Alternatively, a second membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about

95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22.

According to other certain embodiments, the second membrane subunit is a polypeptide according to 5c). According to particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 17 or 22). The alterations may solely be

amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 5b) or 5c) assembles with the proteins CysT, CysA and CysP/Sbp to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18 or 23;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18 or 23; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18 or 23, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the ATP binding subunit is a polypeptide according to 6a). According to particular embodiments, the ATP binding subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, the ATP binding subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 23.

According to other certain embodiments, the ATP binding subunit is a polypeptide according to 6b). Accordingly, a ATP binding subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 80%, such as

at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18.

Alternatively, a ATP binding subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about

95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23.

According to other certain embodiments, the ATP binding subunit is a polypeptide according to 6c). According to particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 18 or 23). The alterations may solely be

amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 6b) or 6c) assembles with the proteins CysT, CysW and CysP/Sbp to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the periplasmic binding protein is CysP.

According to certain embodiments, the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19 or 24;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19 or 24; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19 or 24, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the periplasmic binding protein is a polypeptide according to 7a). According to particular embodiments the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 24.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 7b). Accordingly, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the

amino acid sequence set forth in SEQ ID NO: 19. According to particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19.

Accordingly, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at

least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 7c). According to particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 25, such as about 1 to about

20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 19 or 24). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 7b) or 7c) assembles with the proteins CysT, CysW and CysA to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the periplasmic binding protein is Sbp.

According to certain embodiments, the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20 or 25;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20 or 25; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20 or 25, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the periplasmic binding protein is a polypeptide according to 8a). According to particular embodiments the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 25.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 8b). Accordingly, a periplasmic binding protein may be a polypeptide

comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20.

Alternatively, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the

amino acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 8c). According to particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other

more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 20 or 25). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 8b) or 8c) assembles with the proteins CysT, CysW and CysA to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the sulfate transporter is a member of the SulP family.

The sulfate transporter SulP family (TCDB 2.A.53) is a large and ubiquitous family with members derived from archaea, bacteria, fungi, plants and animals. Many organisms including *Bacillus subtilis*, *Synechocystis sp*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Caenorhabditis elegans* possess multiple SulP family paralogues. Many of these proteins are functionally characterized, and most are inorganic anion uptake transporters or anion:anion exchange transporters. A non-limiting example of a sulfate transporter of the SulP family is that found in *Mycobacterium tuberculosis* (SEQ ID NO: 26; NCBI: NP_216255.1). Zolotarev et al. (Comp Biochem Physiol A Mol Integr Physiol. 2008 Mar; 149(3):255-66) have demonstrate that the overexpression of SulP protein Rv1739c from *M. tuberculosis* in *E. coli* increases sulfate uptake. Another non-limiting example is a SulP protein found in multiple species (NCBI: WP_012536065.1).

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least

about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 9a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 9b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 9c). According to particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 26). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 9b) or 9c) has sulfate transporter activity. More preferably, a polypeptide according to 9b) or 9c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26. With "similar" sulfate transporter activity it is meant that the polypeptide according to 9b) or 9c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 26).

According to certain embodiments, the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

Genes encoding PiT family transporters are widespread throughout the three life domains. A non-limiting example of a CysP transporter of the PiT family is that found in *Bacillus subtilis* (SEQ ID NO: 27; GenBank: CAB13432). Another non-limiting example of a CysP transporter of the PiT family is that found in *Halobacterium salinarum* (NCBI: CAP13497.1).

According to certain embodiments, the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 10a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 10b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%,

or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 10c). According to particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 27). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 10b) or 10c) has sulfate transporter activity. More preferably, a polypeptide according to 10b) or 10c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27.

With "similar" sulfate transporter activity it is meant that the polypeptide according to 10b) or 10c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 27).

According to certain embodiments, the sulfate transporter is an oxyanion permease (PerO).

Oxyanion permeases act as a general oxyanion importer of molybdate, sulfate, tungstate, and vanadate. A non-limiting example of an oxyanion permease is that found in *Rhodobacter capsulatus* (SEQ DI NO: 73).

According to certain embodiments, the sulfate transporter is selected from the group consisting of:

11a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 73;

11b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73; or

11c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 73, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 11a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 11b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 73. According to particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. According to other particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. According to other particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. According to other particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 11c). According to particular embodiments, a polypeptide according to 11c) comprises an amino acid sequence set forth in SEQ ID NO: 73, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 11c) comprises an amino acid sequence set forth in SEQ ID NO: 73, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 11c) comprises an amino acid sequence set forth in SEQ ID NO: 73, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 73). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 11b) or 11c) has sulfate transporter activity. More preferably, a polypeptide according to 11b) or 11c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 73. With "similar" sulfate transporter activity it is meant that the polypeptide according to 11b) or 11c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 73).

Besides external sulfate supply, the sulfation reaction may further depend on the supply of sulfate from 3′ -phosphoadenosine 5′ -phosphosulfate (PAPS) or transferred from another sulfated compound. The inventors have shown that the sulfation reaction can be further enhanced by improving the supply of PAPS (3' -phosphoadenosine 5' -phosphosulfate) and, in addition, by the removal of the product 3′ -phosphoadenosine 5′ -phosphate (PAP). The improved supply is obtained by deregulation, mutation or overexpression of enzymes that increase PAPS concentration or similarly reduce PAP concentration. This is exemplified in Example 2, where an increased production of zosteric acid in Escherichia coli is obtained by increasing the expression of the genes cysD, cysN, and cysC which are responsible for production of PAPS. Without being bound to a specific theory, it is believed that an adenylyl moiety (AMP) of ATP is transferred to sulfate to form activated sulfate, or APS (adenosine 5'-phosphosulfate). This extremely unfavorable reaction is kinetically and energetically linked to the hydrolysis of GTP by the enzyme ATP sulfurylase, which is composed of two types of subunits: an adenylyl transferase (cysD) and a GTPase (cysN). APS is then phosphorylated at the 3'-hydroxyl to form PAPS (3'-phosphoadenosine 5'-phosphosulfate) in a reaction catalysed by APS kinase, which is

encoded by cysC. Furthermore, the inventors have enhanced the production of zosteric acid even more by increasing the expression of the gene cysQ encoding a PAP phosphatase which is responsible for the removal of PAP.

Therefore, in order to further improve the production of a sulfated phenolic compound, such as zosteric acid, a recombinant host cell according to the present invention may be further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; may be further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or may be further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification. By "increased protein expression" it is meant that the amount of the respective protein produced by the thus modified host cell is increased compared an identical host cell that does not carry said modification. More particularly, by "increase expression" it is meant that the amount of respective protein produced by the thus modified host cell is increased by at least 10% , such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared an identical host cell that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry or Western Blotting.

According to certain embodiments, a recombinant host cell employed according to the invention has further been modified to have an increased protein expression an ATP sulfurylase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell employed according to the invention has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell employed according to the invention has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

An increase in protein expression may be achieved by any suitable means well-known to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of the gene or genes encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase) in the host cell, such as by using (e.g., introducing into the host cell) a vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule. An increase in protein expression may also be achieved by integration of at least a second copy of the gene or genes encoding the respective protein into the genome of the host cell. An increase in protein expression may also be achieved by increasing the strength of the promoter(s) operably linked to the gene or genes. An increase in protein expression may also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase). By modifying the sequence of the ribosome binding site the translation initiation rate may be increased, thus increasing the translation efficiency.

ATP sulfurylase encoding genes for use according to the invention may for instance be the cysD and cysN genes from Escherichia coli (encoding SEQ ID NO: 28 and 29, respectively). Alternative ATP sulfurylase encoding genes include the *Arabidopsis thaliana* ATP sulfurylase ASAL gene (GenBank Accession No. U40715, Logan et al. (1996) J Biol Chem 271: 12227); the *Allium cepa* ATP-sulfurylase gene (Gen-Bank Accession No AF21154); the *Lotus japonicus* ATP sulfurylase gene (GenBank Accession No. AW164083); the *Arabidopsis thaliana* met3-1 ATP sulfurylase gene (Gen-Bank Accession No. X79210).

According to certain embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding an ATP sulfurylase.

The ATP sulfurylase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the ATP sulfurylase is

derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the ATP sulfurylase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to certain embodiments, the ATP sulfurylase is a protein constituted by two polypeptides, which are exemplified by the amino acid sequence set forth in SEQ ID NO: 28 and 29, respectively.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and a nucleotide sequence encoding iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28,

and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the MET3 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 30).

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 30 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 30.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide has ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the ATP sulfurylase encoding gene from Bacillus subtilis (encoding SEQ ID NO: 31).

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 31 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 31.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide has ATP sulfurylase activity.

Techniques for determining ATP sulfurylase activity are well known to the skilled person. Exemplary methods have been described, e.g. by Reuveny and Filner (Anal Biochem, 1976, 75(2), 410-428) or Hommes and Moss (Anal Biochem, 1986, 154(1), 100-103).

An APS kinase encoding gene for use according to the invention may for instance be the cysC gene from *Escherichia coli* (encoding SEQ ID NO: 32).

In certain instances a single polypeptide has been shown to possess both an ATP sulfurylase and a 5'-adenylylsulfate kinase activity. For example, an ATP sulfurylase/APS kinase encoding gene has been isolated from mouse (GenBank Accession No. U34883, Li et al. (1995) J Biol Chem)70: 1945), and human (GenBank Accession No. AF033026, Yanagisawa (1998) Biosci Biotechnol Biochem 62: 1037) sources. Other examples of such bifunctional enzyme include 3'-phosphoadenosine 5'-phosphosulfate synthase enzymes (PAPSS) from rat (*Rattus norvegicus*) (SEQ ID NO: 33 or 34).

According to certain embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an APS kinase.

The APS kinase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the APS kinase is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the APS kinase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least

96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the MET14 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 35).

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 35 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 35.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the APS kinase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 36).

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 36 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 36.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide has APS kinase activity.

Alternatively, a polypeptide having both an ATP sulfurylase and a APS kinase activity can be used, such as a 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS).

According to certain embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an 3'-phosphoadenosine 5'-phosphosulfate synthase.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 33 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 33.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 34 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 34.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide has both an ATP sulfurylase and APS kinase activity.

According to particular embodiments, a recombinant host cell employed according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

Techniques for determining APS kinase activity are well known to the skilled person. An exemplary method has been described, e.g. by Burnell and Whatley (Anal Biochem, 1975, 68(1), 281-288).

A PAP phosphatase encoding gene for use according to the invention may for instance be the cysQ gene from Escherichia coli (encoding SEQ ID NO: 37).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an PAP phosphatase.

The PAP phosphatase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the PAP phosphatase is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the PAP phosphatase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the MET22 gene from Saccharomyces cerevisiae (encoding SEQ ID NO: 38).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 38 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 38.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least

about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the PAP phosphatase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 39).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 39 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 39.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at

least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide has PAP phosphatase activity.

Techniques for determining PAP phosphatase activity are well known to the skilled person. An exemplary method has been described, e.g. by Fukuda et al. (Appl Environ Microbiol, 2007, 73(17), 5447-5452).

According to certain embodiments, the nucleotide sequences coding for an ATP sulfurylase, an APS kinase and an PAP phosphatase, respectively, are part of an operon. Accordingly, a recombinant host cell according to the invention may comprise an exogenous nucleic acid molecule (such as vector) which comprises an operon comprising a nucleotide sequence or nucleotide sequences encoding an ATP sulfurylase, a nucleotide sequence encoding an APS kinase, and optionally a nucleotide sequence encoding an PAP phosphatase.

Recombinant host cells in accordance with the invention can be produced from any suitable host organism, including single-celled or multicellular microorganisms such as bacteria, yeast, fungi and algae.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, and algae.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast and fungi.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria and yeast.

Bacterial host cells are selected from Gram-positive and Gram-negative bacteria. Non-limiting examples for Gram-negative bacterial host cells include species from the genera *Escherichia, Erwinia, Klebsiella* and *Citrobacter*. Non-limiting examples of Gram-positive bacterial host cells include species from the genera *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Streptomyces, Streptococcus*, and *Cellulomonas*.

According to certain embodiment, the recombinant host cell is a bacterium of the family selected from the group consisting of *Enterobacteriaceae, Bacillaceae, Lactobacillaceae, Corynebacteriaceae.*

According to certain embodiments, the recombinant host cell is a bacterium of the family *Enterobacteriaceae.*

According to certain embodiments, the recombinant host cell is a bacterium, which may be a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus*, or *Yersinia*.

According to particular embodiments, the recombinant host cell is a bacterium of the genus *Bacillus*. Non-limiting examples of a bacterium of the genus *Bacillus* are *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, and *Bacillus mojavensis*. According to more particular embodiments, the recombinant host cell is *Bacillus subtilis*. According to other more particular embodiments, the recombinant host cell is *Bacillus licheniformis*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Lactococcus*. A non-limiting example of a bacterium of the genus *Lactococcus* is *Lactococcus lactis*. According to more particular embodiments, the recombinant host cell is *Lactococcus lactis.*

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Corynebacterium*. A non-limiting example of a bacterium of the genus *Corynebacterium* is *Corynebacterium glutamicum*. According to more particular embodiments, the recombinant host cell is *Corynebacterium glutamicum*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Streptomyces*. Non-limiting examples of a bacterium of the genus *Streptomyces* are *Streptomyces lividans*, *Streptomyces coelicolor*, or *Streptomyces griseus*. According to more particular embodiments, the recombinant host cell is *Streptomyces lividans*. According to other more particular embodiments, the recombinant host cell is *Streptomyces coelicolor*. According to other more particular embodiments,, the recombinant host cell is *Streptomyces griseus*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Pseudomonas*. A non-limiting example of a bacterium of the genus *Pseudomonas* is *Pseudomonas putida*. According to more particular embodiments, the recombinant host cell is *Pseudomonas putida*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Geobacillus*. A non-limiting examples of a bacterium of the genus *Geobacillus* are *Geobacillus thermoglucosidasius* and *Geobacillus stearothermophilus*. According to more particular embodiments, the recombinant host cell is *Geobacillus thermoglucosidasius*. According to other more particular embodiments,, the recombinant host cell is *Geobacillus stearothermophilus*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Thermoanaerobacterium*. A non-limiting example of a bacterium of the genus *Thermoanaerobacterium* is *Thermoanaerobacterium thermosaccharolyticum*. According to more particular embodiments, the recombinant host cell is *Thermoanaerobacterium thermosaccharolyticum*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Escherichia*. A non-limiting example of a bacterium of the genus *Escherichia* is *Escherichia coli*. According to more particular embodiments, the recombinant host cell is *Escherichia coli*.

Yeast host cells may be derived from e.g., *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to certain embodiments, the recombinant host cell is a yeast, which may be a yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to particular embodiments, the recombinant host cell is a yeast of the genus *Saccharomyces*. A non-limiting example of a yeast of the genus *Saccharomyces* is

Saccharomyces cerevisiae. According to more particular embodiments, the recombinant host cell is *Saccharomyces cerevisiae*.

According to particular embodiments, the recombinant host cell is a yeast of the genus *Pichia*. Non-limiting example of a yeast of the genus *Pichia* are *Pichia pastoris* and *Pichia kudriavzevii*. According to more particular embodiments, the recombinant host cell is *Pichia pastoris*. According to other more particular embodiments, the recombinant host cell is P*ichia kudriavzevii*.

Fungi host cells may be derived from, e.g., *Aspergillus*.

According to certain embodiments, the recombinant host cell is a fungus, such as a fungi of the genus *Aspergillus*. Non-limiting examples of a fungus of the genus *Aspergillus* are *Aspergillus Oryzae, Aspergillus niger* or *Aspergillus awamsii*. According to more particular embodiments, the recombinant host cell is *Aspergillus Oryzae*. According to other more particular embodiments, the recombinant host cell is *Aspergillus niger*. According to other more particular embodiments, the recombinant host cell is *Aspergillus awamsii*.

Algae host cells may be derived from, e.g., *Chlamydomonas, Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

According to certain embodiments, the recombinant host cell is an alga, which may be an algae of the genus *Chlamydomonas, Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

According to particular embodiments, the recombinant host cell is an alga cell of the genus *Chlamydomonas*. A non-limiting example of an alga of the genus *Chlamydomonas* is *Chlamydomonas reinhardtii.*

According to particular embodiments, the recombinant host cell is an alga cell of the genus *Haematococcus*. A non-limiting example of an alga of the genus *Haematococcus* is *Haematococcus pluvialis*.

According to other particular embodiments, the recombinant host cell is an alga cell of the genus *Phaedactylum*. A non-limiting example of an alga of the genus *Phaedactylum* is *Phaedactylum tricornatum*.

Generally, a recombinant host cell according to the invention has been genetically modified to express one or more polypeptides as detailed herein, which means that one or more exogenous nucleic acid molecules, such as DNA molecules, which comprise(s) a nucleotide sequence or nucleotide sequences encoding said polypeptide or polypeptides has been introduced in the host cell. Techniques for introducing exogenous nucleic acid molecule, such as a DNA molecule, into the various host cells are well-known to those of skill in the art, and include transformation (e.g., heat shock or natural transformation), transfection, conjugation, electroporation, microinjection and microparticle bombardment.

Accordingly, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide as detailed herein.

In order to facilitate expression of the polypeptide in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said polypeptide.

Promoters useful in accordance with the invention are any known promoters that are functional in a given host cell to cause the production of an mRNA molecule. Many such promoters are known to the skilled person. Such promoters include promoters normally associated with other genes, and/or promoters isolated from any bacteria, yeast, fungi, alga or plant cell. The use of promoters for protein expression is generally known to those of skilled in the art of molecular biology, for example, see Sambrook et al., Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. , 1989. The promoter employed may be inducible. The term “inducible” used in the context of a promoter means that the promoter only directs transcription of an operably linked nucleotide sequence if a stimulus is present, such as a change in temperature or the presence of a chemical substance (“chemical inducer”). As used herein, “chemical induction” according to the present invention refers to the physical application of a exogenous or endogenous substance (incl. macromolecules, e.g., proteins or nucleic acids) to a host cell. This has the effect of causing the target promoter present in the host cell to increase the rate of transcription. Alternatively, the promoter employed may be constitutive. The term “constitutive” used in the context of a promoter means that the

promoter is capable of directing transcription of an operably linked nucleotide sequence in the absence of stimulus (such as heat shock, chemicals etc.).

Non-limiting examples of promoters functional in bacteria, such as *Bacillus subtilis*, *Lactococcus lactis* or *Escherichia coli*, include both constitutive and inducible promoters such as T7 promoter, the beta-lactamase and lactose promoter systems; alkaline phosphatase (phoA) promoter, a tryptophan (trp) promoter system, tetracycline promoter, lambda-phage promoter, ribosomal protein promoters; and hybrid promoters such as the tac promoter. Other bacterial and synthetic promoters are also suitable.

Non-limiting examples of promoters functional in yeast, such as *Saccharomyces cerevisiae*, include xylose promoter, GAL1 and GAL10 promoters, TEF1 promoter, and pgk1 promoter.

Non-limiting examples of promoters functional in fungi, such as *Aspergillus Oryzae* or *Aspergillus niger*, include promotors derived from the gene encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral α-amylase, *Aspergillus niger* acid stable α-amylase, *Aspergillus niger* or *Aspergillus awamsii* glucoamylase (gluA), *Aspergillus niger* acetamidase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphatase isomerase, *Rhizopus meihei* aspartic proteinase, and *Rhizopus meihei* lipase.

Non-limiting examples of promoters functional in alga, such as *Haematococcus pluvialis*, include the CaMV35S promoter, the SV40 promoter, and promoter of the *Chlamydomonas reinhardtii* RBCS2 gene and the promoter of the *Volvox carteri* ARS gene.

Non-limiting examples of promoters functional in plant cells include the *Lactuca sative* psbA promoter, the tabacco psbA promoter, the tobacco rrn16 PEP+NEP promoter, the CaMV 35S promoter, the 19S promoter, the tomate E8 promoter, the nos promoter, the Mac promoter, and the pet E promoter or the ACT1 promoter.

Besides a promoter, the exogenous nucleic acid molecule may further comprise at least one regulatory element selected from a 5' untranslated region (5'UTR) and 3' untranslated region (3' UTR). Many such 5' UTRs and 3' UTRs derived from prokaryotes and eukaryotes are well known to the skilled person. Such regulatory elements include 5' UTRs and 3' UTRs normally associated with other genes, and/or 5' UTRs and 3' UTRs isolated from any bacteria, yeast, fungi, alga or plant cell.

If the host cell is a prokaryotic organism, the 5' UTR usually contains a ribosome binding site (RBS), also known as the Shine Dalgarno sequence which is usually 3-10 base pairs upstream from the initiation codon. Meanwhile, if the host cell is an eukaryotic organism the 5' UTR usually contains the Kozak consensus sequence. An eukaryotic 5' UTR may also contain cis-acting regulatory elements.

The exogenous nucleic acid molecule may be a vector or part of a vector, such as an expression vector. Normally, such a vector remains extrachromosomal within the host cell which means that it is found outside of the nucleus or nucleoid region of the host cell.

According to certain embodiments, a recombinant host cell employed according to the invention does not express an endogenous PAPS-dependent aryl sulfotransferase.

It is also contemplated by the present invention that the exogenous nucleic acid molecule is stably integrated into the genome of the host cell. Means for stable integration into the genome of a host cell, e.g., by homologous recombination, are well known to the skilled person.

The sulfating step (c) may be carried out during or after the pre-treatment step (a) and/or during or after the hydrolysis step (b). According to certain embodiments, sulfating step (c) is carried out after the pre-treatment step (a). According to certain embodiments, sulfating step (c) is carried out after the hydrolysis step (b). According to certain embodiments, sulfating step (c) is carried out after the pre-treatment step (a) and after the hydrolysis step (b). According to certain embodiments, the pre-treatment step (a) and the sulfating step (c) are carried out simultaneously. According to certain embodiments, hydrolysis step (b) and the sulfating step (c) are carried out simultaneously. According to certain embodiments, the pre-treatment step (a), the hydrolysis step (b) and the sulfating step (c) are carried out simultaneously .

Suitable sulfate donor molecules metabolized by a polypeptide having aryl sulfotransferase activity are well-known to one skilled in the art. Non-limiting examples include 3'-phosphoadenosine 5'-phosphosulfate (PAPS), para-nitrophenyl sulfate (pNPS) and 4-methylumbelliferyl sulfate (MUS). Such sulfate donor molecules may be employed to facilitate the sulfation of phenolic compounds in accordance with the invention.

The medium employed for culturing the recombinant host cell may be any conventional medium suitable for culturing the host cell in question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective host cell, such as carbon and nitrogen sources and other inorganic salts, such as sulfate salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as B. subtilis, L. lactis or E. coli cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others. Suitable media for culturing eukaryotic cells, such as yeast cells, are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular host cell being cultured. The medium for culturing eukaryotic cells may also be any kind of minimal media such as Yeast minimal media.

The hydrolyzed product obtained in accordance to the process as described above may be further fermented to obtain a fermentation product.

Therefore, present invention provides in a further aspect a process for the production of a fermentation product, from a lignocellulose-containing material, comprising the steps:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

(d) fermenting the hydrolyzate obtained in step (b) using a fermenting organism, thereby obtaining a fermentation product.

It is to be understood that the details given above in the context of the first aspect, in particular with respect to steps (a), (b) and (c), apply mutatis mutandis to this further aspect.

According to this further aspect of the invention, the pre-treated and hydrolyzed lignocellulose-containing material (the hydrolyzate) is fermented by at least one fermenting organism capable of fermenting fermentable sugars, such as glucose, xylose, mannose, and galactose directly or indirectly into a desired fermentation product.

Suitable process time, temperature and pH conditions for the fermentation are well-known to one skilled in the art. For example, fermentation may be carried out at a temperature in the range from about 20°C to about 80°C, such as from about 20°C to about 70°C, from about 20°C to about 60°C, from about 20°C to about 50°C, from about 20°C to about 40°C, from about 20°C to about 37°C from about 25°C to about 60°C, from about 25°C to about 50°C, from about 25°C to about 40°C or from about 25°C to about 37°C. According to certain embodiments, the fermentation is carried out at a temperature in the range from about 25°C to about 40°C. According to certain embodiments, the fermentation is carried out at a temperature in the range from about 60°C to about 80°C.

Fermentation may, for example, be carried out at a pH ranging from about pH 1 to about pH 9, but will normally range from about pH 5.0 to about pH 9.0, such as from about pH 5.5 to about pH 8.0, such as from about pH 6 to about 7.5, such as from about pH 6.5 to about pH 7, such as at about pH 6 or pH 7. According to certain embodiments, the fermentation is carried out at a pH ranging from about pH 5.5 to about pH 8.0. According to certain other embodiments, the fermentation is carried out at a pH ranging from about pH 6 to about pH 8, such as from about pH 6.5 to about pH 7.5.

Fermentation may, for example, be carried out for at least about 4 hours, such as for at least about 8 hours, at least about 12 hours, at least about 18 hours, at least about 24 hours, at least about 48 hours, at least about 72 hours, at least about 96 hours, at least about one week, at least about two weeks, at least about one month or at least about 3 months. According to certain embodiments, the fermentation is carried out for at least about 8 hours.

The hydrolysis step (b), the sulfating step (c), and the fermentation step (d) may be carried out simultaneously, sequentially, or as a hybrid. According to certain embodiments, hydrolysis step (b), sulfating step (c) and fermentation step (d) are carried out sequentially.

According to certain embodiments, hydrolysis step (b), sulfating step (c) and fermentation step (d) are carried out simultaneously.

According to certain embodiments, sulfating step (c) is performed simultaneously with fermentation step (d). According to particular embodiments, the fermenting organism employed in step (d) is a recombinant host cell as detailed herein. In this case, fermenting organism not only expresses a polypeptide having aryl sulfotransferase activity, but also converts fermentable sugars, such as glucose, xylose, mannose, and galactose directly or indirectly into a desired fermentation product.

According to certain other embodiments, hydrolysis step (b) and fermentation step (d) are carried out as a separate hydrolysis and fermentation, where the hydrolysis is taken to completion before initiation of fermentation. This is often referred to as "SHF".

According to certain other embodiments, hydrolysis step (b) and fermentation step (d) are carried out as a simultaneous hydrolysis and fermentation, where the fermentation takes place while the pre-treated pre-treated lignocellulose-containing material is hydrolyzed. In general this means that a simultaneous hydrolysis and fermentation is carried out at conditions (e.g., temperature and/or pH) suitable, preferably optimal, for the fermenting organism(s) in question. When the fermentation (e.g., ethanol fermentation using, e.g., a Saccharomyces yeast) is performed simultaneous with hydrolysis, the temperature is suitably in the range from about 26°C to about 35°C, such as from about 30°C to 34°C, such as at about 32°C. However, higher temperatures, such as in the range from about 60°C to about 80°C, are also envisioned if the fermenting organism is a thermophile.

According to certain other embodiments, hydrolysis step (b) and fermentation step (d) are carried out as hybrid hydrolysis and fermentation (HHF), which is a combination of the SHF and SSF processes. HHF typically begins with a separate partial hydrolysis step and ends with a simultaneous hydrolysis and fermentation step. The separate partial hydrolysis step is an enzymatic cellulose saccharification step typically carried out at conditions (e.g., at higher temperatures) suitable, preferably optimal, for the hydrolyzing enzyme(s) in question. The subsequent simultaneous hydrolysis and fermentation step is typically carried out at conditions suitable for the fermenting organism(s) (often at lower temperatures than the separate hydrolysis step).

One or more cellulolytic enzymes as detailed above may be employed in the fermentation step (d). Likewise, one or more hemicellulolytic enzymes as detailed above may be employed in the fermentation step (d).

Subsequent to fermentation the fermentation product may be separated from the fermentation medium/broth. The medium/broth may be distilled to extract the fermentation product or the fermentation product may be extracted from the fermentation medium/broth by micro or membrane filtration techniques. Alternatively, the fermentation product may be recovered by stripping. Recovery methods are well known in the art.

Especially contemplated fermentation products include, but are not limited to, alcohols (e.g., ethanol, methanol, butanol); organic acids (e.g., citric acid, acetic acid, itaconic acid, lactic acid, gluconic acid); ketones (e.g., acetone); amino acids (e.g., glutamic acid); gases (e.g., H2 and CO2); antibiotics (e.g., penicillin and tetracycline); enzymes; vitamins (e.g., riboflavin, B12, beta-carotene); and hormones. Other, non-limiting examples of fermentation products are diamines, diols, triols, carboxylic acids, diacids, aromatic acids, dienes and isoprenoids.

According to certain embodiments, the fermentation product obtained in accordance with the invention is an alcohol, especially ethanol. The fermentation product, such as ethanol, obtained according to the invention, may preferably be fuel alcohol/ethanol. However, in the case of ethanol it may also be used as potable ethanol.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is an organic acid.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a ketone.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is an amino acid.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a diamine.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a diol.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a triol.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a carboxylic acid.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is a diacid.

According to certain other embodiments, the fermentation product obtained in accordance with the invention is an aromatic acid.

The term "fermenting organism" as used herein refers to any organism, including bacterial and fungal organisms, suitable for producing a desired fermentation product. Especially suitable fermenting organisms according to the invention are able to ferment, i.e., convert, sugars, such as glucose, directly or indirectly into the desired fermentation product.

Examples of fermenting organisms include bacteria, yeast, fungi and algae.

According to certain embodiments, a fermenting organism in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, algae and plant.

According to certain other embodiments, a fermenting organism in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, and algae.

According to certain other embodiments, a fermenting organism in accordance with the invention is selected from the group consisting of bacteria, yeast and fungi.

According to certain other embodiments, a fermenting organism in accordance with the invention is selected from the group consisting of bacteria and yeast.

Bacterial host cells are selected from Gram-positive and Gram-negative bacteria. Non-limiting examples for Gram-negative bacterial host cells include species from the genera *Escherichia, Erwinia, Klebsiella* and *Citrobacter*. Non-limiting examples of Gram-positive

bacterial host cells include species from the genera *Bacillus*, *Lactococcus*, *Lactobacillus*, *Clostridium*, *Corynebacterium*, *Streptomyces, Streptococcus*, and *Cellulomonas*.

According to certain embodiments, the fermenting organism is a bacterium, which may be a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus*, or *Yersinia*.

According to particular embodiments, the fermenting organism is a bacterium of the genus *Bacillus*. Non-limiting examples of a bacterium of the genus *Bacillus* are *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, and *Bacillus mojavensis*. According to more particular embodiments, the fermenting organism is *Bacillus subtilis*. According to other more particular embodiments, the fermenting organism is *Bacillus licheniformis*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Lactococcus*. A non-limiting example of a bacterium of the genus *Lactococcus* is *Lactococcus lactis*. According to more particular embodiments, the fermenting organism is *Lactococcus lactis*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Corynebacterium*. A non-limiting example of a bacterium of the genus *Corynebacterium* is *Corynebacterium glutamicum*. According to more particular embodiments, the fermenting organism is *Corynebacterium glutamicum*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Streptomyces*. A non-limiting examples of a bacterium of the genus *Streptomyces* are *Streptomyces lividans*, *Streptomyces coelicolor*, or *Streptomyces griseus*. According to more particular embodiments, the fermenting organism is *Streptomyces lividans*. According to other more particular embodiments, the fermenting organism is *Streptomyces coelicolor*. According to other more particular embodiments, the fermenting organism is *Streptomyces griseus*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Pseudomonas*. A non-limiting example of a bacterium of the genus *Pseudomonas* is *Pseudomonas putida*. According to more particular embodiments, the fermenting organism is *Pseudomonas putida*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Geobacillus*. A non-limiting examples of a bacterium of the genus *Geobacillus* are *Geobacillus thermoglucosidasius* and *Geobacillus stearothermophilus*. According to more particular embodiments, the fermenting organism is *Geobacillus thermoglucosidasius*. According to other more particular embodiments, the fermenting organism is *Geobacillus stearothermophilus*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Thermoanaerobacterium*. A non-limiting example of a bacterium of the genus *Thermoanaerobacterium* is *Thermoanaerobacterium thermosaccharolyticum*. According to more particular embodiments, the fermenting organism is *Thermoanaerobacterium thermosaccharolyticum*.

According to other particular embodiments, the fermenting organism is a bacterium of the genus *Escherichia*. A non-limiting example of a bacterium of the genus *Escherichia* is *Escherichia coli*. According to more particular embodiments, the fermenting organism is *Escherichia coli*.

Yeast host cells may be derived from e.g., *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to certain embodiments, the fermenting organism is a yeast, which may be a yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to particular embodiments, the fermenting organism is a yeast of the genus *Saccharomyces*. A non-limiting example of a yeast of the genus *Saccharomyces* is

Saccharomyces cerevisiae. According to more particular embodiments, the fermenting organism is *Saccharomyces cerevisiae*.

According to particular embodiments, the fermenting organism is a yeast of the genus *Pichia*. Non-limiting example of a yeast of the genus *Pichia* are Pichia pastoris and *Pichia kudriavzevii*. According to more particular embodiments, the fermenting organism is *Pichia pastoris*. According to other more particular embodiments, the fermenting organism is *Pichia kudriavzevii*.

Fungi host cells may be derived from, e.g., *Aspergillus*.

According to certain embodiments, the fermenting organism is a fungus, such as a fungi of the genus *Aspergillus*. Non-limiting examples of a fungus of the genus *Aspergillus* are *Aspergillus Oryzae*, *Aspergillus niger* or *Aspergillus awamsii*. According to more particular embodiments, the fermenting organism is *Aspergillus Oryzae*. According to other more particular embodiments, the fermenting organism is *Aspergillus niger*. According to other more particular embodiments, the fermenting organism is *Aspergillus awamsii*.

Algae host cells may be derived from, e.g., *Chlamydomonas, Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

According to certain embodiments, the fermenting organism is an alga, which may be an algae of the genus *Chlamydomonas, Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

According to particular embodiments, the fermenting organism is an alga cell of the genus *Chlamydomonas*. A non-limiting example of an alga of the genus *Chlamydomonas* is *Chlamydomonas reinhardtii*.

According to particular embodiments, the fermenting organism is an alga cell of the genus *Haematococcus*. A non-limiting example of an alga of the genus *Haematococcus* is *Haematococcus pluvialis.*

According to other particular embodiments, the fermenting organism is an alga cell of the genus *Phaedactylum*. A non-limiting example of an alga of the genus *Phaedactylum* is *Phaedactylum tricornatum*.

As mentioned above, the fermenting organism may a recombinant host cell as detailed herein.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the invention.

It should be noted that embodiments and features described in the context of one of the aspects of the present invention also apply to the other aspects of the invention.

Certain other definitions

“Aryl sulfotransferase activity” as used herein refers to the ability of a polypeptide to catalyze the transfer of a sulfate group from a donor molecule to an aryl acceptor molecule.

“Sulfate transporter” or “sulfate permease” are used herein interchangeably to refer to a protein or protein complex that mediates sulfate uptake by a cell.

“ATP sulfurylase” as used herein refers to an enzyme that catalyzes the reaction: ATP + sulfate = diphosphate + adenosine 5’-phosphosulfate (APS).

“APS kinase” as used herein refers to an enzyme that catalyzes the reaction: ATP + adenosine 5’-phosphosulfate (APS) = ADP + 3’-phosphoadenosine 5’-phosphosulfate (PAPS).

“PAP phosphatase” as used herein refers to an enzyme that catalyzes the reaction: 3'-phosphoadenosine 5'-phosphate (PAP)+ H_2O = AMP + phosphate.

“Polypeptide," or "protein" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-transiational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

"Nucleic acid" or "polynucleotide" are used interchangeably herein to denote a polymer of at least two nucleic acid monomer units or bases (e.g., adenine, cytosine, guanine, thymine) covalently linked by a phosphodiester bond, regardless of length or base modification.

"Recombinant" or "non-naturally occurring" when used with reference to, e.g., a host cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant host cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

“Substitution” or “substituted” refers to modification of the polypeptide by replacing one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a polypeptide sequence is an amino acid substitution.

"Conservative substitution" refers to a substitution of an amino acid residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid, e.g., alanine, valine, leucine, and isoleucine; an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain, e.g., serine and threonine; an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain, e.g., phenylalanine, tyrosine, tryptophan, and histidine; an amino acid with a basic side chain is substituted with another amino acid with a basic side chain, e.g., lysine and arginine; an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain, e.g., aspartic acid or glutamic acid; and a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

"Non-conservative substitution" refers to substitution of an amino acid in a polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and

affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

"Deletion" or “deleted” refers to modification of the polypeptide by removal of one or more amino acids in the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide, in various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

"Insertion" or “inserted” refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. Insertions can comprise addition of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the reference polypeptide.

"Host cell" as used herein refers to a living cell or microorganism that is capable of reproducing its genetic material and along with it recombinant genetic material that has been introduced into it - e.g., via heterologous transformation.

"Expression" includes any step involved in the production of a polypeptide (e.g., encoded enzyme) including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded nucleic acid loop into which additional nucleic acid segments can be ligated. Certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". Certain other vectors are capable of facilitating the insertion of an exogenous nucleic acid molecule into a genome of a host cell. Such vectors are referred to herein as "transformation vectors". In general, vectors of utility in recombinant nucleic acid techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of a vector. Large numbers of suitable vectors are known to those of skill in the art and commercially available.

As used herein, "promoter" refers to a sequence of DNA, usually upstream (5') of the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription. The selection of the promoter will depend upon the nucleic acid sequence of interest. A "promoter functional in a host cell" refers to a "promoter" which is capable of supporting the initiation of transcription in said cell, causing the production of an mRNA molecule.

As used herein, "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence. A promoter sequence is "operably-linked" to a gene when it is in sufficient proximity to the transcription start site of a gene to regulate transcription of the gene.

As used herein, an “operon” is a functioning unit of DNA containing a cluster of genes under the control of a single promoter.

"Percentage of sequence identity," "% sequence identity" and "percent identity" are used herein to refer to comparisons between an amino acid sequence and a reference amino acid sequence. The “% sequence identify”, as used herein, is calculated from the two amino acid sequences as follows: The sequences are aligned using Version 9 of the Genetic Computing Group's GAP (global alignment program), using the default BLOSUM62 matrix (see below) with a gap open penalty of -12 (for the first null of a gap) and a gap extension penalty of -4 (for each additional null in the gap). After alignment, percentage identity is calculated by expressing the number of matches as a percentage of the number of amino acids in the reference amino acid sequence.

The following BLOSUM62 matrix is used:

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

"Reference sequence" or “reference amino acid sequence” refers to a defined sequence to which another sequence is compared. In the context of the present invention a reference amino acid sequence may be any amino acid sequence set forth in SEQ ID NO: 1 to 39.

Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and sub ranges within a numerical limit or range are specifically included as if explicitly written out.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

Examples

Example 1 – Production of zosteric acid in *E. coli*

A range of aryl sulfotransferases including SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1), SULT1A1 *Homo sapiens* (SEQ ID NO: 2), SULT1A1 *Equus caballus* (SEQ ID NO: 3), SULT1A1 *Sus scrofa domesticus* (SEQ ID NO: 4), SULT1A1 Canis lupus familiaris (SEQ ID NO: 5) and SULT1E1 *Gallus gallus domesticus* (SEQ ID NO: 6) were expressed in *Escherichia coli*. The respective genes encoding SEQ ID NO. 1, 3, 4, 5, and 6 were cloned amplified from liver tissue cDNA (Zyagen) by PCR using the primers listed in Table 1. The nucleotide sequence of the gene encoding SEQ ID NO: 2 was codon optimized for expression in *Escherichia coli* (GeneArt, Life Technologies) and amplified by PCR using the primers in Table 1. The pETDuet-1 plasmid was digested with restriction endonucleases *Nco*I and *Sal*I. The PCR products were then individually cloned into the plasmid pETDuet-1 using the Gibson reaction (New England Biolabs). The resulting plasmids were transformed into BL21(DE3)pLysS (Life Technologies). Figure 1 shows the plasmid map of the plasmid encoding SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1).

Table 1: Overview of enzymes and primers for cloning aryl sulfotransferases

SEQ ID NO	Name	Fwd Primer	Rev Primer
1	SULT1A1 Rattus norvegicus	CBJP472	CBJP473
2	SULT1A1 Homo sapiens	CBJP470	CBJP471
3	SULT1A1 Equus caballus	CBJP499	CBJP500
4	SULT1A1 Sus scrofa domesticus	CBJP505	CBJP506
5	SULT1A1 Canis lupus familiaris	CBJP503	CBJP504
6	SULT1E1 Gallus gallus domesticus	CBJP501	CBJP502

The strains were grown in M9 minimal media containing glucose as a carbon source, and 0.1 mM IPTG for induction of gene expression as well as 0.1 mM *p*-coumaric acid (pHCA). After four days of growth, samples were withdrawn by filtration and analyzed by HPLC.

The concentration of p-courmaric acid (pHCA) and zosteric acid in the supernatant was quantified by high performance (HPLC) and compared to chemical standards. HPLC was done on a Thermo setup using a HS-F5 column and mobile phases: 5 mM ammonium formate pH 4.0 (A) and acetonitrile (B) at 1.5 mL min-1, using a gradient elution starting at 5% B. From 0.5 min after injection to 7 min, the fraction of B increased linearly from 5% to 60%, and between 9.5 min and 9.6 the fraction of B decreased back to 5%, and remaining there until 12 min. pHCA and zosteric acid were quantified by measuring absorbance at 277 nm.

Table 2 shows the remaining pHCA and the produced zosteric acid in the culture media. Zosteric acid was formed with an aryl sulfotransferase heterologously expressed in a microorganism exemplified by *E. coli* supplied with the substrate.

Table 2: Production of zosteric acid in E. coli from pHCA through the heterologous expression of sulfotransferases.

Enzyme	pHCA remaining (mM)	Zosteric acid formed (mM)
No enzyme	0.10	Not detectable
SULT1A1 Rattus norvegicus	0.02	0.10
SULT1A1 Homo sapiens	0.08	0.02
SULT1A1 Equus caballus	0.09	0.01
SULT1A1 Sus scrofa domesticus	0.09	0.01
SULT1A1 Canis lupus familiaris	0.10	0.01
SULT1E1 Gallus gallus domesticus	0.08	0.01

Example 2 – Increased production of zosteric acid in *E. coli*

The addition of sulfated groups to targets is dependent on supply of the donor molecule 3'-Phosphoadenosine 5'-phosphosulfate (PAPS). We examined if we could increase the production of zosteric acid by overexpressing enzymes providing PAPS and an enzyme that removes the product 3'-Phosphoadenosine 5'-phosphate (PAP).

Table 3: Cloning of enzymes involved in activating sulfate and product removal.

Genes	Fwd Primer	Rev Primer
cysDNC alone	CBJP491	CBJP492
cysDNC for artificial operon	CBJP491	CBJP497
cysQ for artificial operon	CBJP498	CBJP496

In E. coli, the genes *cysD* and *cysN* encode the two subunits of ATP sulfurylase (EC:2.7.7.4), *cysC* encodes APS kinase (EC:2.7.1.25), and *cysQ* encode a PAP phosphatase.

The cysDNC cluster was amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in table3. The plasmid pRSFDuet-1 (Life Technologies) was digested by the restriction endonucleases NdeI and BglII. The gene cluster was inserted into the digested plasmid using the Gibson reaction (New England Biolabs). Figure 2 shows the resulting plasmid. For the combined expression of cysDNC and cysQ in an artificial operon, cysDNCQ, the two parts were amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in Table 3. Again the parts were inserted into the digested plasmid. Figure 3 shows the resulting plasmids. The plasmid expressing SULT1A1 Homo sapiens (SEQ ID NO: 2) from example 1 was co-transformed into *E. coli* BL21(DE3)pLysS cells (Life Technologies) with either the plasmid expressing cysDNC or cysDNCQ.

Cells were grown as in Example 1 and the supernatants were analyzed for product formation as in example 1. The strain expressing SULT1A1 in combination with cysDNCQ was also grown without the addition of IPTIG for induction. Table 4 shows the concentrations of pHCA and zosteric acid.

Table 4: Concentrations of pHCA and zosteric acid in culture media with *E. coli* expressing an aryl sulfotransferase in combination with cysDNC and cysQ.

Enzymes	**Induction**	**pHCA remaining (mM)**	**Zosteric acid formed (mM)**
SULT1A1 Homo sapiens	0.1 mM IPTG	0.08	0.02
SULT1A1 Homo sapiens, CysDNC	0.1 mM IPTG	0.06	0.06
SULT1A1 Homo sapiens, CysDNCQ	0.1 mM IPTG	0.04	0.09
SULT1A1 Homo sapiens, CysDNCQ	None	0.10	Not detectable

This shows that more of the pHCA is transformed into zosteric acid when the protein expression of cysDNC is increased. Even more zosteric acid is formed when the protein expression cysQ is additionally increased.

Example 3 – Decreased toxicity of sulfated product

E. coli MG1655 was grown in chemically defined M9 minimal media with 0.2% glucose as a carbon source without further addition or with the additions of either 10 mM, 20 mM, 25 mM, 30 mM, 35 mM or 40 mM p-coumaric acid (pHCA), or with 20 mM or 40 mM of the sulfate ester of pHCA (zosteric acid). All media preparations had been adjusted to pH 7. Cells were grown at 37°C with 250 rpm shaking in an orbital shaker. The growth rates were examined by following the optical density at 600 nm. The resulting growth rates in exponential growth phase are shown in Figure 4. Filled squares represent growth rates in media with pHCA. Open squares represent growth rates in media with zosteric acid. And the circle represents the growth rate in media without any of these additions. It is evident that the presence of pHCA is toxic to the cells, while the sulfate ester, zosteric acid is much less so.

Example 4 – Decreased growth inhibition with the expression of an aryl sulfotransferase

SULT1A1 from *Homo sapiens* (SEQ ID NO: 2) mentioned in example 1 was cloned into the plasmid vector pET-28a(+) (Novagen, Life Technologies) as follows: The gene encoding the SULT1A1 was codon optimized for expression in E. coli by GeneArt (LifeTechnologies) and synthesized such that the start codon is an NcoI restriction site. The stop codon is immediately followed by a SalI site. The NcoI and SalI fragment of this DNA piece was cloned into the pET-28a(+) digested with NcoI and SalI using T4 DNA ligase. The resulting plasmid was transformed into *E. coli* BL21(DE3)pLysS (Life Technologies) selecting for resistance to 50 mg mL^{-1} kanamycin.

The two strains BL21(DE3)pLysS and the strain carrying SULT1A1 were grown in M9 minimal media with 0.2% glucose, appropriate antibiotics for maintenance of the plasmids, and 0.1 mM IPTG for induction of expression of the sulfotransferase. We added ferulic acid to final concentration of either 2.5 mM or 5 mM (ferulic acid was dissolved in ethanol, which reached a final concentration of 2% in the media). Ferulic acid is a growth inhibitory compound found in lignocellulotic biomass hydrolysate. Cells were inoculated into the medium to an optical density at 600 nm (OD_{600}) of 0.002 in a 1 cm light path. Table 5 below shows the cell density reached after 17 h of growth at 37°C with orbital shaking at 250 rpm for the two strains with or without SULT1A1. Presence of an aryl sulfotransferase removes the inhibition caused by ferulic acid.

Table 5: Cell densities of E. coli reached in media with the growth inhibitor ferulic acid with or without the expression of a sulfotransferase.

	Optical density at 600 nm	
Concentration of inhibitor	**Without SULT1A1**	**With SULT1A1**
2.5 mM ferulic acid	0.680	1.410
5 mM ferulic acid	0.370	1.340

Example 5 – Production of sulfated products in other hosts

We have shown that zosteric acid can be produced in vivo in *Escherichia coli* by expression of an aryl sulfotransferase. To show that the reaction is possible in other microorganisms, we here show that the yeast *Saccharomyces cerevisiae* can also be used as a host for the production.

The gene encoding aryl sulfotransferase SULT1A (Example 1) was cloned after a TEF1 promoter into an episomal plasmid with a 2-micron origin of replication as follows. The gene was amplified by PCR using primers CBJP633 and CBJP634. Alternatively, the gene was codon-optimized for *E. coli* and synthesized by GeneArt and amplified by primers CBJP635 and CBJP636. The TEF1 promoter (Jensen et al., 2014, *FEMS Yeast Res* 14: 238-248) was amplified by PCR using the primers PTEF1_fw and PTEF1_rv. Plasmid pCfB132 (Jensen et al., supra) was digested by restriction enzymes AsiSI and Nt.BsmI. The three fragments – plasmid, TEF1 promotor and SULT1A1-encoding gene – were assembled using a uracil-excission cloning procedure, resulting in plasmids pCBJ283 and pCBJ284, which was subsequently transformed into the *Saccharomyces cerevisiae* strain CEN.PK102-5B selecting for growth on synthetic dropout media plates lacking uracil. A control strain was also made by transformation of pCfB132 into CEN.PK102-5B.

Table 6: Primers

Oligonucleotide	Gene/promoter	Direction	Sequence
CBJP633	SULT1A1 rat	Forward	AGTGCAGGUAAAACAATGgagttctcccgtcca
CBJP634	SULT1A1 rat	Reverse	CGTGCGAUTCAtagttcacaacgaaacttg
CBJP635	SULT1A1 rat (E. coli)	Forward	ATCTGTCAUAAAACAATGgaattttcacgtccgc
CBJP636	SULT1A1 rat	Reverse	CACGCGAUTCAcagttcacaacgaaatttgaa

	(E. coli)		
PTEF1_fw	PTEF1	Forward	Cacgcgaugcacacaccatagcttc
PTEF1_rv	PTEF1	Reverse	Cgtgcgauggaagtaccttcaaaga

The strains were grown in modified Delft medium (Jensen et al., supra) with 20 mg/mL histidine and 60 mg/mL leucine and 10 mM *p*-coumaric acid overnight at 30°C with aeration. The supernatant was then isolated and examined by HPLC as described in Example 1. The table below shows that zosteric acid (ZA) was produced by the strain expressing SULT1A1 and not the control strain lacking a sulfotransferase.

Table 7: Titers of zosteric acid

Sulfotransferase	**µM ZA** (averages and standard deviations of replicate experiments)
None	0 ± 0
SULT1A1 rat (native)	37.8 ± 5.7
SULT1A1 rat (codon optimized for E. coli)	46.2 ± 3.5

It is evident that zosteric acid is formed only when a sulfotransferase is expressed in yeast, and that the gene encoding this may be natural or encoded by a synthetic gene with a specific codon-optimization. Conclusively, the sulfation reactions shown to be catalyzed by sulfotransferases in *E. coli* are also catalyzed when the sulfotransferases are expressed in other organisms, as demonstrated here for the yeast *S. cerevisiae*. The efficacy of production may be affected by means such as the codon-usage of the genes encoding the sulfotransferase. Thus yeast expressing sulfotransferases may be able to detoxify aromatic compounds such as *p*-coumaric acid, and form sulfated products such as zosteric acid.

Example 6 – A range of compounds are substrates for sulfation in vivo

Here we show that the expression of an aryl sulfotransferase may be able to convert several substrates. Some of these are inhibitors that can be found in biomass hydrolyzate used as a substrate for cell growth and production in biotechnology. The compounds also include some that are of biotechnological interest as products of a cell culture or be some whose sulfate ester is of economic interest.

Different sulfotransferases were examined for their substrate specificities against three substrates. We tested the sulfotransferases mentioned in example 1, as well as additional ones. The genes encoding these were cloned as described in example 1 using the primers shown in the table below from cDNA libraries of the respective organisms, except for the SULT1A1 from rat (*Rattus norvegicus*) codon-optimized for *E. coli* (described above). The resulting vectors were transformed into BL21(DE3)pLysS.

Table 8: Primers

Oligonucleotide	Gene	Direction	Sequence
CBJP517	SULT1C1 *Gallus gallus domesticus*	Forward	TAGAAATAATTTTGTTTAACTTTA AGAAGGAGATATACCatggccctgg ataaaatgg
CBJP518	SULT1C1 *Gallus gallus domesticus*	Reverse	TAAGCATTATGCGGCCGCAAGCT TGtcacaattccatgcgaaaaactag
CBJP533	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Forward	TAGAAATAATTTTGTTTAACTTTA AGAAGGAGATATACCatggaattttc acgtcc
CBJP534	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Reverse	TAAGCATTATGCGGCCGCAAGCT TGttacagttcacaacgaaatttg

The resulting strains were grown in M9 medium containing either 100 µM pHCA, 95 µM resveratrol or 87 µM kaempferol. The cultures were grown overnight at 37°C, 300 rpm. The following day the supernatants were isolated and examined by HPLC as described in example 1. BL21(DE3)pLysS were used as a control strain and did not convert the substrates.

Table 9: Percent conversion of the various substrates

Enzyme	pHCA	resveratrol	kaempferol
	100 µM	95 µM	87 µM

SULT1A1 *Rattus norvegicus*	93%	93%	95%
SULT1C1 *Gallus gallus domesticus*	26%	100%	80%
SULT1A1 *Rattus norvegicus* (Codon-optimized for *E. coli*)	73%	58%	38%
SULT1A1 human	39%	36%	97%
SULT1A1 *Equus caballus*	21%	100%	96%
SULT1E1 *Gallus gallus domesticus*	17%	100%	47%
SULT1A1 *Canis lupus familiaris*	34%	61%	60%
SULT1A1 *Sus scrofa domesticus*	8%	88%	45%

The table shows the percent conversion of the various substrates by cells expressing the different sulfotransferases. The results show that several sulfotransferases, and especially the aryl sulfotransferase from rat (*Rattus norvegicus*), may be employed in the sulfation of phenolic compounds.

To further test the range of substrates that can be sulfated, we used strains carrying plasmids expressing SULT1A1 from rat (*Rattus norvegicus*) and SULT1E1 from chicken (*Gallus gallus domesticus*) (Example 1) cloned into the expression vector pETDuet-1, and cysDNCQ from *E. coli* cloned into expression vector pRSFDuet-1 (Example 2). The plasmids were introduced into the *E. coli* expression strain BL21(DE3)pLysS as described previously, selecting for transformants with appropriate antibiotics, namely 34 µg mL^{-1} chloramphenicol for pLysS, 100 µg mL^{-1} ampicillin for pETDuet-1-based vectors, and 100 µg mL^{-1} kanamycin for pRSFDuet-1-based vectors. The table below shows the combination of over-expressed genes on plasmids. A control strain without a sulfotransferase gene or cysDNCQ operon was also examined.

Table 10: Combination of over-expressed genes on plasmids

E. coli strains	Sulfotransferase	Cys genes
Control strain	-	-
SULT1A1 rat	SULT1A1 rat	-

SULT1E1 chicken	SULT1E1 chicken	-
SULT1A1 rat + CysDNCQ	SULT1A1 rat	CysDNCQ

The strains were precultured in 2xYT medium with appropriate antibiotics. 10 µL of these precultures were used to inoculate M9 media with 1 mM IPTG and none or a single substrate for sulfation. After overnight growth at 37°C, 300 rpm the supernatants were withdrawn and examined by HPLC as described in Example 1. The compounds were detected by UV absorbance. The table below shows the percent reduction in concentration in the strains expressing sulfotransferases alone or in combination with cysDNCQ genes when compared to the control strain.

Table 11: Percent reduction

Compound	Start concentration in µM	SULT1A1	SULT1E1	SULT1A1 + CysDNCQ
Ferulic acid	110	72%	67%	100%
Quercetin	85	75%	74%	81%
4-hydroxybenzoic acid	287	5%	4%	6%
4-acetamidophenol	114	24%	10%	30%
3-Hydroxy-4-methoxycinnamic acid	132	51%	24%	62%
4-Hydroxyphenylpyruvic acid	255	47%	100%	64%
3-(4-Hydroxyphenyl)propionic acid	241	3%	1%	7%
Vanillic acid	173	33%	0%	39%
Luteolin	61	27%	0%	37%
Apigenin	77	41%	98%	99%
fisetin	81	98%	98%	100%

Conclusively, a wide range of phenolic compounds are substrates for sulfotransferases. In the shown examples, the conversion is enhanced by the overexpression of cysDNCQ genes. Some of these compounds and their sulfate esters are of interest in biotechnology. Also, some of these compounds are inhibitors of cell growth and function, and thus conversion by sulfation is of interest for use in biological systems.

Example 7 – Use of sulfotransferases in complex biomass

A preferred source of carbon for cell growth and as substrates for manufacture of chemicals and other products in biotechnology is a complex carbon source, such as biomass hydrolysate. These complex carbon sources often contains phenolic compounds that are toxic to prokaryotic as well as eukaryotic organisms (Adeboye et al., 2014, *AMB Express* **4:** 46-014-0046-7. eCollection 2014)

Here, we examine pre-treated wheat straw biomass hydrolysate in growth medium for a biotechnological relevant organism, exemplified by *E. coli*. Specifically, we used strains carrying plasmids expressing SULT1A1 from rat (Rattus norvegicus) and SULT1E1 from chicken (Gallus gallus domesticus) (Example 1) cloned into the expression vector pETDuet-1, and cysDNCQ from *E. coli* cloned into expression vector pRSFDuet-1 (Example 2) as described previously.

Cells expressing no sulfotransferase or SULT1A1 from rat were grown in M9 media with 1 mM IPTG with or without inhibitors of growth. Biomass hydrolysate was prepared by mixing 100 g biomass hydrolysate with 200 g of water for 2 h at 37°C, followed by centrifugation and filtration through a 0.2 µL-filter. Figure 5 shows the growth curves of cultures without sulfotransferase (solid lines) or with SULT1A1 (dotted lines) growing in M9 without supplements (black), with 8% biomass hydrolysate (BH8%, dark grey), or with 12% biomass hydrolysate (BH12%, light grey).

There is no difference in growth rate between cultures in M9, but with biomass hydrolysate the growth rate of the strain expressing a gene encoding SULT1A1 is significantly higher. Conclusively, cells expressing sulfotransferase have a growth advantage in media containing phenolic compounds that inhibit the growth cells, such as biomass hydrolysate being used as a carbon source for biotechnological production of chemicals.

Example 8 – Increasing uptake of sulfate

E. coli BL21(DE3)-derived strains expressing an aryl sulfotransferase and a sulfate transporter were constructed as follows.

A plasmid (Figure 6) was constructed for the over-expressing of the CysZ (NCBI reference sequence NP_416908.1) from *E. coli* by amplifying the cysZ gene from the chromosome of E. coli MG1655 by PCR using the primers in the table below. The resulting PCR product and the plasmid expressing SULT1A1 from rat mentioned in example 1 (Figure 1) were digested using the restriction enzymes *HindIII* and *NotI*, purified by column purification and ligated together with T4 DNA polymerase, and the ligation reaction was used to transform the *E. coli* cloning strain NEB5α (New England Biolabs). Resulting colonies resistant to ampicillin were tested for correct insert by PCR using primers pET-Upstream (ATGCGTCCGGCGTAGA) and DuetDOWN1 (GATTATGCGGCCGTGTACAA). The correct plasmid was purified and transformed into *E. coli* BL21(DE3) (Life Technologies) together with the plasmid encoding CysDNCQ from example 2 (Figure 3), selecting for both ampicillin and kanamycin.

Similarly a plasmid (Figure 7) was constructed for the over-expressing of the CysP (GenBank AAC75478.1), CysT (GenBank AAC75477.1, CysW (GenBank AAC75476.2) and CysA (Genbank AAC75475.1) from *E. coli*. The cysPTWA (also known as cysPUWA) operon was amplified from the chromosome of E. coli MG1655 by PCR using the primers in the table below. The resulting PCR product and the plasmid expressing CysDNCQ mentioned in example 2 (Figure 3) were digested using restriction enzymes HindIII and NotI, purified by column purification and ligated together with T4 DNA polymerase, and the ligation reaction was used to transform the *E. coli* cloning strain NEB5α (New England Biolabs). Resulting colonies resistant to kanamycin were tested for correct insert by PCR using primers ACYCDuetUP1 (GGATCTCGACGCTCTCCCT) and DuetDOWN1 (GATTATGCGGCCGTGTACAA). The correct plasmid was purified and transformed into *E. coli* BL21(DE3) together with the plasmid encoding SULT1A1 from rat mentioned in example 1 (Figure 1), selecting for both ampicillin and kanamycin.

Table 12: Primers

Oligonucleotide	Gene(s)	Direction	Sequence	Restriction site
CBJP891	*cysZ*	Forward	ttaaaagcttgggattggtcaaaaggagctcatcc	*Hin*dIII

CBJP892	*cysZ*	Reverse	aatagcggccgcttaccgccacatcgcgtgtttat	*Not*I
CBJP893	*cysPTWA*	Forward	ttaaaagcttagaaagtcattaaatttataagggtgcgca	*Hin*dIII
CBJP894	*cysPTWA*	Reverse	aatagcggccgctcaggcgctttgtgcgagagc	*Not*I

Control strains carrying only the emptive plasmids pETDuet-1 and pRSFDuet-1 (Life Technologies), carrying the plasmid encoding SULT1A1 from rat (Figure 1) and pRSFDuet-1, or carrying the plasmid encoding SULT1A1 from rat (Figure 1) and the plasmid encoding cysDNCQ (Figure 3) were used as controls for growth experiments.

The *E. coli* strains were propagated overnight in grown in M9 minimal medium containing 0.2 % (w/v) glucose, 2 mM *p*-coumaric acid, 100 µg/mL ampicillin and 50 µg/mL kanamycin in wells of a 96-well deep-well plate (EnzyScreen) shaking at 300 rpm in an orbital shaker at 37°C. From these cultures, 30 µL was used to inoculate 500 µL of M9 medium with 0.2% (w/v) glucose, 2 mM p-coumaric acid, 100 µg/mL ampicillin, 50 µg/mL kanamycin and further 200 µM IPTG in wells of a 96-well deep-well plate (EnzyScreen) that was left shaking at 300 rpm in an orbital shaker at 37°C overnight. The cell density was then measured by the optical density at 600 nm, and the supernatant was sampled for production of zosteric acid, by two rounds of centrifugation. Zosteric acid in the supernatant was quantified by high performance liquid chromatography (HPLC) and compared to a chemical standard. HPLC was done on a Thermo setup using a HS-F5 column (3 µm) and mobile phases: 5 mM ammonium formate pH 4.0 (A) and acetonitrile (B) at 1.5 mL min^{-1}, using a gradient elution starting at 5% B. From 0.5 min after injection to 7 min, the fraction of B increased linearly from 5% to 60%, and between 9.5 min and 9.6 the fraction of B decreased back to 5%, and remaining there until 12 min. Zosteric acid were quantified by measuring absorbance at 290 nm.

Figure 8 shows that increased titers of zosteric acid were reached when overexpressing CysZ and CysPTWA.

It is clear that the additional expression of a sulfate transporter such as that encoded by *cysZ* (proton symporter) or by *cysPTWA* (ABC transporter) enhanced the sulfation catalyzed by a phenol sulfotransferase, optionally with increased expression of sulfate adenylyltransferase, APS kinase and adenosine-3',5'-bisphosphate nucleotidase. The sulfate

transport activity may be obtained by the activity of transporters belonging to several different families of transporters. As a consequence, the detoxification of biomass inhibitors would also be expected to be improved in strains with increased sulfate transport activity.

Example 9 - increased detoxification upon in strains with increased sulfate uptake

To show that the increased sulfate uptake gives increased efficiency for detoxification of growth inhibitors, we tested this on agar plates as follows. The four strains listed in Example 8 expressing either (1) no genes, (2) SULT1A1 from rat, (3) SULT1A1 from rat in combination with CysDNCQ, or (4) SULT1A1 from rat in combination with CysDNCQ and CysPTWA were used. Additionally, a strain encoding SULT1A1 on one plasmid and encoding CysZ on the plasmid also encoding CysDNCQ was made. The plasmid (Figure 9) for the expression of the CysZ transporter on the plasmid also encoding CysDNCQ was made as follows.

The cysZ gene was amplified by PCR using primers CBJP891 and CBJP892 as described in Example 8. The resulting PCR product and the plasmid carrying the cysDNCQ operon (Figure 3) were digested using restriction enzymes HindIII and NotI, purified by column purification and ligated together with T4 DNA polymerase, and the ligation reaction was used to transform the E. coli cloning strain NEB5α (New England Biolabs). Resulting colonies resistant to kanamycin were tested for correct insert by PCR using primers ACYCDuetUP1 (GGATCTCGACGCTCTCCCT) and DuetDOWN1 (GATTATGCGGCCGTGTACAA). The correct plasmid was purified and transformed into *E. coli* BL21(DE3) together with the plasmid encoding SULT1A1 from rat mentioned in example 1 (Figure 1), selecting for both ampicillin and kanamycin.

The five strains were grown overnight in M9 minimal medium containing 0.2 % (w/v) glucose, 2 mM p-coumaric acid, 100 µg/mL ampicillin and 50 µg/mL kanamycin in wells of a 96-well deep-well plate (EnzyScreen) shaking at 300 rpm in an orbital shaker at 37°C, reaching an optical density of 2.5 measured at 600 nm in over a 1 cm light path. The cultures were diluted in 0.9 % (w/v) NaCl (aq) to 10^{-3}, 10^{-4}, and 10^{-5} of the original concentration of cells. 5 µL of these dilutions were spotted on LB agar plates to which 20 mM MgSO4, 100 µM IPTG and 100 µg/mL ampicillin as well as 0 mM, 5 mM or 8 mM p-coumaric acid from a stock solution of 500 mM p-coumaric acid in 99.9% EtOH.

Additionally, plates were made using 0 mM, 5 mM or 8 mM ferulic acid from a stock solution of 250 mM ferulic acid in 99.9% EtOH, instead of p-coumaric acid. The plates were left in a 37°C incubator overnight. The plates were screened for appearance of colonies, the results of which is shown in Table 13.

Table 13: Colonies on Agar plates containing p-coumaric acid or ferulic acid

	0 mM *p*-coumaric acid			**5 mM *p*-coumaric acid**			**8 mM *p*-coumaric acid**		
Strain carrying plasmids expressing	**10^{-3}**	**10^{-4}**	**10^{-5}**	**10^{-3}**	**10^{-4}**	**10^{-5}**	**10^{-3}**	**10^{-4}**	**10^{-5}**
None	>50	>50	>50	0	0	0	0	0	0
SULT1A1	>50	>50	>50	10	0	0	0	0	0
SULT1A1 + CysDNCQ	>50	>50	>50	12	1	0	0	0	0
SULT1A1 + CysDNCQ + CysZ	>50	>50	>50	>50	4	0	0	0	0
SULT1A1 + CysDNCQ + CysPTWA	>50	>50	>50	>50	16	3	0	0	0
	0 mM ferulic acid			**5 mM ferulic acid**			**8 mM ferulic acid**		
Strain carrying plasmids expressing	**10^{-3}**	**10^{-4}**	**10^{-5}**	**10^{-3}**	**10^{-4}**	**10^{-5}**	**10^{-3}**	**10^{-4}**	**10^{-5}**
None	>50	>50	>50	0	0	0	0	0	0
SULT1A1	>50	>50	>50	6	1	0	0	0	0
SULT1A1 + CysDNCQ	>50	>50	>50	>50	5	0	0	0	0
SULT1A1 + CysDNCQ + CysZ	>50	>50	>50	>50	14	1	0	0	0
SULT1A1 + CysDNCQ + CysPTWA	>50	>50	>50	>50	>50	7	0	0	0

">50" indicates that there were more than 50 colonies, which made it impossible to distinguish individual colonies

It is evident that the expression of SULT1A1 allows for growth in the presence of a phenolic biomass inhibitor. Furthermore, the expression of CysDNCQ and sulfate transporters allows for growth at even higher concentrations of biomass inhibitors.

Example 10 – Different compounds as substrates

Further compounds may be sulfated by sulfotransferases and cells heterologously expressing such, which is shown by the following experiment.

E. coli strain KRX (obtained from Promega) was transformed with either plasmids pETDuet-1 or the derived plasmid encoding SULT1A1 from *Rattus norvegicus* described in Example 1.

M9 medium containing 0.2% glucose, 0.1 mM IPTG, 0.1 % rhamnose, and 100 µg/mL ampicillin was prepared. To aliquots of the medium, phenolic compounds (Table 14) were added from 10 mM stock solutions in 99.9 % ethanol to a final concentration of the compounds of 50 µM.

The strains described above were grown in 2xYT medium with 100 µg/mL ampicillin overnight before they were used to inoculate the media by 50-fold dilution. The cultures were grown overnight with vigorous shaking at 37 °C. The supernatants were isolated by centrifugation and subjected to HPLC analysis as described in example 1. The compounds were detected by UV absorbance except for 4-methylbelliferone, which was measured by fluorescence.

Table 14 shows that the phenolic compounds were all subject to sulfation, when a sulfotransferase is present in the medium. Additional peaks corresponding to more hydrophilic compounds were also the result of the activity of the sulfotransferase on each of the compounds.

Table 14: Percent reduction

Compound	**SULT1A1**
Sinapic acid	31%
Naringenin	60%
4-ethylphenol	100%
4-vinylphenol	100%
4-ethylguaiacol	50%
4-methylbelliferone	100%

4-nitrophenol	30%

Conclusively, a wide range of phenolic compounds are substrates for sulfotransferases. Some of these compounds and their sulfate esters are of interest in biotechnology or as markers of sulfation activity or as donors of sulfate in the reverse reaction. Also, some of these compounds are inhibitors of cell growth and function, and thus conversion by sulfation is of interest for use in biological systems. The example shows that the phenolic acceptor molecule for sulfation may differ by the position of the hydroxyl-group and still remains an active substrate. 4-vinylphenol is a degradation product of *p*-coumaric acid by decarboxylation, and it is still a substrate for sulfation by a sulfotransferase, showing that the side chain can vary significantly, and the compound remains an active substrate.

Example 11 - Different sulfotransferases are active

The sulfotransferases may be of very different sequences. To show this, we tested the sulfotransferases with lower homology to the sequences presented in the examples above against three substrates.

The gene (sequence SEQ ID NO: 76) encoding dmST1 (sequence SEQ ID NO: 77) from *Drosophila melanogaster* was amplified from cDNA using primers listed in Table 15 similarly to the cloning of genes described in Example 1.

Table 15: Primers

Oligonucleotide	Gene	Direction	Sequence
CBJP474	dmST1 *Drosophila melanogaster*	Forward	TAGAAATAATTTTGTTTAACTTTAAGAA GGAGATATAC C ATGCCCCAGTCGAGCTTCTT
CBJP475	dmST1 *Drosophila melanogaster*	Reverse	TAAGCATTATGCGGCCGCAAGCTTG TTACGTGGACGCAAACTTGCT

The gene (sequence SEQ ID NO: 78) encoding SULT1ST1 (sequence SEQ ID NO: 79) from *Danio rerio* was codon-optimized for *E. coli* and synthesized as in example 1.

The gene (sequence SEQ ID NO: 80) encoding SULT6B1 (sequence SEQ ID NO: 81) from *Danio rerio* was codon-optimized for *E. coli* and synthesized as in example 1.

The gene (sequence SEQ ID NO: 82) encoding Hoch_6098 (sequence SEQ ID NO: 83) from the bacterium *Haliangium ochraceum* DSM 14365 was codon-optimized for E. coli and synthesized as in example 1.

The genes were cloned into the vector pETDuet-1 as described in Example 1, and the resulting vectors were transformed into *E. coli* KRX (obtained from Promega). Strains were grown as described above in presence of either 100 µM resveratrol, 20 µM kaempferol or 50 µL 3-hydroxy-4-methoxycinnamic acid, and the supernatant were analyzed as described above, except that for kaempferol the cultures were mixed with an equal volume of methanol before isolation of the supernatants. Table 16 shows the reduction in the concentrations of compounds in the presence sulfotransferases (n.d. = not determined).

Table 16: Percent reduction

	Resveratrol	Kaempferol	3-hydroxy-4-methoxycinnamic acid
D. rerio SULT1ST1	80%	n.d.	12%
D. rerio SULT6B1	n.d.	44%	n.d.
D. melanogaster dmST1	n.d.	n.d.	12%
H. ochraceum DSM 14365 Hoch_6098	7%	100%	n.d.

The example shows that the sulfation reaction may occur in a medium with a cell expressing a heterologous sulfotransferase more distantly related to the sequences in the previous examples. It may even be of non-animal origin, exemplified with the bacterial sulfotransferase from *Haliangium ochraceum*.

Claims

1. A process for producing a hydrolyzed product from a lignocellulose-containing material, comprising the steps of:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b);

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

2. A process for the production of a fermentation product from a lignocellulose-containing material, comprising the steps:

(a) pre-treating a lignocellulose-containing material;

(b) hydrolyzing the pre-treated lignocellulose-containing material to form a hydrolyzate;

(c) enzymatically sulfating phenolic compounds produced during step a) and/or step b).

(d) fermenting the hydrolyzate obtained in step (b) using a fermenting organism, thereby obtaining a fermentation product;

wherein step (c) is performed using a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

3. The process according to claim 2, wherein the fermenting organism in step (d) is the recombinant host cell used in step (c).

4. The process according to claim 2 or 3, wherein the fermentation product is a biofuel or biochemical.

5. The process according to any one of claims 1 to 4, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

1b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has an aryl sulfotransferase activity; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has an aryl sulfotransferase activity.

6. The process according to any one of claims 1 to 5, wherein said recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

7. The process according to claim 6, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter, by modifying the ribosome binding site and/or by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

8. The process according to any one of claims 1 to 7, wherein said recombinant host cell comprises an exogenous nucleic acid molecule comprising one or more nucleotide sequences encoding a sulfate transporter operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

9. The process according to any one of claims 6 to 8, wherein the sulfate transporter is a bacterial sulfate transporter.

10. The process according to any one of claims 6 to 9, wherein the sulfate transporter is selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

11. The process according to any one of claims 1 to 10, wherein the recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

12. The process according to any one of claims 1 to 11, wherein the recombinant host cell is selected from the group consisting of bacteria, yeasts, fungi, and algae.

13. The process according to any one of claims 1 to 12, wherein the recombinant host cell is a bacterium.

14. The process according to any one of claims 1 to 13, wherein in step (b) the pre-treated lignocellulose-containing material is enzymatically hydrolyzed.

15. The process according to any one of claims 1 to 14, wherein in step a) said lignocellulose-containing material is pre-treated chemically, mechanically and/or biologically.

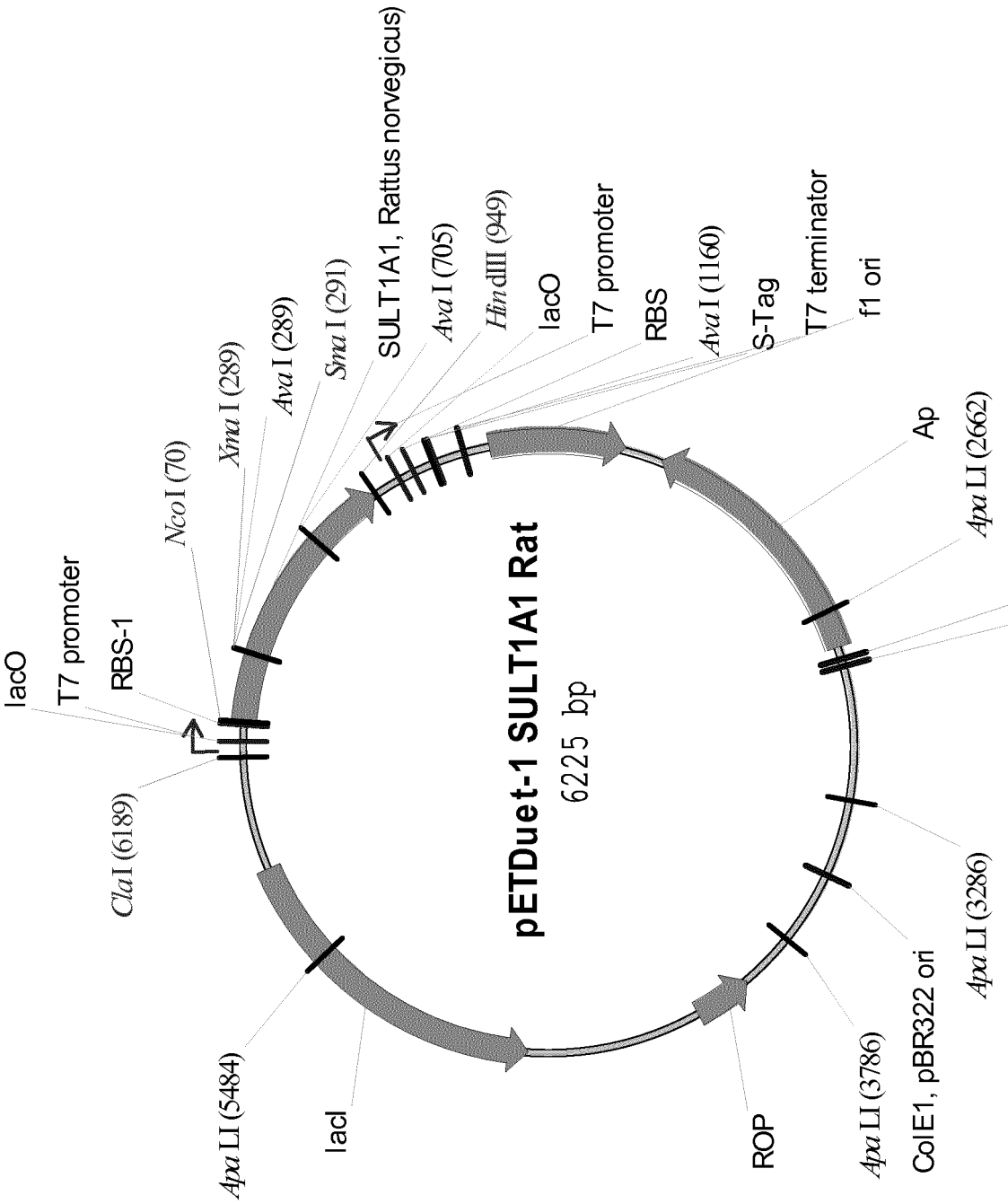
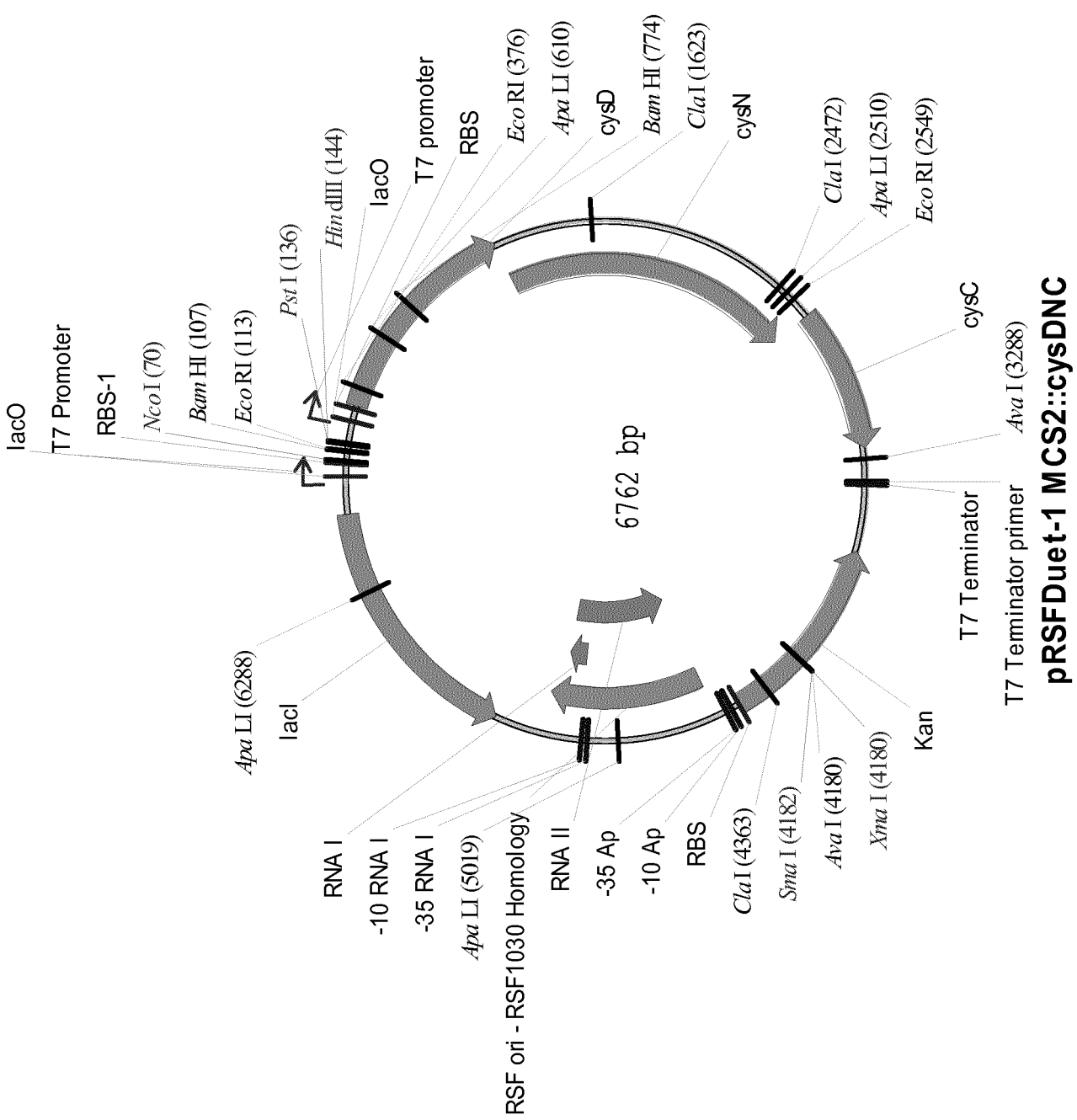
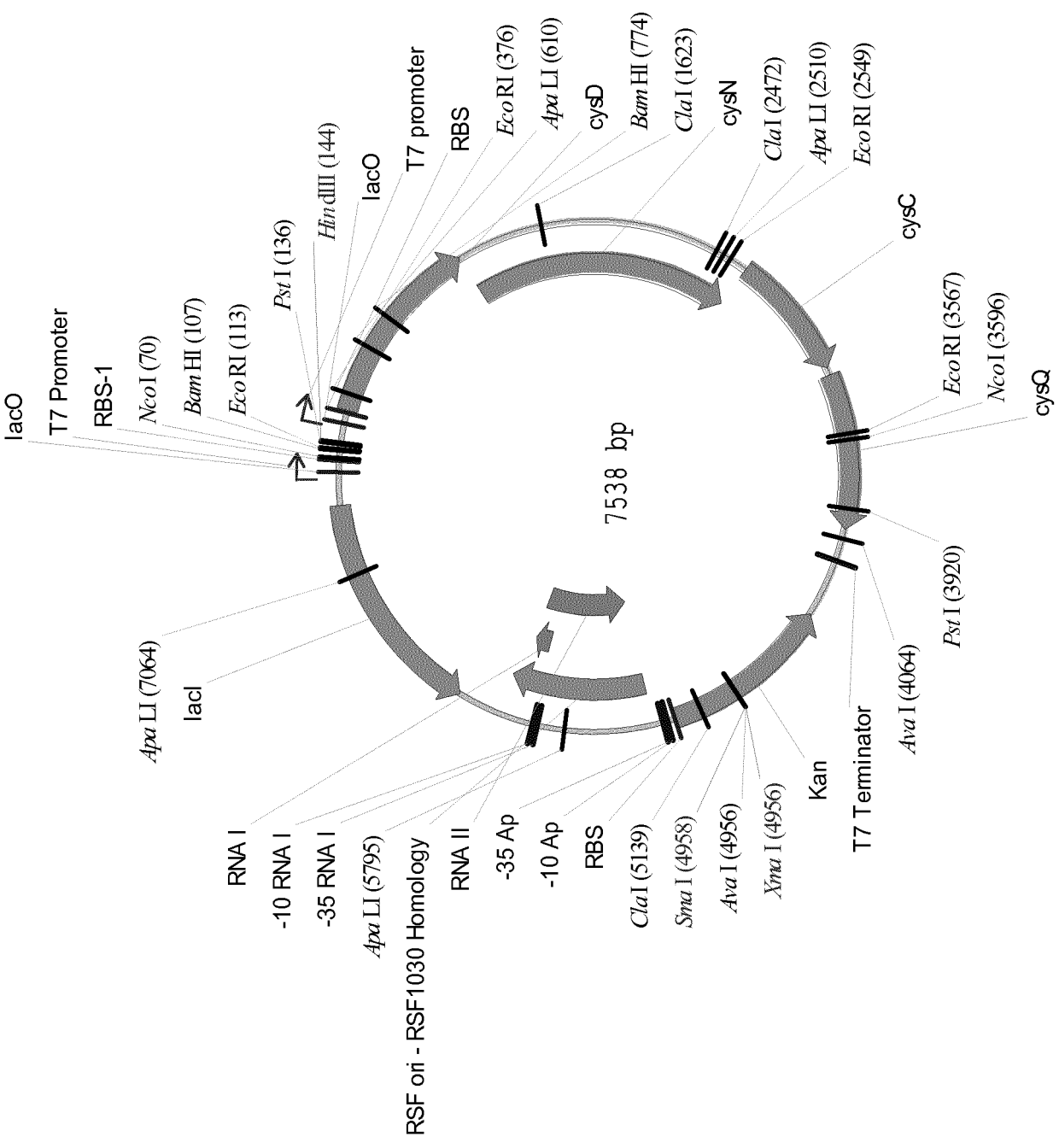

pETDuet-1 SULT1A1 Rat
6225 bp
lacO
T7 promoter
RBS-1
NcoI (70)
XmaI (289)
AvaI (289)
SmaI (291)
SULT1A1, Rattus norvegicus)
AvaI (705)
HindIII (949)
lacO
T7 promoter
RBS
AvaI (1160)
S-Tag
T7 terminator
f1 ori
Ap
ApaLI (2662)
-10 Ap
-35 Ap
ApaLI (3286)
ColE1, pBR322 ori
ApaLI (3786)
ROP
lacI
ApaLI (5484)
ClaI (6189)

Figure 1



pRSFDuet-1 MCS2::cysDNC

Figure 2



pRSFDuet-1 MCS2::cysDNC-Q

Figure 3

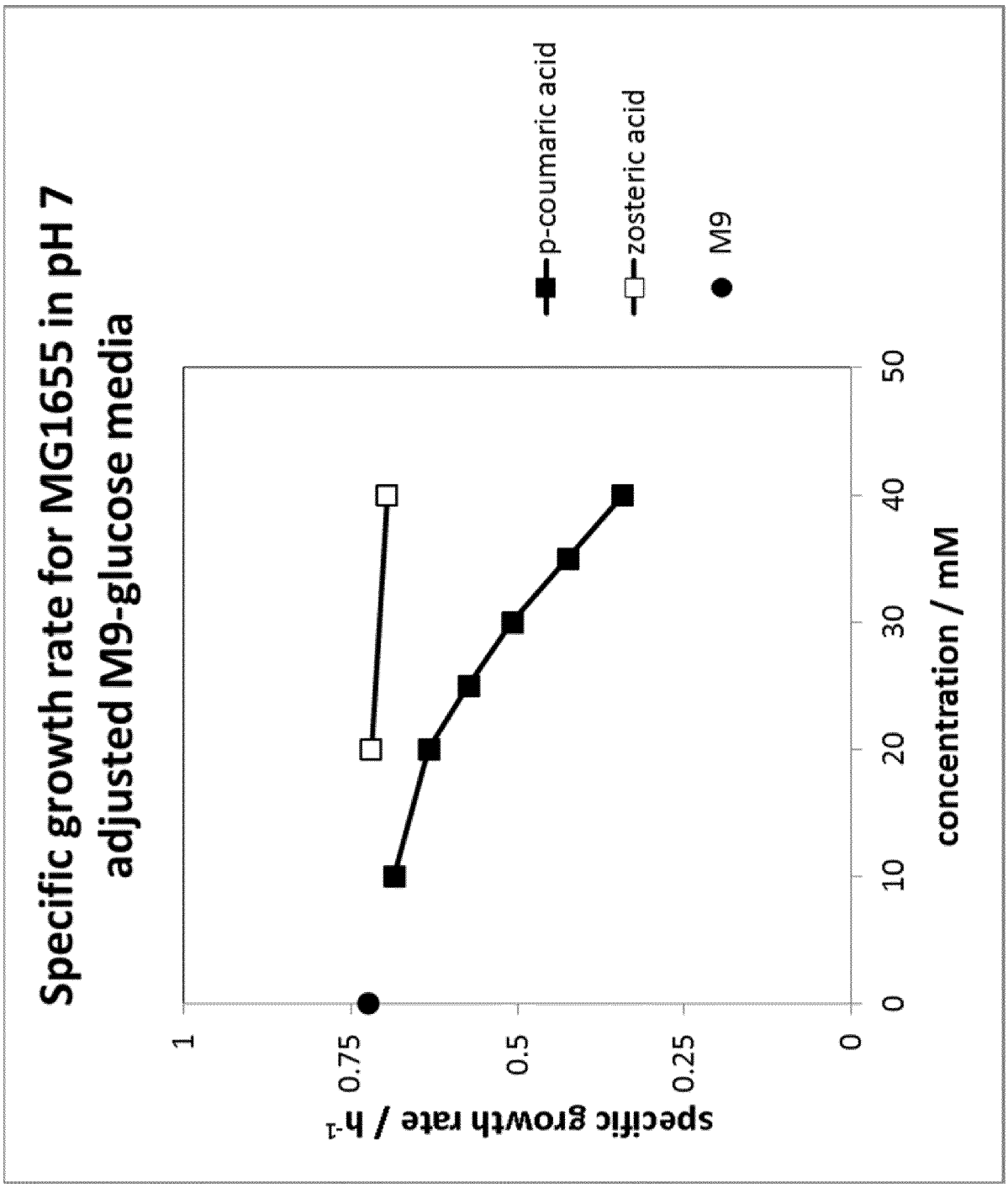

Specific growth rate for MG1655 in pH 7 adjusted M9-glucose media
p-coumaric acid
zosteric acid
M9
specific growth rate / h-1
concentration / mM
0
0.25
0.5
0.75
1
0
10
20
30
40
50

Figure 4

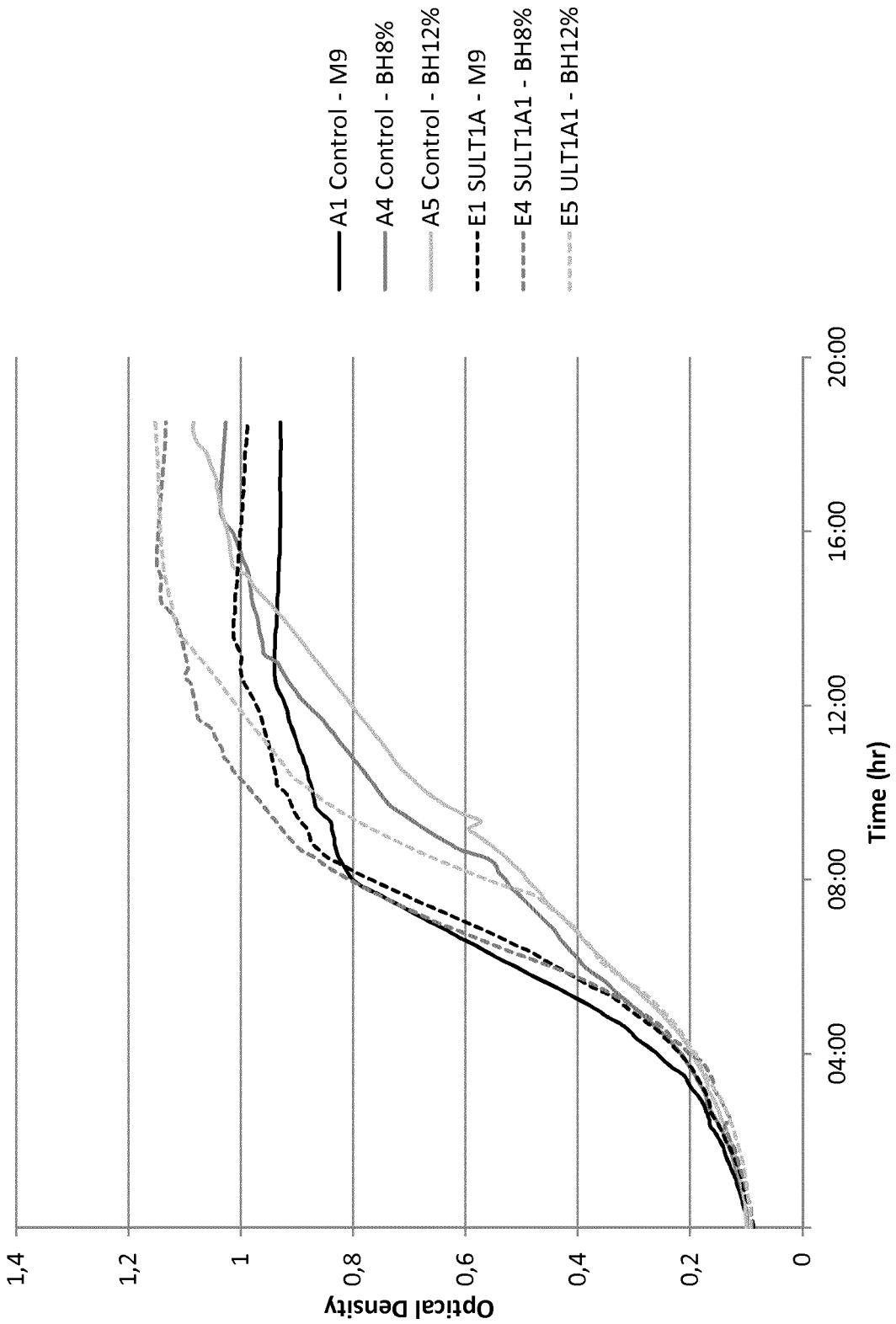

A1 Control - M9
A4 Control - BH8%
A5 Control - BH12%
E1 SULT1A - M9
E4 SULT1A1 - BH8%
E5 ULT1A1 - BH12%
Optical Density
0
0,2
0,4
0,6
0,8
1
1,2
1,4
04:00
08:00
12:00
16:00
20:00
Time (hr)

Figure 5

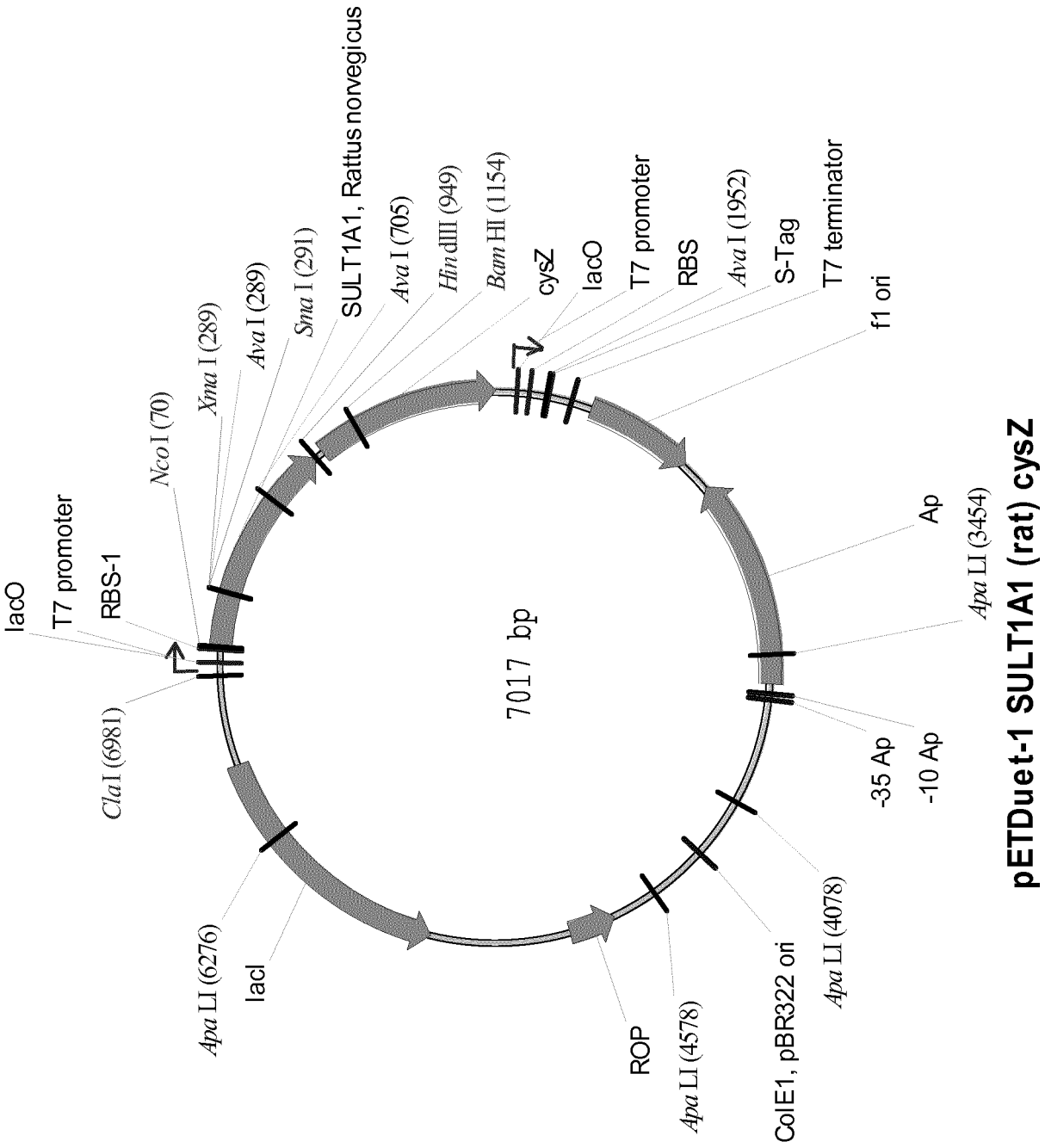
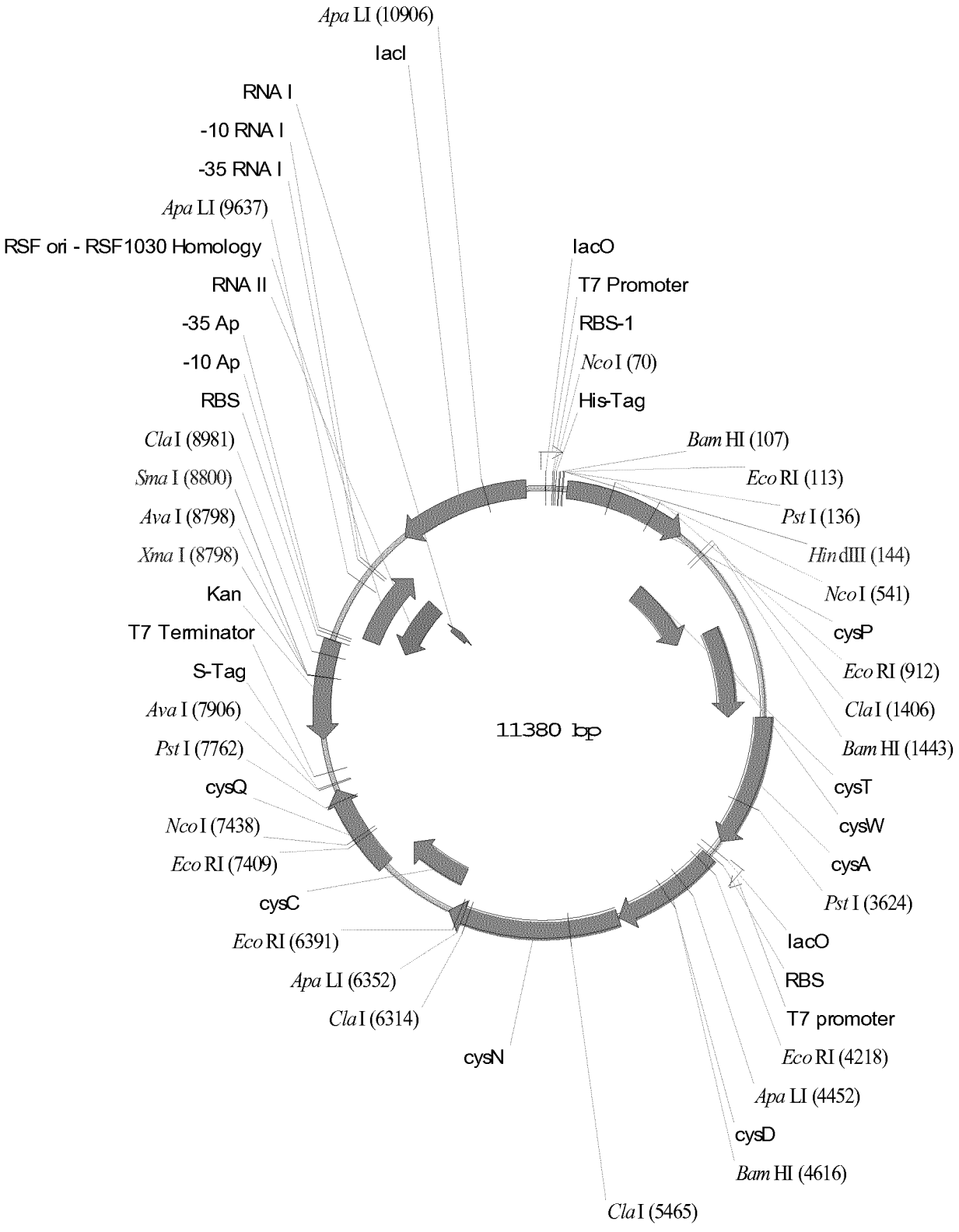

pETDuet-1 SULT1A1 (rat) cysZ
7017 bp
lacO
T7 promoter
RBS-1
NcoI (70)
XmaI (289)
AvaI (289)
SmaI (291)
SULT1A1, Rattus norvegicus
AvaI (705)
HindIII (949)
BamHI (1154)
cysZ
lacO
T7 promoter
RBS
AvaI (1952)
S-Tag
T7 terminator
f1 ori
Ap
ApaLI (3454)
-35 Ap
-10 Ap
ApaLI (4078)
ColE1, pBR322 ori
ApaLI (4578)
ROP
lacI
ApaLI (6276)
ClaI (6981)

Figure 6



pRSFDuet-1 MCS1::cysPTWA MCS2::cysDNCQ

Figure 7

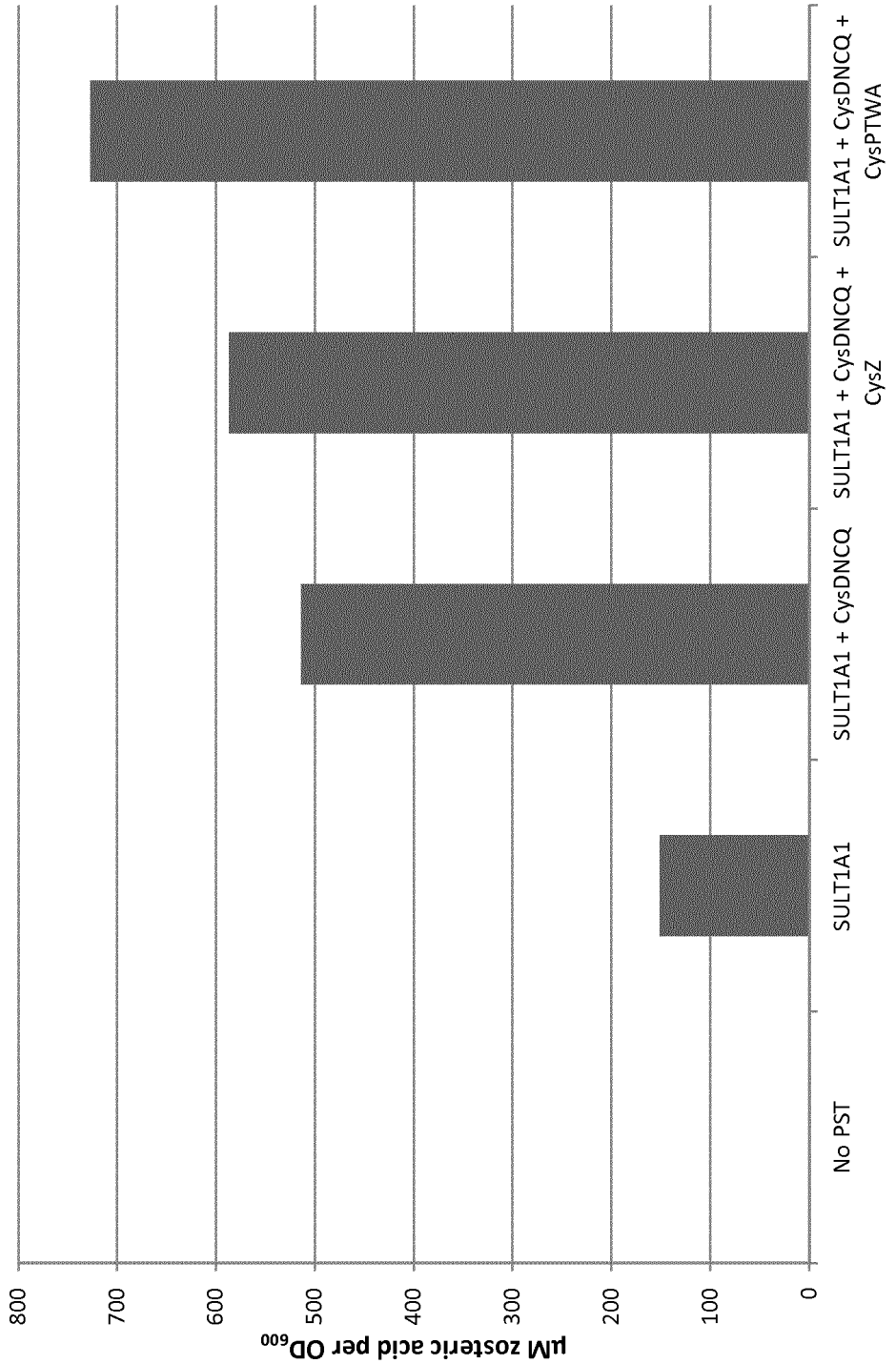
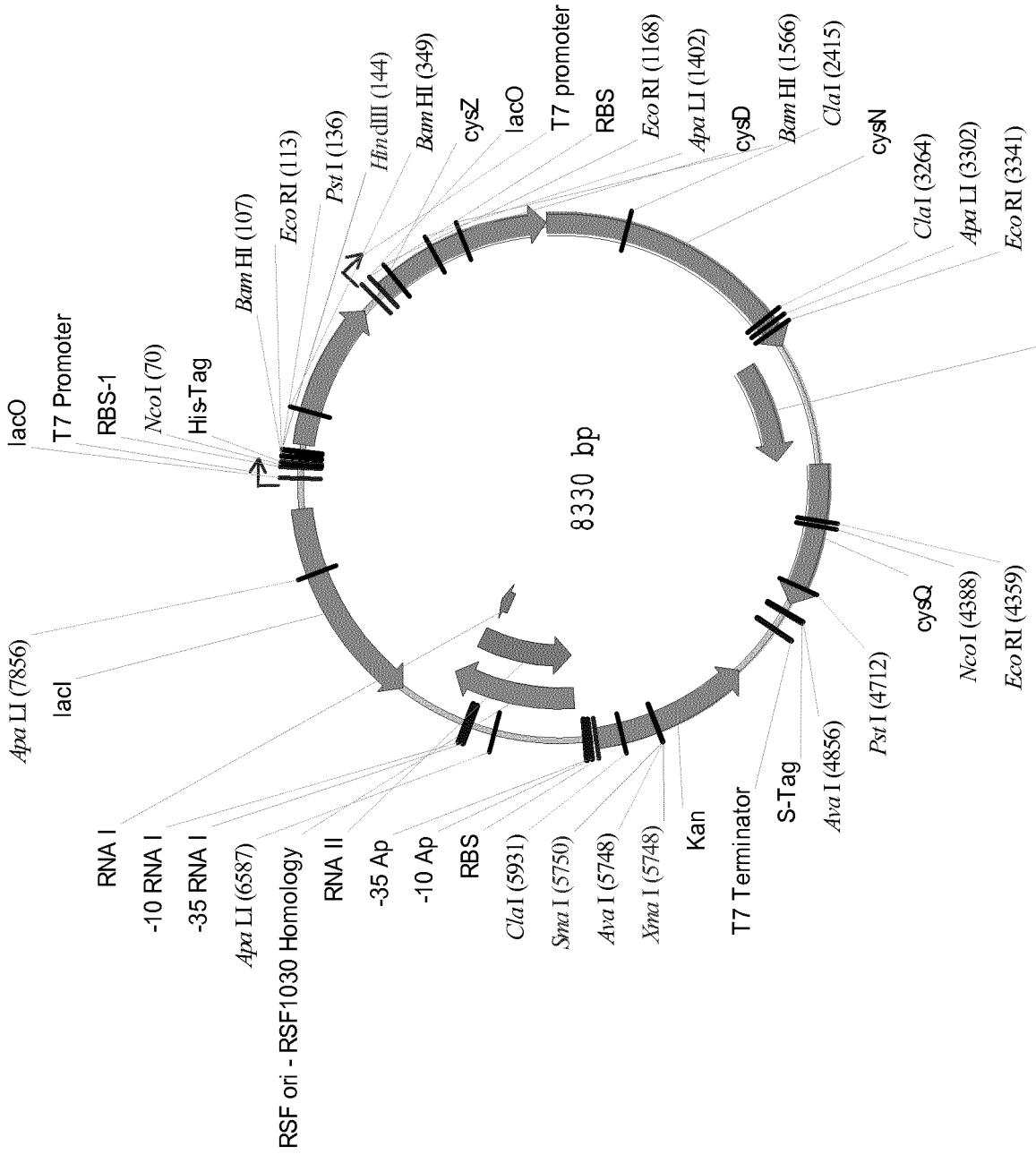

800
700
600
500
400
300
200
100
0
µM zosteric acid per OD600
No PST
SULT1A1
SULT1A1 + CysDNCQ
SULT1A1 + CysDNCQ + CysZ
SULT1A1 + CysDNCQ + CysPTWA

Figure 8



pRSFDuet-1 MCS1::cysZ MCS2::cysDNCQ

Figure 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/054344

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P5/00 C12P11/00 C12N9/10
ADD.

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)
C12P C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98/03636 A1 (BIO TECH RESOURCES [US]; GRUND ALAN DOUGLAS [US]; MAURINA BRUNKER JULI) 29 January 1998 (1998-01-29) the whole document	1-15
A	WO 2008/076738 A2 (NOVOZYMES NORTH AMERICA INC [US]; HOLMES JASON W [US]; DEINHAMMER RAND) 26 June 2008 (2008-06-26) the whole document	1-15

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* 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	Date of mailing of the international search report
24 April 2017	10/05/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Schneider, Patrick

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Xiang Guo ET AL: "R Comparison of methods for detoxification of spruce hydrolysate for bacterial cellulose production", , 12 October 2013 (2013-10-12), XP055165033, Retrieved from the Internet: URL:http://www.microbialcellfactories.com/content/pdf/1475-2859-12-93.pdf [retrieved on 2015-01-27] the whole document -----	1-15
A,P	WO 2016/026977 A1 (UNIV DANMARKS TEKNISKE [DK]) 25 February 2016 (2016-02-25) the whole document -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2017/054344

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9803636 A1	29-01-1998	AU 4326397 A	10-02-1998
		US 6225100 B1	01-05-2001
		WO 9803636 A1	29-01-1998
WO 2008076738 A2	26-06-2008	BR PI0720529 A2	04-02-2014
		EP 2094907 A2	02-09-2009
		US 2008171370 A1	17-07-2008
		WO 2008076738 A2	26-06-2008
WO 2016026977 A1	25-02-2016	NONE	