



(86) Date de dépôt PCT/PCT Filing Date: 2012/03/07
(87) Date publication PCT/PCT Publication Date: 2012/05/18
(85) Entrée phase nationale/National Entry: 2013/09/04
(86) N° demande PCT/PCT Application No.: NL 2012/050141
(87) N° publication PCT/PCT Publication No.: 2012/064195
(30) Priorité/Priority: 2011/03/08 (NL2006359)

(51) Cl.Int./Int.Cl. *C12N 9/00* (2006.01),
C12N 1/15 (2006.01), *C12N 1/21* (2006.01),
C12N 5/10 (2006.01), *C12P 17/04* (2006.01)
(71) Demandeur/Applicant:
PURAC BIOCHEM B.V., NL
(72) Inventeurs/Inventors:
WIERCKX, NICK JOHANNES PETRUS, NL;
ELINK SCHUURMAN, TOM DANIEL, NL;
KUIJPER, SIPKO MAARTEN, NL;
RUIJSSENAARS, HARALD JOHAN, NL
(74) Agent: SMART & BIGGAR

(54) Titre : CELLULE MODIFIEE GENETIQUEMENT ET PROCEDE D'UTILISATION DE CETTE CELLULE
(54) Title: GENETICALLY MODIFIED CELL AND PROCESS FOR USE OF SAID CELL

(57) **Abrégé/Abstract:**

The present invention relates to the field of biotransformation of furanic compounds. More particular the present invention relates to novel genetically modified cells with improved characteristics for biocatalytic transformation of furanic compounds and a vector suitable for the genetic modification of a host cell. Further aspects of the invention are aimed at processes for biotransformation of 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) and its precursors with the use of the cell according to the invention.



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

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number

WO 2012/064195 A3

(43) International Publication Date
18 May 2012 (18.05.2012)

(51) International Patent Classification:

C12N 9/00 (2006.01) *C12N 1/15* (2006.01)
C12N 9/02 (2006.01) *C12N 1/21* (2006.01)
C12P 17/04 (2006.01) *C12N 5/10* (2006.01)

(21) International Application Number:

PCT/NL2012/050141

(22) International Filing Date:

7 March 2012 (07.03.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2006359 8 March 2011 (08.03.2011) NL

(71) Applicants (for all designated States except US): **BIRD ENGINEERING B.V.** [NL/NL]; Westfrankelandsedijk 1, NL-3115 HG Schiedam (NL). **PURAC BIOCHEM B.V.** [NL/NL]; Arkelsedijk 46, NL-4206 AC Gorinchem (NL).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WIERCKX, Nick Johannes Petrus** [NL/NL]; W.H. van Leeuwenlaan 92, NL-2613 ZG Delft (NL). **ELINK SCHUURMAN, Tom Daniël** [NL/NL]; Marmerwit 7, NL- 2718 AN Zoetermeer (NL). **KUIJPER, Sipko Maarten** [NL/NL]; Heemraadstraat 17, NL-2645 KE Delfgauw (NL). **RUIJSSENAARS, Harald Johan** [NL/NL]; De Kievit 120, NL-3972 PL Driebergen (NL).

(74) Agents: **VAN DEN HEUVEL, Henricus Theodorus** et al.; Patentwerk B.V., P.O. Box 1514, NL-5200 BN 's-Hertogenbosch (NL).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a)
- with sequence listing part of description (Rule 5.2(a))

(88) Date of publication of the international search report:

23 August 2012

(54) Title: GENETICALLY MODIFIED CELL AND PROCESS FOR USE OF SAID CELL

(57) Abstract: The present invention relates to the field of biotransformation of furanic compounds. More particular the present invention relates to novel genetically modified cells with improved characteristics for biocatalytic transformation of furanic compounds and a vector suitable for the genetic modification of a host cell. Further aspects of the invention are aimed at processes for biotransformation of 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) and its precursors with the use of the cell according to the invention.

**WO 2012/064195 A3**

GENETICALLY MODIFIED CELL AND PROCESS FOR USE OF SAID CELL**Field of the invention**

5 The present invention in general relates to the field of biotransformation of furanic compounds. Such biotransformations find utility in the production of furan-2,5-dicarboxylic acid (FDCA) and processing of lignocellulose containing material e.g. for the production of biofuels and biochemicals. More particular the present invention relates to novel genetically modified cells with improved characteristics for biocatalytic
10 transformation of furanic compounds. A further aspect of the invention relates to a vector suitable for genetic modification of a host cell to improve its characteristics for biotransformation of furanic compounds. Other aspects of the invention relate to processes for biotransformation of 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) and its precursors with the use of the cell according to the invention.

15

Background of the invention

Biotransformation of furanic compounds is receiving increasing attention. This is both in respect of the bioproduction of furan-2,5-dicarboxylic acid (FDCA), which is a promising value added chemical from biomass (Werpy et al. (2004)), and in respect of
20 their negative role in the fermentative production of biofuels and biochemicals from lignocellulose containing materials (Almeida et al. (2009)).

Recently a furanic compound utilising organism, *Cupriavidus basilensis* HMF14, has been isolated (Koopman et al. (2010a)). This organism is capable of metabolizing furfural and 5-(hydroxymethyl)furan-2-carbaldehyde (HMF). The furfural and HMF degradation
25 pathway of *Cupriavidus basilensis* HMF14 has been disclosed by Koopman et al. (2010a) together with the genes involved.

The functional introduction of the *hmfH* gene from *Cupriavidus basilensis* HMF14 in *Pseudomonas putida* S12 is disclosed by Koopman et al. (2010b). The resulting strain has good FDCA production capabilities using HMF as a substrate. However, the observed
30 accumulation of 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) would require long process times or alternative measures to remove this by-product. Sufficient removal of the HMF-acid by-product is desirable for many of the applications for which the FDCA may be produced and sometimes even is essential.

In search of a solution of the problem of HMF-acid accumulation, the inventors of the present invention have now surprisingly found that expression of certain polypeptides in the *Pseudomonas putida* S12 FDCA production system, effectively reduces HMF-acid accumulation.

5

Summary of the invention

The present findings of the inventors have resulted in the generalized concept that expression in a host of a polypeptide having an amino acid sequence of SEQ ID. NO. 1 or 2 or its analogues/homologues (such as SEQ ID NO: 3 or 4) together with one or more polypeptides capable of conversion of 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid), results in effective HMF-acid bioconversion. Improved HMF-acid bioconversion is beneficial for the elimination of HMF-acid and its furanic precursors from feedstocks wherein furanic compounds are considered to be detrimental, such as feedstocks for ethanologenic fermentations for the production of for example biofuels or for fermentations for the biological production of chemicals. In other applications, improved HMF-acid bioconversion will improve bioproduction of a chemical where HMF-acid is a starting material or an intermediate, such as in FDCA bioproduction.

Accordingly a first object of the invention is a genetically modified cell comprising a first polynucleotide sequence coding for a first polypeptide having an amino acid sequence of SEQ ID. NO: 1, 2, 3 or 4 or its analogues/homologues and a second polynucleotide sequence coding for a second polypeptide having HMF-acid converting activity. The HMF-acid converting polypeptide may be the oxidoreductase encoded by the *Cupriavidus basilensis* HMF14 *hmfH* gene previously described (Koopman et al. 2010a and Koopman et al. 2010b). According to certain embodiments it is preferred that the genetically modified cell comprises a third polynucleotide sequence coding for a third polypeptide having an amino acid sequence of SEQ ID. NO: 19, 20, 21, 22, 23, 24, 25 or its analogues/homologues. Functional expression of the third amino acid sequences results in aldehyde dehydrogenase activity capable of converting furanic aldehydes.

If the second polypeptide is an oxidoreductase, co-expression of the first polypeptide simultaneously with the oxidoreductase may also improve the quality of a whole-cell biocatalyst comprising the oxidoreductase with respect to biocatalytic FDCA production.

The cell according to the invention is genetically modified by functional introduction of at least the first polynucleotide sequence. Preferably the cell is genetically

30

modified by functional introduction of both the first and second polynucleotide sequence. Alternatively the cell is genetically modified by functional introduction of the first and the third polynucleotide sequence. Functional introduction of all three of the first, the second and the third polynucleotide is a further alternative.

5 A further aspect of the invention relates to a vector suitable for genetic modification of a host cell. The vector comprises a first polynucleotide sequence coding for a first polypeptide having an amino acid sequence of SEQ ID. NO. 1, 2, 3, or 4 or its analogues/homologues and a second polynucleotide sequence coding for a second polypeptide having 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) converting
10 activity. Optionally the vector may comprise a third polynucleotide sequence coding for a third polypeptide having an amino acid sequence of SEQ ID. NO: 19, 20, 21, 22, 23, 24, 25 or its analogues/homologues. Such a vector is suitable to obtain a genetically modified cell according to the invention.

Other aspects of the invention relate to a 5-(hydroxymethyl)furan-2-carboxylic
15 acid (HMF-acid) converting process. This process makes use of the cell according to the invention. According to preferred embodiments this process is suitable for the production of FDCA.

A further aspect of the present invention is aimed at the use of a genetically modified cell according to the invention, for the biotransformation of furanic precursors
20 to FDCA.

Brief description of the figures

Figure 1 discloses a schematic representation of oxidative reactions of furanic compounds to 2,5-furan dicarboxylic acid. The following furanic compounds are
25 presented: 1, HMF-alcohol; 2, HMF; 3, HMF-acid; 4, FFA; 5, FDCA.

Figure 2 discloses FDCA production and HMF-acid accumulation in HMF-fed cultures of *P. putida* S12_2642 (figure 2a) and B38 (figure 2b) (◆, FDCA (mM); ■, HMF-acid (mM); ○, cell dry weight (CDW; g/l); dotted line: HMF feed rate (ml 4 M solution/h))

30 Figure 3 discloses specific FDCA productivity at various time points during the fed-batch processes presented in figure 2 (Dark: *P. putida* S12_2642; Light: *P. putida* S12_B38).

Figure 4 discloses FDCA production and HMF-acid accumulation in shake-flask experiments with *P. putida* S12_B38 (figure 4a) and B51 (figure 4b) (▲, HMF (mM); ◆, FDCA (mM); ■, HMF-acid (mM); ○, cell dry weight (CDW; g/l));

Figure 5 discloses FDCA production and HMF-acid accumulation in HMF-fed high-cell density culture of *P. putida* S12_B38 starting at high cell density (□, HMF; ◆, FDCA (mM); ■, HMF-acid (mM); ○, cell dry weight (CDW; g/l); dotted line: HMF feed rate (ml 4 M solution/h)).

Figure 6 discloses the accumulation of FDCA, FFA and HMF-acid in HMF containing shake-flask cultures of *P. putida* S12_B38 (co-expressed HmfH and HmfT1; A); S12_B97 (co-expressed HmfH, HmfT1 and aldehyde dehydrogenase; B); and S12_B101 (co-expressed HmfH and aldehyde dehydrogenase; C). ▲, HMF (mM); ◆, FDCA (mM); ■, HMF-acid (mM); ●, FFA (mM); ○, cell dry weight (CDW; g/l).

Brief description of the sequences

15

SEQ ID NO: 1 sets out the amino acid sequence of HmfT1 from *Cupriavidus basilensis* HMF14. The sequence has GenBank accession number ADE20411.1.

SEQ ID NO: 2 sets out the amino acid sequence of HmfT2 from *Cupriavidus basilensis* HMF14. The sequence has GenBank accession number ADE20404.1.

20

SEQ ID NO: 3 sets out the amino acid sequence of the protein product from the gene with the locus tag *mrads2831_4728* from *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819). The sequence has GenBank accession number ACB26689.1.

SEQ ID NO: 4 sets out the amino acid sequence of the protein product from the *saci_2058* gene from *Sulfolobus acidocaldarius* DSM 639. The sequence has GenBank accession number AAY81352.1.

SEQ ID NO: 5 sets out the amino acid sequence of HmfH from *Cupriavidus basilensis* HMF14. The sequence has GenBank accession number ADE20408.1.

SEQ ID NO: 6 sets out the amino acid sequence of the protein product from the *blr0367* gene from *Bradyrhizobium japonicum* USDA 110. The sequence has GenBank accession number BAC45632.1.

30

SEQ ID NO: 7 sets out the coding sequence of *hmfT1* from *Cupriavidus basilensis* HMF14.

SEQ ID NO: 8 sets out the coding sequence of *hmfT2* from *Cupriavidus basilensis* HMF14.

5 SEQ ID NO: 9 sets out the coding sequence of the gene with the locus tag *mrad2831_4728* from *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819).

SEQ ID NO: 10 sets out the coding sequence of the *saci_2058* gene from *Sulfolobus acidocaldarius* DSM 639.

10 SEQ ID NO: 11 sets out the coding sequence of *hmfH* from *Cupriavidus basilensis* HMF14.

SEQ ID NO: 12 sets out the coding sequence of the *blr0367* gene from *Bradyrhizobium japonicum* USDA 110.

15 SEQ ID NO: 13-18 set out the sequences of various synthetic primers. Restriction locations (underlined), start and stop (reverse complement) codons (italic) and putative ribosome binding sites (lower case) are indicated. The FN23 primer was designed just upstream of the start codon of *hmfH*.

SEQ ID NO: 13 sets out the nucleotide sequence of synthetic DNA primer hmfT1 (f)

20 5'- ACGAATTCAAaggagACAACAATGGAAG-3'

SEQ ID NO: 14 sets out the nucleotide sequence of synthetic DNA primer hmfT1 (r)

5'- AAGCTAGCTGAGCAGTCACCCTCCTC-3'

25 SEQ ID NO: 15 sets out the nucleotide sequence of synthetic DNA primer FN23.

5'-CGGAATTCCACATGACAaggagACCG-3'

SEQ ID NO: 16 sets out the nucleotide sequence of synthetic DNA primer FN24.

5'-CGGAATTCGCTTCGGTCTTCAACTCGGATG-3'

30 SEQ ID NO: 17 sets out the nucleotide sequence of synthetic DNA primer mrad (f).

5'- ACGAATTCggaggAAATCTATGCAGACC -3'

SEQ ID NO: 18 sets out the nucleotide sequence of synthetic DNA primer mrad (r).

5'- AAGCTAGCGCAGAACCGTATCGTCAG -3'

5 SEQ ID NO: 19 sets out the amino acid sequence of the aldehyde dehydrogenase Adh from *Cupriavidus basilensis* HMF14.

SEQ ID NO: 20 sets out the amino acid sequence having Genbank accession number: YP_003609156.1.

SEQ ID NO: 21 sets out the amino acid sequence having Genbank accession number: ZP_02881557.1.

10 SEQ ID NO: 22 sets out the amino acid sequence having Genbank accession number: YP_003451184.1.

SEQ ID NO: 23 sets out the amino acid sequence having Genbank accession number: ACA09737.1.

15 SEQ ID NO: 24 sets out the amino acid sequence having Genbank accession number: YP_530742.1.

SEQ ID NO: 25 sets out the amino acid sequence having Genbank accession number: YP_001541929.1.

SEQ ID NO: 26 sets out the polynucleotide sequence of *adh* encoding the aldehyde dehydrogenase Adh from *Cupriavidus basilensis* HMF14.

20 SEQ ID NO: 27 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: YP_003609156.1.

SEQ ID NO: 28 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: ZP_02881557.1.

25 SEQ ID NO: 29 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: YP_003451184.1.

SEQ ID NO: 30 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: ACA09737.1.

SEQ ID NO: 31 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: YP_530742.1.

30 SEQ ID NO: 32 sets out the polynucleotide sequence encoding the amino acid sequence having Genbank accession number: YP_001541929.1.

SEQ ID NO: 33 sets out the nucleotide sequence of synthetic DNA primer FN13: 5'-ATGCGGCCGCAACAaggagAAGATGGAATGAACG-3' (underlined

sequence: *NotI* restriction site; start codon (ATG) of aldehyde dehydrogenase encoding gene in italics; putative ribosome binding site (RBS) in lower case)

SEQ ID NO: 34 sets out the nucleotide sequence of synthetic DNA primer FN14: 5'-ATGCGGCCGCGCGTCGGGGTCGGTGCTA-3' (underlined sequence: *NotI* restriction site; stop codon (reverse complement strand) in italics).

Detailed description of the invention

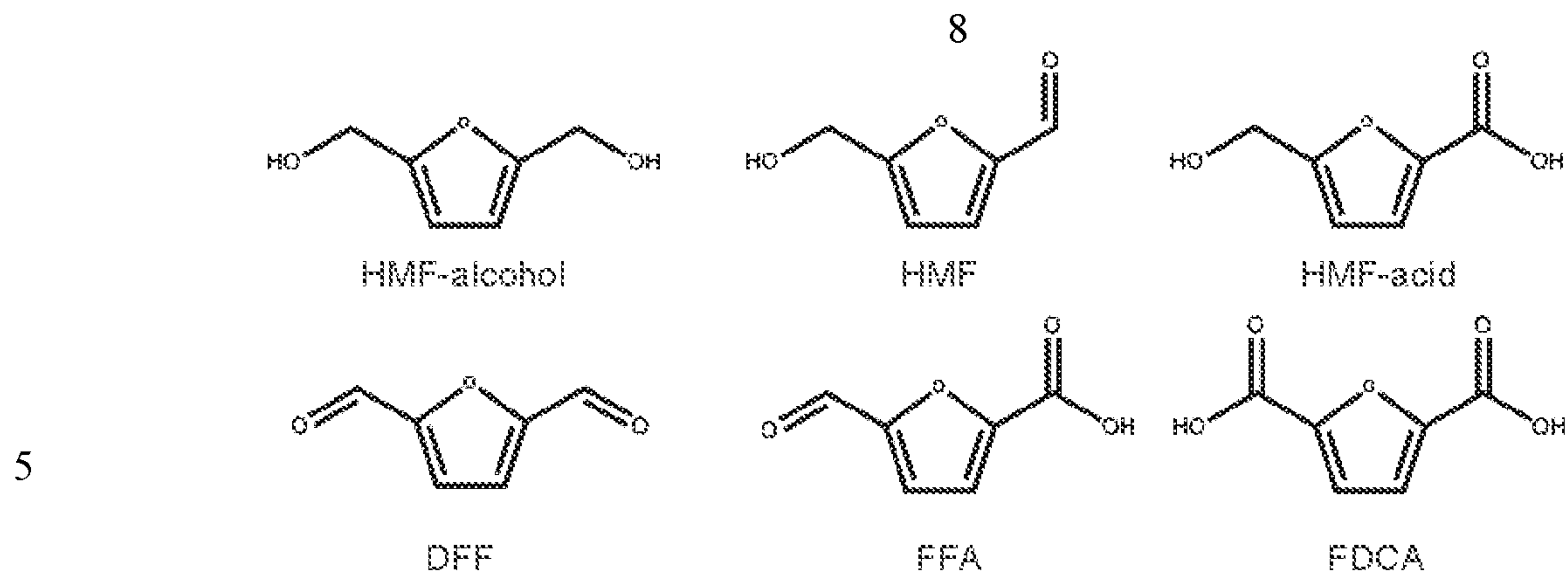
General definitions

10 Throughout the present specification and the accompanying claims, the words "comprise" and "include" and variations such as "comprises", "comprising", "includes" and "including" are to be interpreted inclusively. That is, these words are intended to convey the possible inclusion of other elements or integers not specifically recited, where the context allows.

15 The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to one or at least one) of the grammatical object of the article. By way of example, "an element" may mean one element or more than one element.

The term "a number of" should be understood to have the meaning of one or more.

20 Furanic compounds are herein understood to be any compound having a furan ring. Preferably the furanic compound is a compound that may be oxidized to 2,5-furandicarboxylic acid. Furanic compounds relevant within the context of this invention include [5-(hydroxymethyl)furan-2-yl]methanol ("HMF-alcohol"), 5-(hydroxymethyl)furan-2-carbaldehyde ("HMF"), 5-(hydroxymethyl)furan-2-carboxylic acid ("HMF-acid"), 5-formylfuran-2-carboxylic acid ("FFA"), furan-2,5-dicarbaldehyde (DFF) and furan-2,5-dicarboxylic acid ("FDCA"). The furan ring or any of its substitutable side groups may be substituted, *e.g.*, with OH, C1-C10 alkyl, alkyl, allyl, aryl or RO- ether moiety, including cyclic groups, on any available position in the furan ring. The chemical structures of a number of relevant furanic compounds are presented
30 below.



The term “polynucleotide” includes poly deoxyribonucleic acids (DNA) and poly ribonucleic acids (RNA) and the term may refer to either DNA or RNA. The skilled person will be aware of the differences in stability of DNA and RNA molecules. Thus the skilled person will be able to understand from the context of the use of the term “polynucleotide” which of the forms of polynucleotide (DNA and/or RNA) is suitable.

The term sequence “similarity” as used herein refers to the extent to which individual polynucleotide or protein sequences are alike. The extent of similarity between two sequences is based on the extent of identity combined with the extent of conservative changes. The percentage of “sequence similarity” is the percentage of amino acids or nucleotides which is either identical or conservatively changed viz. “sequence similarity” = (% sequence identity) + (% conservative changes).

For the purpose of this invention “conservative changes” and “identity” are considered to be species of the broader term “similarity”. Thus whenever, the term sequence “similarity” is used it embraces sequence “identity” and “conservative changes”.

The term “sequence identity” is known to the skilled person. In order to determine the degree of sequence identity shared by two amino acid sequences or by two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). Such alignment may be carried out over the full lengths of the sequences being compared. Alternatively, the alignment may be carried out over a shorter comparison length, for example over about 20, about 50, about 100 or more nucleic acids/bases or amino acids.

The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The degree of identity shared

between sequences is typically expressed in terms of percentage identity between the two sequences and is a function of the number of identical positions shared by identical residues in the sequences (i.e., % identity = number of identical residues at corresponding positions/total number of positions x 100). Preferably, the two sequences being
5 compared are of the same or substantially the same length.

The percentage of “conservative changes” may be determined similar to the percentage of sequence identity. However, in this case changes at a specific location of an amino acid or nucleotide sequence that are likely to preserve the functional properties of the original residue are scored as if no change occurred.

10 For amino acid sequences the relevant functional properties are the physico-chemical properties of the amino acids. A conservative substitution for an amino acid in a polypeptide of the invention may be selected from other members of the class to which the amino acid belongs. For example, it is well-known in the art of protein biochemistry that an amino acid belonging to a grouping of amino acids having a particular size or
15 characteristic (such as charge, hydrophobicity and hydrophilicity) can be substituted for another amino acid without altering the activity of a protein, particularly in regions of the protein that are not directly associated with biological activity. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and tyrosine. Polar neutral amino acids include glycine, serine,
20 threonine, cysteine, tyrosine, asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Conservative substitutions include, for example, Lys for Arg and vice versa to maintain a positive charge; Glu for Asp and vice versa to maintain a negative charge; Ser for Thr so that a free -OH is maintained; and
25 Gln for Asn to maintain a free -NH₂.

For nucleotide sequences the relevant functional properties is mainly the biological information that a certain nucleotide carries within the open reading frame of the sequence in relation to the transcription and/or translation machinery. It is common knowledge that the genetic code has degeneracy (or redundancy) and that multiple
30 codons may carry the same information in respect of the amino acid for which they code. For example in certain species the amino acid leucine is coded by UUA, UUG, CUU, CUC, CUA, CUG codons (or TTA, TTG, CTT, CTC, CTA, CTG for DNA), and the amino acid serine is specified by UCA, UCG, UCC, UCU, AGU, AGC (or TCA, TCG,

TCC, TCT, AGT, AGC for DNA). Nucleotide changes that do not alter the translated information are considered conservative changes.

The skilled person will be aware of the fact that several different computer programs, using different mathematical algorithms, are available to determine the identity
5 between two sequences. For instance, use can be made of a computer program employing the Needleman and Wunsch algorithm (Needleman et al. (1970)). According to an embodiment the computer program is the GAP program in the Accelrys GCG software package (Accelrys Inc., San Diego U.S.A). Substitution matrices that may be used are for example a BLOSUM 62 matrix or a PAM250 matrix, with a gap weight of
10 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

According to an embodiment the percent identity between two nucleotide
15 sequences is determined using the GAP program in the Accelrys GCG software package (Accelrys Inc., San Diego U.S.A) A NWSgapdna CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6 is used.

In another embodiment, the percent identity of two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (Meyers et al.
20 (1989)) which has been incorporated into the ALIGN program (version 2.0) (available at the ALIGN Query using sequence data of the Genestream server IGH Montpellier France <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

For the present invention it is most preferred to use BLAST (Basic Local
25 Alignment Tool) to determine the percentage identity and/or similarity between nucleotide or amino acid sequences.

Queries using the BLASTn, BLASTp, BLASTx, tBLASTn and tBLASTx programs of Altschul *et al.* (1990) may be posted via the online versions of BLAST accessible via <http://www.ncbi.nlm.nih.gov/>. Alternatively a standalone version of
30 BLAST (*e.g.*, version 2.2.24 (released 23 August 2010)) downloadable also via the NCBI internet site may be used. Preferably BLAST queries are performed with the following parameters. To determine the percentage identity and/or similarity between amino acid sequences: algorithm: blastp; word size: 3; scoring matrix: BLOSUM62; gap

costs: Existence: 11, Extension: 1; compositional adjustments: conditional compositional score matrix adjustment; filter: off; mask: off. To determine the percentage identity and/or similarity between nucleotide sequences: algorithm: blastn; word size: 11; max matches in query range: 0; match/mismatch scores: 2, -3; gap costs: Existence: 5, 5 Extension: 2; filter: low complexity regions; mask: mask for lookup table only.

The percentage of “conservative changes” may be determined similar to the percentage of sequence identity with the aid of the indicated algorithms and computer programmes. Some computer programmes, e.g., BLASTp, present the number/percentage of positives (= similarity) and the number/percentage of identity. The percentage of conservative changes may be derived therefrom by subtracting the percentage of identity from the percentage of positives/similarity (percentage conservative changes = percentage similarity – percentage identity).

General molecular biological techniques such as hybridization experiments, PCR experiments, restriction enzyme digestions, transformation of hosts etcetera may be performed according to the standard practice known to the skilled person as disclosed in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

20 **First polynucleotide and polypeptide**

The genetically modified cell according to the invention comprises a first polynucleotide coding for a first polypeptide. The first polypeptide comprises an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequence of SEQ ID NO: 1, 2, 3 or 4.

Alternatively the sequence similarity may be expressed as at least, 45%, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % or at least 99 % similarity. According to an embodiment the indicated percentages similarity may be percentages identity. In a particular embodiment the first polypeptide may comprise the amino acid sequence as set out in any of SEQ ID NO: 1, 2, 3 or 4.

Such a polypeptide is capable of providing an improvement with respect to HMF-acid bioconversion.

Without wishing to be bound by any theory it is believed that such a polypeptide has HMF-acid transport capabilities. By transportation of HMF-acid into the cell by the first polypeptide it becomes better available for intracellular conversion. Thus HMF-acid bioconversion may be improved.

5 Such an improvement on HMF-acid bioconversion has been shown in the examples for HmfT1 (SEQ ID NO: 1) and the protein product from the gene with the locus tag *mrاد2831_4728* from *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819) (SEQ ID NO: 3). On the basis of the level of sequence similarity/identity with these sequences it is justified to expect that HmfT2 (SEQ ID NO: 2) and the protein product
10 from the *Saci_2058* gene from *Sulfolobus acidocaldarius* DSM 639 will have similar effects.

HmfT2 (SEQ ID NO: 2) has over 90 % similarity with HmfT1 (SEQ ID NO: 1). The level of identity is 87%.

A similar functionality for the protein product of the *saci_2058* gene from
15 *Sulfolobus acidocaldarius* DSM 639 (SEQ ID NO: 4) may be expected based on a CLUSTALW2 multiple sequence alignment with transporters from different functional families. The *Sulfolobus* transporter forms a cluster with HmfT1 (SEQ ID NO: 1) and Mrad2831_4728 (SEQ ID NO: 3). Moreover, analysis of the *S. acidocaldarius* genome has shown that the transporter gene *saci_2058* is flanked by genes that encode similar
20 functionalities as the *hmf* gene clusters, respectively, the type of activities that are expected in degradation of HMF(-like compounds). Examples: *Saci_2057*, alcohol dehydrogenase; *Saci_2059/2060*, aromatic ring dioxygenase; *Saci_2062*, acyl-CoA synthetase (functionally comparable to *hmfD*); *Saci_2063*, enoyl-CoA hydratase (functionally comparable with *hmfE*); *Saci_2064*, aldehyde oxidase / xanthine
25 dehydrogenase (functionally comparable to *hmfABC*). On the basis of this analysis it is justified to expect that the protein product of the *Saci_2058* gene from *Sulfolobus acidocaldarius* DSM 639 (SEQ ID NO: 4) will have similar effects as HmfT1 (SEQ ID NO: 1) and/or the protein product from the gene with the locus tag *mrاد2831_4728* from *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819) (SEQ
30 ID NO: 3).

The first polypeptide may be encoded by a first polynucleotide sequence having at least 45 %, preferably at least 60 %, such as at least 70 %, more preferably at least 80 %, such as 90 %, most preferably at least 95 % sequence similarity with a polynucleotide

sequence set out in SEQ ID NO: 7, 8, 9 or 10. Suitable alternative levels of similarity of the first polynucleotide with a sequence set out in SEQ ID NO: 7, 8, 9 or 10 may be at least 66%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% similarity. According to an embodiment the indicated percentages
5 similarity may be percentages identity. In a particular embodiment the first polypeptide may be coded by a polynucleotide sequence as set out in SEQ ID NO 7, 8, 9 or 10.

The first polypeptide or the polynucleotide coding for the first polypeptide may be isolated from an organism, preferably a microorganism that expresses the first polypeptide under certain growth conditions. The microorganism may be capable of
10 using HMF-acid or related furanic substances, such as HMF or HMF-alcohol, as carbon source, but this is not necessary.

In a typical approach, gene libraries can be screened to isolate alternative polynucleotides which are suitable. The libraries may be constructed from microorganisms from the superkingdom of Bacteria. These microorganisms may belong
15 to the phylum of Proteobacteria, more specifically to the class of Alphaproteobacteria or Betaproteobacteria. The Alphaproteobacteria may belong to the order of Rhizobiales, to the families of Bradyrhizobiaceae or Methylobacteriaceae. The Bradyrhizobiaceae may belong to the genus of *Bradyrhizobium*, e.g., *Bradyrhizobium japonicum*, or to the genus of *Afipia*. The Methylobacteriaceae may belong to the genus of
20 *Methylobacterium*, e.g., *Methylobacterium nodulans* or *Methylobacterium radiotolerans*. The Betaproteobacteria may belong to the order of Burkholderiales, more specifically the family of Burkholderiaceae. They may belong to the genus *Cupriavidus*, e.g., *Cupriavidus basilensis*; or to the genus *Ralstonia*, e.g., *Ralstonia eutropha*; or to the genus *Burkholderia*, e.g., *Burkholderia phymatum*, *Burkholderia phytofirmans*,
25 *Burkholderia xenovorans*, or *Burkholderia graminis*. The bacteria may also belong to the phylum of Firmicutes, more specifically the class of Bacilli, more specifically the order of Bacillales. The Bacillales may belong to the family of Bacillaceae, more specifically to the genus *Geobacillus*, e.g., *Geobacillus kaustophilus*. Alternatively, the microorganisms may belong to the superkingdom of Archaea, more specifically the
30 phylum of Euryarchaeota, or the phylum of Crenarchaeota. The Euryarchaeota may belong to an unclassified genus, e.g., *Cand. Parvarchaeum acidiphilum*, or to the class of Thermoplasmata, more specifically the order of Thermoplasmatales. The Thermoplasmatales may belong to the family of Thermoplasmataceae, more specifically

the genus *Thermoplasma*, e.g., *Thermoplasma acidophilum* or *Thermoplasma volcanium*. The Crenarchaeota may belong to the class of Thermoprotei, more specifically the order of Sulfolobales. The Sulfolobales may belong to the family of Sulfolobaceae, more specifically the genus *Sulfolobus*, e.g., *Sulfolobus acidocaldarius*,
5 *Sulfolobus islandicus*, *Sulfolobus solfataricus*, or *Sulfolobus tokodaii*; or to the genus of *Metallosphaera*, e.g., *Metallosphaera sedula*. The Thermoprotei may also belong to the order of Thermoproteales, family of Thermoproteaceae. The Thermoproteaceae may belong to the genus *Vulcanisaeta*, e.g., *Vulcanisaeta distributa*, or to the genus *Caldivirga*, e.g., *Caldivirga maquilingensis*.

10 Preferably the first polypeptide and/or the first polynucleotide coding for the first polypeptide is isolated from *Cupriavidus basilensis* HMF14 (Wierckx et al. 2010). According to an alternative embodiment the first polypeptide and/or the first polynucleotide coding for the first polypeptide may be isolated from *Methylobacterium radiotolerans*.

15 Based on the amino acid sequences provided in SEQ. ID. NO: 1, 2, 3 or 4 and/or the nucleotide sequences provided in SEQ. ID. NO: 7, 8, 9 or 10, the skilled person will be able to construct suitable probes and/or primers to isolate a nucleotide sequence coding for the first polypeptide.

Alternatively, based on the amino acid sequences provided in SEQ. ID. NO: 1, 2,
20 3 or 4 and/or the nucleotide sequences provided in SEQ. ID. NO: 7, 8, 9 or 10, the skilled person may obtain synthesized sequences coding for the first polypeptide from commercial sources, as gene synthesis is becoming increasingly available. Synthetic sequences may be purchased for example from Genart A.G. (Regensburg, Germany) or from Genscript USA Inc. (Piscataway, NJ, USA) to name but a few.

25 The cell according to the invention is genetically modified by functional introduction of the first polynucleotide. With functional introduction of a polynucleotide is meant, an introduction of said polynucleotide in a cell, such that said cell acquires the possibility to express a functional polypeptide product of the polynucleotide. Methods and techniques for functional introduction of polynucleotides in host cells are within the
30 general knowledge of the skilled person.

HMF-acid converting polypeptide

The genetically modified cell according to the invention comprises a second

polynucleotide coding for a second polypeptide. The second polypeptide has HMF-acid converting activity and may be selected from any polypeptide capable of converting HMF-acid to a product. The inventors have observed that bioconversion of HMF-acid by the HMF-acid converting second polypeptide, is effectively improved by expression of the first polypeptide in a cell.

The second polynucleotide coding for the second polypeptide may be a natural component of the cell according to the invention viz. the cell need not be genetically modified in respect of the second polynucleotide. However, according to certain embodiments of the invention the cell according to the invention is genetically modified in respect of the second polynucleotide, by functional introduction of the second polynucleotide. The term “functional introduction” has already been explained above in connection to the first polypeptide and first polynucleotide.

According to a preferred embodiment of the invention, the second polypeptide is an HMF-acid converting oxidoreductase. This HMF-acid converting oxidoreductase may comprise an amino acid sequence set out in SEQ ID NO: 5 or 6 or a variant polypeptide thereof having at least 45 %, preferably at least 60 %, such as at least 70 %, more preferably at least 80 %, such as 90 %, most preferably at least 95 % sequence similarity with the amino acid sequence set out in SEQ ID NO 5 or 6.

Alternatively the sequence similarity may be at least, 45%, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % or at least 99 % similarity. According to an embodiment the indicated percentages similarity may be percentages identity. In a particular embodiment the second polypeptide may comprise the amino acid sequence as set out in any of SEQ ID NO: 5 or 6.

The HMF-acid converting oxidoreductase of SEQ ID NO: 5 was previously disclosed by Koopman et al. (2010a) and Koopman et al. (2010b) and designated HmfH. The HMF-acid converting oxidoreductase of SEQ ID NO: 6 may be isolated from *Bradyrhizobium japonicum* USDA 110 and corresponds to the translated protein product of the *blr0367* gene. *B. japonicum* USDA 110 contains homologues for all HMF/furfural utilization genes from *C. basilensis* HMF14 in its genome (Koopman et al. 2010a). The translated product of the *blr0367* gene is the protein of *B. japonicum* USDA 110 that showed highest homology to HmfH from *C. basilensis* HMF14. In view of the fact that *B. japonicum* USDA 110 was shown to utilize HMF as the sole carbon

source (Koopman et al. 2010a) it must harbour a functional HmfH homologue. It is therefore justified to expect that HmfH similar activity arises from *blr0367*.

Although the invention is exemplified with reference to a number of HMF-acid converting oxidoreductases it should be noted that within the invention it is expressly permitted that the second polypeptide is a different HMF-acid converting oxidoreductase or yet a different HMF-acid converting polypeptide not having oxidoreductase activity.

The second polypeptide may be encoded by a second polynucleotide sequence having at least 45 %, preferably at least 60 %, such as at least 70 %, more preferably at least 80 %, such as 90 %, most preferably at least 95 % sequence similarity with a polynucleotide sequence set out in SEQ ID NO: 11 or 12. Suitable alternative levels of similarity of the first polynucleotide with a sequence set out in SEQ ID NO: 11 or 12 may be at least 66%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% similarity. In an embodiment the indicated percentages similarity may be percentages identity. In a particular embodiment the second polypeptide may be coded by a polynucleotide sequence as set out in SEQ ID NO 11 or 12.

Isolation of a suitable polynucleotide coding for the second polypeptide having oxidoreductase activity from *Cupriavidus basilensis* has been disclosed in Koopman et al. (2010a). This gene is designated *hmfH*.

In an approach for isolation of the second polynucleotide, gene libraries can be screened to isolate polynucleotides which are suitable. The libraries may be constructed from microorganisms from the superkingdom of Bacteria. These microorganisms may belong to the phylum of Proteobacteria, more specifically to the class of Alphaproteobacteria, Betaproteobacteria, or Gammaproteobacteria. The Alphaproteobacteria may belong to the order of Rhizobiales, or the order of Sphingomonadales. The Rhizobiales may belong to the family of Methylobacteriaceae, e.g., an organism from the genus *Methylobacterium* such as *M. nodulans* or *M. radiotolerans*, or an organism from the family of Rhizobiaceae. The Rhizobiaceae may belong to the Rhizobium/Agrobacterium group, more specifically to the genus *Rhizobium*, such as *R. leguminosarum* or *R. leguminosarum* bv. *trifolii*. They may also belong to the genus *Agrobacterium*, such as *A. radiobacter*. The Rhizobiaceae may also belong to the family of Bradyrhizobiaceae, more specifically the genus of Bradyrhizobium, such as *B. japonicum*. Sphingomonadales may belong to the family of

Sphingomonadaceae, more specifically to the genus of *Sphingomonas*, such as *S. wittichii* or *S. chlorophenolicum*. The Betaproteobacteria may belong to the order of the Methylophilales, or the order of Burkholderiales. The Methylophilales may belong to the family of Methylophilaceae, *e.g.*, an organism from the genus *Methylovorus*. The

5 Burkholderiales may belong to the family of Burkholderiaceae, *e.g.*, an organism from the genus *Cupriavidus*, such as *Cupriavidus basilensis*. They may also belong to the genus *Burkholderia*, such as *Burkholderia phytofirmans*. *B. phymatum*, *B. graminis*, *B. xenovorans*, or *B. cenocepacia*, or to the family of Oxalobacteraceae, genus of *Janthinobacterium*. The Gammaproteobacteria may belong to the order of

10 Enterobacteriales, family of Enterobacteriaceae, genus of *Yersinia* such as *Yersinia ruckeri*. The microorganisms may furthermore be bacteria of the phylum of Actinobacteria, class of Actinobacteria, subclass of Actinobacteridae, order of Actinomycetales. The Actinomycetales may belong to the suborder of Streptomycineae, Pseudonocardineae, or Micromonosporineae. The Streptomycineae may belong to the

15 family of Streptomycetaceae, more specifically the genus *Streptomyces*, such as *S. violaceusniger*, *S. hygroscopicus*, or *S. clavuligerus*. The Pseudonocardineae may belong to the family of Pseudonocardiaceae, more specifically the genus of *Saccharopolyspora* such as *S. erythraea*; or to the family of Actinosynnemataceae, more specifically to the genus of *Saccharothrix* such as *S. mutabilis*, or *S. mutabilis* subsp.

20 *capreolus*, or to the genus *Actinosynnema*, such as *A. mirum*. The Micromonosporineae may belong to the family of Micromonosporaceae, more specifically the genus of *Micromonospora*.

Based on the amino acid sequences provided in SEQ. ID. NO: 5 or 6 and/or the nucleotide sequences provided in SEQ. ID. NO: 11 or 12, the skilled person will be able

25 to construct suitable probes and/or primers to isolate a nucleotide sequence coding for the second polypeptide.

Alternatively, based on the amino acid sequences provided in SEQ. ID. NO: 5 or 6 and the nucleotide sequences provided in SEQ. ID. NO: 11 or 12, the skilled person may obtain synthesized sequences coding for the second polypeptide from commercial

30 sources as already indicated in the section discussing the first polypeptide.

Third polynucleotide and polypeptide

According to an alternative embodiment the genetically modified cell according to the invention comprises a third polynucleotide coding for a third polypeptide. The third polypeptide comprises an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequence of SEQ ID NO: 19, 20, 21, 22, 23, 24 or 25.

Alternatively the sequence similarity may be expressed as at least, 45%, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % or at least 99 % similarity. According to an embodiment the indicated percentages similarity may be percentages identity. In a particular embodiment the first polypeptide may comprise the amino acid sequence as set out in any of SEQ ID NO: 19, 20, 21, 22, 23, 24 or 25. The amino acid sequence of SEQ ID NO: 19 is a preferred selection of the third polypeptide. This amino acid sequence was recently published in WO2011/026906 (SEQ ID NO:15).

Functional expression of such a third polypeptide results in aldehyde dehydrogenase activity (Adh) capable of converting furanic aldehydes and provides a further improvement with respect to HMF-acid bioconversion and/or FDCA production.

The effects associated with the expression of the third polypeptide have been shown in the examples for the amino acid sequence of SEQ ID NO: 19. On the basis of the level of sequence similarity/identity it is justified to expect that the polypeptides of SEQ ID 20-25 and their analogues/homologues will have similar effects.

The third polypeptide may be encoded by a third polynucleotide sequence having at least 45 %, preferably at least 60 %, such as at least 70 %, more preferably at least 80 %, such as 90 %, most preferably at least 95 % sequence similarity with a polynucleotide sequence set out in SEQ ID NO: 26, 27, 28, 29, 30, 31 or 32. Preferably the third polypeptide is encoded by a polynucleotide sequence set out in SEQ ID NO 26 or a homologue having the indicated sequence similarity with the polynucleotide sequence set out in SEQ ID NO: 26. The polynucleotide sequence of SEQ ID NO: 26 was recently published in WO2011/026906 (SEQ ID NO:16). Suitable alternative levels of similarity of the third polynucleotide with a sequence set out in SEQ ID NO: 26, 27, 28, 29, 30, 31 or 32 may be at least 66%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% similarity. According to an embodiment the indicated percentages similarity may be percentages identity. In a particular embodiment the third

polypeptide may be coded by a polynucleotide sequence as set out in SEQ ID NO 26, 27, 28, 29, 30, 31 or 32.

Isolation and further manipulation of the third polypeptide and the corresponding third polynucleotide may be performed in general as is discussed above and hereafter for
5 the first polypeptide and the corresponding first polypeptide.

Vectors

Another aspect of the invention pertains to vectors, including cloning and expression vectors, comprising the first and second polynucleotide or a functional
10 equivalent thereof and methods of growing, transforming or transfecting such vectors in a suitable host cell, for example under conditions in which expression of a polypeptide of the invention occurs. As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked.

The first and second polynucleotide and optionally the third polynucleotide, can
15 be incorporated into a recombinant replicable vector, for example a cloning or expression vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the
20 host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below.

The vector into which the expression cassette or polynucleotide of the invention is inserted may be any vector which may conveniently be subjected to recombinant DNA procedures, and the choice of the vector will often depend on the host cell into which it
25 is to be introduced.

A vector according to the invention may be an autonomously replicating vector, i.e. a vector which exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and
30 replicated together with the chromosome (s) into which it has been integrated.

One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome.

Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, 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”. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. The terms “plasmid” and “vector” can be used interchangeably herein as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as cosmid, viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), phage vectors and transposons and plasposons, which serve equivalent functions.

The skilled person will be able to construct the vectors according to the invention based on the amino acid and polynucleotide sequences provided, his knowledge of the art and commercially available means.

According to a preferred embodiment the first and second polynucleotide sequence are located on a single vector. This vector optionally further comprising the third polynucleotide sequence. As the skilled person will understand, the use of a vector comprising the first and second polynucleotide sequence (and optionally further comprising the third polynucleotide sequence) greatly simplifies construction of genetically modified cells functionally expressing the first and second polypeptide (optionally together with the third polypeptide). However as the skilled person will also understand, genetically modified cells functionally expressing the first and second polypeptide may be obtained via various other transformation schemes involving alternative vectors. In this respect it should be noted that according to certain embodiments, functional introduction of the second polynucleotide sequence is not a requirement. Also according to certain further alternative embodiments, functional introduction of the third polynucleotide sequence is not a requirement.

30

Host cell

The genetically modified cell according to the invention may be constructed from any suitable host cell. The host cell may be an unmodified cell or may already be

genetically modified. The cell may be a prokaryote cell, a eukaryote cell, a plant cell or an animal cell. In such a cell one or more genes may be deleted, knocked-out or disrupted in full or in part, wherein optionally one or more genes encode for protease. According to an embodiment, the host cell according to the invention is a eukaryotic host cell. Preferably, the eukaryotic cell is a mammalian, insect, plant, fungal, or algal cell. Preferred mammalian cells include, *e.g.*, Chinese hamster ovary (CHO) cells, COS cells, 293 cells, PerC6 cells, and hybridomas. Preferred insect cells include *e.g.* Sf9 and Sf21 cells and derivatives thereof. More preferably, the eukaryotic cell is a fungal cell, *i.e.*, a yeast cell, such as *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* strain. More preferably, the eukaryotic cell is *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Pichia stipitis* and *Pichia pastoris*, or a filamentous fungal cell. In certain embodiments, the eukaryotic cell is a filamentous fungal cell.

"Filamentous fungi" include all filamentous forms of the subdivision *Eumycota* and *Oomycota* (as defined by Hawksworth *et al.*, (1995)). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. Filamentous fungal strains include, but are not limited to, strains of *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Phanerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, and *Trichoderma*.

Preferred filamentous fungal cells belong to a species of an *Aspergillus*, *Chrysosporium*, *Penicillium*, *Talaromyces* or *Trichoderma* genus, and most preferably a species selected from *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Trichoderma reesei* or *Penicillium chrysogenum*..

According to another embodiment, the host cell according to the invention is a prokaryotic cell. Preferably, the prokaryotic host cell is bacterial cell. The term "bacterial cell" includes both Gram-negative and Gram-positive microorganisms. Suitable bacteria may be selected from, *e.g.*, the genera *Escherichia*, *Anabaena*, *Caulobacter*, *Gluconobacter*, *Rhodobacter*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Brevibacterium*,

Corynebacterium, *Rhizobium* (*Sinorhizobium*), *Bradyrhizobium*, *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Staphylococcus*, *Streptomyces*, *Zymomonas*, *Acetobacter*, *Streptococcus*, *Bacteroides*, *Selenomonas*, *Megasphaera*, *Burkholderia*, *Cupriavidus*, *Ralstonia*, *Methylobacterium*,
5 *Methylovorus*, *Rhodopseudomonas*, *Acidiphilium*, *Dinoroseobacter*, *Agrobacterium*, *Sulfolobus* or *Sphingomonas*. Preferably, the bacterial cell is selected from the group consisting of *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus pumilus*, *Bacillus megaterium*, *Bacillus halodurans*, *Bacillus pumilus*, *Gluconobacter oxydans*, *Caulobacter crescentus*, *Methylobacterium extorquens*,
10 *Methylobacterium radiotolerans*, *Methylobacterium nodulans*, *Rhodobacter sphaeroides*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Pseudomonas putida* S12, *Paracoccus denitrificans*, *Escherichia coli*, *Corynebacterium glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *Rhizobium radiobacter*, *Rhizobium leguminosarum*,
15 *Rhizobium leguminosarum* bv. *trifolii*, *Agrobacterium radiobacter*, *Cupriavidus basilensis*, *Cupriavidus necator* (*Ralstonia eutropha*), *Ralstonia pickettii*, *Burkholderia phytofirmans*, *Burkholderia phymatum*, *Burkholderia xenovorans*, *Burkholderia graminis*, *Rhodopseudomonas palustris*, *Acidiphilium cryptum*, *Dinoroseobacter shibae*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*.
20

A highly preferred host cell is *Pseudomonas putida* S12. In this strain functional expression of the *hmfH* gene from *Cupriavidus basilensis* HMF14 has proven effective for introducing HMF oxidative capacity, resulting in FDCA production from this substrate.

25 For specific uses of the cell according to the invention, the selection of the host cell may be made according to such use. Particularly preferred are those hosts that are suitable for conversion of lignocellulosic feed stocks and those which are resistant to the conditions preferred for the production of furanic compounds, such as FDCA. The skilled person will have to his availability suitable means and methods to functionally
30 introduce the first and optionally the second poly nucleotide into any of the mentioned host cells.

HMF-acid biotransformation

The genetically engineered cell according to the invention has an improved HMF-acid biotransformation. Improved HMF-acid bioconversion is beneficial for the elimination of HMF-acid and its furanic precursors from feedstocks wherein furanic compounds are considered to be detrimental, such as feedstocks for ethanologenic fermentations for the production of biofuels and biochemicals. In other applications improved HMF-acid bioconversion will improve bioproduction of chemical where HMF-acid is a starting material or an intermediate, such as in FDCA bioproduction.

If the HMF acid-converting enzyme is a HMF-acid converting oxidoreductase, the cell according to the invention will be capable of performing a biological oxidation reaction. The oxidation reaction is herein one or more reactions of an oxidant with HMF-acid in the presence of the oxidoreductase.

A preferred oxidation reaction is the production of FDCA, wherein HMF-acid is converted to FDCA, by reaction with an oxidant in the presence of a HMF-acid converting oxidoreductase. Bioconversions of furanic compounds to FDCA, wherein HMF-acid is an intermediate, have been disclosed in the prior art, for example using HMF as a starting material (see Koopman et al. 2010a and Koopman et al. 2010b). Such bioconversions will be improved if they are performed by a cell according to the invention.

HMF-acid may be generated *in situ* from one or more furanic precursors by the cell of the invention or any other cell present. With *in situ* generation is meant that the HMF-acid is not added from outside the system. Instead HMF-acid is generated within the system via one or more bioconversions that convert furanic precursors to HMF-acid.

The furanic precursor of HMF-acid may be chosen from the group consisting of 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), furan-2,5-dicarbaldehyde (DFF) and [5-(hydroxymethyl)furan-2-yl]methanol (HMF alcohol) and preferably the furanic precursor is HMF.

HMF may be obtained from one or more hexose sugars by acid-catalyzed dehydration, as is known in the art. The hexose sugars may be obtained from biomass, preferably lignocellulosic biomass.

The oxidation reaction may comprise a number of consecutive oxidation reaction steps resulting in a product e.g. the oxidation of HMF-acid to FFA and further the oxidation of FFA to FDCA. Examples of oxidation reactions are given in Fig. 1.

The oxidation reactions are preferably conducted at relatively mild temperature, i.e. 10-80 °C, more preferably 20-45 °C, most preferably around from 25-40 °C. It is preferred to conduct the reaction at a pH where FDCA is either in a neutral form or in a fully dissociated form, such that salt formation may be controlled. In view of the presence of two acid moieties in FDCA there are two separate preferred pH ranges. The pH during the reaction may be from pH 1 to 6, preferably from pH 1 to 4, most preferably from pH 1 to 3. Alternatively the pH during the reaction may be from pH 5 to 9, preferably from pH 5 to 8, most preferably from pH 5 to 7. The skilled person will understand that the requirements of the host cell will also influence the selection of a suitable pH value for the process. Selection of pH values that are suitable for the various host cells that are suitable within the present invention is within the ambit of the skilled person and may be derived from standard text books. For *Pseudomonas putida*, including *Pseudomonas putida* S12, the preferred pH range is from pH 5 to 7.

The reaction time may be 6 – 150 h, with the addition of oxygen from an oxygen source, such as molecular oxygen, or water, or a different source of oxygen depending on the requirements of the furanic oxidizing enzyme. Air may be used conveniently as a source of molecular oxygen..

The reactor may be any suitable (aerated) bioreactor. It may be operated in batch, continuously or fed-batch operation.

After biotransformation, the cells may be separated from the broth by established methods and re-used. Oxidation products such as FDCA, HMF-acid, etc. may be recovered from the reaction mixture by (acid) precipitation and subsequent cooling crystallisation, and separation of the crystallized oxidation product, e.g., crystallized FDCA. However, other recovery methods are suitable, such as but not limited to acid precipitation and solvent extraction, as known in the art.

For many applications, such as removal of HMF-acid from lignocellulosic feedstocks, the exact way of HMF-conversion is irrelevant. What is important is that the HMF-acid is converted effectively in order to remove it as such, or to prevent its accumulation if it is formed from furanic precursors. For such applications the HMF-acid converting polypeptide may be any polypeptide having HMF-converting activity presently known or yet to be discovered.

A further aspect of the present invention is aimed at the use of a genetically modified cell according to the invention, for the biotransformation of furanic precursors

to FDCA. The furanic precursors may in particular be selected from 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), [5-(hydroxymethyl)furan-2-yl]methanol (HMF alcohol), furan-2,5-dicarbaldehyde (DFF), 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) or 5-formylfuran-2-carboxylic acid (FFA). Preferably HMF is a selected furanic precursor. HMF-acid may be an intermediate in the bioconversion of HMF to FDCA.

The invention will be further illustrated with reference to the following examples.

EXAMPLES

10

General Methodology

Strains and plasmids *Pseudomonas putida* S12 (ATCC 700801) was used as the host for expression of genes from *Cupriavidus basilensis* HMF14 (Wierckx et al., 2010; Koopman et al. 2010a) and *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819; genome sequence available at <http://genome.jgi-psf.org/metra/metra.home.html>). *Escherichia coli* strains DH5 α or TOP10 (Invitrogen) were used for general cloning purposes.

For episomal expression of *C. basilensis* or *M. radiotolerans* genes either the pUCP22-derived pJT'mcs (Koopman et al., 2010a) or pJNNmcs(t) (Wierckx et al., 2008), or the pBBR1MCS-derived pBT'mcs (Koopman et al., 2010a) was used. In pJT'mcs and pBT'mcs the expression of the target gene is driven from the constitutive *tac* promoter. In pJNNmcs(t) the expression is driven from the salicylate inducible NagR/P_{nagAa} expression cassette.

25

Media & Culture conditions Mineral salts medium (MM) was used as a defined medium. MM contained the following (per liter of demineralized water): 3.88 g of K₂HPO₄, 1.63 g of NaH₂PO₄, 2.0 g of (NH₄)₂SO₄, 0.1 g of MgCl₂·6H₂O, 10 mg of EDTA, 2 mg of ZnSO₄·7H₂O, 1 mg of CaCl₂·2H₂O, 5 mg of FeSO₄·7H₂O, 0.2 mg of Na₂MoO₄·2H₂O, 0.2 mg of CuSO₄·5H₂O, 0.4 mg of CoCl₂·6H₂O, and 1 mg of MnCl₂·2H₂O, supplemented with a carbon source as specified. Luria broth (L-broth: 10 g/l Bacto trypton (Difco), 5 g/l yeast extract (Difco), 5 g/l NaCl) was used as a complete medium for propagation of *P. putida* S12 and derivative strains, *C. basilensis*, *M.*

radiotolerans and *E. coli* DH5 α and derivatives. For plate culturing, L-broth was solidified with 1.5 % (w/v) of agar (Difco). Ampicillin (Amp) was added to the media to 100 μ g/ml for selection of *E. coli* transformants carrying pJT'mcs or pJNNmcs(t)-derived plasmids. Gentamicin (Gm) was added to 30 μ g/ml in Luria broth and 10 μ g/ml in mineral salts medium for selection of *P. putida* S12 transformants carrying pJT'mcs or pJNNmcs(t)-derived plasmids. For selection of either *E. coli* or *P. putida* S12 transformants carrying pBT'mcs-derived plasmids, 50 μ g/ml of kanamycin (Km) was added to the media. Antibiotics were purchased from Sigma-Aldrich. *P. putida*, *C. basilensis* and *M. radiotolerans* were cultured at 30 °C; *E. coli* was cultured at 37 °C

Fed batch experiments with *P. putida* S12-derived strains were performed in 2-L vessels controlled either by a Labfors 4 Bioreactor system (Infors Benelux BV) or a BioFlo110 controller (New Brunswick Scientific). Pressurized air or pure oxygen was supplied either in the head space or sparged through the broth. The temperature was controlled at 30 °C and the pH was maintained at 7.0 by automatic addition of either NH₄OH, NaOH or KOH. The batch phase was performed in 2x MM medium, supplemented with strain specific antibiotics and 40 g/l glycerol. For high-cell density cultures, the batch-phase medium was furthermore supplemented with 10 g/l of Yeast Extract (YE). After depletion of the initial glycerol, the feed (4 or 8 M of glycerol in 100 mM MgCl₂, supplemented with 1 mM of Na-salicylate when required) was started and controlled to allow for growth while maintaining glycerol as the limiting substrate in the culture. The HMF feed (4M in demineralized water) was fed via a separate feed pump; the feed rate was adjusted depending on the strain employed and the condition studied. The dissolved oxygen tension (DO) was continuously monitored and the stirring speed was adjusted to maintain sufficient aeration.

25

Assays & Analytical methods Cell dry weight (CDW) measurement: CDW content of bacterial cultures was determined by measuring optical density at 600 nm (OD₆₀₀) using a Biowave Cell Density Meter (WPA Ltd) or a μ Quant MQX200 universal microplate spectrophotometer (Biotek), using flat-bottom 96-well microplates (Greiner). An OD₆₀₀ of 1.0 corresponds to 0.56 g CDW / L (Biowave) or 1.4 g CDW / L (μ Quant) for *P. putida*.

30

HPLC analyses: Furan compounds (FDCA, HMF, HMF-alcohol, HMF-acid and FFA) were analyzed by RP-HPLC as described by Koopman et al. (2010a). Sugars,

alcohols and organic acids were also analyzed by HPLC (Agilent 1100 system) using a refractive index (RI) detector. The column used was a Bio-Rad Aminex HPX-87H (300 x 7.8 mm, hydrogen form, 9 µm particle size, 8 % cross linkage, pH range 1–3) with 5 mM H₂SO₄ as the eluent at a flow rate of 0.6 ml / min.

5

Chemicals

HMF was purchased either at Sigma, Eurolabs Ltd (Poynton, UK) or Yore Chemipharm Co. Ltd. (Ningbo, China). Analytical standards of FDCA and 5-hydroxymethyl-furoic acid (HMF acid) were purchased from Immunosource B.V. (Halle-Zoersel, Belgium), respectively, Matrix Scientific (Columbia SC, USA). All other chemicals were purchased from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands).

Molecular and genetic techniques: Genomic DNA was isolated from *C. basilensis* HMF14 and *M. radiotolerans* JCM 2831 using the DNeasy tissue kit (QIAGEN). Plasmid DNA was isolated with the QIAprep spin miniprep kit (QIAGEN). Agarose-trapped DNA fragments were isolated with the QIAEXII gel extraction kit (QIAGEN).

PCR reactions were performed with Accuprime *Pfx* polymerase (Invitrogen) according to the manufacturer's instructions. Oligonucleotide primers (specified in the examples) were synthesized by MWG Biotech AG (Germany). Plasmid DNA was introduced into electrocompetent cells using a Gene Pulser electroporation device (BioRad). Other standard molecular biology techniques were performed according to Sambrook and Russell (2001).

25

Example I: Co-expression of HmfH and HmfT1 improves FDCA production in *P. putida* S12

The *hmfT1* gene (formerly designated *mfs1* (Koopman et al., 2010a); SEQ ID NO: 7) was amplified from genomic DNA of *Cupriavidus basilensis* HMF14 by PCR using primers hmfT1(f) (SEQ ID NO: 13) and hmfT1(r) (SEQ ID NO: 14). The PCR product was introduced as a 1359-bp *EcoRI-NheI* fragment in pJNNmcs(t) yielding pJNNhmfT1(t). The *hmfH* gene (SEQ ID NO: 11) including its native ribosome binding site (RBS) was amplified by PCR from genomic DNA of *C. basilensis* HMF14 using

30

primers FN23 (SEQ ID NO 15) and FN24 (SEQ ID NO: 16). The PCR product was cloned as a 1777-bp *EcoRI* fragment in pBT'mcs yielding plasmid pBT'hmfH. Plasmids pBT'hmfH and pJNNhmfT1(t) were successively introduced into *P. putida* S12, yielding *P. putida* S12_B38.

5 *P. putida* S12_B38 was cultured in fed-batches as described in the general methodology section. In the batch phase, a cell density of approximately 3 g CDW / l was achieved after which the glycerol feed and the HMF feed were started. A control fed-batch culture was performed with *P. putida* S12_2642 (similar to *P. putida* S12_hmfH (Koopman et al., 2010b)) which does not express HmfT1.

10 Fig. 2 shows the concentrations of FDCA and HMF-acid in HMF-fed cultures of *P. putida* strains S12_2642 and S12_B38. The extensive accumulation of HMF-acid is evident for *P. putida* S12_2642. By contrast, the HMF-acid accumulation was negligible in the *P. putida* S12_B38 culture. The reduced HMF-acid accumulation furthermore allowed increased HMF feed rates, which resulted in higher FDCA titers in a
15 considerably shorter process time. This was also clearly reflected in the specific FDCA productivity for the tested strains (Fig. 3).

Example II: Co-expression of HmfH and a polypeptide from *Methylobacterium radiotolerans* JCM 2831 improves FDCA production in *P. putida* S12

The gene with locus tag *mrad2831_4728* (SEQ ID: 9) was amplified from
 5 genomic DNA of *Methylobacterium radiotolerans* JCM2831 by PCR using primers
 Mrad(f) (SEQ ID NO: 17) and Mrad(r) (SEQ ID NO: 18). The PCR product was
 introduced as a 1381-bp *EcoRI-NheI* fragment in pJNNmcs(t) yielding pJNN_Mrad(t).
 Plasmids pBT'hmfH (see Example I) and pJNN_Mrad(t) were successively introduced
 into *P. putida* S12, yielding *P. putida* S12_B51.

10 *P. putida*_B51 was cultured in shake flasks on mineral salts medium (4x buffer
 strength) supplemented with 1 g/l yeast extract, 80 mM glycerol, 2 mM glucose, 100 μ M
 Na-salicylate, 50 μ g/ml kanamycin and 10 μ g/ml gentamicin. As control strain, *P. putida*
 S12_B38 was used (see Example I). After overnight culturing, the cultures were
 supplemented with 80 mM glycerol and 100 μ M Na-salicylate. Subsequently,
 15 approximately 10 mM HMF was added and FDCA production was assessed.

Fig. 4 shows that the HMF-acid accumulation during FDCA production was
 negligible for both strains, confirming that HmfT1 and the Mrad2831_4728 polypeptide
 exhibited a similar functionality. For *P. putida* S12_B38, the FDCA production showed a
 longer lag phase and a somewhat slower rate, which could be attributed to the lower
 20 initial biomass density (Fig. 4A). The specific maximum FDCA productivity, however,
 was identical for both strains, *i.e.*, 2.36 mmol FDCA/(g CDW, h), indicating that the
 Mrad2831_4728 and HmfT1 polypeptides were equally effective in minimizing HMF-
 acid accumulation and maximizing FDCA production.

25 **Example III: High-level HMF-acid conversion capacity by *P. putida* S12 co-
 expressing HmfH and HmfT1**

As demonstrated in Example I, co-expression of HmfH and HmfT1 in *P. putida*
 S12 considerably improved the specific capacity to oxidize HMF to FDCA. To make
 optimal use of this improved capacity, a fed-batch experiment was performed with *P.*
 30 *putida* S12_B38 starting at a high biomass density.

The HMF feed was started at a high rate (20 ml/h; 4 M HMF feed solution) in
 order to saturate the oxidation capacity of *P. putida* S12_B38 and provoke the
 accumulation of HMF and HMF-acid (Fig. 5). When HMF-acid had accumulated to

approximately 20 mM, the HMF feed rate was lowered to 5 ml/h and the furanics concentrations were monitored. Initially, the HMF-acid concentration increased to approximately 37 mM due to oxidation of residual accumulated HMF, after which it dropped to less than 2 mM within 5 h, at an HMF feed rate of 0.72 mmol/(g CDW.h).

5 These results clearly demonstrate that the HMF-acid oxidation capacity was improved by co-expression of HmfH and HmfT1. The results by Koopman *et al.* (2010b), showed that *P. putida* S12_hmfH (which lacks HmfT1) required over 50 h to reduce the HMF-acid concentration from approximately 50 mM to less than 5 mM, at a much lower HMF feed rate (0.09 mmol/(g CDW.h)). The improved HMF-acid oxidation
10 capacity resulted in a much higher final FDCA titer (152 g/l vs 30.1 g/l by Koopman *et al.* (2010b)) that was furthermore achieved in a shorter process time (94 h vs 115 h by Koopman *et al.* (2010b)).

**EXAMPLE IV: Co-expression of HmfH, HmfT1 and an aldehyde
15 dehydrogenase from *C. basilensis* HMF 14 improves FDCA production in *P. putida* S12**

 The gene encoding an aldehyde dehydrogenase (SEQ ID 26; translated amino acid sequence: SEQ ID 19) associated with the HMF-degradation operon in *Cupriavidus basilensis* HMF14 (Wierckx *et al.*, 2011) was amplified by PCR using primers FN13
20 (SEQ ID 33) and FN14 (SEQ ID 34). The PCR product was introduced as a 1543-bp *NotI* fragment in *Bsp120I*-digested (compatible to *NotI*) pBT'hmfH (see example I) yielding pBT'hmfH-adh. The plasmid variant in which the aldehyde dehydrogenase encoding gene was present in the forward (f) orientation (pBT'hmfH-adh(f)) and pJNNhmfT1(t) (see example I) were successively introduced into *P. putida* S12. The
25 resulting strain, *P. putida* S12_B97, co-expressed HmfH, HmfT1, and the aldehyde dehydrogenase. As a control strain for co-expression of the HmfH oxidoreductase and the aldehyde dehydrogenase (*i.e.*, without the HMF-acid transporter HmfT1), *P. putida* S12_B101 was constructed which only contained pBT'hmfH-adh(f). *P. putida* S12_B38 (see example I) was used as control strain for co-expression of HmfT1 and HmfH
30 without the aldehyde dehydrogenase.

P. putida strains S12_B38, S12_B97 and S12_B101 were cultured in shake flasks on mineral salts medium (4x buffer strength) supplemented with 1 g/l yeast extract, 80 mM glycerol, 2 mM glucose, 50 µg/ml kanamycin and 10 µg/ml gentamicin (note: for

strain B101 only kanamycin was added). Na-salicylate (1 μ M) was added for induction of *hmfT1* in the precultures only. After addition of approximately 10 mM HMF, the accumulation of FDCA and HMF-acid was assessed.

5 In the strain that co-expressed HmfH (oxidoreductase) and HmfT1 (HMF-acid transporter) (strain S12_B38; Fig. 6A), FDCA production only commenced after FFA had accumulated to a substantial level. HMF acid accumulated transiently to low amounts, as observed previously (see example II). When the aldehyde dehydrogenase was co-expressed with HmfH and HmfT1 (strain S12_B97; Fig. 6B), FDCA formation commenced without delay, and both FFA and HMF-acid were observed only in trace
10 amounts. Co-expression of the aldehyde dehydrogenase and HmfH without HmfT1 (strain S12_B101; Fig. 6C), resulted in extensive accumulation of HMF-acid whereas only small amounts of FFA and FDCA were produced.

The results demonstrated that the oxidation of HMF to HMF-acid is significantly enhanced by expressing the aldehyde dehydrogenase. HmfT1 must be co-expressed,
15 however, to enable efficient biotransformation of the HMF-acid produced. The aldehyde dehydrogenase furthermore improved the oxidation of the intermediate product FFA to FDCA. Thus, simultaneous expression of the aldehyde dehydrogenase and HmfT1 considerably improves the overall potential for, and rate of, HMF oxidation via HMF-acid to the final product.

REFERENCES

- Almeida et al.(2009) Metabolic effects of furaldehydes and impacts on biotechnological processes. *Applied Microbiology and Biotechnology* , 82 (4): 625-638.
- 5 Altschul et al., (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25 (17): 3389-3402.
- Altschul, et al., (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-10.
- Ausubel et al. (eds.) (1995), *Current Protocols in Molecular Biology*, (John Wiley & Sons, N.Y..
- 10 Hawksworth et al., (1995) In: Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, CAB International, University Press, Cambridge, UK.
- Koopman et al., (2010a) Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of *Cupriavidus basilensis* HMF14. *PNAS*
- 15 (2010a), Vol. 107, p. 4919-4924.
- Koopman et al., (2010b) Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid. *Bioresour. Technol.* (2010b), Vol. 101, p. 6291-6296.
- 20 Meyers et al., (1988) Optimal alignments in linear space *CABIOS*, 4:11-17.
- Needleman et al., (1970) A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* (48): 444-453.
- Nichols NN, Mertens JA. Identification and transcriptional profiling of *Pseudomonas*
- 25 *putida* genes involved in furoic acid metabolism (2008) *FEMS Microbiol Lett* 284: 52-57
- Sambrook et al., (1989), *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y..
- Sambrook et al., (2001) *Molecular cloning – a laboratory manual*. Third edition. Cold
- 30 Spring Harbor Laboratory Press, Cold Sprong Harbor, NY.
- Werpy et al., (2004) *Top Value-Added Chemicals from Biomass, Volume I - Results of screening for potential Candidates from Sugars and Synthesis gas.*

Wierckx et al., (2008) Transcriptome analysis of a phenol producing *Pseudomonas putida* S12 construct: genetic and physiological basis for improved production. J Bacteriol 190: 2822-2830.

- 5 Wierckx et al. (2010) Isolation and characterization of *Cupriavidus basilensis* HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate. Microb Biotechnol 3: 336-343.

- 10 Wierckx et al., (2011) Microbial degradation of furanic compounds: biochemistry, genetics, and impact. Appl Microbiol Biotechnol 92:1095–1105

CLAIMS

1. Genetically modified cell comprising:

- (i) a first polynucleotide sequence coding for a first polypeptide comprising an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequence set out in SEQ ID NO: 1, 2, 3 or 4; and
- (ii) a second polynucleotide sequence coding for a second polypeptide having 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) converting activity;

characterized in that said cell is genetically modified by functional introduction of at least the first polynucleotide sequence, and preferably by functional introduction of both the first and second polynucleotide sequence.

2. Genetically modified cell according to claim 1, wherein the second polypeptide is a 5-(hydroxymethyl)furan-2-carboxylic acid oxidoreductase, preferably a polypeptide having an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequences set out in SEQ ID NO: 5 or 6.

3. A genetically modified cell according to any of the claims 1-2, wherein the first polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60%, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 7, 8, 9, or 10.

4. A genetically modified cell according to any of the claims 1-3, wherein the second polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 11 or 12.

5. A genetically modified cell according to any of the claims 1-4, comprising a third polynucleotide sequence coding for a third polypeptide comprising an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequence set out in SEQ ID NO: 19, 20, 21, 22, 23, 24, or 25.

6. A genetically modified cell according to any of the claims 1-5, wherein the third polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 26, 27, 28, 29, 30, 31 or 32.

7. A genetically modified cell according to any of the claims 1-6, wherein the cell is a prokaryote cell such as a bacterial cell, or a eukaryote cell, such as a yeast cell, a fungal cell, a plant cell or an animal cell.

8. A genetically modified cell according to any of the claims 1-7, wherein the cell is a prokaryotic cell, preferably a bacterial cell selected from, the genera *Escherichia*, *Anabaena*, *Caulobacter*, *Gluconobacter*, *Rhodobacter*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Bradyrhizobium*, *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Staphylococcus*, *Streptomyces*, *Zymomonas*, *Acetobacter*, *Streptococcus*, *Bacteroides*, *Selenomonas*, *Megasphaera*, *Burkholderia*, *Cupriavidus*, *Ralstonia*, *Methylobacterium*, *Methylovorus*, *Rhodopseudomonas*, *Acidiphilium*, *Dinoroseobacter*, *Agrobacterium*, *Sulfolobus* or *Sphingomonas*, more preferably selected from the group consisting of *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus pumilis*, *Bacillus megaterium*, *Bacillus halodurans*, *Bacillus pumilus*, *Gluconobacter oxydans*, *Caulobacter crescentus*, *Methylobacterium extorquens*, *Methylobacterium radiotolerans*, *Methylobacterium nodulans*, *Rhodobacter sphaeroides*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Pseudomonas putida* S12, *Paracoccus*

denitrificans, *Escherichia coli*, *Corynebacterium glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *Rhizobium radiobacter*, *Rhizobium leguminosarum*, *Rhizobium leguminosarum* bv. *trifolii*, *Agrobacterium radiobacter*, *Cupriavidus basilensis*,
5 *Cupriavidus necator* (*Ralstonia eutropha*), *Ralstonia pickettii*, *Burkholderia phytofirmans*, *Burkholderia phymatum*, *Burkholderia xenovorans*, *Burkholderia graminis*, *Rhodopseudomonas palustris*, *Acidiphilium cryptum*, *Dinoroseobacter shibae*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*.

10

9. A genetically modified cell according to any of the claims 1-8, wherein the cell is a eukaryotic host cell, for example a mammalian, insect, plant, fungal, or algal cell, such as a mammalian cell selected from Chinese hamster ovary (CHO) cells, COS cells, 293 cells, PerC6 cells, or hybridomas, an insect cell selected
15 from Sf9 or Sf21 cells and derivatives thereof, preferably an eukaryotic cell selected from fungal or yeast cells, such as a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* species, with more preference a eukaryotic cell selected from *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*,
20 *Pichia stipitis* or *Pichia pastoris*, or a filamentous fungal cell from a species selected from the subdivision *Eumycota* and *Oomycota* or alternatively a species from the genus *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*,
25 *Paecilomyces*, *Penicillium*, *Piromyces*, *Phanerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, or *Trichoderma* such as *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Trichoderma reesei* or
30 *Penicillium chrysogenum*.

10. A genetically modified cell according to any of the claims 1-9, wherein the first and second polynucleotide sequence are located on a single vector, optionally a vector further comprising the third polynucleotide sequence.

5 11. A vector comprising:

- (i) a first polynucleotide sequence coding for a first polypeptide comprising an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with the amino acid sequence set out in SEQ ID NO: 1, 2, 3 or 4; and
- 10 (ii) a second polynucleotide sequence coding for a second polypeptide having 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) converting activity.

15 12. A vector according to claim 11, wherein the second polypeptide is a 5-hydroxymethylfuroic acid oxidoreductase, preferably a polypeptide having an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequences set out in SEQ ID NO: 5

20 or 6.

13. A vector according to any of the claims 11-12, wherein the first polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60%, such as at least 70%, more preferably at least 80%, such

25 as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 7, 8, 9 or 10.

14. A vector according to any of the claims 11-13, wherein the second polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such

30 as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 11 or 12.

15. A vector according to any of the claims 11-14, comprising a third polynucleotide sequence coding for a third polypeptide comprising an amino acid sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with an amino acid sequence set out in SEQ ID NO: 19, 20, 21, 22, 23, 24, or 25.

16. A vector according to any of the claims 11-15, wherein the third polynucleotide sequence comprises a nucleotide sequence having at least 45%, preferably at least 60 %, such as at least 70%, more preferably at least 80%, such as 90%, most preferably at least 95% sequence similarity with a nucleotide sequence set out in SEQ ID NO: 26, 27, 28, 29, 30, 31, or 32.

17. 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) converting process comprising incubating a cell according to any of the claims 1-10 in the presence of HMF-acid under conditions suitable for the conversion by said cell of HMF-acid.

18. Process according to claim 17, wherein HMF-acid is formed *in situ* from one or more furanic precursors of HMF-acid under conditions suitable for the conversion by said cell of the one or more furanic precursors to HMF-acid.

19. Process according to any of the claims 17-18, wherein the second polypeptide of the cell is an HMF-acid oxidoreductase and further comprising subjecting the cell to conditions suitable for the conversion of HMF-acid to FDCA.

20. Process according to any of the claims 17-19, wherein a furanic precursor is selected from the group comprising 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), furan-2,5-dicarbaldehyde (DFF), [5-(hydroxymethyl)furan-2-yl]methanol (HMF alcohol) and preferably is HMF.

21. Process according to any of the claims 17-20, wherein a furanic precursor is obtained from one or more hexose sugars, preferably one or more hexose sugars obtained from lignocellulosic biomass, for example by acid-catalyzed dehydration.

5

22. Process according to any of the claims 17-21, wherein FDCA is recovered from the reaction mixture by a process comprising acid precipitation followed by cooling crystallization, or alternatively followed by solvent extraction.

10

23. Use of a genetically modified cell according to any of the claims 1-10, for the biotransformation of a number of furanic precursors to FDCA.

15

24. Use according to claim 23, wherein the number of furanic precursors are selected from the group comprising 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), [5-(hydroxymethyl)furan-2-yl]methanol (HMF alcohol), 5-(hydroxymethyl)furan-2-carboxylic acid (HMF-acid) furan-2,5-dicarbaldehyde (DFF) or 5-formylfuran-2-carboxylic acid (FFA).

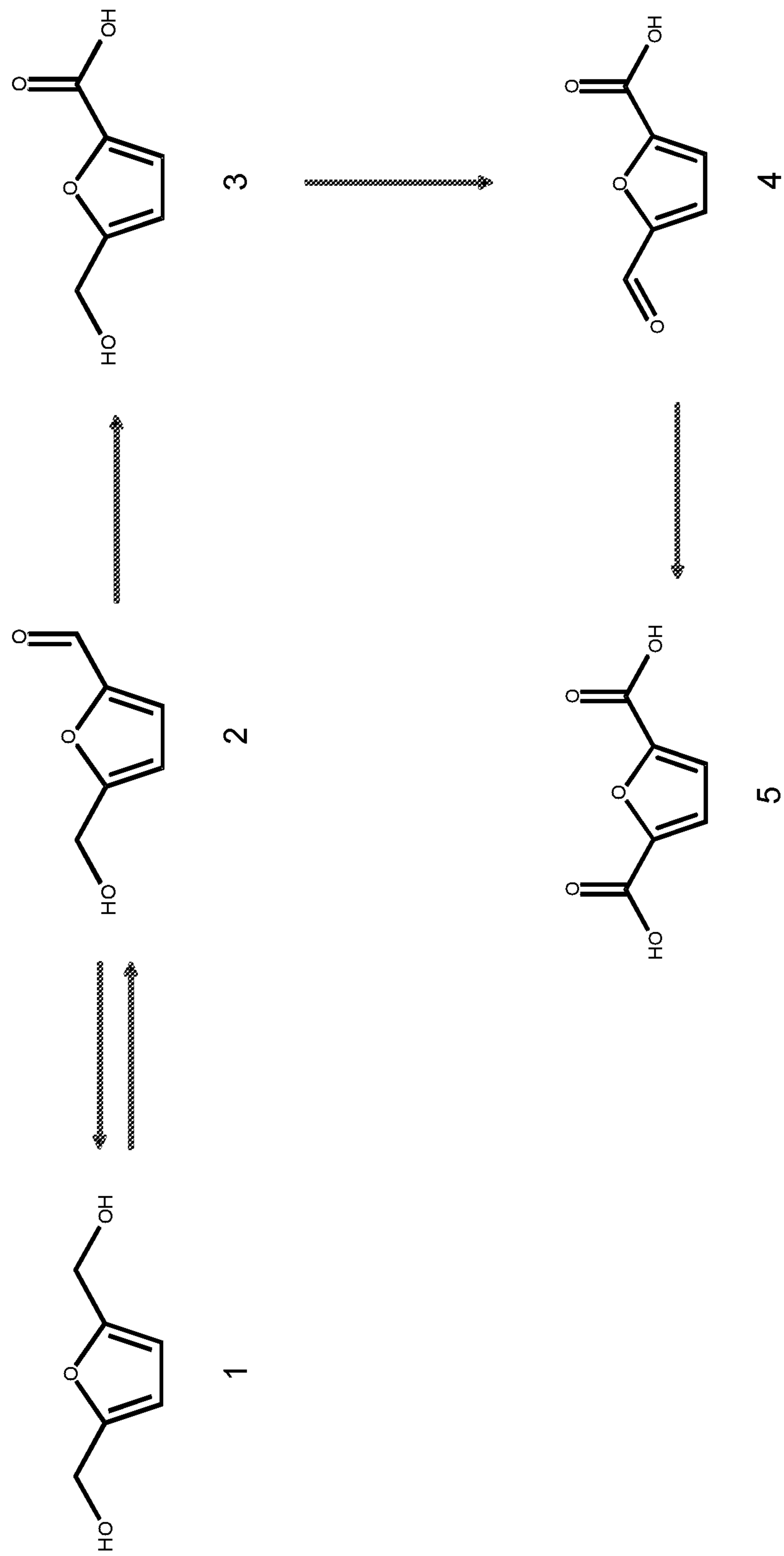


FIG. 1

2/7

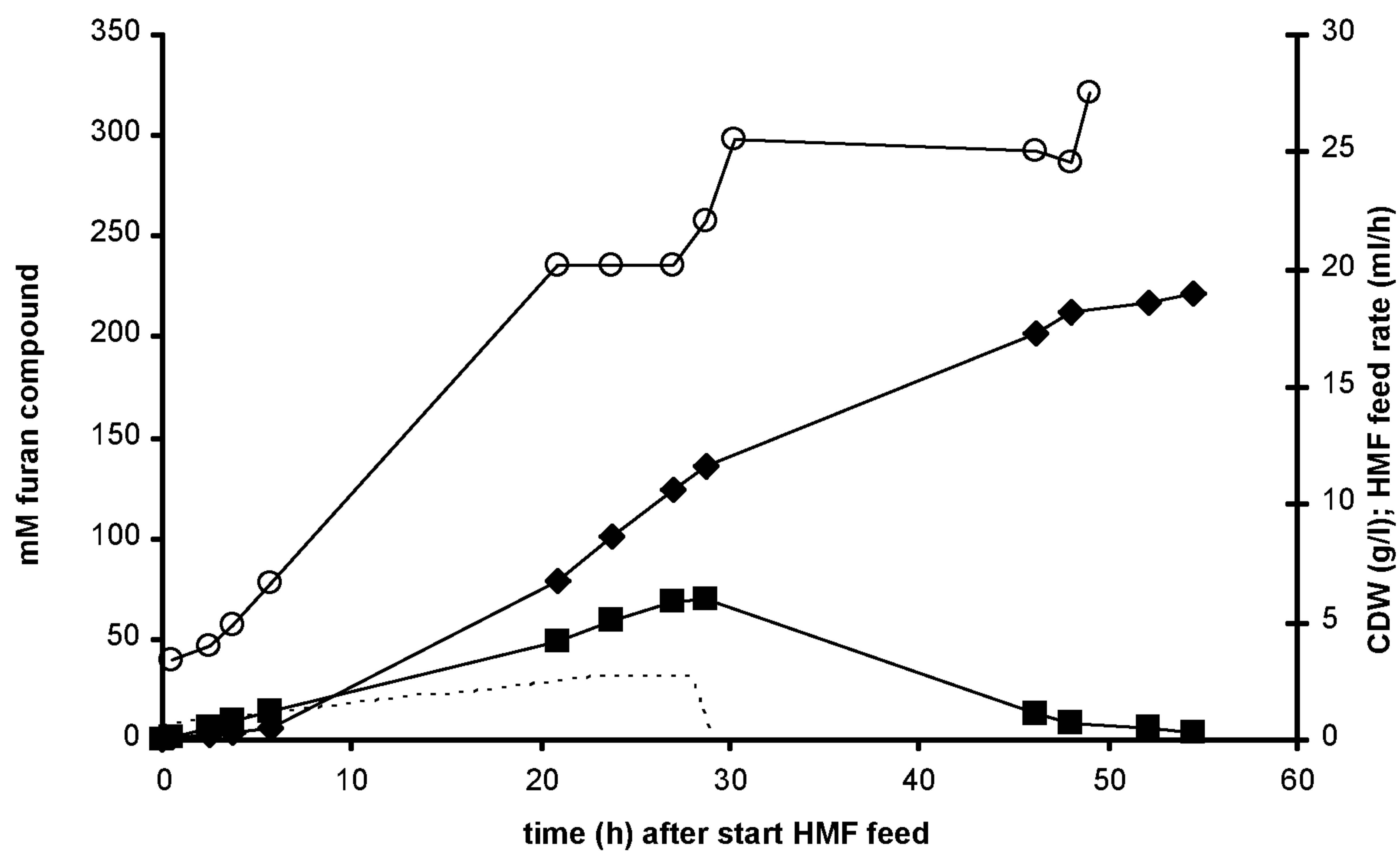


FIG. 2a

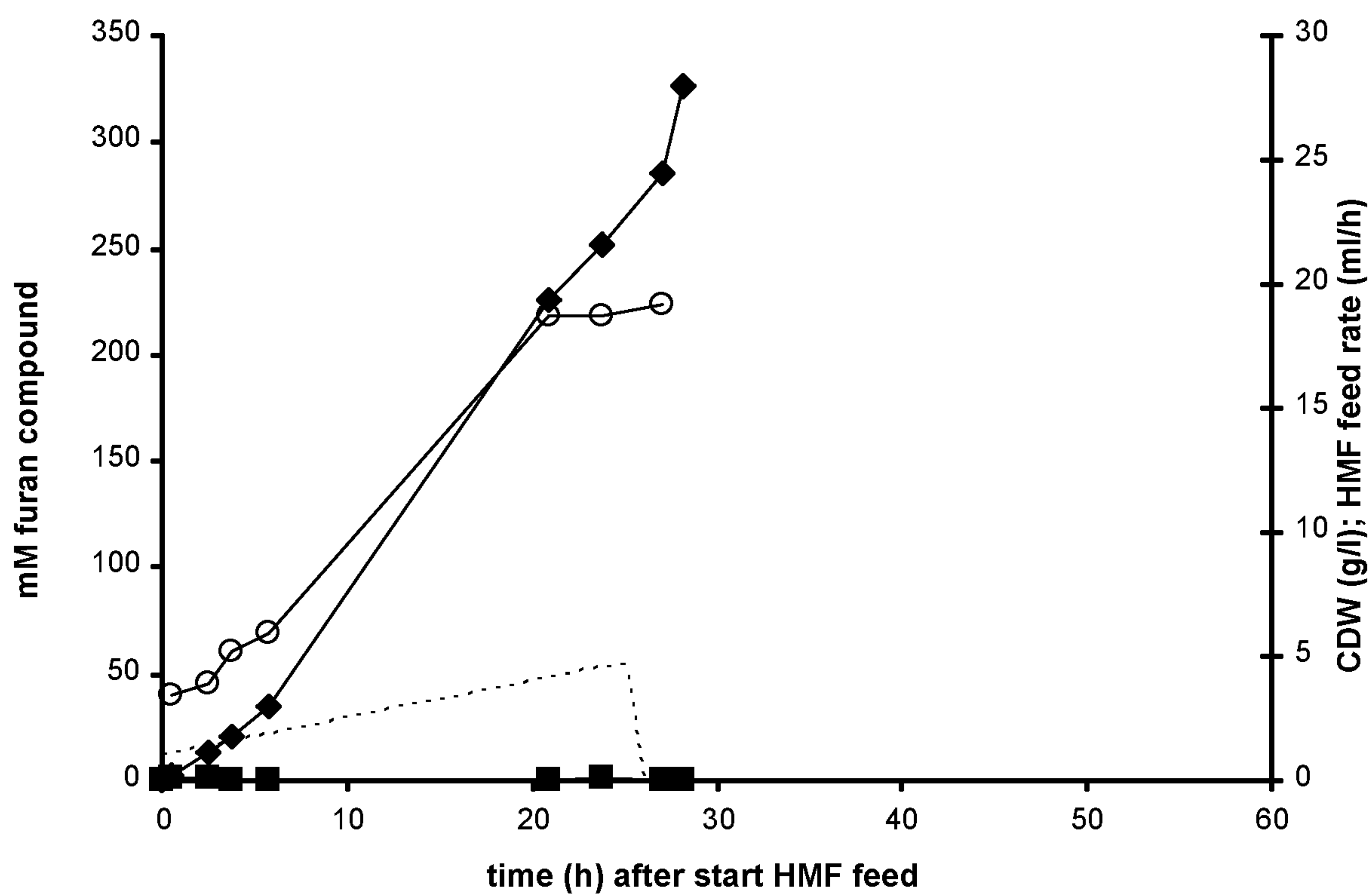


FIG. 2b

3/7

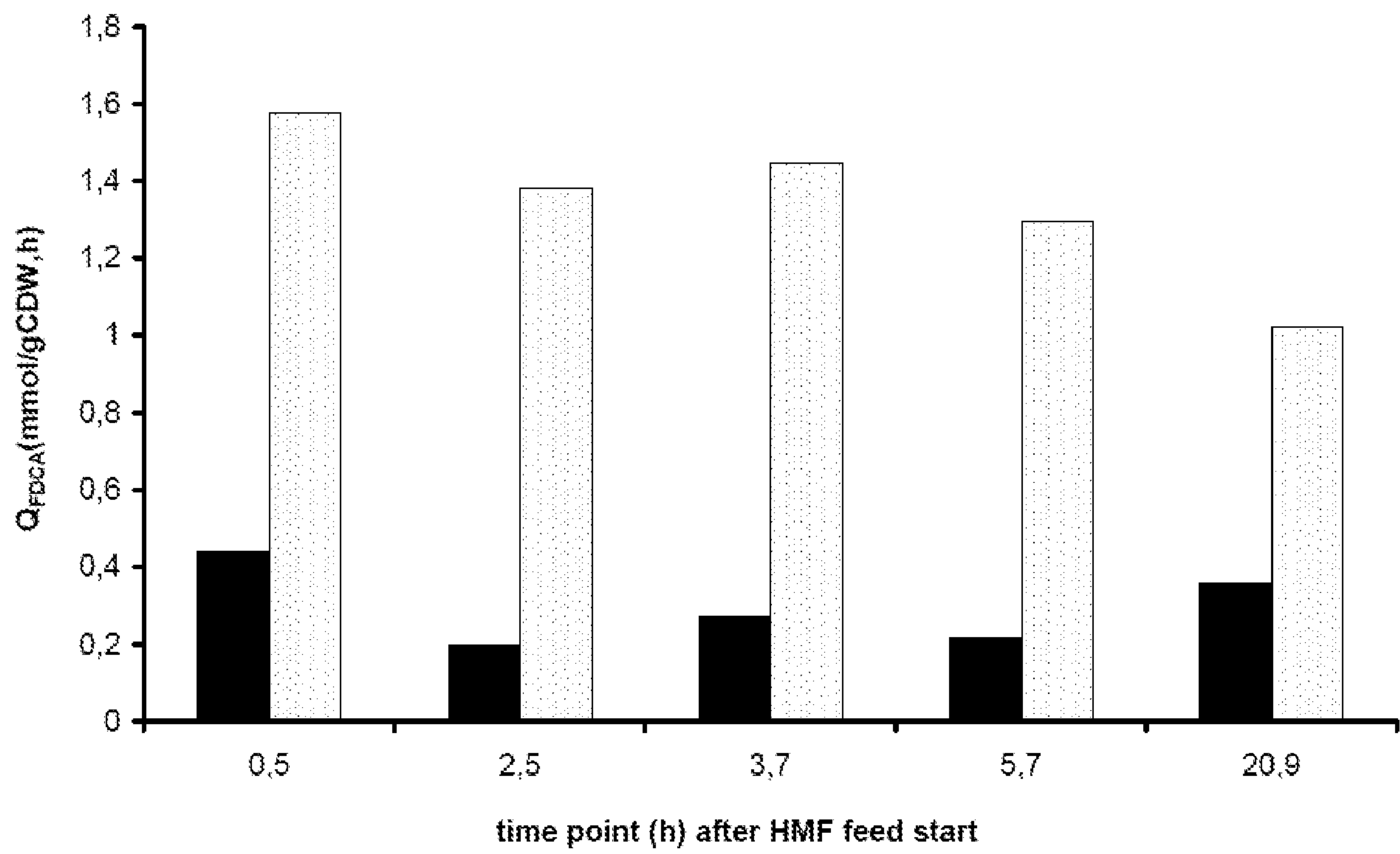


FIG. 3

4/7

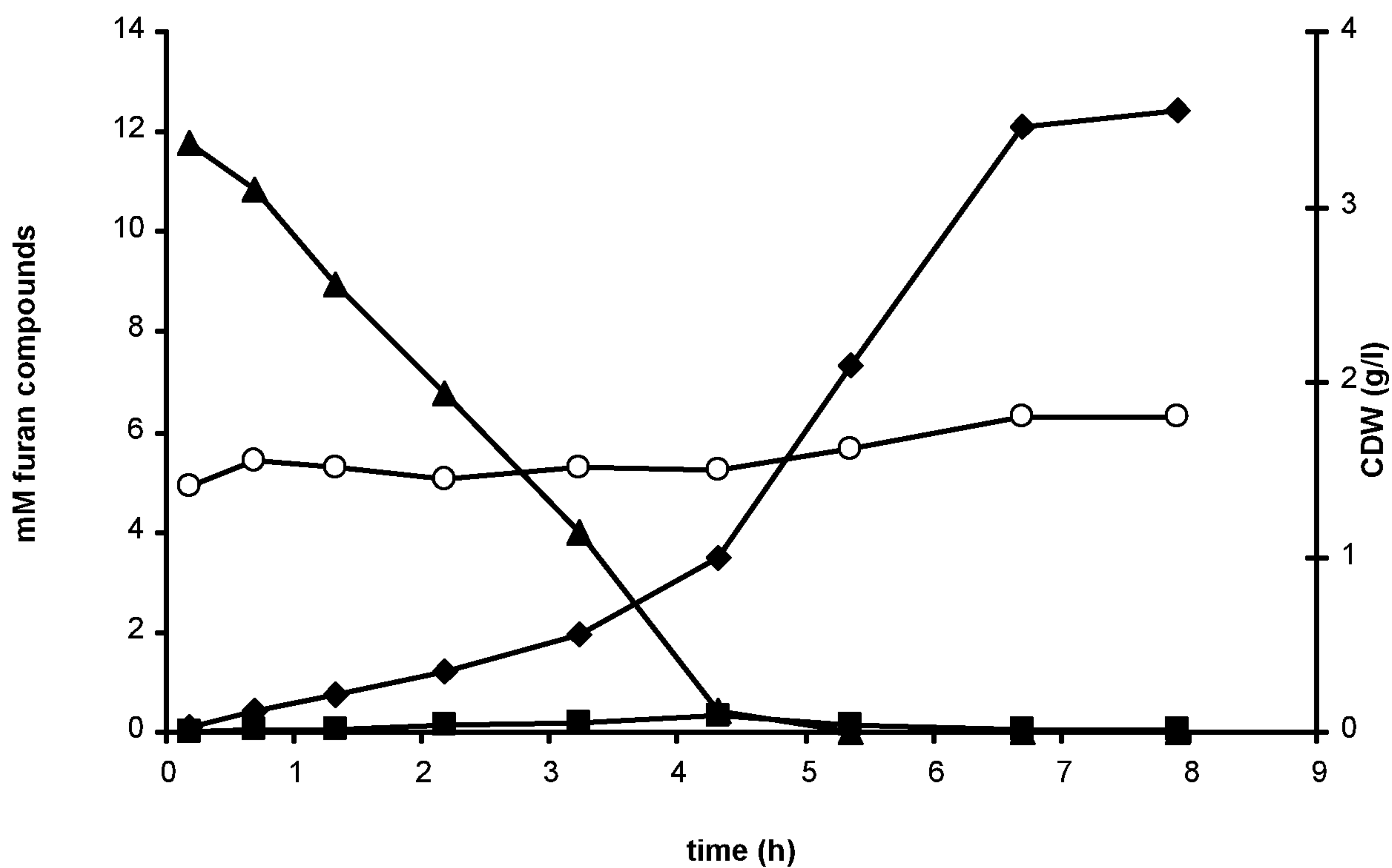


FIG. 4a

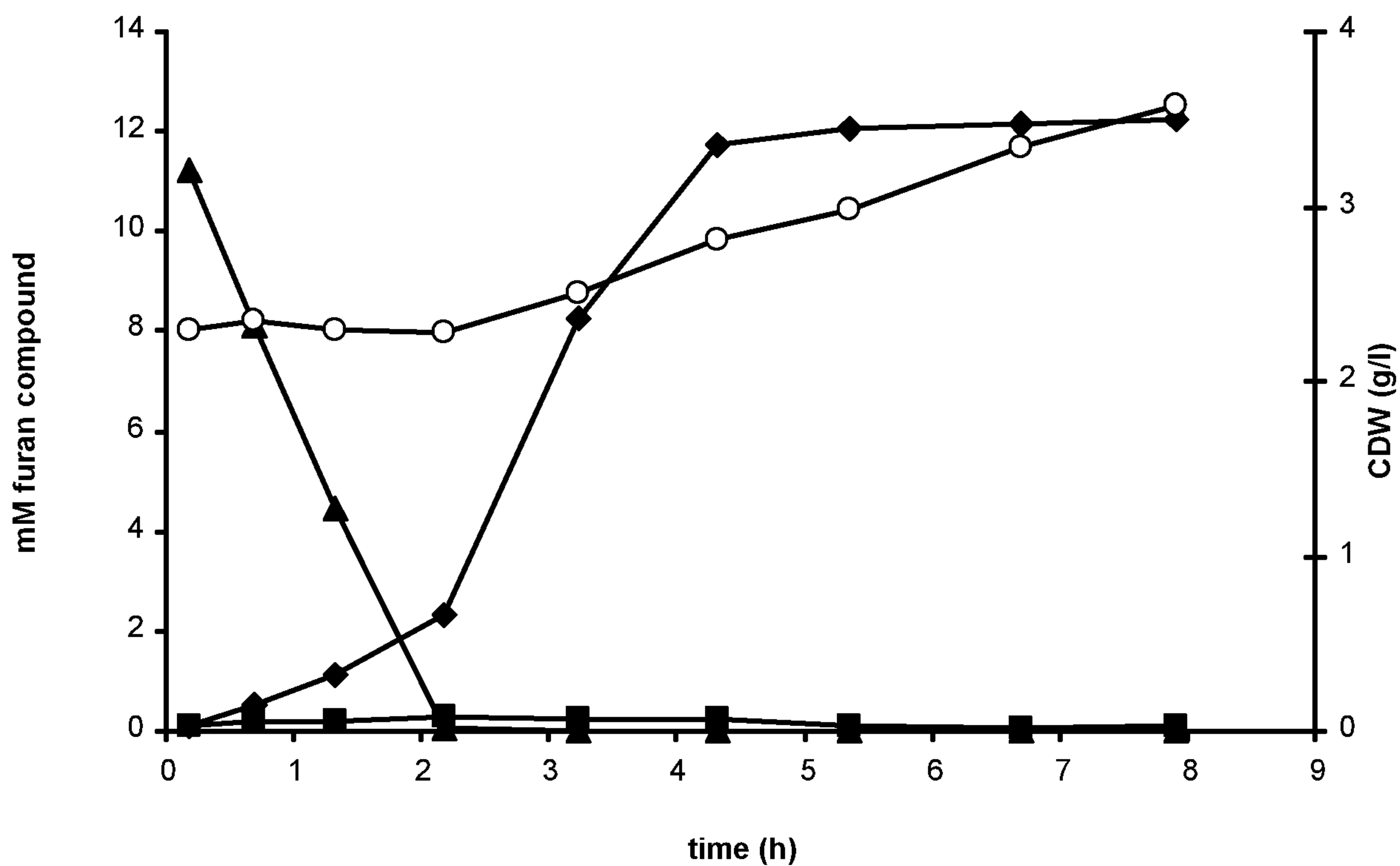


FIG. 4b

5/7

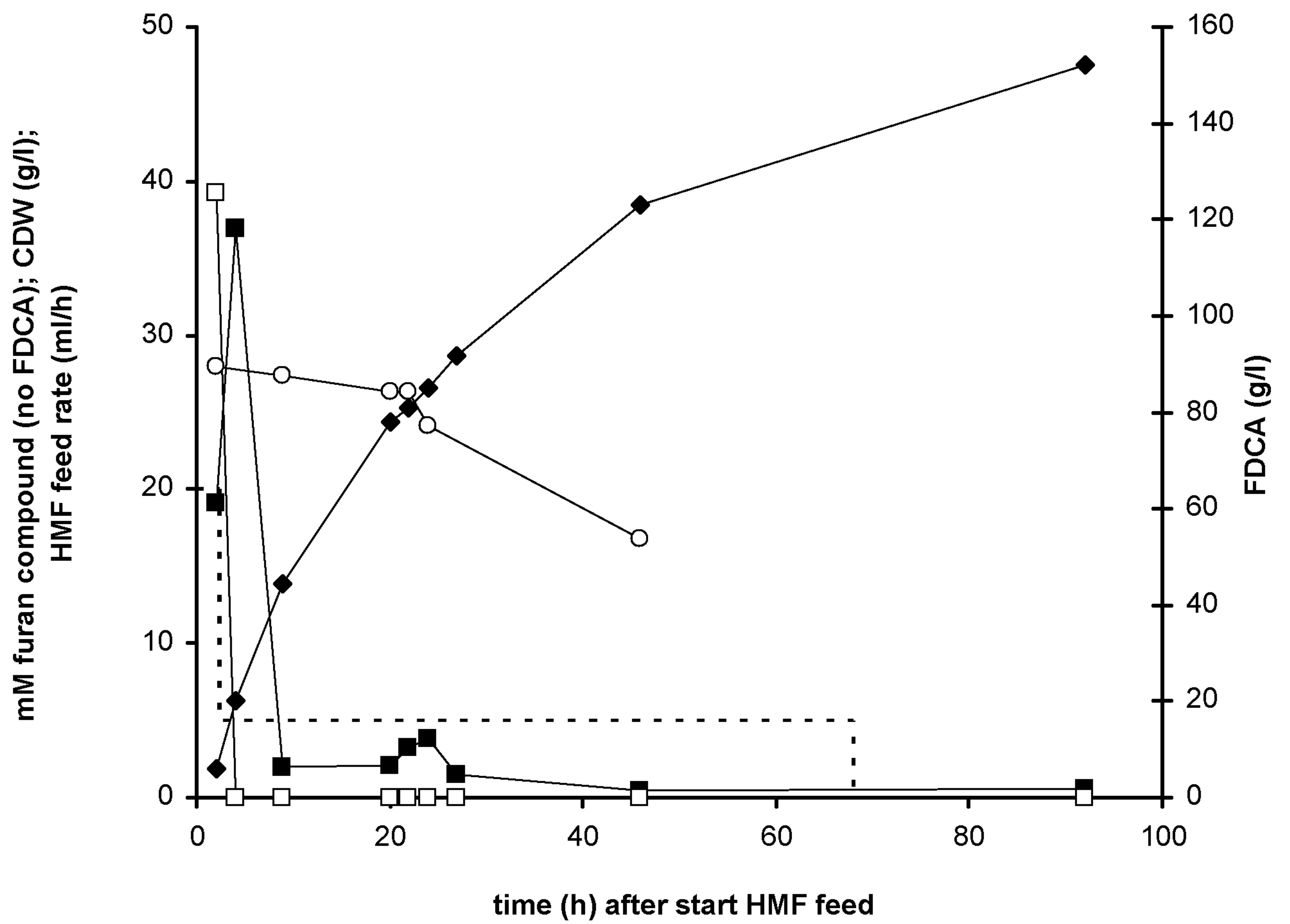


FIG. 5

6/7

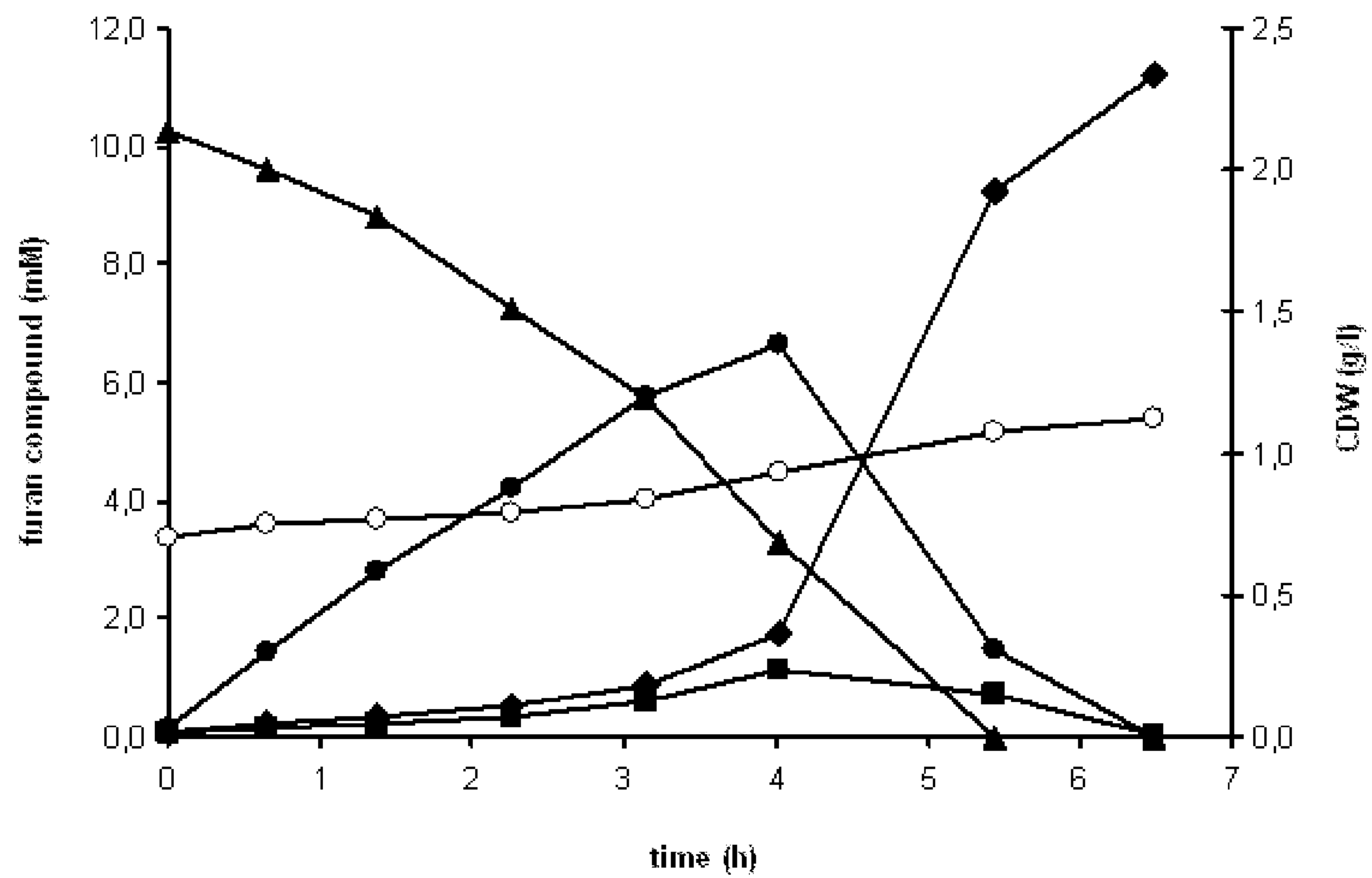


FIG. 6a

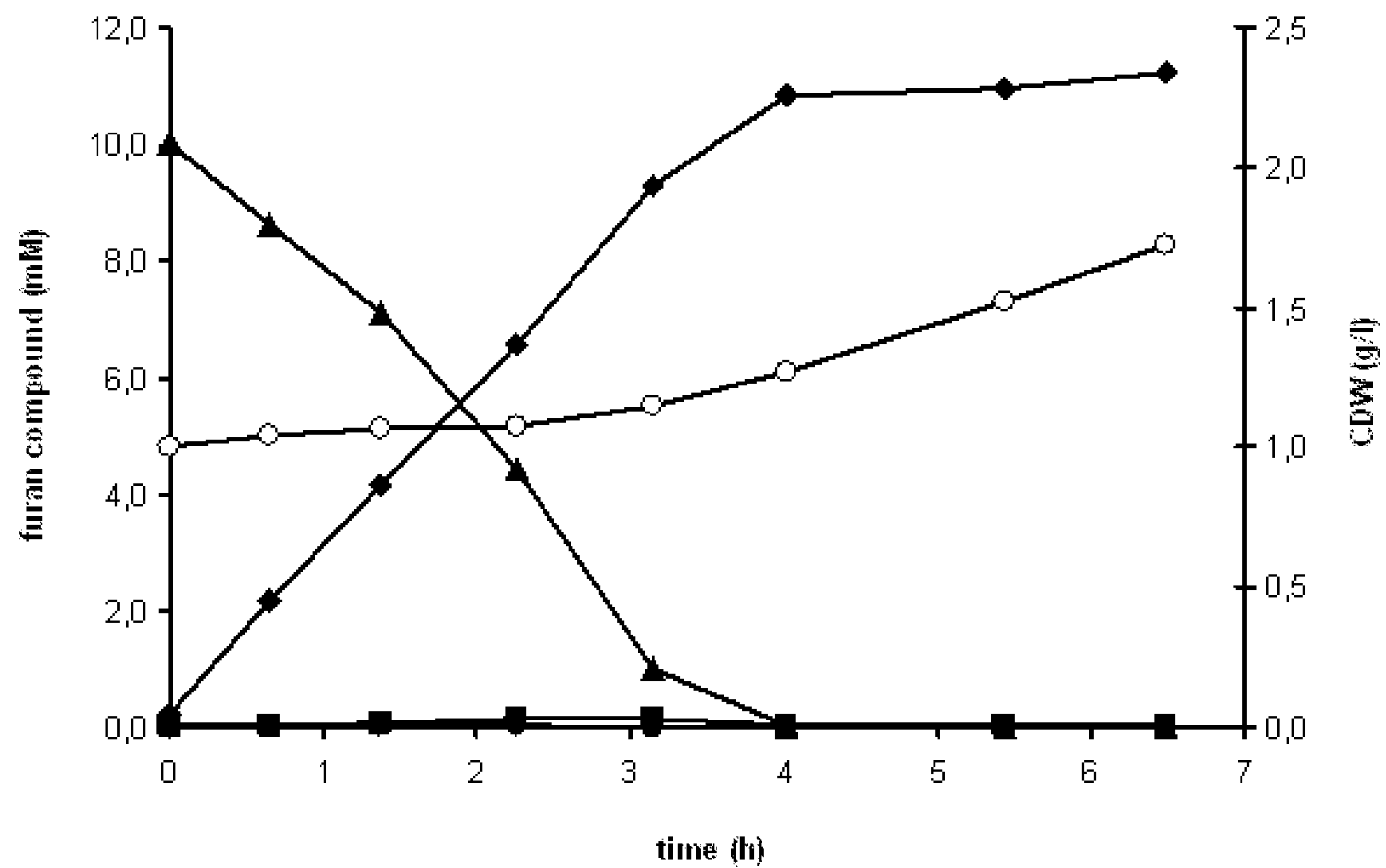


FIG. 6b

7/7

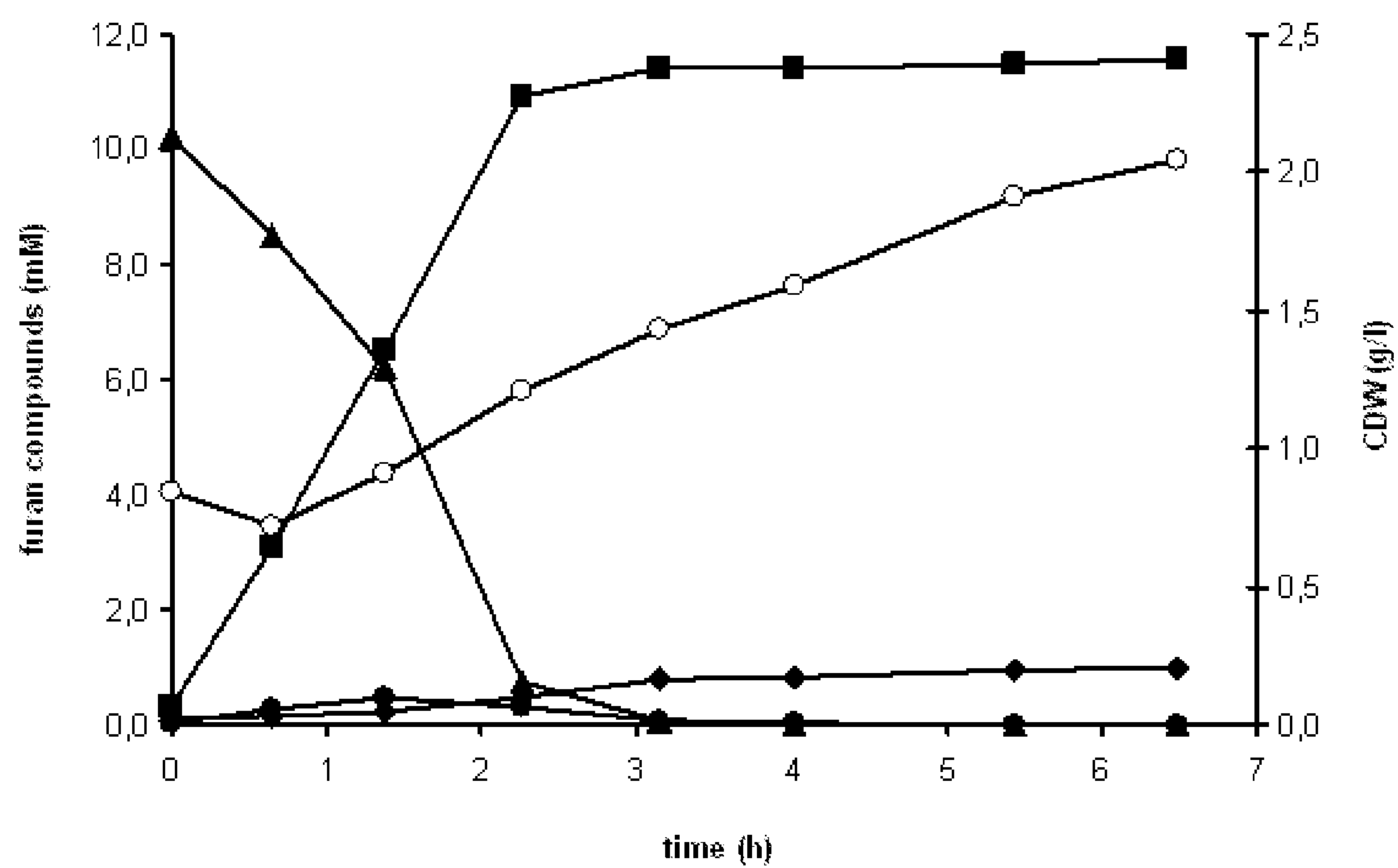


FIG. 6c