



(86) Date de dépôt PCT/PCT Filing Date: 2008/05/09  
(87) Date publication PCT/PCT Publication Date: 2008/11/20  
(85) Entrée phase nationale/National Entry: 2009/10/29  
(86) N° demande PCT/PCT Application No.: US 2008/063237  
(87) N° publication PCT/PCT Publication No.: 2008/141174  
(30) Priorité/Priority: 2007/05/09 (US60/916,978)

(51) Cl.Int./Int.Cl. *C07H 21/02* (2006.01),  
*C12N 1/20* (2006.01), *C12N 15/00* (2006.01)

(71) Demandeur/Applicant:  
MASCOMA CORPORATION, US

(72) Inventeurs/Inventors:  
HOGSETT, DAVID A., US;  
RAJGARHIA, VINEET B., US

(74) Agent: RICHES, MCKENZIE & HERBERT LLP

(54) Titre : ORGANISMES THERMOPHILES ET MESOPHILES ASSOCIES A L'INACTIVATION GENIQUE, ET  
PROCEDES POUR LEUR UTILISATION

(54) Title: GENE KNOCKOUT MESOPHILIC AND THERMOPHILIC ORGANISMS, AND METHODS OF USE THEREOF

(57) **Abrégé/Abstract:**

One aspect of the invention relates to a genetically modified thermophilic or mesophilic microorganism, wherein a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain embodiments, the aforementioned microorganism further comprises a first non-native gene, which first non-native gene encodes a first non-native enzyme involved in the metabolic production of ethanol. Another aspect of the invention relates to a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism.



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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 November 2008 (20.11.2008)

PCT

(10) International Publication Number  
**WO 2008/141174 A3**

## (51) International Patent Classification:

**C07H 21/02** (2006.01) **C12N 15/00** (2006.01)  
**C12N 1/20** (2006.01)

## (21) International Application Number:

PCT/US2008/063237

(22) International Filing Date: 9 May 2008 (09.05.2008)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

60/916,978 9 May 2007 (09.05.2007) US

(71) Applicant (for all designated States except US): **MAS-COMA CORPORATION** [US/US]; 1380 Soldiers Field Road, Boston, MA 02135 (US).

## (72) Inventors; and

(75) Inventors/Applicants (for US only): **HOGSETT, David, A.** [US/US]; Six Winter Hill, Grantham, NH 03753 (US).  
**RAJGARHIA, Vineet, B.** [US/US]; 21 Spencer Street #201, Lebanon, NH 03766 (US).

(74) Agents: **GORDON, Dana, M.** et al.; Foley Hoag LLP, Seaport World Trade Center West, 155 Seaport Blvd., Boston, MA 02210-2600 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Published:

- with international search report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

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

12 March 2009

(54) Title: GENE KNOCKOUT MESOPHILIC AND THERMOPHILIC ORGANISMS, AND METHODS OF USE THEREOF

(57) Abstract: One aspect of the invention relates to a genetically modified thermophilic or mesophilic microorganism, wherein a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain embodiments, the aforementioned microorganism further comprises a first non-native gene, which first non-native gene encodes a first non-native enzyme involved in the metabolic production of ethanol. Another aspect of the invention relates to a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism.



**WO 2008/141174 A3**

---

***Gene Knockout Mesophilic and Thermophilic  
Organisms, and Methods of Use Thereof***

---

RELATED APPLICATIONS

5           This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/916,978, filed May 9, 2007; the entire contents of which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

10           Energy conversion, utilization and access underlie many of the great challenges of our time, including those associated with sustainability, environmental quality, security, and poverty. New applications of emerging technologies are required to respond to these challenges. Biotechnology, one of the most powerful of the emerging technologies, can give rise to important new energy conversion processes. Plant biomass and derivatives thereof are a resource for the biological conversion of energy to forms useful to humanity.

15           Among forms of plant biomass, lignocellulosic biomass ('biomass') is particularly well-suited for energy applications because of its large-scale availability, low cost, and environmentally benign production. In particular, many energy production and utilization cycles based on cellulosic biomass have near-zero greenhouse gas emissions on a life-cycle basis. The primary obstacle impeding the more widespread production of energy from  
20 biomass feedstocks is the general absence of low-cost technology for overcoming the recalcitrance of these materials to conversion into useful fuels. Lignocellulosic biomass contains carbohydrate fractions (*e.g.*, cellulose and hemicellulose) that can be converted into ethanol. In order to convert these fractions, the cellulose and hemicellulose must ultimately be converted or hydrolyzed into monosaccharides; it is the hydrolysis that has  
25 historically proven to be problematic.

Biologically mediated processes are promising for energy conversion, in particular for the conversion of lignocellulosic biomass into fuels. Biomass processing schemes involving enzymatic or microbial hydrolysis commonly involve four biologically mediated transformations: (1) the production of saccharolytic enzymes (cellulases and

hemicellulases); (2) the hydrolysis of carbohydrate components present in pretreated biomass to sugars; (3) the fermentation of hexose sugars (*e.g.*, glucose, mannose, and galactose); and (4) the fermentation of pentose sugars (*e.g.*, xylose and arabinose). These four transformations occur in a single step in a process configuration called consolidated  
5 bioprocessing (CBP), which is distinguished from other less highly integrated configurations in that it does not involve a dedicated process step for cellulase and/or hemicellulase production.

CBP offers the potential for lower cost and higher efficiency than processes featuring dedicated cellulase production. The benefits result in part from avoided capital  
10 costs, substrate and other raw materials, and utilities associated with cellulase production. In addition, several factors support the realization of higher rates of hydrolysis, and hence reduced reactor volume and capital investment using CBP, including enzyme-microbe synergy and the use of thermophilic organisms and/or complexed cellulase systems. Moreover, cellulose-adherent cellulolytic microorganisms are likely to compete  
15 successfully for products of cellulose hydrolysis with non-adhered microbes, *e.g.*, contaminants, which could increase the stability of industrial processes based on microbial cellulose utilization. Progress in developing CBP-enabling microorganisms is being made through two strategies: engineering naturally occurring cellulolytic microorganisms to improve product-related properties, such as yield and titer; and engineering non-cellulolytic  
20 organisms that exhibit high product yields and titers to express a heterologous cellulase and hemicellulase system enabling cellulose and hemicellulose utilization.

Many bacteria have the ability to ferment simple hexose sugars into a mixture of acidic and pH-neutral products via the process of glycolysis. The glycolytic pathway is abundant and comprises a series of enzymatic steps whereby a six carbon glucose molecule  
25 is broken down, via multiple intermediates, into two molecules of the three carbon compound pyruvate. This process results in the net generation of ATP (biological energy supply) and the reduced cofactor NADH.

Pyruvate is an important intermediary compound of metabolism. For example, under aerobic conditions pyruvate may be oxidized to acetyl coenzyme A (acetyl CoA),  
30 which then enters the tricarboxylic acid cycle (TCA), which in turn generates synthetic precursors, CO<sub>2</sub> and reduced cofactors. The cofactors are then oxidized by donating hydrogen equivalents, via a series of enzymatic steps, to oxygen resulting in the formation

of water and ATP. This process of energy formation is known as oxidative phosphorylation.

Under anaerobic conditions (no available oxygen), fermentation occurs in which the degradation products of organic compounds serve as hydrogen donors and acceptors. Excess NADH from glycolysis is oxidized in reactions involving the reduction of organic substrates to products, such as lactate and ethanol. In addition, ATP is regenerated from the production of organic acids, such as acetate, in a process known as substrate level phosphorylation. Therefore, the fermentation products of glycolysis and pyruvate metabolism include a variety of organic acids, alcohols and CO<sub>2</sub>.

The majority of facultative anaerobic bacteria do not produce high yields of ethanol under either aerobic or anaerobic conditions. Most facultative anaerobes metabolize pyruvate aerobically via pyruvate dehydrogenase (PDH) and the tricarboxylic acid cycle (TCA). Under anaerobic conditions, the main energy pathway for the metabolism of pyruvate is via pyruvate-formate-lyase (PFL) pathway to give formate and acetyl-CoA. Acetyl-CoA is then converted to acetate, via phosphotransacetylase (PTA) and acetate kinase (ACK) with the co-production of ATP, or reduced to ethanol via acetaldehyde dehydrogenase (AcDH) and alcohol dehydrogenase (ADH). In order to maintain a balance of reducing equivalents, excess NADH produced from glycolysis is re-oxidized to NAD<sup>+</sup> by lactate dehydrogenase (LDH) during the reduction of pyruvate to lactate. NADH can also be re-oxidized by AcDH and ADH during the reduction of acetyl-CoA to ethanol, but this is a minor reaction in cells with a functional LDH. Theoretical yields of ethanol are therefore not achieved since most acetyl CoA is converted to acetate to regenerate ATP and excess NADH produced during glycolysis is oxidized by LDH.

Metabolic engineering of microorganisms could also result in the creation of a targeted knockout of the genes encoding for the production of enzymes, such as lactate dehydrogenase. In this case, "knock out" of the genes means partial, substantial, or complete deletion, silencing, inactivation, or down-regulation. If the conversion of pyruvate to lactate (the salt form of lactic acid) by the action of LDH was not available in the early stages of the glycolytic pathway, then the pyruvate could be more efficiently converted to acetyl CoA by the action of pyruvate dehydrogenase or pyruvate-ferredoxin oxidoreductase. If the further conversion of acetyl CoA to acetate (the salt form of acetic acid) by phosphotransacetylase and acetate kinase was also not available, *i.e.*, if the genes

encoding for the production of PTA and ACK were knocked out, then the acetyl CoA could be more efficiently converted to ethanol by AcDH and ADH. Accordingly, a genetically modified strain of microorganism with such targeted gene knockouts, which eliminates the production of organic acids, would have an increased ability to produce ethanol as a fermentation product.

Ethanologenic organisms, such as *Zymomonas mobilis*, *Zymobacter palmae*, *Acetobacter pasteurianus*, or *Sarcina ventriculi*, and some yeasts (e.g., *Saccharomyces cerevisiae*), are capable of a second type of anaerobic fermentation, commonly referred to as alcoholic fermentation, in which pyruvate is metabolized to acetaldehyde and CO<sub>2</sub> by pyruvate decarboxylase (PDC). Acetaldehyde is then reduced to ethanol by ADH regenerating NAD<sup>+</sup>. Alcoholic fermentation results in the metabolism of one molecule of glucose to two molecules of ethanol and two molecules of CO<sub>2</sub>. If the conversion of pyruvate to undesired organic acids could be avoided, as detailed above, then such a genetically modified microorganism would have an increased ability to produce ethanol as a fermentation product.

#### SUMMARY OF THE INVENTION

One aspect of the invention relates to an isolated nucleic acid molecule comprising the nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof. Another aspect of the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence which shares at least 80% identity to a nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof. In certain embodiments, the invention relates to the aforementioned nucleic acid molecule which shares at least about 95% sequence identity to the nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof.

Another aspect of the present invention relates to a genetic construct comprising any one of SEQ ID NOS:1-5, 30-31, and 47-61 operably linked to a promoter expressible in a thermophilic or mesophilic bacterium. The present invention also relates to a recombinant thermophilic or mesophilic bacterium comprising the aforementioned genetic construct.

The present invention also encompasses a vector comprising any one of the aforementioned nucleic acid molecules. The present invention also encompasses a host cell comprising any one of the aforementioned nucleic acid molecules. In certain embodiments,

the invention relates to the aforementioned host cell, wherein said host cell is a thermophilic or mesophilic bacterial cell.

Another aspect of the invention relates to a genetically modified thermophilic or mesophilic microorganism, wherein a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a Gram-negative bacterium or a Gram-positive bacterium. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a species of the genera *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, or *Anoxybacillus*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a bacterium selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, and *Anaerocellum thermophilum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Thermoanaerobacterium saccharolyticum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic

microorganism with a native ability to metabolize a hexose sugar; (b) a thermophilic or mesophilic microorganism with a native ability to metabolize a pentose sugar; and (c) a thermophilic or mesophilic microorganism with a native ability to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the  
5 aforementioned genetically modified microorganism, wherein said microorganism has a native ability to metabolize a hexose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Clostridium straminisolvens* or *Clostridium thermocellum*. In certain embodiments, the present invention relates to the aforementioned genetically modified  
10 microorganism, wherein said microorganism has a native ability to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Clostridium cellulolyticum*, *Clostridium kristjanssonii*, or *Clostridium stercorarium subsp. leptosaprartum*. In certain embodiments, the present invention relates to the  
15 aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically  
20 modified microorganism, wherein said microorganism has a native ability to metabolize a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is selected from the group consisting of *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium polysaccharolyticum*, and *Thermoanaerobacterium*  
25 *thermosaccharolyticum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a hexose sugar. In  
30 certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is selected from the group consisting of lactic acid and acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic

acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is lactate dehydrogenase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is acetate kinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a second native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is acetate kinase or phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is lactate dehydrogenase.

Yet another aspect of the invention relates to a genetically modified thermophilic or mesophilic microorganism, wherein (a) a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the metabolic production of ethanol, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or

mesophilic microorganism to metabolize a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH). In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second non-native enzyme is xylose isomerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene corresponds to SEQ ID NOS:6, 10, or 14. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is xylulokinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NOS:7, 11, or 15. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is L-arabinose isomerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NOS:8 or 12. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is L-ribulose-5-phosphate 4-epimerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NO:9 or 13. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said

microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose; (b) a thermophilic or mesophilic microorganism with a native ability to hydrolyze xylan; and (c) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose and xylan. In certain  
5 embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze cellulose and xylan. In certain embodiments, the present invention relates to the  
10 aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze xylan. In certain embodiments, the present invention relates to  
15 the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is selected from the group consisting of lactic acid and acetic acid. In certain embodiments, the  
20 present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is  
25 selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is lactate dehydrogenase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is  
30 acetate kinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a second native gene is

partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said  
5 second native enzyme is acetate kinase or phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is lactate dehydrogenase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein (a) a first native gene is partially, substantially, or completely deleted, silenced,  
10 inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the hydrolysis of a polysaccharide, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain  
15 embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-  
20 native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose; and a second non-native gene is  
25 inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose and xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the  
30 aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH). In certain embodiments, the

present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol.

In certain embodiments, the present invention relates to any of the aforementioned  
5 genetically modified microorganisms, wherein said microorganism is mesophilic. In certain embodiments, the present invention relates to any of the aforementioned genetically modified microorganisms, wherein said microorganism is thermophilic.

Another aspect of the invention relates to a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with any one of the  
10 aforementioned genetically modified thermophilic or mesophilic microorganisms. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls,  
15 barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, and combinations thereof. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is selected from the group consisting of corn stover, sugarcane bagasse,  
20 switchgrass, and poplar wood. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is corn stover. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is sugarcane bagasse. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is  
25 switchgrass. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is poplar wood. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is willow. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is paper sludge.

30

#### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** depicts the glycolysis pathway.

**Figure 2** depicts pentose and glucuronate interconversions and highlights the enzymes, xylose isomerase (XI or 5.3.1.5) and xylulokinase (XK or 2.7.1.17), in the D-xylose to ethanol pathway.

**Figure 3** depicts pentose and glucuronate interconversions and highlights the enzymes, L-arabinose isomerase (5.3.1.4) and L-ribulose-5-phosphate 4-epimerase (5.1.3.4), in the L-arabinose utilization pathway.

**Figure 4** depicts pentose and glucuronate interconversions and shows that the genes for xylose isomerase, xylulokinase, L-arabinose isomerase, and L-ribulose-5-phosphate 4-epimerase are present in *C. cellulolyticum*.

**Figure 5** depicts pentose and glucuronate interconversions and shows that xylose isomerase and xylulokinase are present, while L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase are absent in *C. phytofermentans*.

**Figure 6** shows an alignment of *Clostridium thermocellum* (SEQ ID NO: 77), *Clostridium cellulolyticum* (SEQ ID NO: 78), *Thermoanaerobacterium saccharolyticum* (SEQ ID NO: 79), *C. stercorarium* (SEQ ID NO: 80), *C. stercorarium II* (SEQ ID NO: 81), *Caldiscellulosiruptor kristjanssonii* (SEQ ID NO: 82), *C. phytofermentans* (SEQ ID NO: 83), indicating about 73-89% homology at the level of the 16S rDNA gene.

**Figure 7** shows the construction of a double crossover knockout vector for inactivation of the *ack* gene in *Clostridium thermocellum* based on the plasmid pIKM1.

**Figure 8** shows the construction of a double crossover knockout vector for inactivation of the *ack* gene in *Clostridium thermocellum* based on the replicative plasmid pNW33N.

**Figure 9** shows the construction of a double crossover knockout vector for inactivation of the *ldh* gene in *Clostridium thermocellum* based on the plasmid pIKM1.

**Figure 10** shows the construction of a double crossover knockout vector for inactivation of the *ldh* gene in *Clostridium thermocellum* based on the replicative plasmid vector pNW33N.

**Figure 11** shows the construction of a double crossover suicide vector for inactivation of the *ldh* gene in *Clostridium thermocellum* based on the plasmid pUC19.

**Figures 12A and 12B** show product formation and OD<sub>600</sub> for *C. straminisolvens* grown on cellobiose and Avicel®, respectively.

**Figures 13A and 13B** show product formation and OD<sub>600</sub> for *C. thermocellum* grown on cellobiose and Avicel®, respectively.

5        **Figures 14A and 14B** show product formation and OD<sub>600</sub> for *C. cellulolyticum* grown on cellobiose and Avicel®, respectively.

**Figures 15A and 15B** show product formation and OD<sub>600</sub> for *C. stercorarium* subs. *leptospartum* grown on cellobiose and Avicel®, respectively.

10       **Figures 16A and 16B** show product formation and OD<sub>600</sub> for *Caldicellulosiruptor kristjanssonii* grown on cellobiose and Avicel®, respectively.

**Figures 17A and 17B** show product formation and OD<sub>600</sub> for *Clostridium phytofermentans* grown on cellobiose and Avicel®, respectively.

**Figure 18** shows total metabolic byproducts after 48 hours of fermentation of 2.5 g/L xylan and 2.5 g/L cellobiose.

15       **Figure 19** shows a map of the *ack* gene and the region amplified by PCR for gene disruption.

**Figure 20** shows a map of the *ldh* 2262 gene and the region amplified by PCR for gene disruption.

20       **Figure 21** shows an example of *C. cellulolyticum* (*C. cell.*) *ldh* (2262) double crossover knockout fragment.

**Figure 22** shows a map of the *ack* gene of *Clostridium phytofermentans* and the region amplified by PCR for gene disruption.

**Figure 23** shows an example of a putative double crossover knockout construct with the m<sub>LS</sub> gene as a selectable marker in *Clostridium phytofermentans*.

25       **Figure 24** shows a map of the *ldh* 1389 gene and the region amplified by PCR for gene disruption.

**Figure 25** shows an example of a putative double crossover knockout construct with the m<sub>Ls</sub> gene as a selectable marker.

**Figure 26** is a diagram representing bp 250-550 of pMOD<sup>TM</sup>-2<MCS> (SEQ ID NO: 84).

5 **Figure 27** shows the product concentration profiles for 1% Avicel® using *C. straminisolvens*. The ethanol-to-acetate ratio is depicted as E/A and the ratio of ethanol-to-total products is depicted as E/T.

**Figure 28** shows an example of a vector for retargeting the Ll.LtrB intron to insert in *C. cell*. ACK gene (SEQ ID NO:21).

10 **Figure 29** shows an example of vector for retargeting the Ll.LtrB intron to insert in *C. cell*. LDH2744 gene (SEQ ID NO:23).

**Figure 30** shows an alignment of *T. pseudoethanolicus* 39E (SEQ ID NO: 85), *T. sp* strain 59 (SEQ ID NO: 86), *T. saccharolyticum* B6A-RI (SEQ ID NO: 87), *T. saccharolyticum* YS485 (SEQ ID NO: 88) and consensus (SEQ ID NO: 89) at the level of  
15 the 16S rDNA gene.

**Figure 31** shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 36), *T. pseudoethanolicus* (SEQ ID NO: 35), *T. saccharolyticum* B6A-RI (SEQ ID NO: 38), *T. saccharolyticum* YS485 (SEQ ID NO: 32) and consensus (SEQ ID NO: 90) at the level of the pta gene.

20 **Figure 32** shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 37), *T. pseudoethanolicus* (SEQ ID NO: 34), *T. saccharolyticum* B6A-RI (SEQ ID NO: 39), *T. saccharolyticum* YS485 (SEQ ID NO: 33) and consensus (SEQ ID NO: 91) at the level of the ack gene.

**Figure 33** shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 41), *T. pseudoethanolicus* 39E (SEQ ID NO: 42), *T. saccharolyticum* B6A-RI (SEQ ID NO: 43), *T. saccharolyticum* YS485 (SEQ ID NO: 40) and consensus (SEQ ID NO: 92) at the level  
25 of the ldh gene.

**Figure 34** shows a schematic of the glycolysis/fermentation pathway.

**Figure 35** shows an example of a pMU340 plasmid.

**Figure 36** shows an example of a pMU102 *Z. mobilis* PDC-ADH plasmid.

**Figure 37** shows an example of a pMU102 *Z. palmarum* PDC, *Z. mobilis* ADH plasmid.

5        **Figure 38** shows the plasmid map of pMU360. The DNA sequence of pMU360 is set forth as SEQ ID NO:61.

**Figure 39** shows the lactate levels in nine colonies of thiamphenicol-resistant transformants.

10        **Figure 40** shows an example of a *T. sacch.* pfl KO single crossover plasmid (SEQ ID NO:47).

**Figure 41** shows an example of a *T. sacch.* pfl KO double crossover plasmid (SEQ ID NO:48).

**Figure 42** shows an example of a *C. therm.* pfl KO single crossover plasmid (SEQ ID NO:49).

15        **Figure 43** shows an example of a *C. therm.* pfl KO double crossover plasmid (SEQ ID NO:50).

**Figure 44** shows an example of a *C. phyto.* pfl KO single crossover plasmid (SEQ ID NO:51).

20        **Figure 45** shows an example of a *C. phyto.* pfl KO double crossover plasmid (SEQ ID NO:52).

**Figure 46** shows an example of a *T. sacch.* #59 L-ldh KO single crossover plasmid (SEQ ID NO:53).

**Figure 47** shows an example of a *T. sacch.* #59 L-ldh KO double crossover plasmid (SEQ ID NO:54).

25        **Figure 48** shows an example of a *T. sacch.* #59 pta/ack KO single crossover plasmid (SEQ ID NO:55).

**Figure 49** shows an example of a *T. sacch.* #59 pta/ack KO double crossover plasmid (SEQ ID NO:56).

**Figure 50** shows an example of a *T. pseudo.* L-ldh KO single crossover plasmid (SEQ ID NO:57).

5 **Figure 51** shows an example of a *T. pseudo.* L-ldh KO double crossover plasmid (SEQ ID NO:58).

**Figure 52** shows an example of a *T. pseudo.* ack KO single crossover plasmid (SEQ ID NO:59).

10 **Figure 53** shows an example of a *T. pseudo.* pta/ack KO double crossover plasmid (SEQ ID NO:60).

#### BRIEF DESCRIPTION OF THE TABLES

**Table 1** summarizes representative highly cellulolytic organisms.

**Table 2** summarizes representative native cellulolytic and xylanolytic organisms.

15 **Table 3** shows a categorization of bacterial strains based on their substrate utilization.

**Table 4** shows insertion location and primers to retarget Intron to *C. cellulolyticum* acetate kinase.

**Table 5** shows insertion location and primers to retarget Intron to *C. cellulolyticum* lactate dehydrogenase.

20 **Table 6** shows fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains.

#### DETAILED DESCRIPTION OF THE INVENTION

25 Aspects of the present invention relate to the engineering of thermophilic or mesophilic microorganisms for use in the production of ethanol from lignocellulosic biomass. The use of thermophilic bacteria for ethanol production offers many advantages over traditional processes based upon mesophilic ethanol producers. For example, the use

of thermophilic organisms provides significant economic savings over traditional process methods due to lower ethanol separation costs, reduced requirements for external enzyme addition, and reduced processing times.

Aspects of the present invention relate to a process by which the cost of ethanol  
5 production from cellulosic biomass-containing materials can be reduced by using a novel processing configuration. In particular, the present invention provides numerous methods for increasing ethanol production in a genetically modified microorganism.

In certain other embodiments, the present invention relates to genetically modified thermophilic or mesophilic microorganisms, wherein a gene or a particular polynucleotide  
10 sequence is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which gene or polynucleotide sequence encodes for an enzyme that confers upon the microorganism the ability to produce organic acids as fermentation products, thereby increasing the ability of the microorganism to produce ethanol as the major fermentation product. Further, by virtue of a novel integration of processing steps, commonly known as  
15 consolidated bioprocessing, aspects of the present invention provide for more efficient production of ethanol from cellulosic-biomass-containing raw materials. The incorporation of genetically modified thermophilic or mesophilic microorganisms in the processing of said materials allows for fermentation steps to be conducted at higher temperatures, improving process economics. For example, reaction kinetics are typically proportional to  
20 temperature, so higher temperatures are generally associated with increases in the overall rate of production. Additionally, higher temperature facilitates the removal of volatile products from the broth and reduces the need for cooling after pretreatment.

In certain embodiments, the present invention relates to genetically modified or recombinant thermophilic or mesophilic microorganisms with increased ability to produce  
25 enzymes that confer the ability to produce ethanol as a fermentation product, the presence of which enzyme(s) modify the process of metabolizing lignocellulosic biomass materials to produce ethanol as the major fermentation product. In one aspect of the invention, one or more non-native genes are inserted into a genetically modified thermophilic or mesophilic microorganism, wherein said non-native gene encodes an enzyme involved in the metabolic  
30 production of ethanol, for example, such enzyme may confer the ability to metabolize a pentose sugar and/or a hexose sugar. For example, in one embodiment, the enzyme may be involved in the D-xylose or L-arabinose pathway, thereby allowing the microorganism to

metabolize a pentose sugar, *i.e.*, D-xylose or L-arabinose. By inserting (*e.g.*, introducing or adding) a non-native gene that encodes an enzyme involved in the metabolism or utilization of D-xylose or L-arabinose, the microorganism has an increased ability to produce ethanol relative to the native organism.

5           The present invention also provides novel compositions that may be integrated into the microorganisms of the invention. In one embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of a nucleotide sequence shown in any one of SEQ ID NOS:1-76. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a  
10 complement of a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, is one which is sufficiently complementary to a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, such that it can hybridize to a nucleotide  
15 sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%,  
20 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the nucleotide sequences (*e.g.*, to the entire length of the nucleotide sequence) shown in any one of SEQ ID NOS:1-76, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecules of the invention may comprise only a portion of the nucleic acid sequence of any one of SEQ ID NOS:1-76, or the coding region thereof;  
25 for example, the nucleic acid molecule may be a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a protein. In another embodiment, the nucleic acid molecules may comprise at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of any one of SEQ ID NOS:1-76.

30   *Definitions*

The term “heterologous polynucleotide segment” is intended to include a polynucleotide segment that encodes one or more polypeptides or portions or fragments of polypeptides. A heterologous polynucleotide segment may be derived from any source, *e.g.*, eukaryotes, prokaryotes, viruses, or synthetic polynucleotide fragments.

5           The terms “promoter” or “surrogate promoter” is intended to include a polynucleotide segment that can transcriptionally control a gene-of-interest that it does not transcriptionally control in nature. In certain embodiments, the transcriptional control of a surrogate promoter results in an increase in expression of the gene-of-interest. In certain  
10           embodiments, a surrogate promoter is placed 5’ to the gene-of-interest. A surrogate promoter may be used to replace the natural promoter, or may be used in addition to the natural promoter. A surrogate promoter may be endogenous with regard to the host cell in which it is used, or it may be a heterologous polynucleotide sequence introduced into the host cell, *e.g.*, exogenous with regard to the host cell in which it is used.

          The terms “gene(s)” or “polynucleotide segment” or “polynucleotide sequence(s)”  
15           are intended to include nucleic acid molecules, *e.g.*, polynucleotides which include an open reading frame encoding a polypeptide, and can further include non-coding regulatory sequences, and introns. In addition, the terms are intended to include one or more genes that map to a functional locus. In addition, the terms are intended to include a specific gene for a selected purpose. The gene may be endogenous to the host cell or may be  
20           recombinantly introduced into the host cell, *e.g.*, as a plasmid maintained episomally or a plasmid (or fragment thereof) that is stably integrated into the genome. In addition to the plasmid form, a gene may, for example, be in the form of linear DNA. In certain embodiments, the gene of polynucleotide segment is involved in at least one step in the bioconversion of a carbohydrate to ethanol. Accordingly, the term is intended to include  
25           any gene encoding a polypeptide, such as the enzymes acetate kinase (ACK), phosphotransacetylase (PTA), and/or lactate dehydrogenase (LDH), enzymes in the D-xylose pathway, such as xylose isomerase and xylulokinase, enzymes in the L-arabinose pathway, such as L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase. The term gene is also intended to cover all copies of a particular gene, *e.g.*, all of the DNA  
30           sequences in a cell encoding a particular gene product.

The term “transcriptional control” is intended to include the ability to modulate gene expression at the level of transcription. In certain embodiments, transcription, and thus

gene expression, is modulated by replacing or adding a surrogate promoter near the 5' end of the coding region of a gene-of-interest, thereby resulting in altered gene expression. In certain embodiments, the transcriptional control of one or more gene is engineered to result in the optimal expression of such genes, *e.g.*, in a desired ratio. The term also includes  
5 inducible transcriptional control as recognized in the art.

The term "expression" is intended to include the expression of a gene at least at the level of mRNA production.

The term "expression product" is intended to include the resultant product, *e.g.*, a polypeptide, of an expressed gene.

10 The term "increased expression" is intended to include an alteration in gene expression at least at the level of increased mRNA production and, preferably, at the level of polypeptide expression. The term "increased production" is intended to include an increase in the amount of a polypeptide expressed, in the level of the enzymatic activity of the polypeptide, or a combination thereof.

15 The terms "activity," "activities," "enzymatic activity," and "enzymatic activities" are used interchangeably and are intended to include any functional activity normally attributed to a selected polypeptide when produced under favorable conditions. Typically, the activity of a selected polypeptide encompasses the total enzymatic activity associated with the produced polypeptide. The polypeptide produced by a host cell and having  
20 enzymatic activity may be located in the intracellular space of the cell, cell-associated, secreted into the extracellular milieu, or a combination thereof. Techniques for determining total activity as compared to secreted activity are described herein and are known in the art.

The term "xylanolytic activity" is intended to include the ability to hydrolyze glycosidic linkages in oligopentoses and polypentoses.

25 The term "cellulolytic activity" is intended to include the ability to hydrolyze glycosidic linkages in oligohexoses and polyhexoses. Cellulolytic activity may also include the ability to depolymerize or debranch cellulose and hemicellulose.

As used herein, the term “lactate dehydrogenase” or “LDH” is intended to include the enzyme capable of converting pyruvate into lactate. It is understood that LDH can also catalyze the oxidation of hydroxybutyrate.

As used herein the term “alcohol dehydrogenase” or “ADH” is intended to include  
5 the enzyme capable of converting acetaldehyde into an alcohol, advantageously, ethanol.

The term “pyruvate decarboxylase activity” is intended to include the ability of a polypeptide to enzymatically convert pyruvate into acetaldehyde (*e.g.*, “pyruvate decarboxylase” or “PDC”). Typically, the activity of a selected polypeptide encompasses the total enzymatic activity associated with the produced polypeptide, comprising, *e.g.*, the  
10 superior substrate affinity of the enzyme, thermostability, stability at different pHs, or a combination of these attributes.

The term “ethanologenic” is intended to include the ability of a microorganism to produce ethanol from a carbohydrate as a fermentation product. The term is intended to include, but is not limited to, naturally occurring ethanologenic organisms, ethanologenic  
15 organisms with naturally occurring or induced mutations, and ethanologenic organisms which have been genetically modified.

The terms “fermenting” and “fermentation” are intended to include the enzymatic process (*e.g.*, cellular or acellular, *e.g.*, a lysate or purified polypeptide mixture) by which ethanol is produced from a carbohydrate, in particular, as a product of fermentation.

20 The term “secreted” is intended to include the movement of polypeptides to the periplasmic space or extracellular milieu. The term “increased secretion” is intended to include situations in which a given polypeptide is secreted at an increased level (*i.e.*, in excess of the naturally-occurring amount of secretion). In certain embodiments, the term “increased secreted” refers to an increase in secretion of a given polypeptide that is at least  
25 about 10% or at least about 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, as compared to the naturally-occurring level of secretion.

The term “secretory polypeptide” is intended to include any polypeptide(s), alone or in combination with other polypeptides, that facilitate the transport of another polypeptide from the intracellular space of a cell to the extracellular milieu. In certain embodiments,  
30 the secretory polypeptide(s) encompass all the necessary secretory polypeptides sufficient

to impart secretory activity to a Gram-negative or Gram-positive host cell. Typically, secretory proteins are encoded in a single region or locus that may be isolated from one host cell and transferred to another host cell using genetic engineering. In certain embodiments, the secretory polypeptide(s) are derived from any bacterial cell having secretory activity. In  
5 certain embodiments, the secretory polypeptide(s) are derived from a host cell having Type II secretory activity. In certain embodiments, the host cell is a thermophilic bacterial cell.

The term “derived from” is intended to include the isolation (in whole or in part) of a polynucleotide segment from an indicated source or the purification of a polypeptide from an indicated source. The term is intended to include, for example, direct cloning, PCR  
10 amplification, or artificial synthesis from or based on a sequence associated with the indicated polynucleotide source.

By “thermophilic” is meant an organism that thrives at a temperature of about 45°C or higher.

By “mesophilic” is meant an organism that thrives at a temperature of about 20-  
15 45°C.

The term “organic acid” is art-recognized. The term “lactic acid” refers to the organic acid 2-hydroxypropionic acid in either the free acid or salt form. The salt form of lactic acid is referred to as “lactate” regardless of the neutralizing agent, *i.e.*, calcium carbonate or ammonium hydroxide. The term “acetic acid” refers to the organic acid  
20 methanecarboxylic acid, also known as ethanoic acid, in either free acid or salt form. The salt form of acetic acid is referred to as “acetate.”

Certain embodiments of the present invention provide for the “insertion,” (*e.g.*, the addition, integration, incorporation, or introduction) of certain genes or particular polynucleotide sequences within thermophilic or mesophilic microorganisms, which  
25 insertion of genes or particular polynucleotide sequences may be understood to encompass “genetic modification(s)” or “transformation(s)” such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be “genetically modified” or “transformed.” In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

Certain embodiments of the present invention provide for the “inactivation” or “deletion” of certain genes or particular polynucleotide sequences within thermophilic or mesophilic microorganisms, which “inactivation” or “deletion” of genes or particular polynucleotide sequences may be understood to encompass “genetic modification(s)” or “transformation(s)” such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be “genetically modified” or “transformed.” In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

The term “CBP organism” is intended to include microorganisms of the invention, *e.g.*, microorganisms that have properties suitable for CBP.

In one aspect of the invention, the genes or particular polynucleotide sequences are inserted to activate the activity for which they encode, such as the expression of an enzyme. In certain embodiments, genes encoding enzymes in the metabolic production of ethanol, *e.g.*, enzymes that metabolize pentose and/or hexose sugars, may be added to a mesophilic or thermophilic organism. In certain embodiments of the invention, the enzyme may confer the ability to metabolize a pentose sugar and be involved, for example, in the D-xylose pathway and/or L-arabinose pathway.

In one aspect of the invention, the genes or particular polynucleotide sequences are partially, substantially, or completely deleted, silenced, inactivated, or down-regulated in order to inactivate the activity for which they encode, such as the expression of an enzyme. Deletions provide maximum stability because there is no opportunity for a reverse mutation to restore function. Alternatively, genes can be partially, substantially, or completely deleted, silenced, inactivated, or down-regulated by insertion of nucleic acid sequences that disrupt the function and/or expression of the gene (*e.g.*, P1 transduction or other methods known in the art). The terms “eliminate,” “elimination,” and “knockout” are used interchangeably with the term “deletion.” In certain embodiments, strains of thermophilic or mesophilic microorganisms of interest may be engineered by site directed homologous recombination to knockout the production of organic acids. In still other embodiments, RNAi or antisense DNA (asDNA) may be used to partially, substantially, or completely silence, inactivate, or down-regulate a particular gene of interest.

In certain embodiments, the genes targeted for deletion or inactivation as described herein may be endogenous to the native strain of the microorganism, and may thus be

understood to be referred to as “native gene(s)” or “endogenous gene(s).” An organism is in “a native state” if it has not been genetically engineered or otherwise manipulated by the hand of man in a manner that intentionally alters the genetic and/or phenotypic constitution of the organism. For example, wild-type organisms may be considered to be in a native  
5 state. In other embodiments, the gene(s) targeted for deletion or inactivation may be non-native to the organism.

### *Biomass*

The terms “lignocellulosic material,” “lignocellulosic substrate,” and “cellulosic biomass” mean any type of biomass comprising cellulose, hemicellulose, lignin, or  
10 combinations thereof, such as but not limited to woody biomass, forage grasses, herbaceous energy crops, non-woody-plant biomass, agricultural wastes and/or agricultural residues, forestry residues and/or forestry wastes, paper-production sludge and/or waste paper sludge, waste-water-treatment sludge, municipal solid waste, corn fiber from wet and dry mill corn ethanol plants, and sugar-processing residues.

15 In a non-limiting example, the lignocellulosic material can include, but is not limited to, woody biomass, such as recycled wood pulp fiber, sawdust, hardwood, softwood, and combinations thereof; grasses, such as switch grass, cord grass, rye grass, reed canary grass, miscanthus, or a combination thereof; sugar-processing residues, such as but not limited to sugar cane bagasse; agricultural wastes, such as but not limited to rice  
20 straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, and corn fiber; stover, such as but not limited to soybean stover, corn stover; and forestry wastes, such as but not limited to recycled wood pulp fiber, sawdust, hardwood (e.g., poplar, oak, maple, birch, willow), softwood, or any combination thereof. Lignocellulosic material may comprise one species of fiber; alternatively, lignocellulosic  
25 material may comprise a mixture of fibers that originate from different lignocellulosic materials. Particularly advantageous lignocellulosic materials are agricultural wastes, such as cereal straws, including wheat straw, barley straw, canola straw and oat straw; corn fiber; stovers, such as corn stover and soybean stover; grasses, such as switch grass, reed canary grass, cord grass, and miscanthus; or combinations thereof.

30 Paper sludge is also a viable feedstock for ethanol production. Paper sludge is solid residue arising from pulping and paper-making, and is typically removed from process

wastewater in a primary clarifier. At a disposal cost of \$30/wet ton, the cost of sludge disposal equates to \$5/ton of paper that is produced for sale. The cost of disposing of wet sludge is a significant incentive to convert the material for other uses, such as conversion to ethanol. Processes provided by the present invention are widely applicable. Moreover, the  
5 saccharification and/or fermentation products may be used to produce ethanol or higher value added chemicals, such as organic acids, aromatics, esters, acetone and polymer intermediates.

*Pyruvate formate lyase (PFL)*

Pyruvate formate lyase (PFL) is an important enzyme (found in *Escherichia coli* and  
10 other organisms) that helps regulate anaerobic glucose metabolism. Using radical chemistry, it catalyzes the reversible conversion of pyruvate and coenzyme-A into formate and acetyl-CoA, a precursor of ethanol. Pyruvate formate lyase is a homodimer made of 85 kDa, 759-residue subunits. It has a 10-stranded beta/alpha barrel motif into which is inserted a beta finger that contains major catalytic residues. The active site of the enzyme,  
15 elucidated by x-ray crystallography, holds three essential amino acids that perform catalysis (Gly734, Cys418, and Cys419), three major residues that hold the substrate pyruvate close by (Arg435, Arg176, and Ala272), and two flanking hydrophobic residues (Trp333 and Phe432).

Studies have found structural similarities between the active site of pyruvate  
20 formate lyase and that of Class I and Class III ribonucleotide reductase (RNR) enzymes. The roles of the 3 catalytic residues are as follows: Gly734 (glycyl radical) – transfers the radical on and off Cys418, via Cys419; Cys418 (thiyl radical) – performs acylation chemistry on the carbon atom of the pyruvate carbonyl; Cys419 (thiyl radical) – performs hydrogen-atom transfers.

25 The proposed mechanism for pyruvate formate lyase begins with radical transfer from Gly734 to Cys418, via Cys419. The Cys418 thiyl radical adds covalently to C2 (second carbon atom) of pyruvate, generating an acetyl-enzyme intermediate (which now contains the radical). The acetyl-enzyme intermediate releases a formyl radical that undergoes hydrogen-atom transfer with Cys419. This generates formate and a Cys419  
30 radical. Coenzyme-A undergoes hydrogen-atom transfer with the Cys419 radical to generate a coenzyme-A radical. The coenzyme-A radical then picks up the acetyl group

from Cys418 to generate acetyl-CoA, leaving behind a Cys418 radical. Pyruvate formate lyase can then undergo radical transfer to put the radical back onto Gly734. Each of the above mentioned steps are also reversible.

Two additional enzymes regulate the “on” and “off” states of pyruvate formate lyase to regulate anaerobic glucose metabolism: PFL activase (AE) and PFL deactivase (DA). Activated pyruvate formate lyase allows formation of acetyl-CoA, a small molecule important in the production of energy, when pyruvate is available. Deactivated pyruvate formate lyase, even with substrates present, does not catalyze the reaction. PFL activase is part of the radical SAM (S-adenosylmethionine) superfamily.

The enzyme turns pyruvate formate lyase “on” by converting Gly734 (G-H) into a Gly734 radical (G\*) via a 5'-deoxyadenosyl radical (radical SAM). PFL deactivase (DA) turns pyruvate formate lyase “off” by quenching the Gly734 radical. Furthermore, pyruvate formate lyase is sensitive to molecular oxygen (O<sub>2</sub>), the presence of which shuts the enzyme off.

#### *Xylose metabolism*

Xylose is a five-carbon monosaccharide that can be metabolized into useful products by a variety of organisms. There are two main pathways of xylose metabolism, each unique in the characteristic enzymes they utilize. One pathway is called the “Xylose Reductase-Xylitol Dehydrogenase” or XR-XDH pathway. Xylose reductase (XR) and xylitol dehydrogenase (XDH) are the two main enzymes used in this method of xylose degradation. XR, encoded by the XYL1 gene, is responsible for the reduction of xylose to xylitol and is aided by cofactors NADH or NADPH. Xylitol is then oxidized to xylulose by XDH, which is expressed through the XYL2 gene, and accomplished exclusively with the cofactor NAD<sup>+</sup>. Because of the varying cofactors needed in this pathway and the degree to which they are available for usage, an imbalance can result in an overproduction of xylitol byproduct and an inefficient production of desirable ethanol. Varying expression of the XR and XDH enzyme levels have been tested in the laboratory in the attempt to optimize the efficiency of the xylose metabolism pathway.

The other pathway for xylose metabolism is called the “Xylose Isomerase” (XI) pathway. Enzyme XI is responsible for direct conversion of xylose into xylulose, and does not proceed via a xylitol intermediate. Both pathways create xylulose, although the

enzymes utilized are different. After production of xylulose both the XR-XDH and XI pathways proceed through enzyme xylulokinase (XK), encoded on gene XKS1, to further modify xylulose into xylulose-5-P where it then enters the pentose phosphate pathway for further catabolism.

5           Studies on flux through the pentose phosphate pathway during xylose metabolism have revealed that limiting the speed of this step may be beneficial to the efficiency of fermentation to ethanol. Modifications to this flux that may improve ethanol production include a) lowering phosphoglucose isomerase activity, b) deleting the GND1 gene, and c) deleting the ZWF1 gene (Jeppsson et al., 2002). Since the pentose phosphate pathway  
10 produces additional NADPH during metabolism, limiting this step will help to correct the already evident imbalance between NAD(P)H and NAD<sup>+</sup> cofactors and reduce xylitol byproduct. Another experiment comparing the two xylose metabolizing pathways revealed that the XI pathway was best able to metabolize xylose to produce the greatest ethanol yield, while the XR-XDH pathway reached a much faster rate of ethanol production  
15 (Karhumaa et al., 2007).

### *Microorganisms*

The present invention includes multiple strategies for the development of microorganisms with the combination of substrate-utilization and product-formation properties required for CBP. The “native cellulolytic strategy” involves engineering  
20 naturally occurring cellulolytic microorganisms to improve product-related properties, such as yield and titer. The “recombinant cellulolytic strategy” involves engineering natively non-cellulolytic organisms that exhibit high product yields and titers to express a heterologous cellulase system that enables cellulose utilization or hemicellulose utilization or both.

### 25 *Cellulolytic Microorganisms*

Several microorganisms reported in the literature to be cellulolytic or have cellulolytic activity have been characterized by a variety of means, including their ability to grow on microcrystalline cellulose as well as a variety of other sugars. Additionally, the organisms may be characterized by other means, including but not limited to, their ability to  
30 depolymerize and debranch cellulose and hemicellulose. *Clostridium thermocellum* (strain DSMZ 1237) was used to benchmark the organisms of interest. As used herein, *C.*

*thermocellum* may include various strains, including, but not limited to, DSMZ 1237, DSMZ 1313, DSMZ 2360, DSMZ 4150, DSMZ 7072, and ATCC 31924. In certain embodiments of the invention, the strain of *C. thermocellum* may include, but is not limited to, DSMZ 1313 or DSMZ 1237. In another embodiment, particularly suitable organisms of interest for use in the present invention include cellulolytic microorganisms with a greater than 70% 16S rDNA homology to *C. thermocellum*. Alignment of *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Thermoanaerobacterium saccharolyticum*, *C. stercorarium*, *C. stercorarium II*, *Caldicellulosiruptor kristjanssonii*, *C. phytofermentans* indicate a 73 – 85% homology at the level of the 16S rDNA gene (Figure 6).

*Clostridium straminisolvens* has been determined to grow nearly as well as *C. thermocellum* on Avicel®. Table 1 summarizes certain highly cellulolytic organisms.

Table 1

| Strain                             | DSMZ No. | T optimum; or range | pH optimum; or range | Gram Stain | Aero-tolerant | Utilizes              | Products                                                                 |
|------------------------------------|----------|---------------------|----------------------|------------|---------------|-----------------------|--------------------------------------------------------------------------|
| <i>Clostridium thermocellum</i>    | 1313     | 55-60               | 7                    | positive   | No            | cellobiose, cellulose | acetic acid<br>lactic acid<br>ethanol, H <sub>2</sub><br>CO <sub>2</sub> |
| <i>Clostridium straminisolvens</i> | 16021    | 50-55;<br>45-60     | 6.5-6.8;<br>6.0-8.5  | positive   | Yes           | cellobiose, cellulose | acetic acid<br>lactic acid<br>ethanol, H <sub>2</sub><br>CO <sub>2</sub> |

Organisms were grown on 20 g/L cellobiose or 20 g/L Avicel®. *C. thermocellum* was grown at 60°C and *C. straminisolvens* was grown at 55°C. Both were pre-cultured from -80°C freezer stock (origin DSMZ) on M122 with 50mM MOPS. During mid to late log growth phase pre-cultures were used to inoculate the batch cultures in 100 mL serum bottles to a working volume of 50 mL. Liquid samples were removed periodically for HPLC analysis of metabolic byproducts and sugar consumption. OD<sub>600</sub> was taken at each of these time points. Figures 12A and 12B show product formation and OD<sub>600</sub> for *C. straminisolvens* on cellobiose and Avicel®, respectively. Substantial cellobiose (37%) was consumed with 48 hours before OD dropped and product formation leveled off. Figures 13A and 13B show product formation and OD<sub>600</sub> for *C. thermocellum* on cellobiose and Avicel®, respectively. *C. thermocellum* consumed ~60% of cellobiose within 48 hours, at which point product formation leveled out. Inhibition due to formation of organic acids caused incomplete utilization of substrates.

Certain microorganisms, including, for example, *C. thermocellum* and *C. straminisolvens*, cannot metabolize pentose sugars, such as D-xylose or L-arabinose, but are able to metabolize hexose sugars. Both D-xylose and L-arabinose are abundant sugars in biomass with D-xylose accounting for approximately 16 – 20% in soft and hard woods and L-arabinose accounting for approximately 25% in corn fiber. Accordingly, one object of the invention is to provide genetically-modified cellulolytic microorganisms, with the ability to metabolize pentose sugars, such as D-xylose and L-arabinose, thereby to enhance their use as biocatalysts for fermentation in the biomass-to-ethanol industry.

### *Cellulolytic and Xylanolytic Microorganisms*

Several microorganisms determined from literature to be both cellulolytic and xylanolytic have been characterized by their ability to grow on microcrystalline cellulose and birchwood xylan as well as a variety of other sugars. *Clostridium thermocellum* was used to benchmark the organisms of interest. Of the strains selected for characterization *Clostridium cellulolyticum*, *Clostridium stercorarium* subs. *leptospartum*, *Caldicellulosiruptor kristjanssonii* and *Clostridium phytofermentans* grew weakly on Avicel® and well on birchwood xylan. Table 2 summarizes some of the native cellulolytic and xylanolytic organisms.

**Table 2**

| Strain                                                    | Source/<br>No. | T<br>optimum;<br>or range | pH<br>optimum;<br>or range | Gram Stain | Aero-<br>tolerant | Utilizes                                                                                | Products                                                            |
|-----------------------------------------------------------|----------------|---------------------------|----------------------------|------------|-------------------|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| <i>Clostridium cellulolyticum</i>                         | DSM 5812       | 34                        | 7.2                        | negative   | no                | Cellulose, xylan, arabinose, mannose, galactose, xylose, glucose, cellobiose            | acetic acid, lactic acid, ethanol, H <sub>2</sub> , CO <sub>2</sub> |
| <i>Clostridium stercorarium</i> subs. <i>leptospartum</i> | DSM 9219       | 60-65                     | 7.0-7.5                    | negative   | no                | Cellulose, cellobiose, lactose, xylose, melibiose, raffinose, ribose, fructose, sucrose | acetic acid, lactic acid, ethanol, H <sub>2</sub> , CO <sub>2</sub> |
| <i>Caldicellulosiruptor kristjanssonii</i>                | DSM 12137      | 78; 45-82                 | 7; 5.8-8.0                 | negative   | No                | cellobiose, glucose, xylose, galactose, mannose, cellulose                              | acetic acid, CO <sub>2</sub> , lactate, acetate, ethanol, formate   |

|                                        |                |            |            |                                        |    |                                                                                                                               |                                                                                               |
|----------------------------------------|----------------|------------|------------|----------------------------------------|----|-------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| <i>Clostridium<br/>phytofermentans</i> | ATCC<br>700394 | 37; 5 - 45 | 8.5; 6 - 9 | Negative<br>(gram<br>type<br>positive) | no | Cellulose,<br>xylan,<br>cellobiose,<br>fructose,<br>galactose,<br>glucose, lactose,<br>maltose,<br>mannose,<br>ribose, xylose | acetic acid,<br>H <sub>2</sub> , CO <sub>2</sub> ,<br>lactic acid,<br>ethanol,<br>formic acid |
|----------------------------------------|----------------|------------|------------|----------------------------------------|----|-------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|

Organisms were grown on 20 g/L cellobiose, 20 g/L Avicel® or 5 g/L birchwood xylan. *C. cellulolyticum* was grown at 37°C, *C. stercorarium* subs. *leptospartum* was  
5 grown at 60°C, *Caldicellulosiruptor kristjanssonii* was grown at 75°C and *Clostridium phytofermentans* was grown at 37°C. All were pre-cultured from -80°C freezer stock in M122c supplemented with 50mM MOPS. During mid to late log growth phase pre-cultures were used to inoculate the batch cultures in 100 mL serum bottles to a working volume of 50 mL. Liquid samples were removed periodically for HPLC analysis of metabolic  
10 byproducts and sugar consumption. OD<sub>600</sub> was taken at each of these time points. Figures 14A-17B show product formation and OD<sub>600</sub> for growth on cellobiose and Avicel®.

In a separate experiment organisms were grown on 2.5 g/L single sugars including cellobiose, glucose, xylose, galactose, arabinose, mannose and lactose as well as 5 g/L Avicel® and birchwood xylan. In Figure 18 product formation is compared on cellobiose  
15 and birchwood xylan after two days. Table 3 summarizes how bacterial strains may be categorized based on their substrate utilization.

**Table 3**

|                                                    | cellobiose | glucose | xylose | galactose | arabinose | mannose | lac |
|----------------------------------------------------|------------|---------|--------|-----------|-----------|---------|-----|
| <i>C.cellulolyticum</i>                            | ×          | ×       | ×      | ×         | ×         |         |     |
| <i>C.stercorarium</i><br><i>subs. leptospartum</i> | ×          | ×       | ×      | ×         | ×         | ×       |     |
| <i>C.kristjanssonii</i>                            | ×          | ×       | ×      | ×         |           | ×       |     |
| <i>C. phytofermentans</i>                          | ×          | ×       | ×      | ×         |           | ×       |     |

## 20 Transgenic Conversion of Microorganisms

The present invention provides compositions and methods for the transgenic conversion of certain microorganisms. When genes encoding enzymes involved in the metabolic pathway of ethanol, including, for example, D-xylose and/or L-arabinose, are introduced into a bacterial strain that lacks one or more of these genes, for example, *C.*

*thermocellum* or *C. straminisolvans*, one may select transformed strains for growth on D-xylose or growth on L-arabinose. It is expected that genes from other Clostridial species should be expressed in *C. thermocellum* and *C. straminisolvans*. Target gene donors may include microorganisms that confer the ability to metabolize hexose and pentose sugars,  
5 *e.g.*, *C. cellulolyticum*, *Caldicellulosiruptor kristjanssonii*, *C. phytofermentans*, *C. stercorarium*, and *Thermoanaerobacterium saccharolyticum*.

The genomes of *T. saccharolyticum*, *C. cellulolyticum*, and *C. phytofermentans* are available. Accordingly, the present invention provides sequences which correspond to xylose isomerase and xylulokinase in each of the three hosts set forth above. In particular,  
10 the sequences corresponding to xylose isomerase (SEQ ID NO:6), xylulokinase (SEQ ID NO:7), L-arabinose isomerase (SEQ ID NO:8), and L-ribulose-5-phosphate 4-epimerase (SEQ ID NO:9) from *T. saccharolyticum* are set forth herein. Similarly, the sequences corresponding to xylose isomerase (SEQ ID NO:10), xylulokinase (SEQ ID NO:11), L-arabinose isomerase (SEQ ID NO:12), and L-ribulose-5-phosphate 4-epimerase (SEQ ID  
15 NO:13) from *C. cellulolyticum* are provided herein. *C. phytofermentans* utilizes the D-xylose pathway and does not utilize L-arabinose. Accordingly, the sequences corresponding to xylose isomerase (SEQ ID NO:14) and xylulokinase (SEQ ID NO:15) from *C. phytofermentans* are set forth herein.

*C. kristjanssonii* does metabolize xylose. To this end, the xylose isomerase (SEQ  
20 ID NO:71) and xylulokinase (SEQ ID NO:70) genes of *C. kristjanssonii* have been sequenced and are provided herein. *C. straminisolvans* has not been shown to grow on xylose, however it does contain xylose isomerase (SEQ ID NO:73) and xylulokinase (SEQ ID NO:72) genes, which may be functional after adaptation on xylose as a carbon source.

*C. thermocellum* and *C. straminisolvans* may lack one or more known genes or  
25 enzymes in the D-xylose to ethanol pathway and/or the L-arabinose utilization pathway. Figures 2 and 3 depict two key enzymes that are missing in each of these pathways in *C. thermocellum*. *C. straminisolvans* has xylose isomerase and xylulokinase, but the functionality of these enzymes is not known. Genomic sequencing has not revealed a copy of either L-arabinose isomerase or L-ribulose-5-phosphate 4-epimerase in *C.*  
30 *straminisolvans*.

*C. thermocellum* and *C. straminsolvens* are unable to metabolize xylulose which could reflect the absence (*C. thermocellum*) or lack of activity and/or expression (*C. straminsolvens*) of genes for xylose isomerase (referred to in Figure 2 as “XI” or 5.3.1.5), which converts D-xylose to D-xylulose, and xylulokinase (also referred to in Figure 2 as “XK” or 2.7.1.1), which converts D-xylulose to D-xylulose-5-phosphate. Furthermore, transport of xylose may be a limitation for *C. straminsolvens*. This potential limitation could be overcome by expression sugar transport genes from xylose utilizing organisms such as *T. saccharolyticum* and *C. kristjanssonii*.

*C. thermocellum* and *C. straminsolvens* are also unable to metabolize L-arabinose which could reflect the absence of genes for L-arabinose isomerase (also referred to in Figure 3 as 5.3.1.4) and L-ribulose-5-phosphate 4-epimerase (also referred to in Figure 3 as 5.1.3.4).

The four genes described above, e.g., xylose isomerase, xylulokinase, L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase, are present in several Clostridial species and *Thermoanaerobacterium saccharolyticum* species, including, but not limited to, *Clostridium cellulolyticum* (see Figure 4), *Thermoanaerobacterium saccharolyticum*, *C. stercorarium*, *Caldicellulosiruptor kristjanssonii*, and *C. phytofermentans*; these strains are good utilizers of these sugars. It will be appreciated that the foregoing bacterial strains may be used as donors of the genes described herein.

*C. phytofermentans* express the two xylose pathway genes described above (xylose isomerase and xylulokinase), but lack or do not express the arabinose pathway genes described above (L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase ) (see Figure 5).

Accordingly, it is an object of the invention to modify some of the above-described bacterial strains so as to optimize sugar utilization capability by, for example, introducing genes for one or more enzymes required for the production of ethanol from biomass-derived pentoses, e.g., D-xylose or L-arabinose metabolism. Promoters, including the native promoters of *C. thermocellum* or *C. straminsolvens*, such as triose phosphate isomerase (TPI), GAPDH, and LDH, may be used to express these genes. The sequences that correspond to native promoters of *C. thermocellum* include (TPI) (SEQ ID NO:16), GAPDH (SEQ ID NO:17), and LDH (SEQ ID NO:18). Once the gene has been cloned,

codon optimization may be performed before expression. Cassettes containing, for example, the native promoter, a xylanolytic gene or arabinolytic gene, and a selectable marker may then be used to transform *C. thermocellum* or *C. straminisolvens* and select for D-xylose and L-arabinose growth on medium containing D-xylose or L-arabinose as the  
5 sole carbohydrate source.

### *Transposons*

To select for foreign DNA that has entered a host it is preferable that the DNA be stably maintained in the organism of interest. With regard to plasmids, there are two processes by which this can occur. One is through the use of replicative plasmids. These  
10 plasmids have origins of replication that are recognized by the host and allow the plasmids to replicate as stable, autonomous, extrachromosomal elements that are partitioned during cell division into daughter cells. The second process occurs through the integration of a plasmid onto the chromosome. This predominately happens by homologous recombination and results in the insertion of the entire plasmid, or parts of the plasmid, into the host  
15 chromosome. Thus, the plasmid and selectable marker(s) are replicated as an integral piece of the chromosome and segregated into daughter cells. Therefore, to ascertain if plasmid DNA is entering a cell during a transformation event through the use of selectable markers requires the use of a replicative plasmid or the ability to recombine the plasmid onto the chromosome. These qualifiers cannot always be met, especially when handling organisms  
20 that do not have a suite of genetic tools.

One way to avoid issues regarding plasmid-associated markers is through the use of transposons. A transposon is a mobile DNA element, defined by mosaic DNA sequences that are recognized by enzymatic machinery referred to as a transposase. The function of the transposase is to randomly insert the transposon DNA into host or target DNA. A  
25 selectable marker can be cloned onto a transposon by standard genetic engineering. The resulting DNA fragment can be coupled to the transposase machinery in an *in vitro* reaction and the complex can be introduced into target cells by electroporation. Stable insertion of the marker onto the chromosome requires only the function of the transposase machinery and alleviates the need for homologous recombination or replicative plasmids.

30 The random nature associated with the integration of transposons has the added advantage of acting as a form of mutagenesis. Libraries can be created that comprise

amalgamations of transposon mutants. These libraries can be used in screens or selections to produce mutants with desired phenotypes. For instance, a transposon library of a CBP organism could be screened for the ability to produce more ethanol, or less lactic acid and/or less acetate.

5 *Native cellulolytic strategy*

Naturally occurring cellulolytic microorganisms are starting points for CBP organism development via the native strategy. Anaerobes and facultative anaerobes are of particular interest. The primary objective is to engineer product yields and ethanol titers to satisfy the requirements of an industrial process. Metabolic engineering of mixed-acid  
10 fermentations in relation to these objectives has been successful in the case of mesophilic, non-cellulolytic, enteric bacteria. Recent developments in suitable gene-transfer techniques allow for this type of work to be undertaken with cellulolytic bacteria.

*Recombinant cellulolytic strategy*

Non-cellulolytic microorganisms with desired product-formation properties (*e.g.*,  
15 high ethanol yield and titer) are starting points for CBP organism development by the recombinant cellulolytic strategy. The primary objective of such developments is to engineer a heterologous cellulase system that enables growth and fermentation on pretreated lignocellulose. The heterologous production of cellulases has been pursued primarily with bacterial hosts producing ethanol at high yield (engineered strains of *E. coli*,  
20 *Klebsiella oxytoca*, and *Zymomonas mobilis*) and the yeast *Saccharomyces cerevisiae*. Cellulase expression in strains of *K. oxytoca* resulted in increased hydrolysis yields – but not growth without added cellulase – for microcrystalline cellulose, and anaerobic growth on amorphous cellulose. Although dozens of saccharolytic enzymes have been functionally expressed in *S. cerevisiae*, anaerobic growth on cellulose as the result of such expression  
25 has not been definitively demonstrated.

Aspects of the present invention relate to the use of thermophilic or mesophilic microorganisms as hosts for modification via the native cellulolytic strategy. Their potential in process applications in biotechnology stems from their ability to grow at relatively high temperatures with attendant high metabolic rates, production of physically  
30 and chemically stable enzymes, and elevated yields of end products. Major groups of thermophilic bacteria include eubacteria and archaeobacteria. Thermophilic eubacteria

include: phototropic bacteria, such as cyanobacteria, purple bacteria, and green bacteria; Gram-positive bacteria, such as *Bacillus*, *Clostridium*, Lactic acid bacteria, and Actinomyces; and other eubacteria, such as *Thiobacillus*, Spirochete, *Desulfotomaculum*, Gram-negative aerobes, Gram-negative anaerobes, and *Thermotoga*. Within archaeobacteria are considered Methanogens, extreme thermophiles (an art-recognized term), and *Thermoplasma*. In certain embodiments, the present invention relates to Gram-negative organotrophic thermophiles of the genera *Thermus*, Gram-positive eubacteria, such as genera *Clostridium*, and also which comprise both rods and cocci, genera in group of eubacteria, such as *Thermosipho* and *Thermotoga*, genera of Archaeobacteria, such as *Thermococcus*, *Thermoproteus* (rod-shaped), *Thermofilum* (rod-shaped), *Pyrodictium*, *Acidianus*, *Sulfolobus*, *Pyrobaculum*, *Pyrococcus*, *Thermodiscus*, *Staphylothermus*, *Desulfurococcus*, *Archaeoglobus*, and *Methanopyrus*. Some examples of thermophilic or mesophilic (including bacteria, procaryotic microorganism, and fungi), which may be suitable for the present invention include, but are not limited to: *Clostridium thermosulfurogenes*, *Clostridium cellulolyticum*, *Clostridium thermocellum*, *Clostridium thermohydrosulfuricum*, *Clostridium thermoaceticum*, *Clostridium thermosaccharolyticum*, *Clostridium tartarivorum*, *Clostridium thermocellulaseum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Thermoanaerobacterium thermosaccarolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermobacteroides acetoethylicus*, *Thermoanaerobium brockii*, *Methanobacterium thermoautotrophicum*, *Anaerocellum thermophilum*, *Pyrodictium occultum*, *Thermoproteus neutrophilus*, *Thermofilum librum*, *Thermothrix thioparus*, *Desulfovibrio thermophilus*, *Thermoplasma acidophilum*, *Hydrogenomonas thermophilus*, *Thermomicrobium roseum*, *Thermus flavas*, *Thermus ruber*, *Pyrococcus furiosus*, *Thermus aquaticus*, *Thermus thermophilus*, *Chloroflexus aurantiacus*, *Thermococcus litoralis*, *Pyrodictium abyssi*, *Bacillus stearothermophilus*, *Cyanidium caldarium*, *Mastigocladus laminosus*, *Chlamydothrix calidissima*, *Chlamydothrix penicillata*, *Thiothrix carnea*, *Phormidium tenuissimum*, *Phormidium geysericola*, *Phormidium subterraneum*, *Phormidium bijahensi*, *Oscillatoria filiformis*, *Synechococcus lividus*, *Chloroflexus aurantiacus*, *Pyrodictium brockii*, *Thiobacillus thiooxidans*, *Sulfolobus acidocaldarius*, *Thiobacillus thermophilica*, *Bacillus stearothermophilus*, *Cercosulcifer hamathensis*, *Vahlkampfia reichi*, *Cyclidium citrullus*, *Dactylaria gallopava*, *Synechococcus lividus*, *Synechococcus elongatus*, *Synechococcus minervae*, *Synechocystis aquatilis*, *Aphanocapsa thermalis*, *Oscillatoria terebriformis*,

*Oscillatoria amphibia*, *Oscillatoria germinata*, *Oscillatoria okenii*, *Phormidium*  
*laminosum*, *Phormidium parparasiens*, *Symploca thermalis*, *Bacillus acidocaldarias*,  
*Bacillus coagulans*, *Bacillus thermocatenalatus*, *Bacillus licheniformis*, *Bacillus pamilas*,  
*Bacillus macerans*, *Bacillus circulans*, *Bacillus laterosporus*, *Bacillus brevis*, *Bacillus*  
5 *subtilis*, *Bacillus sphaericus*, *Desulfotomaculum nigrificans*, *Streptococcus thermophilus*,  
*Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Bifidobacterium thermophilum*,  
*Streptomyces fragmentosporus*, *Streptomyces thermonitrificans*, *Streptomyces*  
*thermovulgaris*, *Pseudonocardia thermophila*, *Thermoactinomyces vulgaris*,  
*Thermoactinomyces sacchari*, *Thermoactinomyces candidas*, *Thermomonospora curvata*,  
10 *Thermomonospora viridis*, *Thermomonospora citrina*, *Microbispora thermodiastatica*,  
*Microbispora aerata*, *Microbispora bispore*, *Actinobifida dichotomica*, *Actinobifida*  
*chromogena*, *Micropolyspora caesia*, *Micropolyspora faeni*, *Micropolyspora cectivugida*,  
*Micropolyspora cabrobrunea*, *Micropolyspora thermovirida*, *Micropolyspora viridinigra*,  
*Methanobacterium thermoautothropicum*, *Caldicellulosiruptor acetigenus*,  
15 *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*,  
*Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, variants thereof,  
and/or progeny thereof.

In certain embodiments, the present invention relates to thermophilic bacteria  
selected from the group consisting of *Fervidobacterium gondwanense*, *Clostridium*  
20 *thermolacticum*, *Moorella* sp., and *Rhodothermus marinus*.

In certain embodiments, the present invention relates to thermophilic bacteria of the  
genera *Thermoanaerobacterium* or *Thermoanaerobacter*, including, but not limited to,  
species selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*,  
*Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*,  
25 *Thermoanaerobacterium zae*, *Thermoanaerobacterium xylanolyticum*,  
*Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*,  
*Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter*  
*thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brockii*,  
variants thereof, and progeny thereof.

30 In certain embodiments, the present invention relates to microorganisms of the  
genera *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, and *Anoxybacillus*,  
including, but not limited to, species selected from the group consisting of: *Geobacillus*

*thermoglucoasidarius*, *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, variants thereof, and progeny thereof.

5 In certain embodiments, the present invention relates to mesophilic bacteria selected from the group consisting of *Saccharophagus degradans*; *Flavobacterium johnsoniae*; *Fibrobacter succinogenes*; *Clostridium hungatei*; *Clostridium phytofermentans*; *Clostridium cellulolyticum*; *Clostridium aldrichii*; *Clostridium termitididis*; *Acetivibrio cellulolyticus*; *Acetivibrio ethanoligignens*; *Acetivibrio multivorans*; *Bacteroides*  
10 *cellulosolvens*; and *Alkalibacter saccharofomentans*, variants thereof and progeny thereof.

#### *Methods of the Invention*

During glycolysis, cells convert simple sugars, such as glucose, into pyruvic acid, with a net production of ATP and NADH. In the absence of a functioning electron transport system for oxidative phosphorylation, at least 95% of the pyruvic acid is  
15 consumed in short pathways which regenerate NAD<sup>+</sup>, an obligate requirement for continued glycolysis and ATP production. The waste products of these NAD<sup>+</sup> regeneration systems are commonly referred to as fermentation products.

Microorganisms produce a diverse array of fermentation products, including organic acids, such as lactate (the salt form of lactic acid), acetate (the salt form of acetic acid),  
20 succinate, and butyrate, and neutral products, such as ethanol, butanol, acetone, and butanediol. End products of fermentation share to varying degrees several fundamental features, including: they are relatively nontoxic under the conditions in which they are initially produced, but become more toxic upon accumulation; and they are more reduced than pyruvate because their immediate precursors have served as terminal electron  
25 acceptors during glycolysis. Aspects of the present invention relate to the use of gene knockout technology to provide novel microorganisms useful in the production of ethanol from lignocellulosic biomass substrates. The transformed organisms are prepared by deleting or inactivating one or more genes that encode competing pathways, such as the  
30 based selection for mutants with improved performance for producing ethanol as a fermentation product.

In certain embodiments, a thermophilic or mesophilic microorganism, which in a native state contains at least one gene that confers upon the microorganism an ability to produce lactic acid as a fermentation product, is transformed to decrease or eliminate expression of said at least one gene. The gene that confers upon said microorganism an ability to produce lactic acid as a fermentation product may code for expression of lactate dehydrogenase. The deletion or suppression of the gene(s) or particular polynucleotide sequence(s) that encode for expression of LDH diminishes or eliminates the reaction scheme in the overall glycolytic pathway whereby pyruvate is converted to lactic acid; the resulting relative abundance of pyruvate from these first stages of glycolysis should allow for the increased production of ethanol.

In certain embodiments, a thermophilic or mesophilic microorganism, which in a native state contains at least one gene that confers upon the microorganism an ability to produce acetic acid as a fermentation product, is transformed to eliminate expression of said at least one gene. The gene that confers upon the microorganism an ability to produce acetic acid as a fermentation product may code for expression of acetate kinase and/or phosphotransacetylase. The deletion or suppression of the gene(s) or particular polynucleotide sequence(s) that encode for expression of ACK and/or PTA diminishes or eliminates the reaction scheme in the overall glycolytic pathway whereby acetyl CoA is converted to acetic acid (Figure 1); the resulting relative abundance of acetyl CoA from these later stages of glycolysis should allow for the increased production of ethanol.

In certain embodiments, the above-detailed gene knockout schemes can be applied individually or in concert. Eliminating the mechanism for the production of lactate (*i.e.*, knocking out the genes or particular polynucleotide sequences that encode for expression of LDH) generates more acetyl CoA; it follows that if the mechanism for the production of acetate is also eliminated (*i.e.*, knocking out the genes or particular polynucleotide sequences that encode for expression of ACK and/or PTA), the abundance of acetyl CoA will be further enhanced, which should result in increased production of ethanol.

In certain embodiments, it is not required that the thermophilic or mesophilic microorganisms have native or endogenous PDC or ADH. In certain embodiments, the genes encoding for PDC and/or ADH can be expressed recombinantly in the genetically modified microorganisms of the present invention. In certain embodiments, the gene knockout technology of the present invention can be applied to recombinant

microorganisms, which may comprise a heterologous gene that codes for PDC and/or ADH, wherein said heterologous gene is expressed at sufficient levels to increase the ability of said recombinant microorganism (which may be thermophilic) to produce ethanol as a fermentation product or to confer upon said recombinant microorganism (which may be thermophilic) the ability to produce ethanol as a fermentation product.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 70% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 80% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 90% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, substantial or complete elimination of organic acid production from microorganisms in a native state may be achieved using one or more site-directed DNA homologous recombination events.

Operating either a simultaneous saccharification and co-fermentation (SSCF) or CBP process at thermophilic temperatures offers several important benefits over conventional mesophilic fermentation temperatures of 30-37 °C. In particular, costs for a process step dedicated to cellulase production are substantially reduced (*e.g.*, 2-fold or more) for thermophilic SSCF and are eliminated for CBP. Costs associated with fermentor cooling and also heat exchange before and after fermentation are also expected to be reduced for both thermophilic SSCF and CBP. Finally, processes featuring thermophilic biocatalysts may be less susceptible to microbial contamination as compared to processes featuring conventional mesophilic biocatalysts.

The ability to redirect electron flow by virtue of modifications to carbon flow has broad implications. For example, this approach could be used to produce high ethanol yields in strains other than *T. saccharolyticum* and/or to produce solvents other than ethanol, for example, higher alcohols (*i.e.*, butanol).

*Metabolic engineering through antisense oligonucleotide (asRNA) strategies*

Fermentative microorganisms such as yeast and anaerobic bacteria ferment sugars to ethanol and other reduced organic end products. Theoretically, carbon flow can be directed to ethanol production if the formation of competing end-products, such as lactate and acetate, can be suppressed. The present invention provides several genetic engineering approaches designed to remove such competing pathways in the CBP organisms of the invention. The bulk of these approaches utilize knock-out constructs (for single crossover recombination) or allele-exchange constructs (for double crossover recombination) and target the genetic loci for *ack* and *ldh*. Although these tools employ “tried and true” strain development techniques, there are several potential issues that could stall progress: (i) they are dependent on the host recombination efficiency which in all cases is unknown for the CBP organisms; (ii) they can be used to knock out only one pathway at a time, so successive genetic alterations are incumbent upon having several selectable markers or a recyclable marker; (iii) deletion of target genes may be toxic or have polar effects on downstream gene expression.

The present invention provides additional approaches towards genetic engineering that do not rely on host recombination efficiency. One of these alternative tools is called antisense RNA (asRNA). Although antisense oligonucleotides have been used for over twenty-five years to inhibit gene expression levels both *in vitro* and *in vivo*, recent advances in mRNA structure prediction has facilitated smarter design of asRNA molecules. These advances have prompted a number of groups to demonstrate the usefulness of asRNA in metabolic engineering of bacteria.

The benefits of using asRNA over knock-out and allele-exchange technology are numerous: (i) alleviates the need for multiple selectable markers because multiple pathways can be targeted by a single asRNA construct; (ii) attenuation level of target mRNA can be adjusted by increasing or decreasing the association rate between asRNA; (iii) pathway inactivation can be conditional if asRNA transcripts are driven by conditional promoters. Recently, this technology has been used to increase solventogenesis in the Gram positive mesophile, *Clostridium acetobutylicum* (Tummala et al. (2003)). Although the exact molecular mechanism of how asRNA attenuates gene expression is unclear, the likely mechanism is triggered upon hybridization of the asRNA to the target mRNA. Mechanisms may include one or more of the following: (i) inhibition of translation of

mRNA into protein by blocking the ribosome binding site from properly interacting with the ribosome, (ii) decreasing the half-life of mRNA through dsRNA-dependent RNases, such as RNase H, that rapidly degrade duplex RNA, and (iii) inhibition of transcription due to early transcription termination of mRNA.

#### 5 *Design of antisense sequences*

asRNAs are typically 18-25 nucleotides in length. There are several computation tools available for rational design of RNA-targeting nucleic acids (Sfold, Integrated DNA Technologies, STZ Nucleic Acid Design) which may be used to select asRNA sequences. For instance, the gene sequence for *Clostridium thermocellum ack* (acetate kinase) can be  
10 submitted to a rational design server and several asRNA sequences can be culled. In brief, the design parameters select for mRNA target sequences that do not contain predicted secondary structure.

#### *Design of delivery vector*

A replicative plasmid will be used to deliver the asRNA coding sequence to the  
15 target organism. Vectors such as, but not limited to, pNW33N, pJIR418, pJIR751, and pCTC1, will form the backbone of the asRNA constructs for delivery of the asRNA coding sequences to inside the host cell. In addition to extra-chromosomal (plasmid based) expression, asRNAs may be stably inserted at a heterologous locus into the genome of the microorganism to get stable expression of asRNAs. In certain embodiments, strains of  
20 thermophilic or mesophilic microorganisms of interest may be engineered by site directed homologous recombination to knockout the production of organic acids and other genes of interest may be partially, substantially, or completely deleted, silenced, inactivated, or down-regulated by asRNA.

#### *Promoter choice*

25 To ensure expression of asRNA transcripts, compatible promoters for the given host will be fused to the asRNA coding sequence. The promoter-asRNA cassettes are constructed in a single PCR step. Sense and antisense primers designed to amplify a promoter region will be modified such that the asRNA sequence (culled from the rational design approach) is attached to the 5' end of the antisense primer. Additionally, restriction  
30 sites, such as *EcoRI* or *BamHI*, will be added to the terminal ends of each primer so that the

final PCR amplicon can be digested directly with restriction enzymes and inserted into the vector backbone through traditional cloning techniques.

With respect to microorganisms that do not have the ability to metabolize pentose sugars, but are able to metabolize hexose sugars as described herein, it will be appreciated  
5 that the *ack* and *ldh* genes of *Clostridium thermocellum* and *Clostridium straminisolvens*, for example, may be targeted for inactivation using antisense RNA according to the methods described herein.

With respect to microorganisms that confer the ability to metabolize pentose and hexose sugars as described herein, it will be appreciated that the *ack* and *ldh* genes of  
10 *Clostridium cellulolyticum*, *Clostridium phytofermentans* and *Caldicellulosiruptor kristjanssonii*, for example, may be targeted for inactivation using antisense according to the methods described herein.

In addition to antibiotic selection for strains expressing the asRNA delivery vectors, such strains may be selected on conditional media that contains any of the several toxic  
15 metabolite analogues such as sodium fluoroacetate (SFA), bromoacetic acid (BAA), chloroacetic acid (CAA), 5-fluoroorotic acid (5-FOA) and chlorolactic acid. Use of chemical mutagens including, but not exclusively, ethane methyl sulfonate (EMS) may be used in combination with the expression of antisense oligonucleotide (asRNA) to generate strains that have one or more genes partially, substantially, or completely deleted, silenced,  
20 inactivated, or down-regulated.

## EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration  
25 of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

## EXAMPLE 1

**Generation of Custom Transposons For Mesophilic and Thermophilic Cellulolytic, Xylanolytic Organisms**

The present invention provides methods for generating custom transposons for  
5 cellulolytic and/or xylanolytic and/or thermophilic organisms. To do this, a native  
promoter from the host organism will be fused to a selectable marker which has been  
determined to work in this organism. This fragment will be cloned into the EZ-Tn5<sup>TM</sup>  
transposon that is carried on the vector pMOD<sup>TM</sup>-2<MCS> (Epicenter®Biotechnologies).  
For example, the *C. thermocellum* the *gapDH* promoter will be fused to the mLs drug  
10 marker, as well as the *cat* gene and then subcloned into vector pMOD<sup>TM</sup>-2<MCS>.

Commercial transposons are lacking in thermostable drug markers and native  
promoters of cellulolytic and/or xylanolytic and/or thermophilic organisms. The mLs and  
*cat* markers have functioned in thermophilic bacteria and the *gapDH* promoter regulates a  
key glycolytic enzyme and should be constantly expressed. The combination of the above  
15 drug markers and the *gapDH* promoter will greatly enhance the probability of generating a  
functional transposon. This approach may be applied to other cellulolytic and/or  
xylanolytic and/or thermophilic organisms.

*Experimental Design*

Figure 26 is a diagram taken from the Epicenter®Biotechnologies user manual,  
20 which is incorporated herein by reference, representing bp 250-550 of pMOD<sup>TM</sup>-2<MCS>.  
In the top portion, the black arrowheads labeled ME denote 19 bp mosaic ends that define  
the transposon. The *EcoRI* and *HindIII* sites define the multi-cloning site, which is  
represented by the black box labeled MCS. In the bottom portion, the DNA sequence and  
the restriction enzymes associated with the MCS are shown.

25 The following primers will be used to amplify promoter fusion fragments from  
pMQ87-*gapDH*-*cat* and pMQ87-*gapDH*-mLs: GGCGgaattc CTT GGT CTG ACA ATC  
GAT GC (SEQ ID NO:19); GGCGgaattc TATCAGTTATTACCCACTTTTCG (SEQ ID  
NO:20). The lower case letters denote engineered *EcoRI* restriction sites. The size of the  
amplicon generated will be ~1.9 kb. Standard molecular procedures will allow the  
30 amplicon to be digested with *EcoRI* and cloned into the unique *EcoRI* site of pMOD<sup>TM</sup>-

2<MCS>. The transposon and subsequent transpososome will be generated and introduced into host organisms as described by the manufacturer.

## EXAMPLE 2

### Constructs for Engineering Cellulolytic and Xylanolytic Strains

5       The present invention provides compositions and methods for genetically engineering an organism of interest to CBP by mutating genes encoding key enzymes of metabolic pathways which divert carbon flow away from ethanol. Single crossover knockout constructs are designed so as to insert large fragments of foreign DNA into the gene of interest to partially, substantially, or completely delete, silence, inactivate, or down-  
10 regulate it. Double crossover knockout constructs are designed so as to partially, substantially, or completely delete, silence, inactivate, or down-regulate the gene of interest from the chromosome or replace the gene of interest on the chromosome with a mutated copy of the gene, such as a form of the gene interrupted by an antibiotic resistance cassette.

15       The design of single crossover knockout vectors requires the cloning of an internal fragment of the gene of interest into a plasmid based system. Ideally, this vector will carry a selectable marker that is expressed in the host strain but will not replicate in the host strain. Thus, upon introduction into the host strain the plasmid will not replicate. If the cells are placed in a conditional medium that selects for the marker carried on the plasmid, only those cells that have found a way to maintain the plasmid will grow. Because the  
20 plasmid is unable to replicate as an autonomous DNA element, the most likely way that the plasmid will be maintained is through recombination onto the host chromosome. The most likely place for the recombination to occur is at a region of homology between the plasmid and the host chromosome.

25       Alternatively, replicating plasmids can be used to create single crossover interruptions. Cells that have taken up the knockout vector can be selected on a conditional medium, then passaged in the absence of selection. Without the positive selection provided by the conditional medium, many organisms will lose the plasmid. In the event that the plasmid is inserted onto the host chromosome, it will not be lost in the absence of selection. The cells can then be returned to a conditional medium and only those that have retained  
30 the marker, through chromosomal integration, will grow. A PCR based method will be devised to screen for organisms that contain the marker located on the chromosome.

The design of double crossover knockout vectors requires at least cloning the DNA flanking (~ 1 kb) the gene of interest into a plasmid and in some cases may include cloning the gene of interest. A selectable marker may be placed between the flanking DNA or if the gene of interest is cloned the marker is placed internally with respect to the gene. Ideally  
5 the plasmid used is not capable of replicating in the host strain. Upon the introduction of the plasmid into the host and selection on a medium conditional to the marker, only cells that have recombined the homologous DNA onto the chromosome will grow. Two recombination events are needed to replace the gene of interest with the selectable marker.

Alternatively, replicating plasmids can be used to create double crossover gene  
10 replacements. Cells that have taken up the knockout vector can be selected on a conditional medium, then passaged in the absence of selection. Without the positive selection provided by the conditional medium, many organisms will lose the plasmid. In the event that the drug marker is inserted onto the host chromosome, it will not be lost in the absence of selection. The cells can then be returned to a conditional medium and only those that have  
15 retained the marker, through chromosomal integration, will grow. A PCR based method may be devised to screen for organisms that contain the marker located on the chromosome.

In addition to antibiotic selection schemes, several toxic metabolite analogues such as sodium fluoroacetate (SFA), bromoacetic acid (BAA), chloroacetic acid (CAA), 5-fluoroorotic acid (5-FOA) and chlorolactic acid may be used to select mutants arising from  
20 either homologous recombinations, or transposon-based strategies. Use of chemical mutagens including, but not exclusively, ethane methyl sulfonate (EMS) may be used in combination with the directed mutagenesis schemes that employ homologous recombinations, or transposon-based strategies.

### *C. cellulolyticum Knockout Constructs*

25 *Acetate kinase (gene 131 from C. cellulolyticum published genome):*

#### *Single Crossover*

The acetate kinase gene of *C. cellulolyticum* is 1,110 bp in length. A 662 bp internal fragment (SEQ ID NO:21) spanning nucleotides 91-752 was amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers.  
30 Selectable markers may include those that provide erythromycin and chloramphenicol

resistance. These plasmids will be used to disrupt the *ack* gene. A map of the *ack* gene and the region amplified by PCR for gene disruption are shown in Figure 19. The underlined portions of SEQ ID NO:21 set forth below correspond to the sites that are *EcoRI* sites that flank the knockout fragment.

5 gaattctgcgacagaatagggattgacaattcctttataaagcaatcaaggggttcagaagaggctgttatttgaataaagagctaaagaat  
 cacaagatgcaatagaggctgttatttgcactgactgacgataatggcggtataaaaaacatgtccgaaatatcagcagtgggaca  
 cagaatagtacacggcgggtgaaaaattcaacagttctgtagttatagatgaaaacggtatgaatgcagtaagagagtgtatagacgttgac  
 cgcttcataatccgccgaatattataggtatagaggcttgccagcagattatgccaatatacctatggtagctgtattgataccactttccaca  
 gctccatgcctgattatgcatacctttacgcattgccatgaactttatgaaaagtagcgtataagaaaatatggtttccacggaacatcacac  
 10 aaatatgttcagaaagagcttctgcaatgcttgataagtctttgaacgaattaaagataattacatgccatcttggaacggttcaagtattgt  
 gctgttaacaagggtaaatcaattgatacttccatgggctttacaccttgcagggacttgcaatgggtacaagaagcgggtacaatagaccct  
 gaagttgttacgaattc

These sites were engineered during the design of the “ack KO primers” and will allow subsequent cloning of the fragment into numerous vectors.

#### 15 *Double crossover*

To construct a double crossover vector for the *ack* gene of *C. cellulolyticum* ~1 kb of DNA flanking each side of the *ack* gene will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers may include those that provide erythromycin and chloramphenicol resistance. The 3' flanking region of the *ack* gene is not  
 20 available in the available draft genome. To acquire this DNA, a kit such as GenomeWalker from Clontech will be used.

*Lactate dehydrogenase (genes 2262 and 2744 of C. cellulolyticum published genome):*

#### *Single crossover*

The *ldh* genes of *C. cellulolyticum* are 951 bp (for gene 2262) (SEQ ID NO:22) and  
 25 932 bp (for gene 2744) (SEQ ID NO:23) in length. A ~500 bp internal fragment near the 5' end of each gene will be amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers. Selectable markers may include those that provide drug resistance, such as erythromycin and chloramphenicol. These plasmids will be used to disrupt the *ldh* 2262 and *ldh* 2744 genes. As an example, a map of the *ldh* 2262  
 30 gene and the region amplified by PCR for gene disruption are shown in Figure 20.

*Double Crossover*

To construct a double crossover knockout vector for the *ldh* gene(s) of *C. cellulolyticum* ~1 kb of DNA flanking each side of the *ldh* gene(s) will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers may include those that provide drug resistance, such as erythromycin and chloramphenicol. Figure 21 provides an example of *C. cellulolyticum* *ldh* (2262) double crossover knock out fragment.

In the sequence set forth below (SEQ ID NO:24) the mLs gene (selectable marker) is underlined and the flanking DNA is the remaining sequence. During primer design, restriction sites will be engineered and the 5' and 3' ends of the above fragment so that it can be cloned into a number of replicative and non-replicative vectors. The same strategy will be used to create a vector to delete *ldh* 2744.

gacgcatacaggttgtaacacccatttcccttagctttcgggagatgaataaaacaaactttccgggtcctttaccacacccgcccacataaa  
 gagctatgccgcatgaaagaaacgatatgttatcatttttcgtaaactgttattccgaacccggataaagctttaccatattattaactgctgcc  
 15 gtccttgcatgtgtacacccataaccactattttcatatacatcctccttggttgcttgtaaataatatacccatataaccacctaataatatttataa  
 acaaattcgggtatcattctttggtaaataaaaagtacatccgatattagaatgtacctaaaaaaattattttattgtatgctttatctgttt  
 cattatatggttgctatccattctacggtaaaatcaagtaattccattaagtagctgctgctgctatcctgctataatccgtattactgattt  
 ctcaataaaatcatggtgttcaactttgtgggagagaagcttgcgatatcctatgctatgcatgtattcttctcataggtaaaatgaaagacagt  
 gtaatcttttagttccgtaattagccgtacaatttcatcatatttgtctgtaataagctgattttcgtggcctcataaattccgaagcaatctggaat  
 20 agtttctatgctgttcgtcgattttctcaattccaagaataaattcgtctctcattctatcatatggaccctcctaaattgtaatgtataccaagatta  
 tacatacttctagaatataaacaatacaaggataaaatttaatatcgtatacctacataaatgactaacttaagctctctaaaacttcttttta  
 ttatttctatactactaaaatcaaaaatattctctaaagtatttctacaaatgttgttttgcaacaaagtagtatactttgcacccagaatgttttgta  
 taacttacaattaggggtatattatagtaaataactaaatggaagagtaggatattgattatgaacgagaaaaatataaaacacagtcaaa  
 25 actttattacttcaaaacataatatagataaaaataatgacaaatataagattaaatgaacatgataatctttgaaatcggctcaggaaaagg  
gcattttacccttgaattagtagagaggtgtaatttcgtaactgccattgaaatagaccataaattatgcaaaactacagaaaaataaacttgttg  
atcacgataattccaagttttaacaaggatataattgcagtttaaatctctaaaaaccaatcctataaaatatttggaatataccttataacata  
agtagcgatataatacgcataaattgttttgatagtagctgtagagattttaaactcgtggaatacgggttgctaaaagattataaatacaa  
aacgctcattggcattattttaatggcagaagttgatatttctatattaagtaggttccaagagaatatttcatcctaaacctaaagtgaatagct  
cacttatcagattaaatagaaaaaatcaagaatatcacacaaagataaacagaagtataattttcgttatgaaatgggttaacaaagaa  
 30 tacaagaaaaatattacaaaaaatcaatttaacaattccttaaaacatgcaggaattgacgatttaacaaatattagctttgaacaattcttatct  
ctttcaatagctataaattatttaataagatcccctttactcggatgcatgccgcaggcaggcatccgaagtagtttctcattatacaagtattc  
tcttgagtacgtcgtcgtctcagcagctgcttgcctttccctgtttccggcacaatggagataagtgtagtctgtaggcttaatagtggtgccaat  
gtcaattgcctttcgaagtcctgctccttattttaagggttccacaaaattgataaaacccgtatcagtcagaaattttactaccgctgatatct  
gtgttcttgaaccctgctcataagataggttgcaatccaacctgaattccatgaagctgaggtgtctccagcagcttatctaaagcatgagat  
 35 attagatgctcactaccgctggctggagcactgctgtctgctatctgcatggcaattccgctcattgtcagagagcttaccatttctttaaaaag  
aagtttctgtaacctgtgtgtagggcatccttacaatactgtttactgacttttagcaatcattgcagcaaaatcgtcaacctttgccgcattgttc  
ctttctcaaaataaccagtcatacacagccgtaatttggatattatgtctccgagacctgaataaataaatttcataggtgcatttttaatacatct  
aaatccactaatattccaaatggcatcgaggcatgtacggaagtagcctgacattataatcaaagagcagcctgagctggaaaaacat  
cgttgagggtgatgtaggtatactgataaaaggaagctgtttaaaaaagctatatatttggctgcatcaagcacctttcctcctactccgac  
 40 cactgcatcgggttttgagggaatagtaaaagccttgagcataagatttcaagctttatgtcatcatagtcgtaagttcaagtactgcaagag  
attttcttgactttatggaatccagaatctttcaccaaataagtcacgtattccctctcaaaaagtactacaacattactaattcctgccccttcaa  
 tatgtgc

*C. phytofermentans Knockout Constructs*

*For acetate kinase (gene 327 from C. phytofermentans published genome):*

*Single crossover*

The acetate kinase gene of *C. phytofermentans* is 1,244 bp in length. A 572 bp  
 5 internal fragment spanning nucleotides 55-626 will be amplified by PCR and cloned into  
 suicide vectors and replicating vectors that have different selectable markers. Selectable  
 markers to use will include those that provide drug resistance to *C. phytofermentans*. These  
 plasmids will be used to disrupt the *ack* gene. A map of the *ack* gene and the region  
 amplified by PCR for gene disruption are shown in Figure 22. Restriction sites will be  
 10 engineered during the design of the “ack KO primers” and will allow subsequent cloning of  
 the fragment into numerous vectors. The sequence of the knockout fragment described  
 above is set forth as SEQ ID NO:25.

*Double crossover*

To construct a double crossover knockout vector for the *ack* gene of *C.*  
 15 *phytofermentans* ~1 kb of DNA flanking each side of the *ack* gene will be cloned. A  
 selectable marker will be inserted between the flanking DNA. Selectable markers to use  
 will include those that provide drug resistance to this strain. An example of a putative  
 double crossover knockout construct with the mLS gene as a putative selectable marker is  
 shown in Figure 23.

20 The sequence that corresponds to the fragment depicted in Figure 23 (SEQ ID  
 NO:26) is set forth below. The mLS gene (putative selectable marker) is underlined and the  
 remainder of the sequence corresponds to the flanking DNA. During primer design,  
 restriction sites will be engineered and the 5' and 3' ends of the above fragment so that it  
 can be cloned into a number of replicative and non-replicative vectors.

25 ctgagtgaatgtaaaaaaggatgcctcaagtattctgaaacatccttatattatactacaaaatcataaagtaaattactcagctgtagcaat  
 gatctctttttgtgtaagatccacaagctttacaaactctatgaggcatcataagtgcaccacacttgctgcatttcactaagttggagcagtc  
 atctccagtttgacgacgactatctctctagctttggaatgtttattctttggacaaaatagctccattgattacacctcctaaactgttaaaaa  
 tatctcgatagcagacattcttgggtctagttctgtacgggtcacacccgcactctccttcatttaggttagcaccgcagaccttgacagattcctt  
 acagtcttcttgacagaaaccttcattgggaaaccaatcaagacttctcatagataagtttatctacgtctaaatcataccggaaacaaaatt  
 30 tgttcatctaaatcctcggtacgctgttctctgtttcgatacatcaatctctgtagccacgtcgatgtctgttgatggtttctccttcaacaac  
 gatcgcaaggaacggctaacgctaatttcgttttgctccaccagaattttcgccacctagattagtttaactaagtttaaccggttctttag  
 gtaatagaataaccgacaccatttaattcgaatatatcaaatcaatcggtgcagtgtattctttgagaccattaggaacattcatgactcaga  
 catttgatcagcataagtaactctgtctaaaaaacgcataatgtaagcgcccaaaaattcacactgtagtattataaacgcttaaaatag

gttgtcaactcctaactgttaaaaatgtcagaattgtgtaaccatattttctcttcattatcggttctcccttattaaataatttatagctattgaaaaga  
 gataagaattgttcaaagctaattgtttaaatcgtaattcctgcatgtttaaggaattgttaaattgatttttgtaaataattttctgtattctttgtta  
 acccatttcataacgaaataattatacttctgtttatctttgtgtgatatcttgattttttctatttaacttgataagtgagctattcactttagggttaggat  
 gaaaatattctcttggaaccatacttaatatagaaatatcaacttctgccattaaaaataatgccaatgagcggtttgtatttaataatcttttagca  
 5 aaccggtattccacgattaaataaatctcatcagctatactatcaaaaaacaatttgcgattatatccgtacttatgttataagggtatattaccaaa  
 tttttataggattggttttaggaaatttaaactgcaatatacctgtttaaaacttggaattatcgtgatcaacaagtttattttctgtagttttgcat  
 aatttatgggtctattcaatggcagttacgaaattacacctctgtactaattcaagggtaaaatgcccttttctgagccgatttcaaagatattatc  
 atgttcatttaattcttataattgtcattattttatctatattatgtttgaagtaataaagttttgactgtgttttatattttctcggttcattgtattttctcctataat  
 gttcttaaatcatttatcacggggcaacttaatatatccgaaatatagttcttctatatcggtcccccagttataatgattattatactatttaattcttcaa  
 10 cttacaattggagtttccagttaagaaataataatttaattgcaaagcggtatctcgcaatccgcttacgctacttgctcataacctcaacagg  
 caatgaagctaagttaattattactctgtgcctgaacagcagtgattgcaacaacaccaacgatatcatcagaagaacaacctcttgataaa  
 tcatttactggagctgcaataccctgagttaattggtccataagcttctgcctttgcaagacgctgtgttaacttatccaatgttaccagcatcaa  
 ggtctgggaagattaatacgttagctttccagcaatatcactaccaggagctttgaagcacctacactaggaacgattgctgcatctaactg  
 gaactcgccgtcgatcttataattctgggtataattcatttgcaatcttagttgcttctacaaccttatcaacatctgcatgctttgcgcttccctttgtga  
 15 atgagaaagcatagctacgataggttcagagccaactaattgttcaaaactcttcgctgtggaaccagcgattgctgctaactcttcagcattt  
 ggattctgatttaaaccagcatcagagaaaaggaaagttccatttgcgccatatcacaattagggtactaccattacgaagaaagcagaaaa  
 ctaacttagtatttgagcagttttaaaatctgaagacatggtcttaaggtatctgctgtagagtgaagcaccagatactaaaccatctgca  
 tcgcccatcttaaccatcattacaccgtatgtaatgtagtctgtgttaaaagctcttttgccttttcaggggtcatgcctttgcctgtctaagttctaca  
 agcttgtaatgtaagc

20

*For Lactate dehydrogenase (genes 1389 and 2971 of C. phytofermentans published genome)*

#### *Single crossover*

The *ldh* genes of *C. phytofermentans* are 978 bp (for gene 1389) (SEQ ID NO:27)  
 25 and 960 bp (for gene 2971) (SEQ ID NO:28) in length. A ~500 bp internal fragment near  
 the 5' end of each gene will be amplified by PCR and cloned into suicide vectors and  
 replicating vectors that have different selectable markers. Selectable markers to use will  
 include those that provide drug resistance. These plasmids will be used to disrupt the *ldh*  
 1389 and *ldh* 2971 genes. As an example, a map of the *ldh* 1389 gene and the region  
 30 amplified by PCR for gene disruption are shown in Figure 24.

#### *Double crossover*

To construct a double crossover knockout vector for the *ldh* gene(s) of *C.*  
*phytofermentans* ~1 kb of DNA flanking each side of the *ldh* gene(s) will be cloned. A  
 selectable marker will be inserted between the flanking DNA. Selectable markers to use  
 35 will include those that provide drug resistance to this strain. An example of a putative  
 double crossover knockout construct with the mLS gene as a putative selectable marker is  
 shown in Figure 25.

The sequence that corresponds to the fragment depicted in Figure 25 is set forth below as SEQ ID NO:29. The *mLs* gene (selectable marker) is underlined and the remaining portion of the sequence corresponds to the flanking DNA. During primer design, restriction sites will be engineered and the 5' and 3' ends of the above fragment so that it can be cloned into a number of replicative and non-replicative vectors. The same strategy will be used to create a vector to delete *ldh* 2971.

5      tggaatctcactatgcaccaatgtggtactaaattatatctttatctatggaaaattaggtttccgcgaatggagatagagggagctgccattgc  
       tactttaattgtagaattcttgagagtatttagttgttatttatgtataaggggtgagaaggtacttaagatgagactttctatattttaagagatct  
       aaacagtatcttcgctctttggctcgttatagtcgccagtgcttatgagtgaggtaactgggggcttgggattgctgttcagtcgcaatcattgg  
 10     gcgatgggtgtagttttctacagcccgccagctcattaatgtagtacaacagtttagccggaatcattctgattggtattggtgtgggttcgagc  
       attataatagggaaattgattggtgagggaaaagagcatgaggcgagaatgctagccaataagttaatacgtatcagtatgatactcggagg  
       aattgttgctttgcagtaacttactacgtccaatcgctcctaactttattgaggcgctcaaggaaacagcggatttaattcgtcagatgctatttgtt  
       tcggcttacctcttattctccaagccttatctgtattaactatggccggaatattacgtggtgcaggggatacccttactgtgcaacctttgatgtt  
       tgacctatgggtactaaaacttgaggagggtttgcttgaaccatagtagtctcctccacctgtatgggtttactttatcttaagtagcgatgag  
 15     tgtgttaaagcgctatttacgggtaccgcggtcttaaggggacgttggaatcatgatacaacactgcattaagattcatatgtccagatattttg  
       cacagtagcataattactagagcttattcctataatattcataggttttgatggtccattttacgttacgatagcataatattacataaaaaccaattct  
       atataagatgaggttatagtagaacgagagaaaaataaaaaacacagtcaaaactttattacttcaaaacataatatagataaaaataatgaca  
       aatataagattaaatgaacatgataatactttgaaatcggctcaggaaaaagggcattttacccttgaatttagtacagaggtgtaatttcgtaac  
       tgccattgaaatagaccataaattatgcaaaactacagaaaaataaactgttgatcacgataatttccaagttttaacaaggatatattgcagt  
 20     ttaaatttcctaaaaaccaatcctataaaaattttggaatataccttataacataagtagcgataataacgcaaaattgttttgatagtagct  
       gatgagatttattaatcgtggaatacgggtttgctaaaagattattaaatacaaaacgctcattggcatttttaattgacagaaagttgatatttct  
       atattaagtaggttccaagagaatatttcatcctaaacctaagtgaaatagctcacttatcagattaaatagaaaaaaatcaagaatatcac  
       acaaagataaacagaagtataattttcgttatgaaatgggttaacaaagaatacaagaaaatatttacaaaaaatcaatttaacaattcctt  
       aaaacatgcagggaattgacgatttaacaatattagctttgaacaattcttatcttttcaatagctataaatttttaataagaagtaataggaa  
 25     ataatactcgaattattctgcaatctgttctaaaaaataaaattaagaaattactatagcaagccagggttaaataactagcttgctattttgtgca  
       ttagtagcagttttgattattaagaataaaatttaataactattttgcaataagttattgactattcacaagttagtgtactatacaagtagaaata  
       aagatacataaaaaaataaataatgaacataaattcatgacatgcggaatagaatgaaagaatattatgtcggttcctaataactaaatg  
       gatataacaatctattgaaacacttatggggtgtaagtgtggagagaatttctaaagcgccaaaagactctacatatgaaattctaaagcttca  
       cacgggaataatctaatttatgtatcttattatcataattcaggaaggtagtgtgaaaataaaaaattagtttctgtttcattcaggcagtagca  
 30     tttcttaacaaaatttgctatgcattgggtgttatctgaaaaacaaaaagcaattttctcacaactattttgaacaacaatggtattaaaaatttg  
       gaggaggattttactatgaaaaaaacggtaacattactgttggtctgaccatgggtggaagcttatttgcagcatgtggttaagaaaaatggat  
       caagcgaaaccggcacaaaagatcctgtggcaacaagcgggtgcaaaagaacctgacaaacaagatccaggcaataaagagcctga  
       aaaacaagacctgttaaaatcaagatttactctgataatgaaccttaccatttaagaagattgggttagttataaaggaagctgagaa  
       gagatttaattgttattcgatttcgaagtaattccaattgcagattatcaacaaaagtttcttaacattaaatacaggaaataacgctccagat  
 35     gtcatcctttatcagtcaacgcaggagagagaatgcatct

*Cald. kristjanssonii* and *C. stercorarium* subs *leptospartum*

To the best of our knowledge, genome sequencing of the above organisms has not occurred and if it has, it has not been made available to the public. Based on our experimental results these organisms are cellulolytic and xylanolytic. The DNA sequences of genes encoding key metabolic enzymes are needed from these organisms in order to genetically engineer them and divert carbon flow to ethanol. These include such enzymes as acetate kinase and lactate dehydrogenase. In order to obtain the sequences of these genes, the genomes of these organisms will be sequenced.

With access to genome sequences, the conserved nature of the above enzymes may be used to find the encoding genes and flanking DNA. These sequences will be used to design constructs for targeted mutagenesis employing both single and double crossover strategies. These strategies will be identical to those described above. We will also  
5 determine which antibiotics can be used as selectable markers in these organisms and which protocols for transformation work best.

### EXAMPLE 3

#### **Transformation of *C. cellulolyticum***

Cells were grown in 50 mL of GS media with 4g/l cellobiose to an OD of 0.8 in  
10 anaerobic conditions, incubated at 34 degrees C. After harvesting they were washed 3 times in equal volumes with a wash buffer containing 500mM sucrose and 5mM MOPS with pH adjusted to 7. After the final wash, the cell pellet was resuspended in an equal volume of wash buffer 10ul aliquots of the cell suspension were placed in a standard electroporation cuvette with a 1mm electrode spacing. 1ul plasmid DNA was added. The  
15 concentration of the plasmid DNA was adjusted to ensure between a 1:1 and 10:1 molar ratio of plasmid to cells. A 5ms pulse was applied with a field strength of 7kV/cm (measured) across the sample. A custom pulse generator was used. The sample was immediately diluted 1000:1 with the same media used in the initial culturing and allowed to recover until growth resumed, and was determined via an increase in the OD (24-48h). The  
20 recovered sample was diluted 50:1 and placed in selective media with either 15ug/mL erythromycin or 15ug/mL chloramphenicol and allowed to grow for 5-6 days. Samples exhibiting growth in selective media were tested to confirm that they were in fact *C. cellulolyticum* and that they had the plasmid.

### EXAMPLE 4

#### **Constructs for Engineering Cellulolytic Strains**

25 Cellulose is one of the main components of biomass, which can be potentially used as a substrate for generation of fuel ethanol by fermentation with *Clostridium thermocellum*. However, in this process, much energy and carbon sources are used to form by-product acetate and lactate. Engineering of the metabolic pathways of cellulose

utilization in *Clostridium thermocellum* is necessary to minimize the lactate and acetate production and make energy and carbon flows favorable to ethanol formation.

Acetate kinase is an important enzyme in the metabolic pathway of cellulose utilization to form acetate in *Clostridium thermocellum*, which is encoded by the *ack* gene.

5 Inactivation of the *ack* gene may interrupt acetate kinase, leading to reduction or elimination of acetate.

Lactate dehydrogenase is an important enzyme in the metabolic pathway of cellulose utilization to form lactate in *Clostridium thermocellum*, which is encoded by the *ldh* gene. Inactivation of the *ldh* gene may interrupt lactate dehydrogenase, leading to  
10 reduction or elimination of lactate generation.

*Inactivation of the ack Gene in C. thermocellum based on the plasmid pIKM1*

To knock out the *ack* gene, a vector is constructed on the multiple cloning sites (MCS) of the plasmid pIKM1, in which the *cat* gene, encoding chloramphenicol acetyltransferase, is inserted into a DNA fragment of 3055 bp, involving the *ack* and the *pta*  
15 genes (encoding phosphotransacetylase), leading to knockout of 476 bp of the *ack* gene and 399 bp of the *pta* gene, and forming 1025 bp and 1048 bp flanking regions on both sides of the *mLs* gene respectively (Figure 7). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and *Staphylococcus aureus*, and is anticipated to be stably replicated in Gram positive strains of bacteria, including *Clostridium thermocellum*. The  
20 sequence of the *ack* knockout vector constructed on plasmid pIKM1 is set forth as SEQ ID NO:1.

*Inactivation of the ack Gene in C. thermocellum based on the replicative plasmid pNW33N*

To knock out the *ack* gene, a vector is constructed on the multiple cloning sites (MCS) of the replicative plasmid pNW33N, in which the macrolide, lincosamide, and  
25 streptogramin B (MLS<sub>B</sub>) resistant gene *mLs* is inserted into a DNA fragment of 3345 bp, which includes the *ack* gene, the *pta* gene (encoding phosphotransacetylase) and an unknown upstream gene, leading to knockout of 855 bp of the *ack* gene and formation of flanking regions of 1195 bp and 1301 bp on either side of the *mLs* gene (Figure 8). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and  
30 *Staphylococcus aureus*, and is anticipated to be stably replicated in Gram positive strains of

bacteria, including *Clostridium thermocellum*. The sequence of the *ack* knockout vector constructed on plasmid pNW33N is set forth as SEQ ID NO:2.

*Inactivation of the ldh Gene in C. thermocellum based on the plasmid pIKM1*

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the plasmid pIKM1, in which the *cat* gene, encoding chloramphenicol acetyltransferase, is inserted into a DNA fragment of 3188 bp, involving the *ldh* and the *mdh* gene (encoding malate dehydrogenase), leading to knockout of a DNA fragment of 1171 bp, including part of the *ldh* and *mdh* genes, and forming 894 bp and 1123 bp flanking regions on both sides of the *mLs* gene, respectively (Figure 9). The sequence of the *ldh* knockout vector constructed on plasmid pIKM1 is set forth as SEQ ID NO:3.

*Inactivation of the ldh Gene in C. thermocellum based on plasmid pNW33N*

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the replicative plasmid pNW33N, in which the macrolide, lincosamide, and streptogramin B (MLS<sub>B</sub>) resistant gene *mLs* is inserted into a DNA fragment of 2523 bp, which includes the *ldh* gene and the *mdh* gene (encoding malate dehydrogenase), leading to knocking out of a fragment of 489 bp of the *ldh* gene and formation of flanking regions of 1034 bp and 1000 bp on either side of the *mLs* gene (Figure 10). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and *Staphylococcus aureus*, and is anticipated to be stably replicated in other Gram positive strains of bacteria, including *Clostridium thermocellum*. The sequence of the *ldh* knockout vector constructed on plasmid pNW33N is set forth as SEQ ID NO:4.

*Inactivation of the ldh Gene in Clostridium thermocellum based on plasmid pUC19*

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the pUC19 plasmid, in which a gene encoding chloramphenicol acetyltransferase (the *cat* gene) is inserted into a *ldh* gene fragment of 717 bp, leading to a flanking region of 245 bp and 255 bp on either side of the *cat* gene (Figure 11). pUC19 is an *E. coli* plasmid vector, containing pMB1 origin, which cannot be amplified in Gram positive strains of bacteria, including *Clostridium thermocellum*. A similar vector may be constructed, in which the *mLs* gene is flanked by the *ldh* gene fragments. The sequence of the *ldh* knockout vector constructed on plasmid pUC19 is set forth as SEQ ID NO:5.

*Expression of xylose isomerase and xylulose kinase in C. thermocellum and C. straminisolvans (prophetic example)*

For expression of *xylose isomerase* and *xylulose kinase* in *C. thermocellum*, the *xylose isomerase* and *xylulose kinase* genes were cloned from *T. saccharolyticum* and placed under control of the *C. thermocellum gapDH* promoter. This cassette is harbored in a *C. thermocellum* replicative plasmid based on the pNW33N backbone, resulting in pMU340 (Fig. 35) SEQ ID NO:74. Upon transfer into *C. thermocellum*, the resulting transformation can be assayed for the ability to grow on xylose. Analogous constructs can be created using the *C. kristajanssonii xylose isomerase* and *xylulose kinase* genes. These constructs can be tested for functionality in *C. straminsolvans* as well.

*Expression of pyruvate decarboxylase and alcohol dehydrogenase in C. thermocellum and C. straminisolvans (prophetic example)*

For expression of *pyruvate decarboxylase* and *alcohol dehydrogenase* in *C. thermocellum*, the *pyruvate decarboxylase* genes are cloned from sources *Z. mobilis* and *Z. palmarum* and the *alcohol dehydrogenase* gene is cloned from source *Z. mobilis*. These genes (*pdc* and *adh*) will be expressed as an operon from the *C. thermocellum pta-ack* promoter. This cassette is harbored in a *C. thermocellum* replicative plasmid based on the pNW33N backbone (Figures 36 and 37), SEQ ID NOS:75 and 76. Upon transfer into *C. thermocellum*, the resulting transformation can be screened for enhanced ethanol production and/or aldehyde production to measure the functionality of the expressed enzymes. These constructs will be tested for functionality in *C. straminsolvans* as well.

## EXAMPLE 5

### Fermentation of Avicel® using *C. straminisolvans*

*C. straminisolvans* was used to ferment 1% Avicel® in serum bottles containing CTFUD medium. The product concentration profile and the ratios are shown in Figure 27. About 2 g/L of total products was generated in 3 d with ethanol constituting about 50% of the total products. Figure 27 shows the product concentration profiles for 1% Avicel® using *C. straminisolvans*. The ethanol to acetate ratio is depicted as E/A and the ratio of ethanol to total products is depicted as E/T.

## EXAMPLE 6

**Engineered Group II Introns For Mesophilic and Thermophilic Cellulolytic,  
Xylanolytic Organisms**

Mobile group II introns, found in many bacterial genomes, are both catalytic RNAs  
5 and retrotransposable elements. They use a mobility mechanism known as  
retrotransposition in which the excised intron RNA reverse splices directly into a DNA  
target site and is then reverse transcribed by an intron-encoded protein. The mobile  
*Lactococcus lactis* L1.LtrB group II intron has been developed into genetic tools known as  
Targetron<sup>TM</sup> vectors, which are commercially available from Sigma Aldrich (Catalog #  
10 TA0100). This product and its use are the subject of one or more of U.S. Patent Nos.  
5,698,421, 5,804,418, 5,869,634, 6,027,895, 6,001,608, and 6,306,596 and/or other pending  
U.S. and foreign patent applications controlled by InGex, LLC.

Targetrons cassettes (Figures 28 and 29) which contain all the necessary sequences  
for retro-transposition may be sub-cloned into vectors capable of replication in mesophilic  
15 or thermophilic cellulolytic organisms. The Targetron cassette may be modified by  
replacing the lac promoter with any host- or species-specific constitutive or inducible  
promoters. The cassettes may be further modified through site-directed mutagenesis of the  
native recognition sequences such that the Group II intron is retargeted to insert into genes  
of interest creating genetic knockouts. For example, the group II intron could be redesigned  
20 to knockout lactate dehydrogenase or acetate kinase in any mesophilic or thermophilic  
cellulolytic organism. Table 4 depicts an example of insertion location and primers to  
retarget Intron to *C. cellulolyticum* acetate kinase (SEQ ID NO:21). Table 5 depicts an  
example of insertion location and primers to retarget Intron to *C. cellulolyticum* lactate  
dehydrogenase (SEQ ID NO:21).

25 An example of a vector for retargeting the L1.Ltrb intron to insert in *C. cell.* ack  
gene (SEQ ID NO:21) is depicted in Figure 28. The vector sequence of pMU367 (*C. cell.*  
acetate kinase KO vector) is SEQ ID NO:30.

An example of a vector for retargeting the L1.Ltrb intron to insert in *C. cell.*  
LDH2744 gene (SEQ ID NO:23) is depicted in Figure 29. The vector sequence of pMU367  
30 (*C. cell.* lactate dehydrogenase KO vector) is set for as SEQ ID NO:31.

Table 4

|                                             |                                                                       |
|---------------------------------------------|-----------------------------------------------------------------------|
| Predicted Insertion location (SEQ ID NO:62) | ATTTACCTGGCTGGGAATACTGAGACATAT - intron - GTCATTGAGGCCGTA             |
| IBS1 mutagenic primer (SEQ ID NO:63)        | AAAAAAGCTTATAATTATCCTTA <b>ATTTCTACTAC</b> GTGCGCCCAGATAGGGTG         |
| EBS1d mutagenic primer (SEQ ID NO:64)       | CAGATTGTACAAATGTGGTGATAACAGATAAGTC <b>TACTACTG</b> TAACTTACCTTTCTTTGT |
| EBS2 mutagenic primer (SEQ ID NO:65)        | TGAACGCAAGTTTCTAATTTTCG <b>GTTGAAATC</b> CGATAGAGGAAAGTGTCT           |

Table 5

|                                             |                                                                       |
|---------------------------------------------|-----------------------------------------------------------------------|
| Predicted Insertion location (SEQ ID NO:66) | TTAAATGTTGATAAGGAAGCTCTTTTCAAT - intron - GAAGTTAAGGTAGCA             |
| IBS1 mutagenic primer (SEQ ID NO:67)        | AAAAAAGCTTATAATTATCCTTAG <b>GCTCTCTTCAAT</b> GTGCGCCCAGATAGGGTG       |
| EBS1d mutagenic primer (SEQ ID NO 68)       | CAGATTGTACAAATGTGGTGATAACAGATAAGTC <b>TTCAATG</b> AATACTTACCTTTCTTTGT |
| EBS2 mutagenic primer (SEQ ID NO:69)        | TGAACGCAAGTTTCTAATTTTCG <b>ATTAGAGCT</b> CGATAGAGGAAAGTGTCT           |

5

EXAMPLE 7

Transformation of *Thermoanaerobacter* and *Thermoanaerobacterium* strains (prophetic example)

*Thermoanaerobacter pseudoethanolicus* 39E, *Thermoanaerobacterium saccharolyticum* JW/SL-YS485, *Thermoanaerobacterium saccharolyticum* B6A-RI, and *Thermoanaerobacter* sp. strain 59 will be transformed with the following protocol. Cells are grown at 55 °C in 40 mL of DSMZ M122 media ([http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php)) with the following modifications: 5 g/L cellobiose instead of cellulose, 1.8 g/L K<sub>2</sub>HPO<sub>4</sub>, no glutathione, and 0.5 g/L L-cystiene-

10

HCl until an optical density of 0.6 to 0.8. Cells are then harvested and washed twice with 40 mL 0.2 M cellobiose at room temperature. Cells are re-suspended in 0.2 M cellobiose in aliquots of 100 uL and 0.1 to 1 ug plasmid DNA is added to the sample in a 1 mm gap-width electroportation cuvette. An exponential pulse (Bio-Rad Instruments) of 1.8 kV, 25  $\mu$ F, 200 $\Omega$ , ~ 3-6 ms is applied to the cuvette, and cells are diluted 100-200 fold in fresh M122 and incubated for 12-16 hours at 55°C. The recovered cells are then diluted 25-100 fold in petri-plates with fresh agar-containing media containing a selective agent, such as 200  $\mu$ g/mL kanamycin. Once the media has solidified, plates incubated at 55°C for 24-72 hours for colony formation. Colonies can be tested by PCR for evidence of site-specific recombination.

### EXAMPLE 8

#### Fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains

Table 6 depicts the fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains. Cultures were grown for 24 hours in M122 at 55°C without shaking. The following abbreviations are used in Table 6: Cellobiose (CB), glucose (G), lactic acid (LA), acetic acid (AA), and ethanol (Etoh). Values are in grams per liter. YS485 – *Thermoanaerobacterium saccharolyticum* JW/SL-YS485, B6A-RI – *Thermoanaerobacterium saccharolyticum* B6A-RI, 39E – *Thermoanaerobacter pseudoethanolicus* 39E.

**Table 6**

| Fermentation sample                               | CB   | G    | LA   | AA   | Etoh |
|---------------------------------------------------|------|------|------|------|------|
| YS485 wildtype                                    | 0    | 0    | 0.77 | 1.04 | 1.40 |
| YS485 $\Delta$ L-ldh                              | 0    | 0    | 0    | 0.92 | 1.73 |
| YS485 $\Delta$ pta/ack                            | 2.51 | 0    | 0.75 | 0.06 | 0.62 |
| YS485 $\Delta$ L-ldh, $\Delta$ pta/ack            | 0    | 0    | 0    | 0    | 2.69 |
| B6A-RI wildtype                                   | 0    | 0    | 0    | 1.0  | 1.76 |
| B6A-RI $\Delta$ L-ldh, $\Delta$ pta/ack strain #1 | 0    | 0    | 0    | 0    | 2.72 |
| B6A-RI $\Delta$ L-ldh, $\Delta$ pta/ack strain #2 | 0.45 | 0    | 0    | 0    | 2.49 |
| 39E wildtype                                      | 0.51 | 0    | 1.51 | 0.15 | 1.87 |
| Media                                             | 5.10 | 0.25 | 0    | 0    | 0    |

## EXAMPLE 9

**Construct for Engineering Cellulolytic and Xylanolytic Strains -****Antisense RNA technology example**

A replicative plasmid (Figure 38) carrying an antisense RNA cassette targeting a *C. thermocellum* gene coding for lactate dehydrogenase (Cthe\_1053) was transferred to *C. thermocellum* 1313 by electroporation and thiamphenicol selection. The transformation efficiency observed for this plasmid was equal to that of the parent vector, pMU102. The sequence of the plasmid is shown in SEQ ID NO: 61. The asRNA cassette is depicted in Figure 38 and is organized as follows: (i) the entire 1827 bp cassette is cloned into the multicloning site of pMU102 in the orientation shown in Figure 38, (ii) the native promoter region is contained within the first 600 bp of the cassette, (iii) the first 877 bp of the *ldh* open reading frame are fused to the native promoter in the antisense orientation, (iv) approximately 300 additional bp are included downstream of the asRNA *ldh* region.

The resulting thiamphenicol resistant colonies were screened for altered end product formation by growing standing cultures on M122C media in the presence of 6 ug/mL thiamphenicol (to maintain the plasmid), as shown in Figure 39. A preliminary screen of 9 randomly selected thiamphenicol-resistant transformants showed that 4 cultures exhibited low levels of lactate production relative to wild type. Additionally, a construct carrying antisense RNA directed to both *ldh* genes are to be constructed in order to partially, substantially, or completely delete, silence, inactivate, or down-regulate both genes simultaneously.

## EXAMPLE 10

SEQ ID NOS:44, 45, and 46 are the pyruvate-formate-lyase (aka formate acetyltransferase, EC. 2.3.1.54, pfl) genes from *Thermoanaerobacterium saccharolyticum* YS485, *Clostridium thermocellum* ATCC 27405, and *Clostridium phytofermentans*. Pfl catalyzes the conversion of pyruvate to Acetyl-CoA and formate (Figure 34). Deletion of pfl will result in the elimination of formate production, and could result in a decrease in acetic acid yield in some thermophilic strains, with a resulting increase in ethanol yield.

SEQ ID NOS:47-52, depicted in Figures 40-45, show pfl knockout plasmids, two each for the three organisms listed above. Each organism has a single crossover and double crossover plasmid designed to partially, substantially, or completely delete, silence, inactivate, or down-regulate the pfl enzyme. Single crossover plasmids are designed with a  
5 single DNA sequence (400 bp to 1000 bp) homologous to an internal section of the pfl gene, double crossover plasmids are designed with two DNA sequences (400 to 1000 bp) homologous to regions upstream (5') and downstream (3') to the pfl gene. All plasmids are designed to use the best available antibiotic markers for selection in the given organism. Plasmids can be maintained in *E. coli* and constructed through a DNA synthesis contract  
10 company, such as Codon Devices or DNA 2.0.

#### INCORPORATION BY REFERENCE

All of the U.S. patents and U.S. published patent applications cited herein are hereby incorporated by reference.

#### EQUIVALENTS

15 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

We claim:

1. An isolated nucleic acid molecule comprising the nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof.
2. An isolated nucleic acid molecule comprising a nucleotide sequence which shares at least 80% identity to a nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof.
3. The nucleic acid molecule of claim 2 having at least about 95% sequence identity to the nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof.
4. A genetic construct comprising any one of SEQ ID NOS:1-5, 30-31, and 47-61 operably linked to a promoter expressible in a thermophilic or mesophilic bacterium.
5. A recombinant thermophilic or mesophilic bacterium comprising the genetic construct of claim 4.
6. A vector comprising the nucleic acid molecule of any one of claims 1-3.
7. A host cell comprising the nucleic acid molecule of any one of claims 1-3.
8. The host cell of claim 7, wherein the host cell is a thermophilic or mesophilic bacterial cell.
9. A genetically modified thermophilic or mesophilic microorganism, wherein a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.
10. The genetically modified microorganism according to claim 9, wherein said microorganism is a Gram-negative bacterium or a Gram-positive bacterium.
11. The genetically modified microorganism according to claim 9, wherein said microorganism is a species of the genera *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, or *Anoxybacillus*.

12. The genetically modified microorganism according to claim 9, wherein said microorganism is a bacterium selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zaeae*,  
5 *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*,  
10 *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*,  
15 *Caldicellulosiruptor lactoaceticus*, and *Anaerocellum thermophilum*.
13. The genetically modified microorganism according to claim 9, wherein said microorganism is *Thermoanaerobacterium saccharolyticum*.
14. The genetically modified microorganism according to claim 9, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or  
20 mesophilic microorganism with a native ability to metabolize a hexose sugar; (b) a thermophilic or mesophilic microorganism with a native ability to metabolize a pentose sugar; and (c) a thermophilic or mesophilic microorganism with a native ability to metabolize a hexose sugar and a pentose sugar.
15. The genetically modified microorganism according to claim 9, wherein said  
25 microorganism has a native ability to metabolize a hexose sugar.
16. The genetically modified microorganism according to claim 15, wherein said microorganism is *Clostridium straminisolvans* or *Clostridium thermocellum*.
17. The genetically modified microorganism according to claim 9, wherein said  
30 microorganism has a native ability to metabolize a hexose sugar and a pentose sugar.

18. The genetically modified microorganism according to claim 17, wherein said microorganism is *Clostridium cellulolyticum*, *Clostridium kristjanssonii*, or *Clostridium stercorarium subsp. leptosaprartum*.
19. The genetically modified microorganism of claim 15, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a pentose sugar.
20. The genetically modified microorganism according to claim 9, wherein said microorganism has a native ability to metabolize a pentose sugar.
21. The genetically modified microorganism according to claim 20, wherein said microorganism is selected from the group consisting of *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium polysaccharolyticum*, and *Thermoanaerobacterium thermosaccharolyticum*.
22. The genetically modified microorganism of claim 20, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a hexose sugar.
23. The genetically modified microorganism according to any one of claims 9-22, wherein said organic acid is selected from the group consisting of lactic acid and acetic acid.
24. The genetically modified microorganism according to any one of claims 9-22, wherein said organic acid is lactic acid.
25. The genetically modified microorganism according to any one of claims 9-22, wherein said organic acid is acetic acid.
26. The genetically modified microorganism according to any one of claims 9-22, wherein said first native enzyme is selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase.
27. The genetically modified microorganism according to any one of claims 9-22, wherein said first native enzyme is lactate dehydrogenase.

28. The genetically modified microorganism according to any one of claims 9-22, wherein said first native enzyme is acetate kinase.
29. The genetically modified microorganism according to any one of claims 9-22, wherein said first native enzyme is phosphotransacetylase.
- 5 30. The genetically modified microorganism according to any one of claims 9-29, wherein a second native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof.
- 10 31. The genetically modified microorganism according to claim 30, wherein said second native enzyme is acetate kinase or phosphotransacetylase.
32. The genetically modified microorganism according to claim 30, wherein said second native enzyme is lactate dehydrogenase.
- 15 33. A genetically modified thermophilic or mesophilic microorganism, wherein (a) a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the metabolic production of ethanol, thereby  
20 allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.
34. The genetically modified microorganism of claim 33, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to  
25 metabolize a hexose sugar.
35. The genetically modified microorganism of claim 33, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a pentose sugar.
- 30 36. The genetically modified microorganism of claim 33, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a

hexose sugar; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar and a pentose sugar.

- 5 37. The genetically modified microorganism according to any one of claims 33-36, wherein said organic acid is lactic acid.
38. The genetically modified microorganism according to any one of claims 33-36, wherein said organic acid is acetic acid.
39. The genetically modified microorganism according to claim 33, 34 or 36, wherein  
10 said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH).
40. The genetically modified microorganism according to claim 33 or 35, wherein said second non-native enzyme is xylose isomerase.
41. The genetically modified microorganism according to claim 40, wherein said first  
15 non-native gene corresponds to SEQ ID NOS:6, 10, or 14.
42. The genetically modified microorganism according to claim 33, 35 or 36, wherein said non-native enzyme is xylulokinase.
43. The genetically modified microorganism according to claim 42, wherein said non-native gene corresponds to SEQ ID NOS:7, 11, or 15.
- 20 44. The genetically modified microorganism according to claim 33, 35, or 36, wherein said non-native enzyme is L-arabinose isomerase.
45. The genetically modified microorganism according to claim 44, wherein said non-native gene corresponds to SEQ ID NOS:8 or 12.
46. The genetically modified microorganism according to claim 33, 35 or 36, wherein  
25 said non-native enzyme is L-ribulose-5-phosphate 4-epimerase.
47. The genetically modified microorganism according to claim 46, wherein said non-native gene corresponds to SEQ ID NO:9 or 13.
48. The genetically modified microorganism according to any one of claims 9-47, wherein said microorganism is able to convert at least 60% of carbon from  
30 metabolized biomass into ethanol.

49. The genetically modified microorganism according to claim 9, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose; (b) a thermophilic or mesophilic microorganism with a native ability to hydrolyze xylan;  
5 and (c) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose and xylan.
50. The genetically modified microorganism according to claim 9 or 33, wherein said microorganism has a native ability to hydrolyze cellulose.
51. The genetically modified microorganism according to claim 9 or 33, wherein said  
10 microorganism has a native ability to hydrolyze cellulose and xylan.
52. The genetically modified microorganism of claim 50, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan.
53. The genetically modified microorganism according to claim 9 or 33, wherein said  
15 microorganism has a native ability to hydrolyze xylan.
54. The genetically modified microorganism of claim 53, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose.
55. The genetically modified microorganism according to any one of claims 49-54,  
20 wherein said organic acid is selected from the group consisting of lactic acid and acetic acid.
56. The genetically modified microorganism according to any one of claims 49-54, wherein said organic acid is lactic acid.
57. The genetically modified microorganism according to any one of claims 49-54,  
25 wherein said organic acid is acetic acid.
58. The genetically modified microorganism according to any one of claims 49-54, wherein said first native enzyme is selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase.
59. The genetically modified microorganism according to any one of claims 49-54,  
30 wherein said first native enzyme is lactate dehydrogenase.

60. The genetically modified microorganism according to any one of claims 49-54, wherein said first native enzyme is acetate kinase.
61. The genetically modified microorganism according to any one of claims 49-54, wherein said first native enzyme is phosphotransacetylase.
- 5 62. The genetically modified microorganism according to any one of claims 50-61, wherein a second native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof.
- 10 63. The genetically modified microorganism according to claim 62, wherein said second native enzyme is acetate kinase or phosphotransacetylase.
64. The genetically modified microorganism according to claim 62, wherein said second native enzyme is lactate dehydrogenase.
- 15 65. A genetically modified thermophilic or mesophilic microorganism, wherein (a) a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the hydrolysis of a polysaccharide, thereby  
20 allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.
- 25 66. The genetically modified microorganism of claim 65, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose.
67. The genetically modified microorganism of claim 65, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze xylan.
- 30 68. The genetically modified microorganism of claim 65, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze

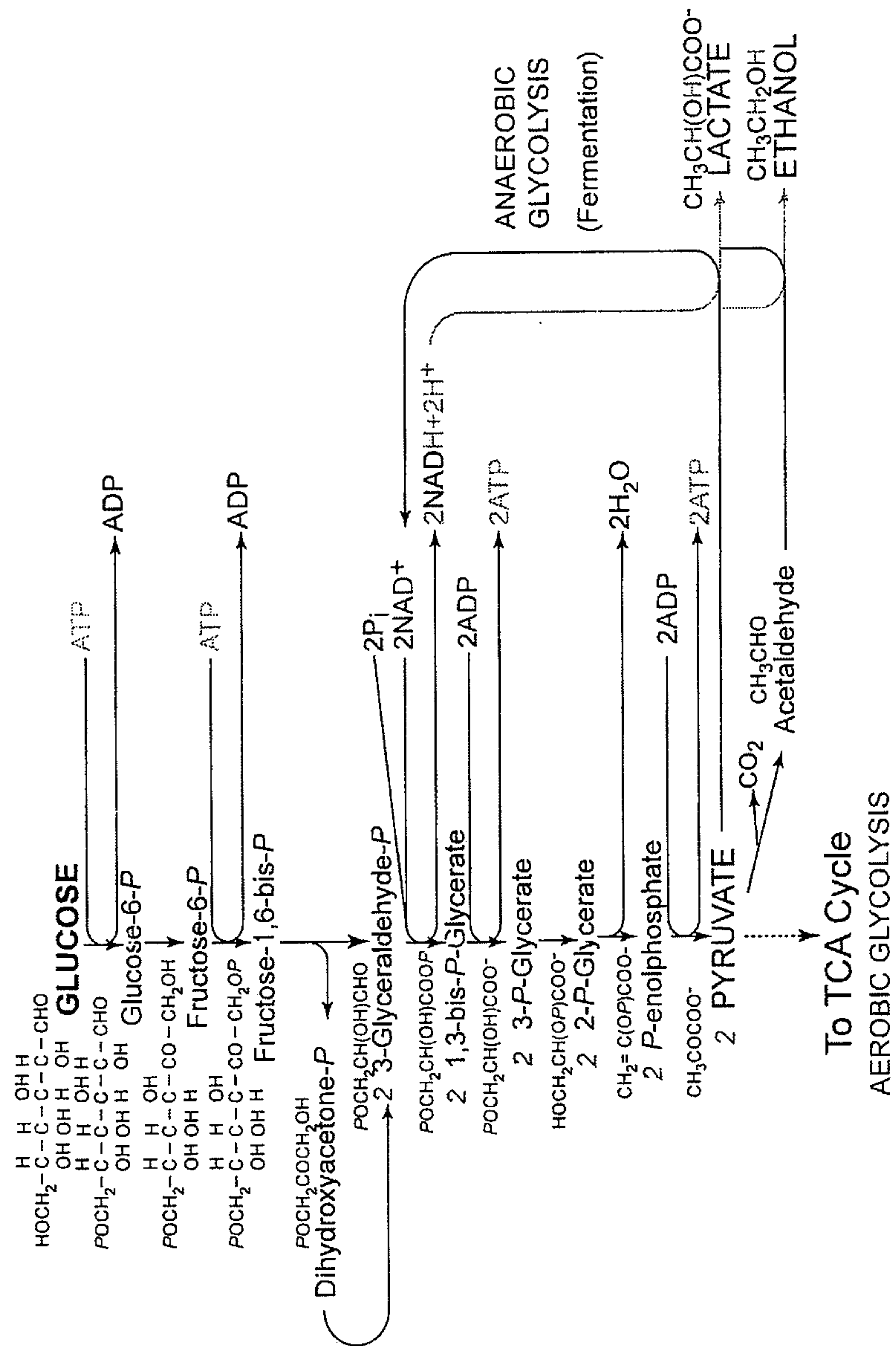
cellulose; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose and xylan.

- 5 69. The genetically modified microorganism according to any one of claims 65-68, wherein said organic acid is lactic acid.
70. The genetically modified microorganism according to any one of claims 65-68, wherein said organic acid is acetic acid.
71. The genetically modified microorganism according to claim 65, 66, or 68, wherein  
10 said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH).
72. The genetically modified microorganism according to any one of claims 49-71, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol.
- 15 73. The genetically-modified microorganism according to any one of claims 9, 11, 12, 14, 16, 18, 19, 22, 33, 34, 35, 36, 49, 65, 66, 67, and 68, wherein said microorganism is mesophilic.
74. The genetically-modified microorganism according to any one of claims 9, 11, 12, 13, 14, 19, 21, 22, 33, 34, 35, 36, 49, 65, 66, 67, and 68, wherein said  
20 microorganism is thermophilic.
75. A process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to any one of claims 9-48.
76. The process of claim 75, wherein said lignocellulosic biomass is selected from the  
25 group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge,  
30 sawdust, hardwood, softwood, and combinations thereof.

77. The process of claim 75, wherein said lignocellulosic biomass is selected from the group consisting of corn stover, sugarcane bagasse, switchgrass, and poplar wood.
78. The process of claim 75, wherein said lignocellulosic biomass is corn stover.
79. The process of claim 75, wherein said lignocellulosic biomass is sugarcane bagasse.
- 5 80. The process of claim 75, wherein said lignocellulosic biomass is switchgrass.
81. The process of claim 75, wherein said lignocellulosic biomass is poplar wood.
82. The process of claim 75, wherein said lignocellulosic biomass is willow.
83. The process of claim 75, wherein said lignocellulosic biomass is paper sludge.
84. A process for converting lignocellulosic biomass to ethanol, comprising contacting  
10 lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to any one of claims 49-74.
85. The process of claim 84, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse,  
15 sugarcane straw agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, and combinations thereof.
86. The process of claim 84, wherein said lignocellulosic biomass is selected from the  
20 group consisting of corn stover, sugarcane bagasse, switchgrass, and poplar wood.
87. The process of claim 84, wherein said lignocellulosic biomass is corn stover.
88. The process of claim 84, wherein said lignocellulosic biomass is sugarcane bagasse.
89. The process of claim 84, wherein said lignocellulosic biomass is switchgrass.
90. The process of claim 84, wherein said lignocellulosic biomass is poplar wood.
- 25 91. The process of claim 84, wherein said lignocellulosic biomass is willow.
92. The process of claim 84, wherein said lignocellulosic biomass is paper sludge.

Figure 1

## GLYCOLYSIS







[illegible]



## Figure 6

16 Apr 2007

## Alignment Results

Alignment: Global DNA alignment against reference molecule  
Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)  
Reference molecule: Clostridium thermocellum, Region 1 to 1509  
Number of sequences to align: 7  
Total length of aligned sequences with gaps: 1767 bps

| Sequence                   | Start | End  | #Match | NonMatch | %Match |
|----------------------------|-------|------|--------|----------|--------|
| Clostridium thermocellum   | 1     | 1509 |        |          |        |
| Clostridium cellulolyticum | 1     | 1642 | 1372   | 274      | 83     |
| Thermoanaerobacterium sacc | 1     | 1552 | 1286   | 278      | 82     |
| C. stercorarium            | 1     | 1519 | 1374   | 155      | 89     |
| C. stercorarium II         | 1     | 1500 | 1339   | 194      | 87     |
| Caldicellulosiruptor krist | 1     | 1508 | 1250   | 300      | 80     |
| C. phytofermentans         | 1     | 1371 | 1115   | 403      | 73     |

16 Apr 2007

## Alignment Results

Alignment: Global DNA alignment against reference molecule

Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

Reference molecule: Clostridium thermocellum, Region 1 to 1509

Number of sequences to align: 7

Total length of aligned sequences with gaps: 1767 bps

Settings: Similarity significance value cutoff: &gt;= 60%

## Summary of Percent Matches:

|      |                                       |           |             |     |     |
|------|---------------------------------------|-----------|-------------|-----|-----|
| Ref: | Clostridium thermocellum              | 1 to 1509 | ( 1509 bps) | --  |     |
| 2:   | Clostridium cellulolyticum            | 1 to 1642 | ( 1642 bps) | 83% |     |
| 3:   | Thermoanaerobacterium saccharolyticum | 1 to 1552 | ( 1552 bps) |     | 82% |
| 4:   | C. stercoararium                      | 1 to 1519 | ( 1519 bps) | 89% |     |
| 5:   | C. stercoararium II                   | 1 to 1500 | ( 1500 bps) | 87% |     |
| 6:   | Caldicellulosiruptor kristjanssonii   | 1 to 1508 | ( 1508 bps) |     | 80% |
| 7:   | C. phytofermentans                    | 1 to 1371 | ( 1371 bps) | 73% |     |

```

Clostridium      1 tttgatecctgggtcaggacgaacgcctggcgccgctgacctaacacatgcaagtgcgagcg---
Clostridium      1 --tgatcctgngacaggnegagggcctgngggcgctgacctaacacatggcgagtcgagcg---
Thermoanaero     1 tttgatcccggtcaggacgaacgcctggcgccgctgacctaacacatgcaagtgcgagcgatc
C. stercoar      1 tttgatcctgggtcaggacgaacgcctggcgccgctgacctaacacatgcaagtgcgaacg---
C. stercoar      1 -----cctgggtcaggacgaacgcctggcgccgctgacctaacacatgcaagtgcgaacg---
Caldicellulo     1 -----gggtcaggacgaacgcctggcgccgctgacctaacgcatgcaagtgcgagcg---
C. phytoferm     1 -----

```

```

Clostridium      58 -gg----gata-----
Clostridium      56 -ga----gttaccttttagcncctgagtattcttgganatgatgctgncccgacagcgctcat
Thermoanaero     61 cgg----nact-----
C. stercoar      58 -ggatccgtgt-----
C. stercoar      52 -ggatccgtgt-----
Caldicellulo     49 -ga----gatg-----
C. phytoferm     1 -----

```

```

Clostridium      64 -----taccg--gaa-----g-gttt---a-
Clostridium      111 ccnnnaacaaccttat--gaaatatattagttggagttttgcatcacgcg-tttt---a-
Thermoanaero     68 -----caat--taa-----gcgctt---a-
C. stercoar      68 -----taccg--gag-----g-tcttttggg-
C. stercoar      62 -----taccg--gag-----g-tcttcggg-
Caldicellulo     55 -----gtagctgaa-----g-gtga---tg
C. phytoferm     1 -----

```

```

Clostridium      77 --ccgga-----a-----g-ta-----
Clostridium      164 --tcaaagtgtcaacacataata-----g-tagaagagaatgttcagtg
Thermoanaero     82 --cagaa-----aaagagagagaaantgag-ta-----
C. stercoar      84 --ccgaa-----g-----t-gg-----
C. stercoar      78 --ccgaa-----g-----t-gg-----
Caldicellulo     71 agctggg-----a-----gcta-----
C. phytoferm     1 -----

```

```

Clostridium      86 -----t-atcc-----t---agcgaggcaggggtgagtaacgcggtgg
Clostridium      205 ctgaaggt-aact-----t---agcgaggcaggggtgagtaacgcggtgg
Thermoanaero     107 -----a-acgcaaagttgagtgccggat---agcgaggcaggggtgagtaacgcggtgg
C. stercoar      93 -----c-atgg-----tgagagtggcggaagggcgagtaacgcggtga
C. stercoar      87 -----c-atgg-----tgagagtggcggaagggcgagtaacgcggtga
Caldicellulo     83 -----tcatct-----t---agcgaggcaggggtgagtaacgcggtga
C. phytoferm      1 -----ct-----t---agtggcggaaggggtgagtaacgcggtgg

```

|              |     |                                                                                 |
|--------------|-----|---------------------------------------------------------------------------------|
| Clostridium  | 119 | gtaaectacctcatacagggggataaacaaggggaaactgtgcttaataccgcata-----                   |
| Clostridium  | 245 | gcaacgtgectgttaacagggggataaacaaggggaaactgtgcttaataccgcata-----                  |
| Thermoanaero | 156 | acaatgtacccctgtagtttggggttaacacccctcgaaagggttgcttaataccgcata-----               |
| C. stercorar | 129 | gcaacgtgectgttaggtgggggataaacaaggggaaactgtgcttaataccgcata-----                  |
| C. stercorar | 123 | gcaacgtgectgttaggtgggggataaacaaggggaaactgtgcttaataccgcata-----                  |
| Caldicellulo | 117 | gcaacgtacccctcagcacggggataaacaagctcgaaagggttgcttaataccgcata-----                |
| C. phytoferm | 31  | gtaaectgectcatacagggggataaacaagctcgaaagggttgcttaataccgcata-----                 |
| Clostridium  | 174 | a-----cggg-----gcg-----gcaacgtcctggttaacaaaggaga-----a                          |
| Clostridium  | 300 | a-----caca-----acgaagaagcatttctnttgggtgcaaaaggagc-----a                         |
| Thermoanaero | 211 | a-----tgtcaagaagtgt-----gcaacacttttgaagaaggaga-----a                            |
| C. stercorar | 184 | agaccacagt-----gac-----gcaatgtacagtggt-aaagctg-----                             |
| C. stercorar | 178 | agaccacagt-----gac-----gcaatgtcacagtggtaaaagctg-----                            |
| Caldicellulo | 177 | a-----cggc-----atc-----gcaatggtgctgtggtgaaagggttagccnagaggcta                   |
| C. phytoferm | 86  | a-----tata-----gcgaaccgcatgattttgctatcaat-att-----t                             |
| Clostridium  | 207 | atccggtatgagatgggcccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag                 |
| Clostridium  | 338 | atccggtgacagatgggcccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag                 |
| Thermoanaero | 249 | atccggtataggaatgaggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag               |
| C. stercorar | 218 | agggcgcataggaatggggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag               |
| C. stercorar | 214 | agggcgcataggaatggggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag               |
| Caldicellulo | 222 | tnccggtggtgggtgggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag                 |
| C. phytoferm | 123 | at-aaggtatgagatgggcccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag                |
| Clostridium  | 267 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| Clostridium  | 398 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| Thermoanaero | 309 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| C. stercorar | 278 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| C. stercorar | 274 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| Caldicellulo | 282 | gggacgatcgggtgagccgaactgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag |
| C. phytoferm | 182 | gggacgatcagtagccggttgagaggttggtccgggtccgaatagctaggttggtgaggttaacgggtcaaccaag    |
| Clostridium  | 327 | agactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| Clostridium  | 458 | agactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| Thermoanaero | 369 | agactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| C. stercorar | 338 | agactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| C. stercorar | 334 | agactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| Caldicellulo | 342 | acactccctacgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| C. phytoferm | 242 | nnactnctnccgggagggagcagcagtggtggaatattggcgaatgggggaaaccctgacccagc               |
| Clostridium  | 387 | aacggcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                   |
| Clostridium  | 518 | aacggcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                   |
| Thermoanaero | 429 | gaagcgcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                  |
| C. stercorar | 398 | gaagcgcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                  |
| C. stercorar | 394 | gaagcgcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                  |
| Caldicellulo | 402 | gaagcgcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                  |
| C. phytoferm | 302 | gaagcgcggtggaagggaagggttgggggttgtaaaacttcttggattgggggacgaaggga                  |
| Clostridium  | 447 | agtgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |
| Clostridium  | 578 | aatgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |
| Thermoanaero | 486 | agtgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |
| C. stercorar | 456 | -atgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |
| C. stercorar | 452 | -atgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |
| Caldicellulo | 461 | gaaggtaggtacccgttttaaaagccaggggttaactacgtgccagcagccgggttaatacgg                 |
| C. phytoferm | 359 | aatgaagggtacccagaagaacaaagccaggggttaactacgtgccagcagccgggttaatacgg               |

|              |     |                                                                   |
|--------------|-----|-------------------------------------------------------------------|
| Clostridium  | 507 | taggtggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcggga-tg-c-a    |
| Clostridium  | 638 | taggtggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcgggaatg-t-a    |
| Thermoanaero | 544 | tagggggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcgggaatg-t-a    |
| C. stercorar | 515 | taggtggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcgggaatg-t-a    |
| C. stercorar | 511 | taggtggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcgggaatg-t-a    |
| Caldicellulo | 521 | taggtggcgagcgttgcgcgggaattactgggtgttaaagggcgctaaagcgggaatg-t-a    |
| C. phytoferm | 419 | tannnnnnnagcgttatccgggattactgggtgttaaagggcgctaaagcgggaatg-t-a     |
| Clostridium  | 564 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| Clostridium  | 696 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| Thermoanaero | 602 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| C. stercorar | 573 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| C. stercorar | 569 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| Caldicellulo | 579 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| C. phytoferm | 477 | agtcagatgtgaaatcccggggttaaccccggggctgcattctgaaactgtatctcttga      |
| Clostridium  | 624 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| Clostridium  | 756 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| Thermoanaero | 662 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| C. stercorar | 633 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| C. stercorar | 629 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| Caldicellulo | 639 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| C. phytoferm | 537 | gtgctggagaggaaggggaattcctaggtgtagcgggtgaaatgggtagatattaggagga     |
| Clostridium  | 684 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| Clostridium  | 816 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| Thermoanaero | 722 | ataccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| C. stercorar | 693 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| C. stercorar | 689 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| Caldicellulo | 699 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| C. phytoferm | 597 | acaccagtgccgaagggcggttctggcaggttaactgaagctgagggcggaaggggtggg      |
| Clostridium  | 744 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| Clostridium  | 876 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| Thermoanaero | 782 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| C. stercorar | 753 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| C. stercorar | 749 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| Caldicellulo | 759 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| C. phytoferm | 657 | gagcaaacaggattagataaccctgggtaggtccagcccgtaaacgatggatattagggttag   |
| Clostridium  | 804 | gaggtatcg-acccttctgtgcccggagtttaacacaataagttatccacccctggggagttac  |
| Clostridium  | 936 | gaggtatcg-acccttctgtgcccggagtttaacacaataagttatccacccctggggagttac  |
| Thermoanaero | 842 | gtg----aa-gcatatccggcccggagtttaacacaataagttatccacccctggggagttac   |
| C. stercorar | 813 | gaggtatcg-acccttctgtgcccggagtttaacacaataagttatccacccctggggagttac  |
| C. stercorar | 809 | gaggtatcg-acccttctgtgcccggagtttaacacaataagttatccacccctggggagttac  |
| Caldicellulo | 818 | ggggagaag-aactcttccgtgggttagtttaacacaataagttatccacccctggggagttac  |
| C. phytoferm | 717 | ggggtcnnacaggggttccgtggggcacgttaacacaataagttatccacccctggggagttac  |
| Clostridium  | 863 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| Clostridium  | 995 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| Thermoanaero | 897 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| C. stercorar | 872 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| C. stercorar | 868 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| Caldicellulo | 877 | ggccgcaaggttgaactcaaaaggaattgacggggggcccgcaaaagcagtggaagtatgtg    |
| C. phytoferm | 777 | gttcgcaaggaatgaactcaaaaggaattgacggggggannccgcaaaagcagtggaagtatgtg |

|              |      |                                                                   |
|--------------|------|-------------------------------------------------------------------|
| Clostridium  | 923  | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatccctctgacagctcta   |
| Clostridium  | 1055 | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatataagggaatncggca   |
| Thermoanaero | 957  | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatccacagaatcaggta    |
| C. stercorar | 932  | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatcccccctgacggatgta  |
| C. stercorar | 928  | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatcccccctgacggatgta  |
| Caldicellulo | 937  | gtttaattcgaagcaacggcgaagaaaccttaccaggggtttgacatgccc---gggaacccctg |
| C. phytoferm | 837  | gtttaattcgaannaacggcgaagaaaccttaccaggtctttgacatccctctgacaaccgag   |
| Clostridium  | 983  | gagatag--g---gc---tccctcg---gg-----gcaga-----g---g--agacag        |
| Clostridium  | 1115 | gagatgt--c---gtagccctcg---gg-----actgc-----t---atcacag            |
| Thermoanaero | 1016 | gaaataccag---agt-gcctcga---aa-----gagga-----gctgtg--agacag        |
| C. stercorar | 992  | gagatac--a---tct---ctccgcaaggga-----gcagg-----g---g--agacag       |
| C. stercorar | 988  | gagatac--a---tct---ctccgcaaggga-----gcagg-----g---g--agacag       |
| Caldicellulo | 994  | ccgaaag--gcgggggt--gcctggt---tggttaagaggaaggagcccg---g--acacag    |
| C. phytoferm | 897  | tacgtc--g---gnn--tctctcg---gg-----ncaga-----g---g--ngacag         |
| Clostridium  | 1016 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| Clostridium  | 1150 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| Thermoanaero | 1055 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| C. stercorar | 1028 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| C. stercorar | 1024 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| Caldicellulo | 1042 | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| C. phytoferm | 930  | gtggtgcatggttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc     |
| Clostridium  | 1076 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| Clostridium  | 1210 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| Thermoanaero | 1115 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| C. stercorar | 1088 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| C. stercorar | 1084 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| Caldicellulo | 1102 | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| C. phytoferm | 990  | gcaacccttgcctcagctcctgtgctgagatggttgggttaagttcccgcaacgagc         |
| Clostridium  | 1136 | acaagtccgaaggaaggtggggacgacotcaaa--aatcatgccccttatgtccctgggctac   |
| Clostridium  | 1268 | acaatccgaaggaaggtggggacgacotcaaa--aatcatgccccttatgtccctgggctac    |
| Thermoanaero | 1173 | agaacacgggaaggaaggtggggatgacotcaaa--aatcatgccccttatgtccctgggctac  |
| C. stercorar | 1146 | agaagtccgaaggaaggtggggatgacotcaaa--aatcatgccccttatgtccctgggctac   |
| C. stercorar | 1142 | agaagtccgaaggaaggtggggatgacotcaaa--aatcatgccccttatgtccctgggctac   |
| Caldicellulo | 1162 | atgagggcgaaggaaggtggggatgacotcaaa--aatcatgccccttatgtccctgggctac   |
| C. phytoferm | 1049 | ataacctggaaggaagcgggaatgaggttnnaatcaatcagggcccnnaatgattcgggctac   |
| Clostridium  | 1196 | acacgtaactaaatggtgctacaaaagggaaggatadcgcgagggtgagcaaatcccca       |
| Clostridium  | 1328 | acacgtaactaaatggtgctacaaaagggaaggatagctgcaaatgaggcaaatcccca       |
| Thermoanaero | 1233 | acacgtaactaaatggtgctacaaaagggaaggatcccgaggagcgaatcccgag           |
| C. stercorar | 1206 | acacgtgctaaatggtgctacaaaagggaaggatcccgaggagcgaatcccgag            |
| C. stercorar | 1202 | acacgtgctaaatggtgctacaaaagggaaggatcccgaggagcgaatcccgag            |
| Caldicellulo | 1222 | acacgtgctaaatggtgctacaaaagggaaggatcccgaggagcgaatcccgag            |
| C. phytoferm | 1109 | acacgtgctaaatggtgctacaaaagggaaggatcccgaggagcgaatcccgag            |
| Clostridium  | 1256 | aaa--gcag--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct   |
| Clostridium  | 1388 | aaa--atag--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct   |
| Thermoanaero | 1293 | aaa--acagg--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct  |
| C. stercorar | 1266 | aag--gtcg--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct   |
| C. stercorar | 1262 | aag--gtcg--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct   |
| Caldicellulo | 1282 | aaaagcac--cccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct     |
| C. phytoferm | 1169 | aaa--ggtca--tcccagttccggatggcagggtgaaatccgcccgaatgaagtccgaattgct  |

|              |      |                                                                 |
|--------------|------|-----------------------------------------------------------------|
| Clostridium  | 1314 | agtaatggcagggtcagcctaactgccctgaataaggttcccgggccttgtacacacccgccc |
| Clostridium  | 1446 | agtaatggtaggtcagctatactgtcctgaataaggttcccgggccttgtacacacccgccc  |
| Thermoanaero | 1352 | agtaatcggggatcagcctagccgggggaataacgttcccgggccttgtacacacccgccc   |
| C. stercorar | 1324 | agtaatggcagggtcagcctaactgccctgaataaggttcccgggccttgtacacacccgccc |
| C. stercorar | 1320 | agtaatggcagggtcagcctaactgccctgaataaggttcccgggccttgtacacacccgccc |
| Caldicellulo | 1341 | agtaatcggggatcagcctagccgggggaataacgttcccgggccttgtacacacccgccc   |
| C. phytoferm | 1228 | agtaatcgggaacagaaagtccgggggaataacgttcccgggccttgtacacacccgccc    |
| Clostridium  | 1374 | tcacaccatgagagctctgcaaacccgaagtc-a-tagtc-t-aaccgcaaggaggggcgc   |
| Clostridium  | 1506 | tcacaccatgagagctctgcaaacccgaagtcga-tagtc-t-aaccgcaaggaggaggct   |
| Thermoanaero | 1412 | tcacaccacgagagctttagaacacccgaagtc-a-gtgacct-aaccgcaaggaggaggc   |
| C. stercorar | 1384 | tcacaccatgagagctctgcaaacccgaagtc-g-tagtc-t-aaccgagaggaggggcgc   |
| C. stercorar | 1380 | tcacaccatgagagctctgcaaacccgaagtc-gtagtc-t-aaccgagaggaggggcgc    |
| Caldicellulo | 1401 | tcacaccatgagagctcagcaaacctgaagac-a-cagga-t-atctg-----           |
| C. phytoferm | 1288 | tcactccatgggagctaggttaacgcaggaagtc-a-gtgac-cyaaccgtaaggaggggagc |
| Clostridium  | 1430 | tcacgaaggtggggcagatgattgggggtgaantcgttaaca---aggtagccgtatcggaa  |
| Clostridium  | 1563 | cgccgaaggtggggccgataattgggtctgaagtccgtaca---aggtagccgtatcggaa   |
| Thermoanaero | 1469 | tcacgaaggtggggtaaatcattgggggtgaagtccgtaca---aggtagccgtatcggaa   |
| C. stercorar | 1440 | cgtcgaaggtggggcaggtcattgggggtgaagtccgtaca---aggtagccgtatcggaa   |
| C. stercorar | 1437 | cgtcgaaggtggggcaccggaagccgt---agcctaaccgagagggggggcgccgtcgaa    |
| Caldicellulo | 1445 | tgttgaaggtggggctgattgattgggggtgaagtccgtaca---aggtagccgtacgggaa  |
| C. phytoferm | 1345 | tcacgaaggtggggatctataactgggg-----                               |
| Clostridium  | 1487 | ggggggctggatcacctccttt----                                      |
| Clostridium  | 1620 | ggggggctggatcacctccttt----                                      |
| Thermoanaero | 1526 | ggggggctggatcacctcctttccct                                      |
| C. stercorar | 1497 | ggggggctggatcacctccttt----                                      |
| C. stercorar | 1494 | gggggg-----                                                     |
| Caldicellulo | 1502 | cggggg-----                                                     |
| C. phytoferm |      | -----                                                           |

Figure 7

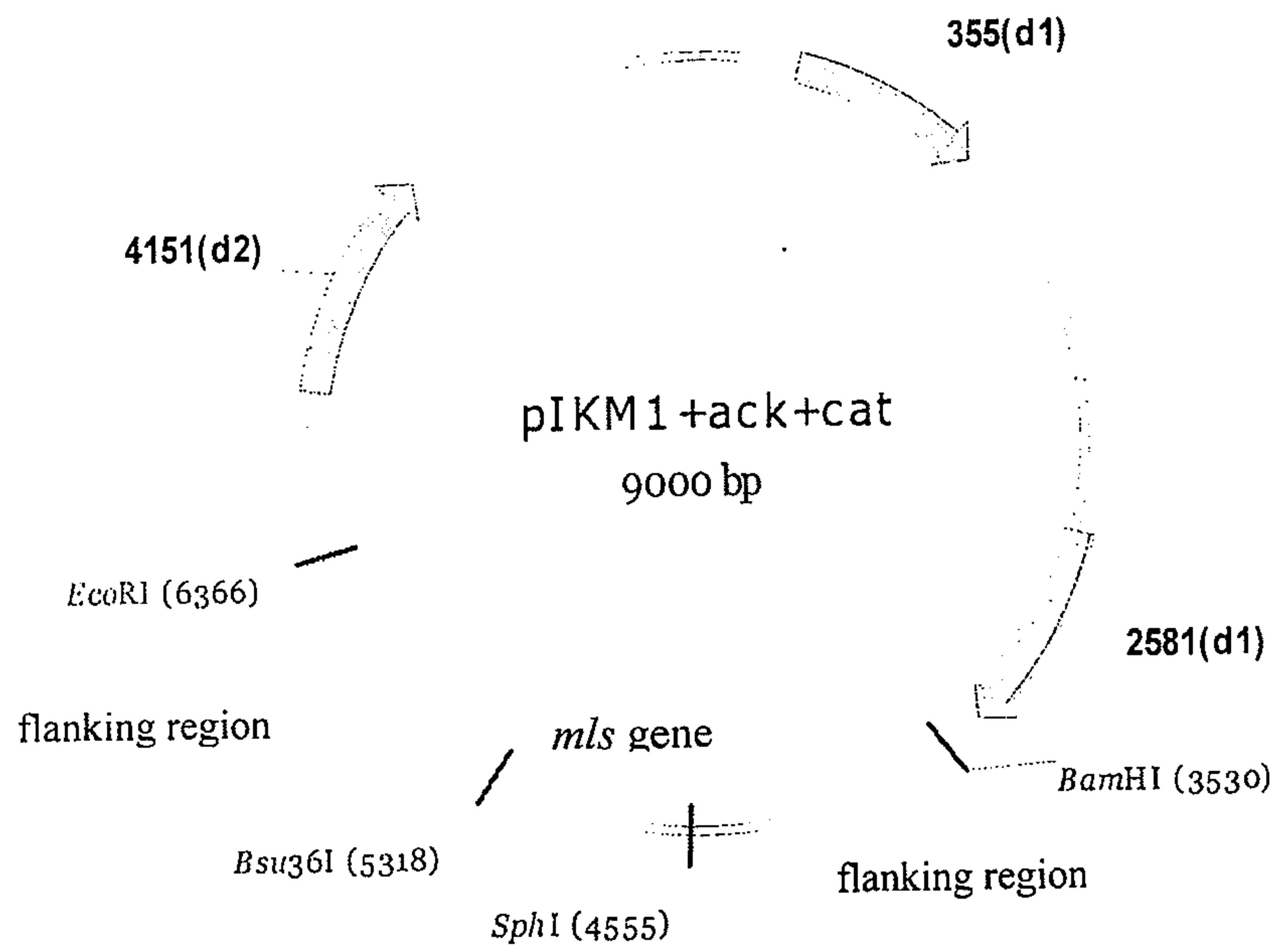


Figure 8

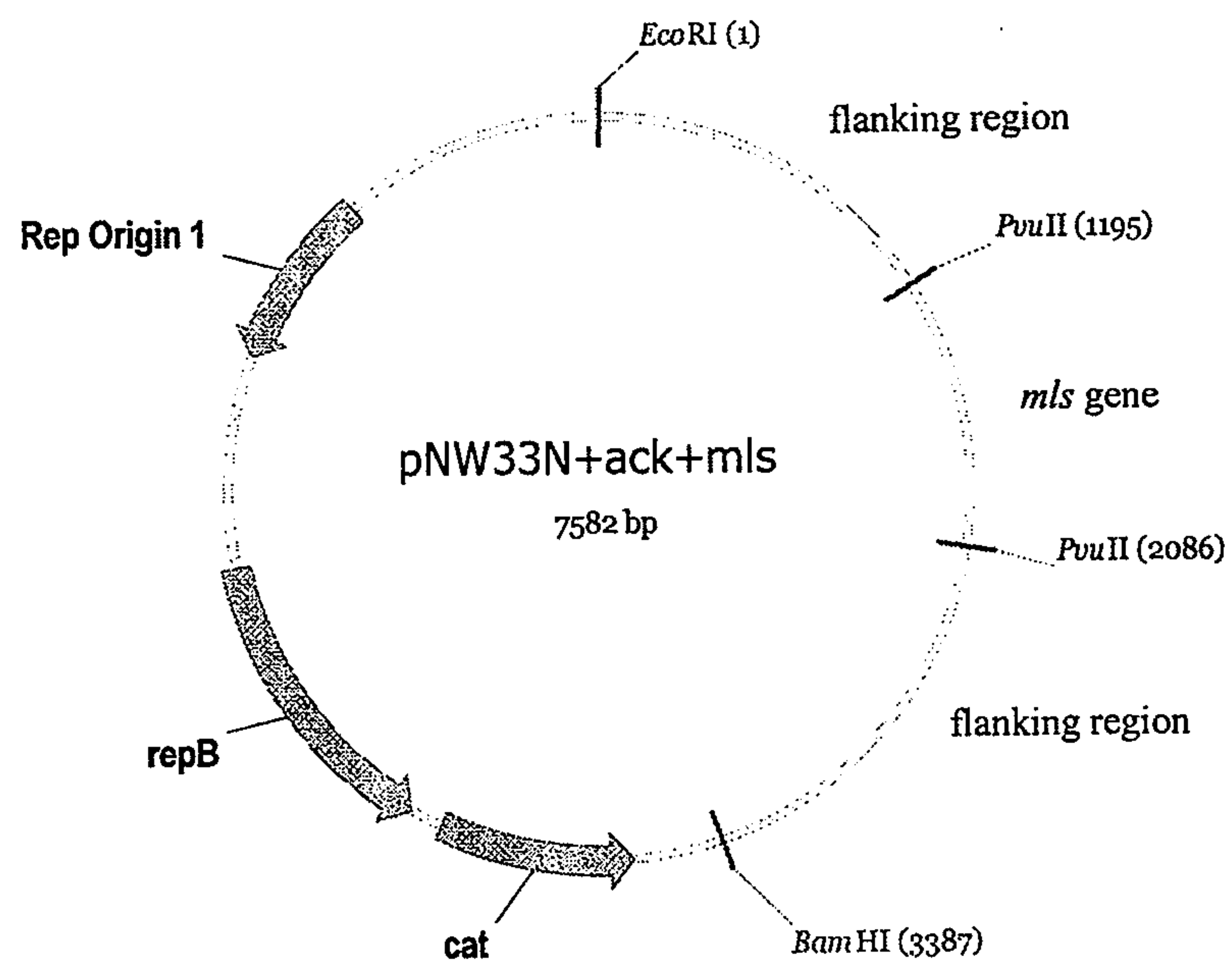


Figure 9

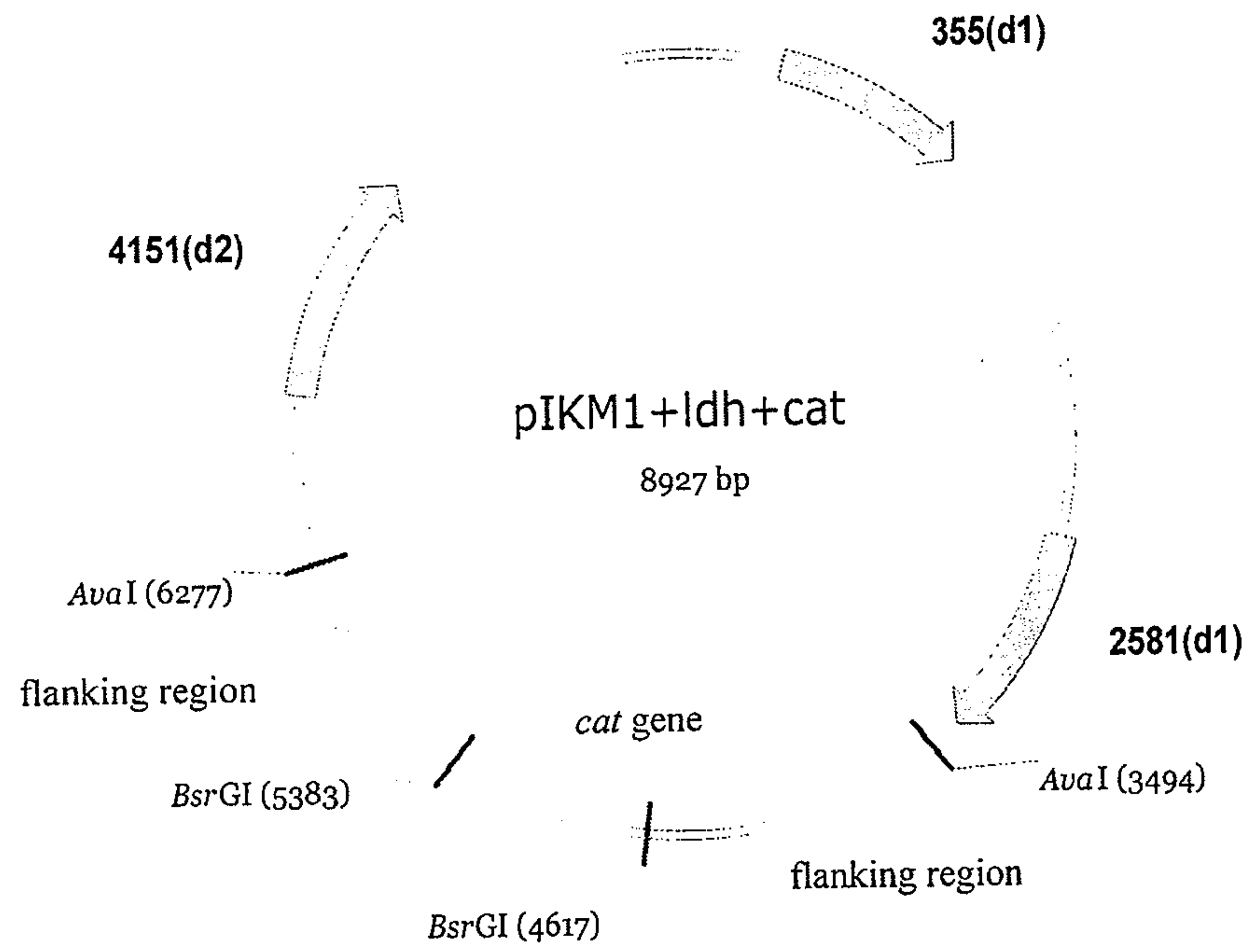


Figure 10

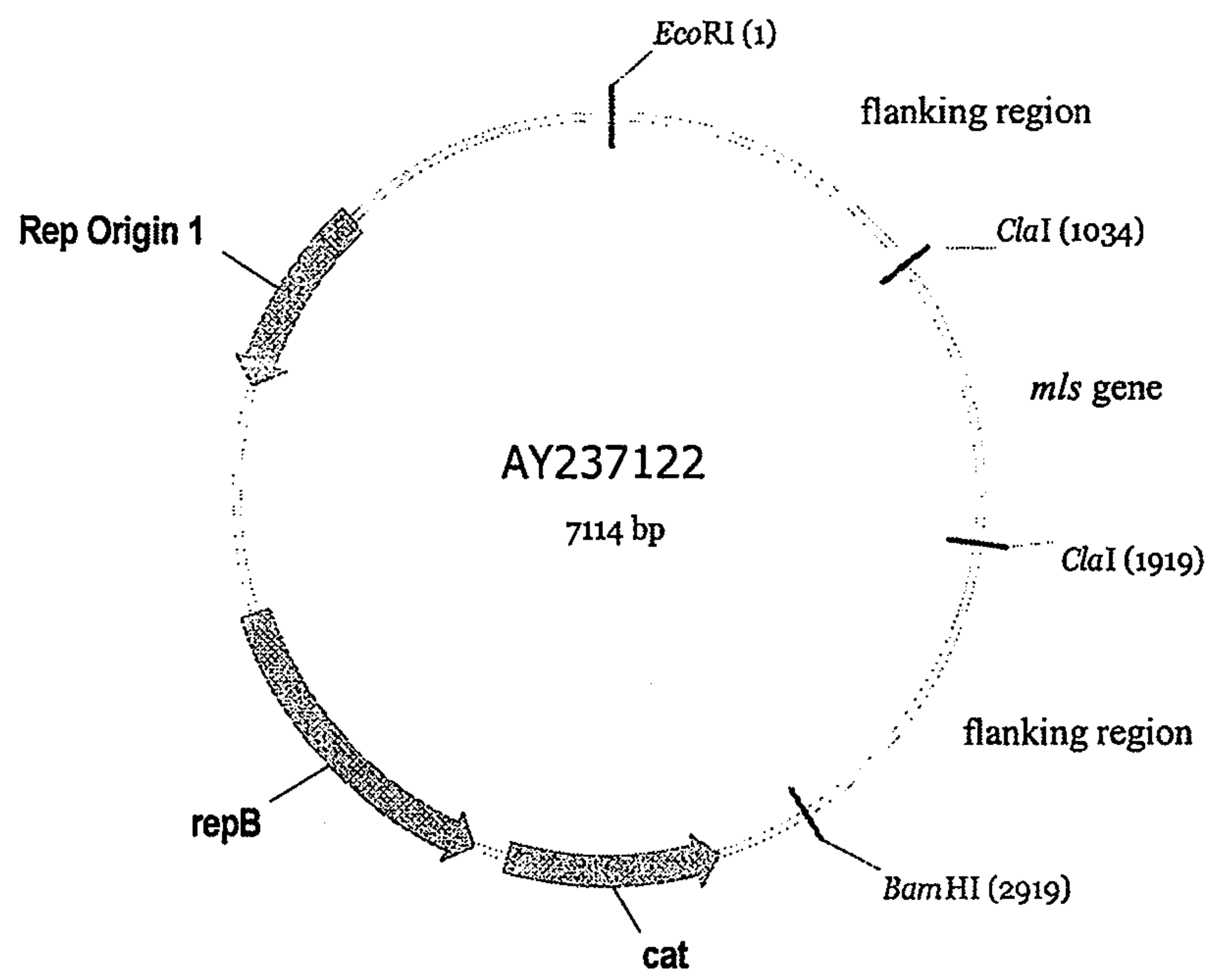


Figure 11

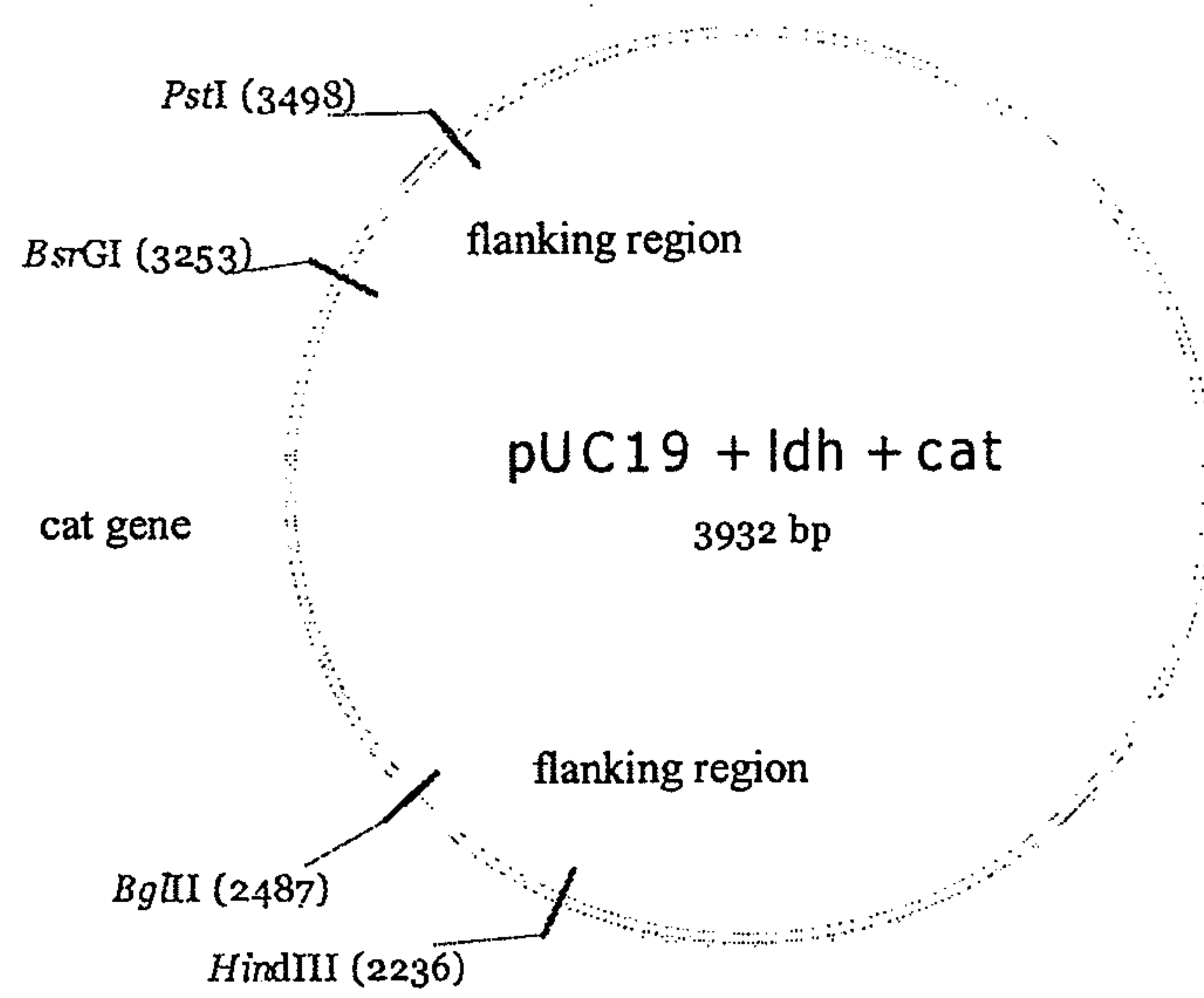


Figure 12A

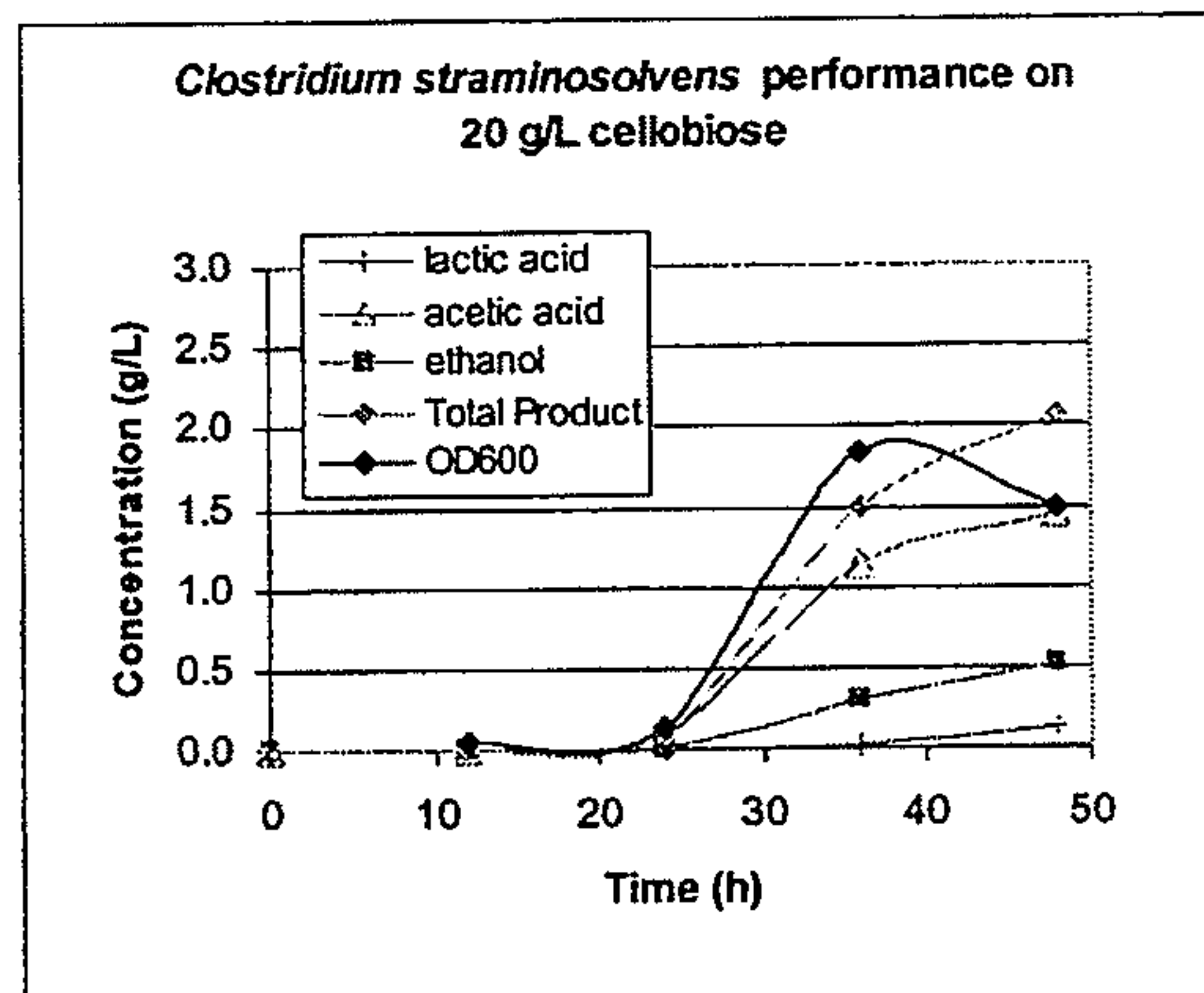


Figure 12B

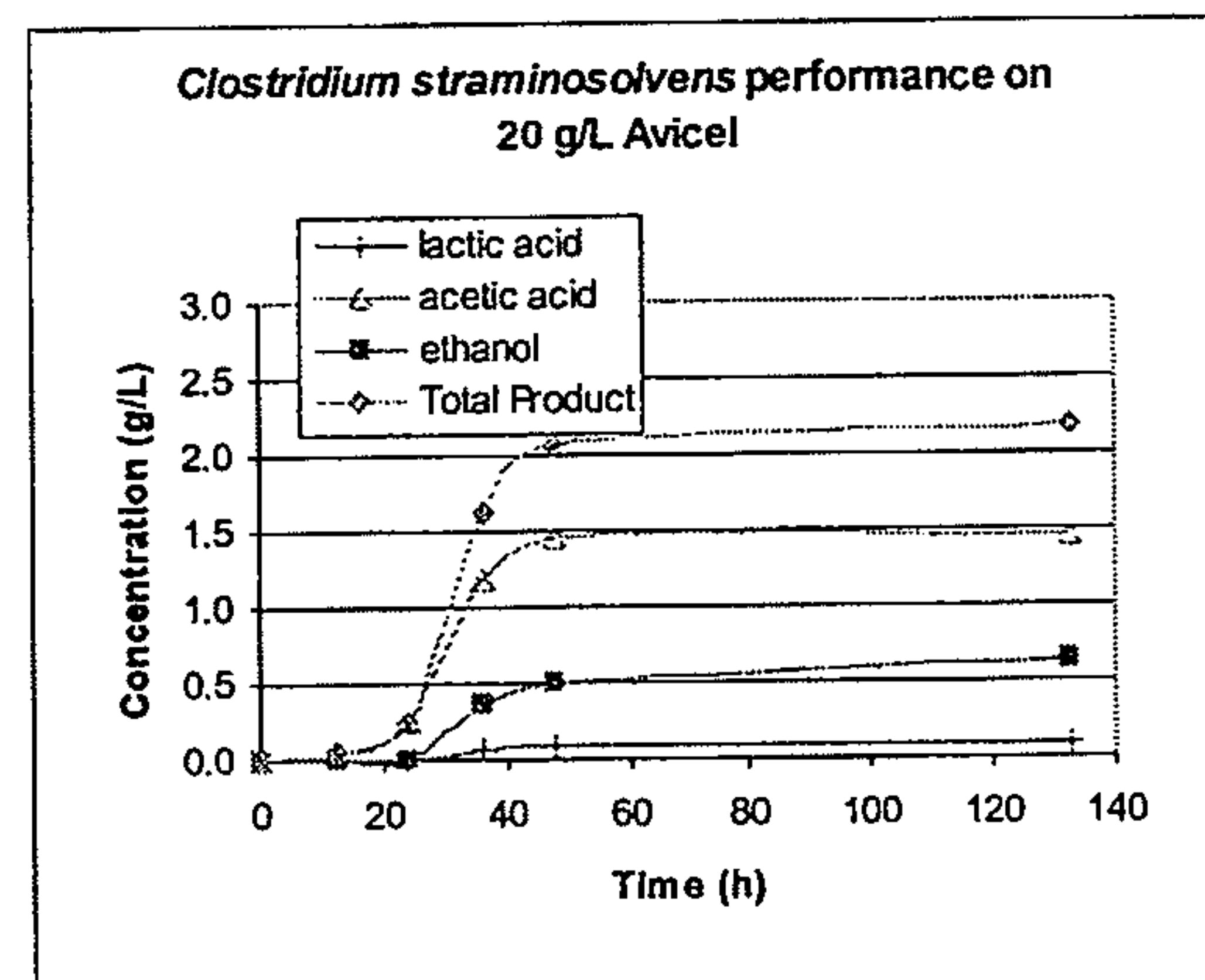


Figure 13A

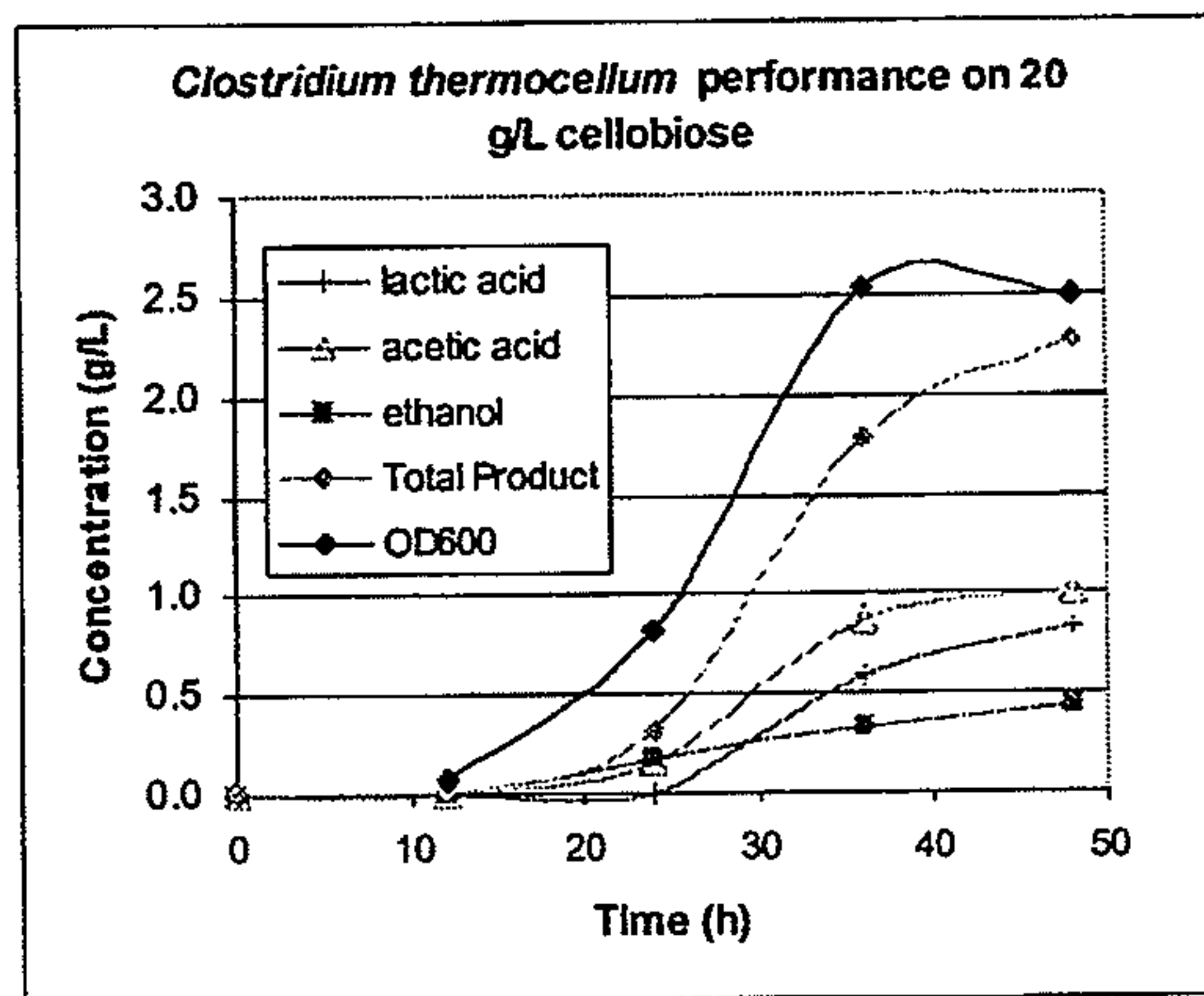


Figure 13B

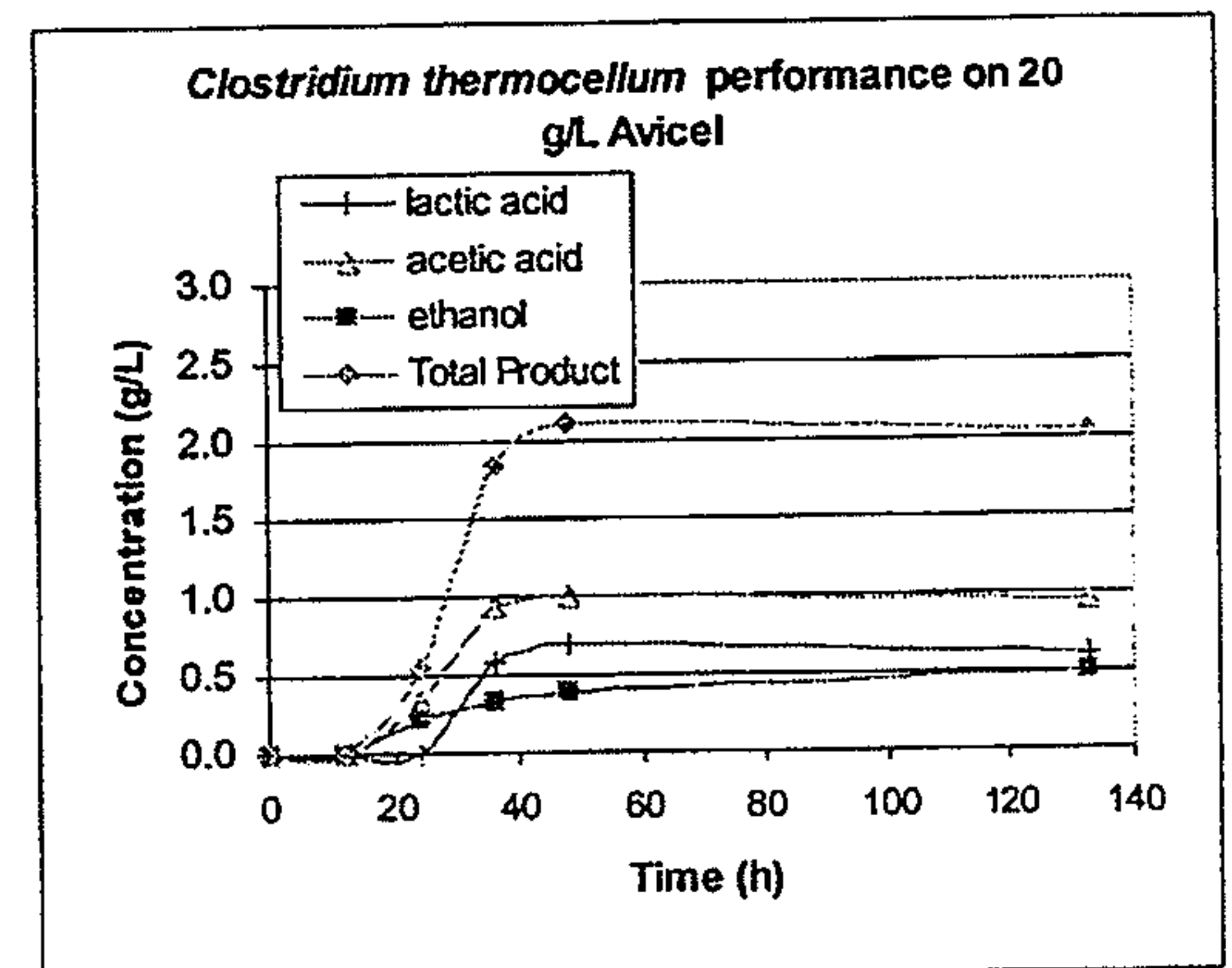


Figure 14A

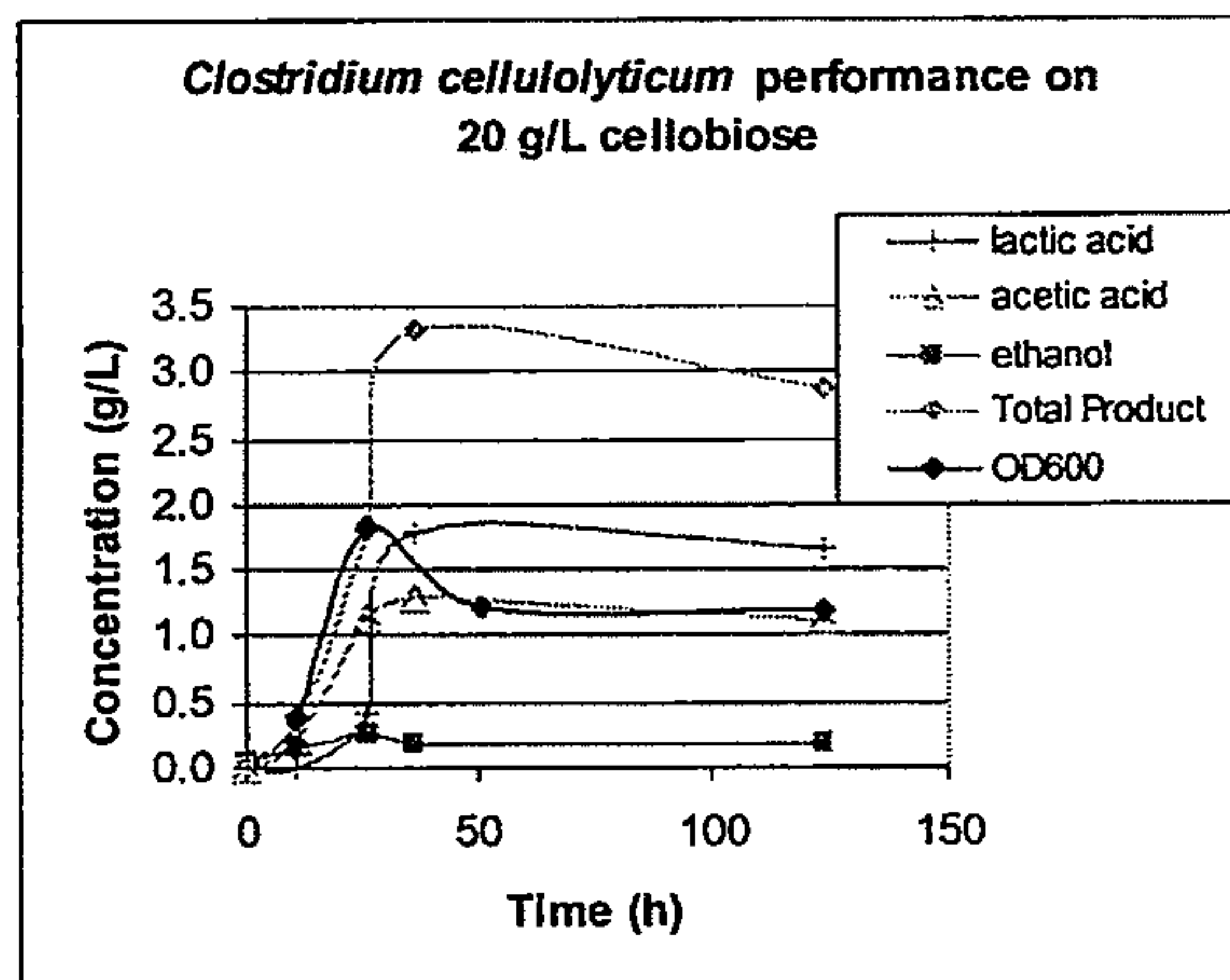


Figure14B

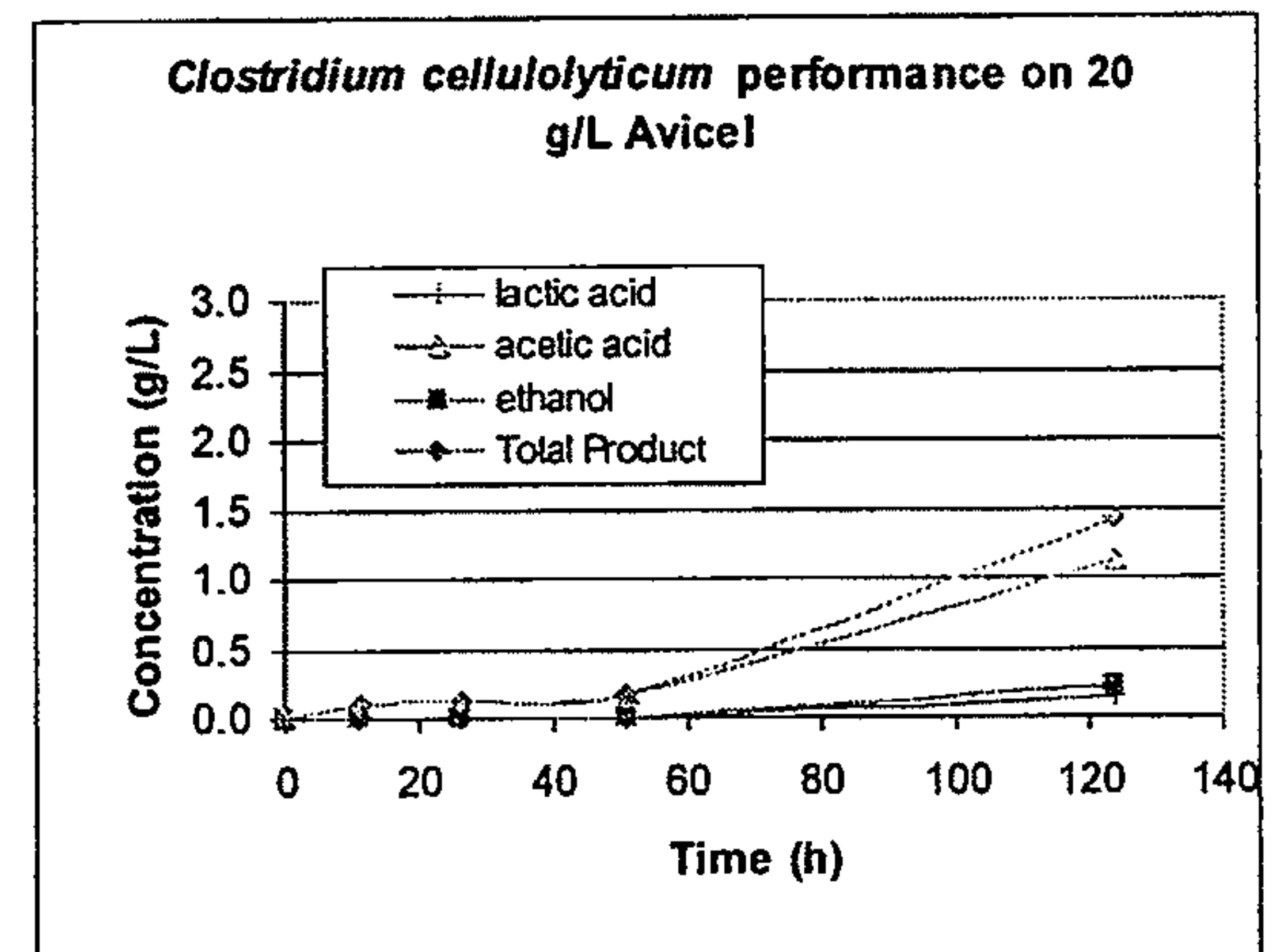


Figure 15A

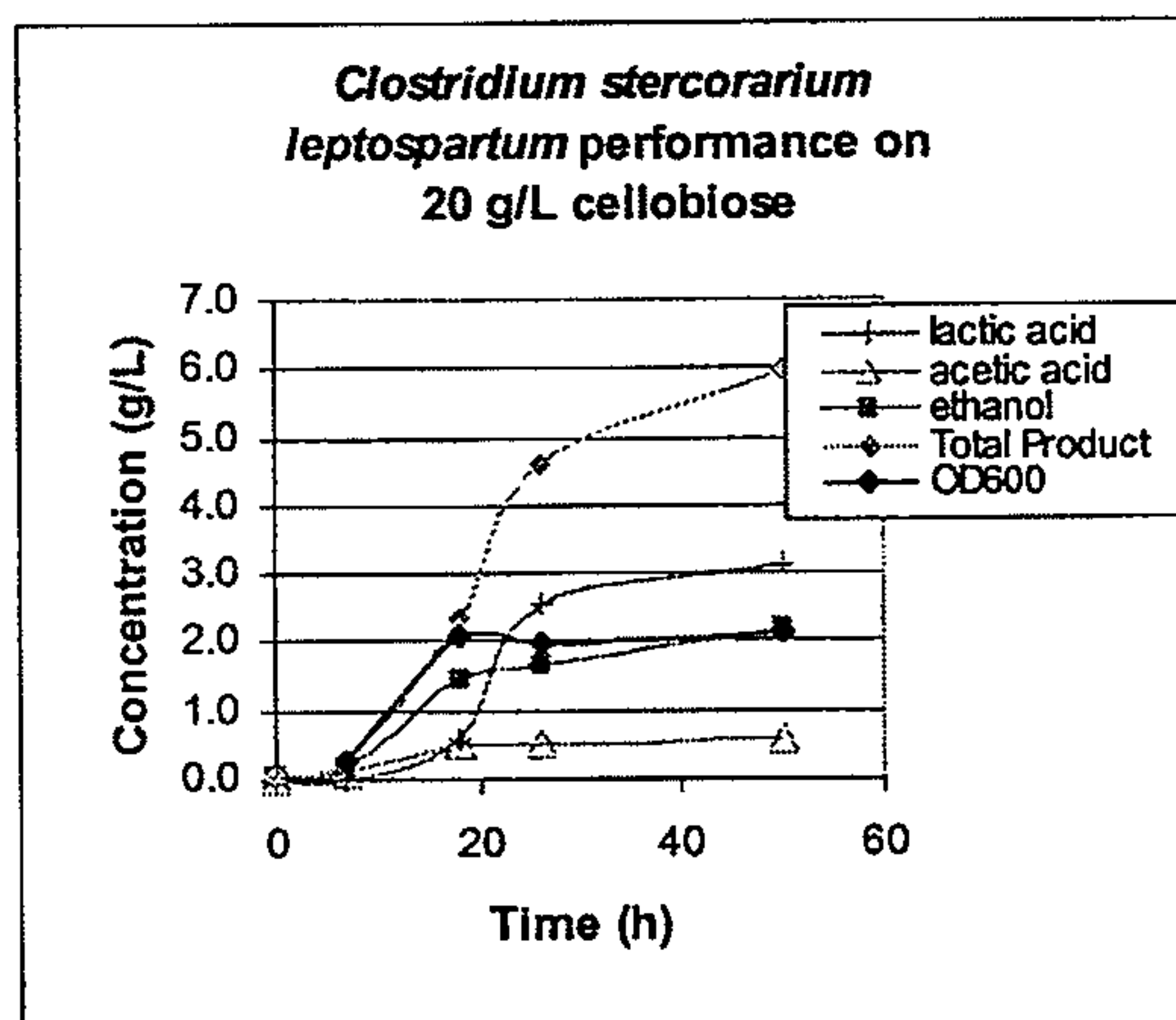


Figure 15B

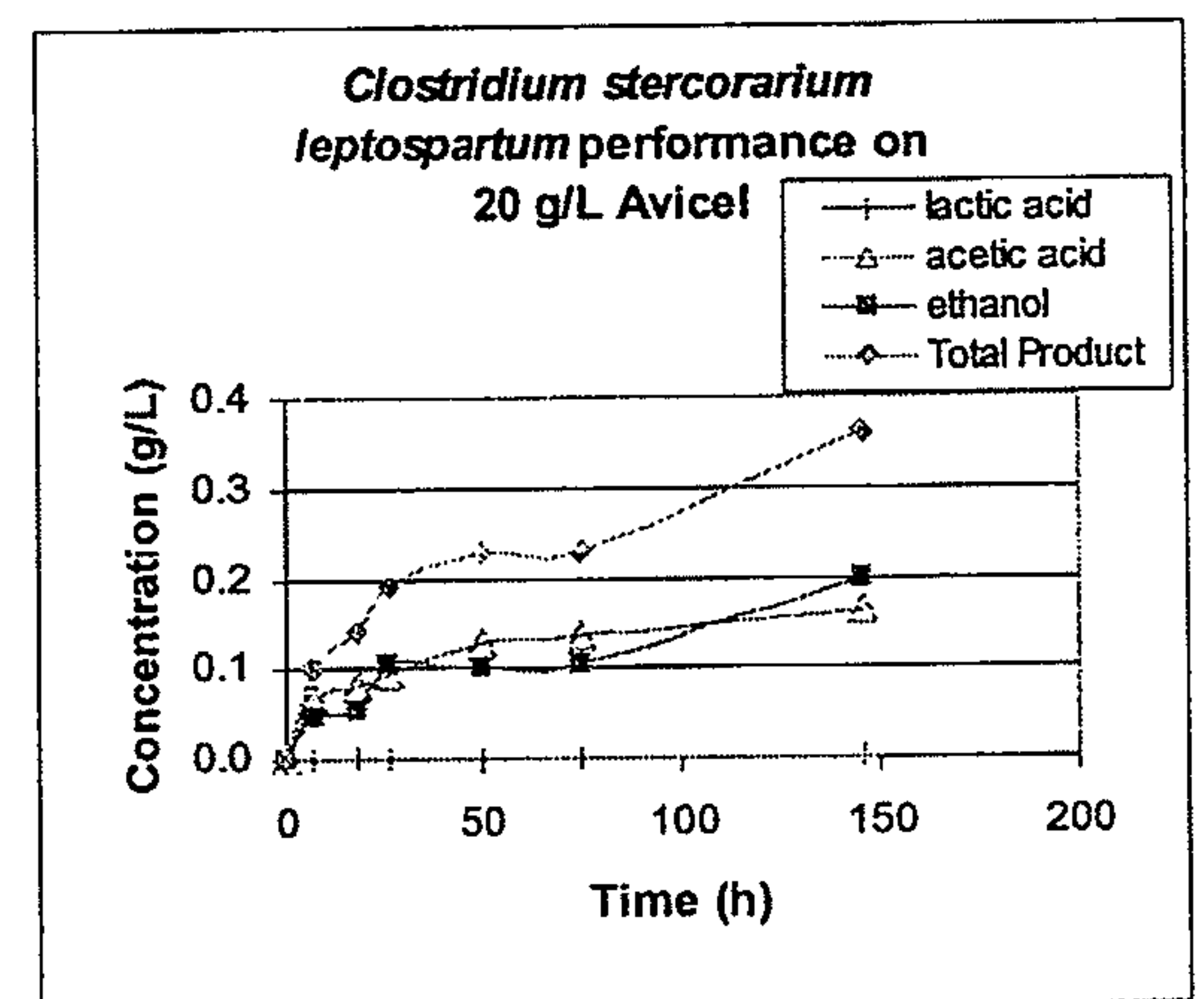


Figure 16A

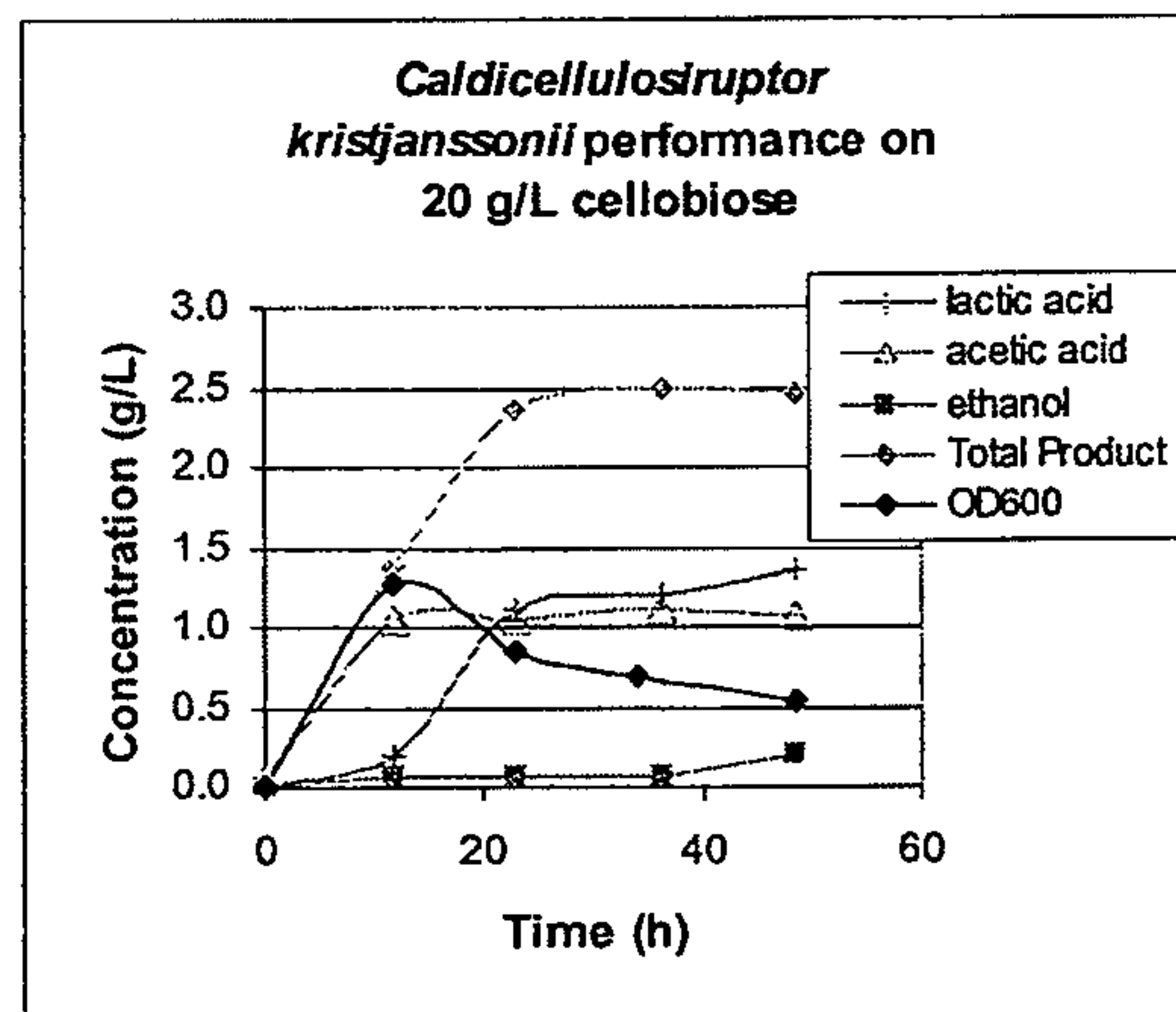


Figure 16B

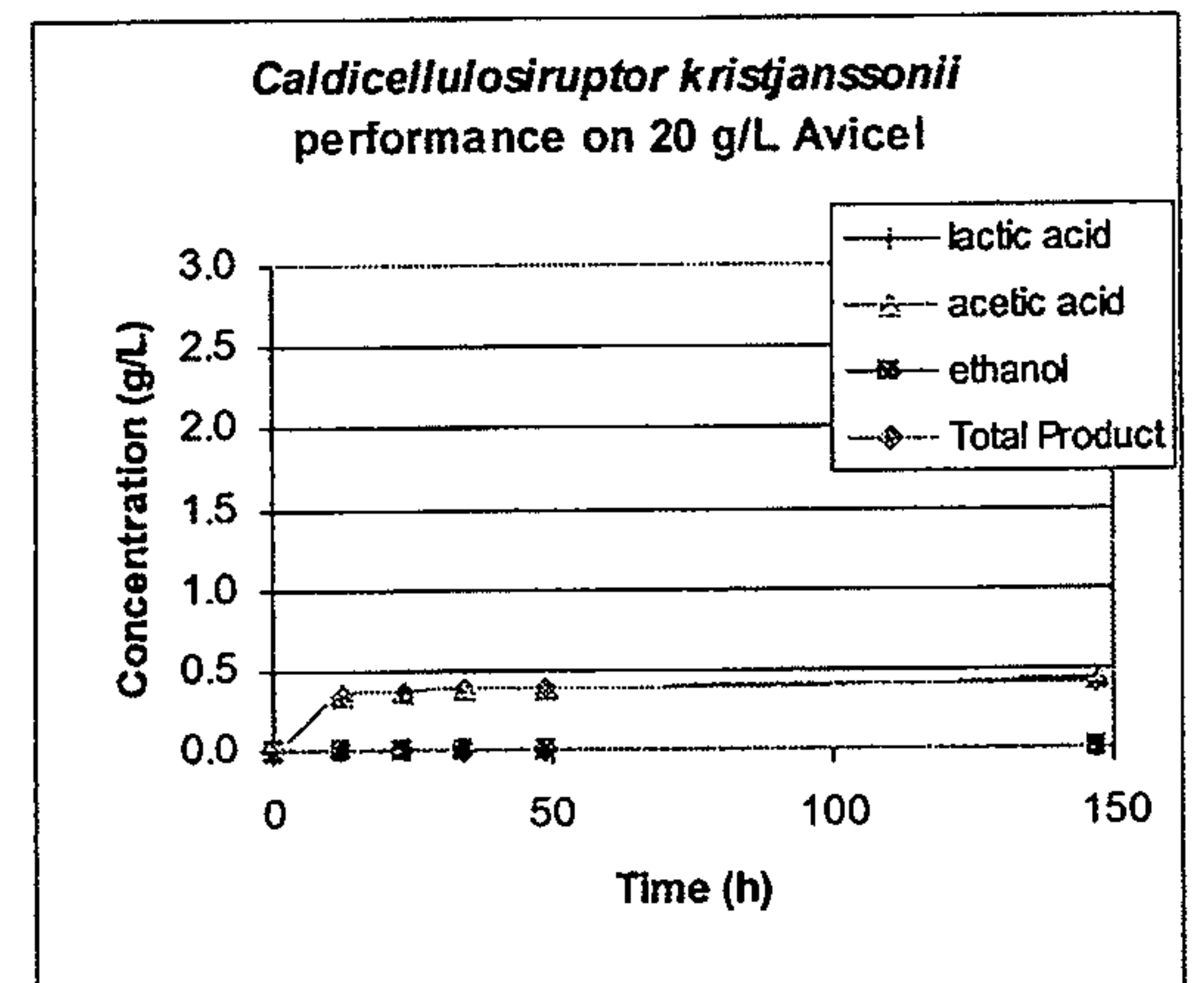


Figure 17A

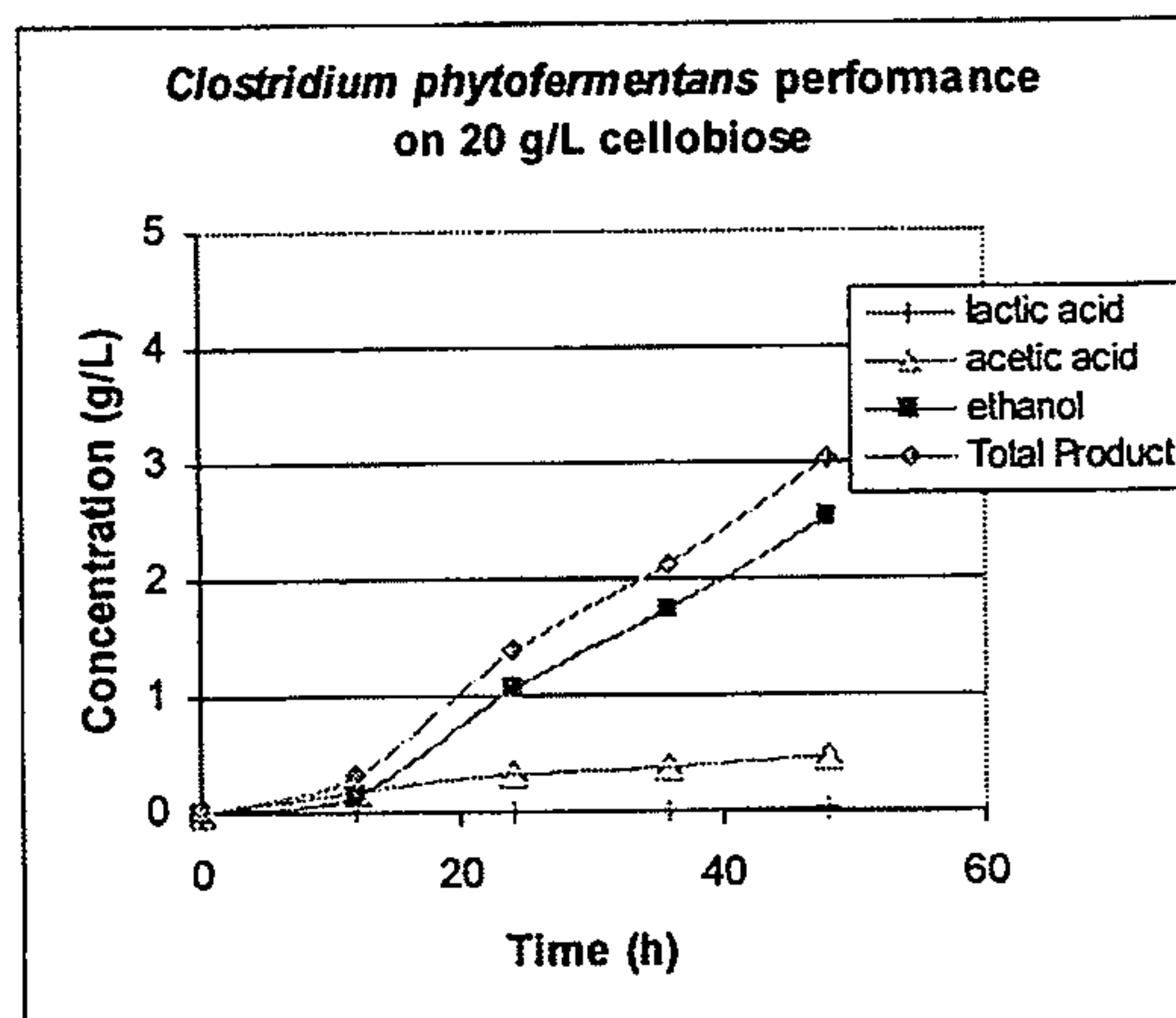


Figure 17B

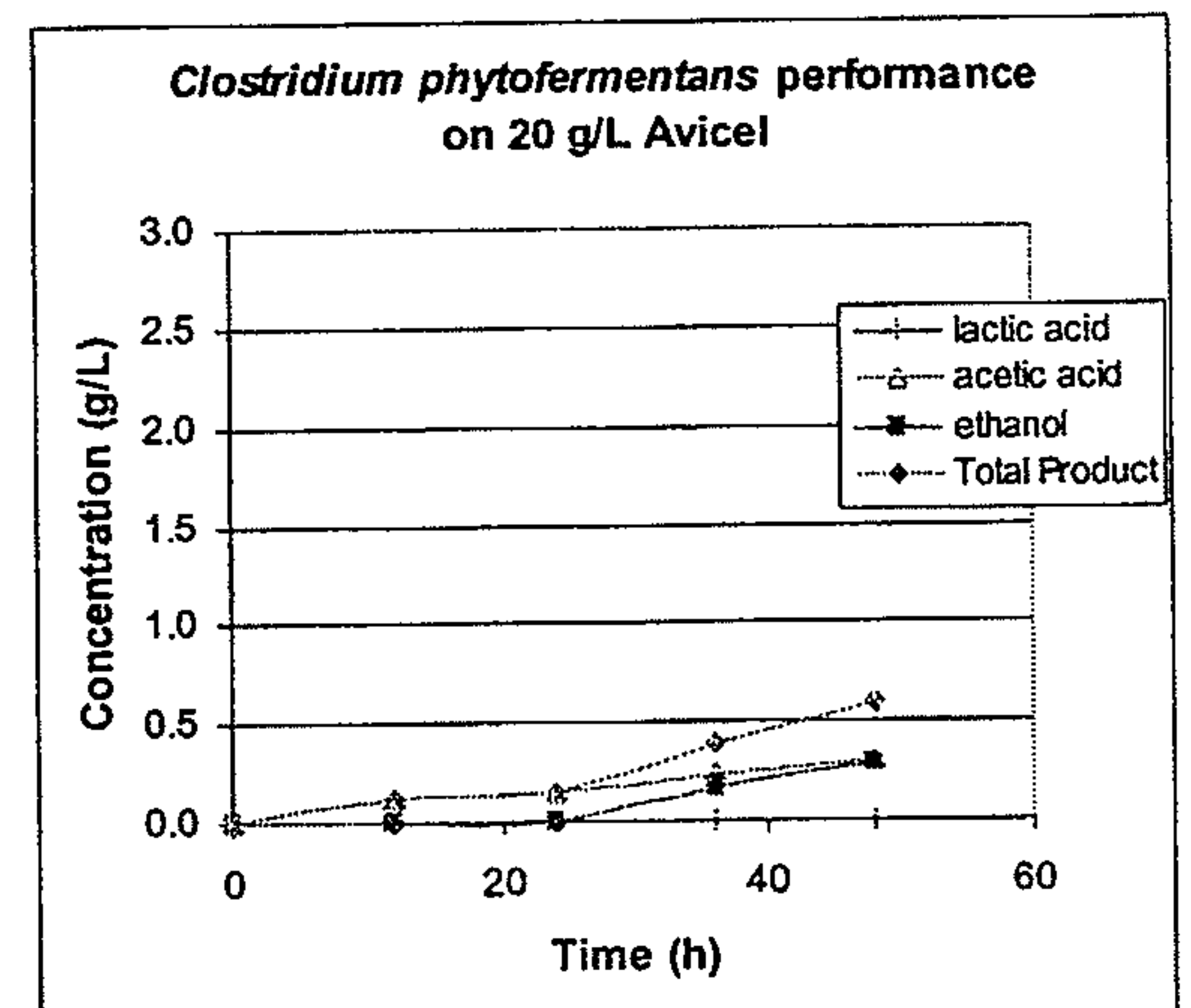


Figure 18

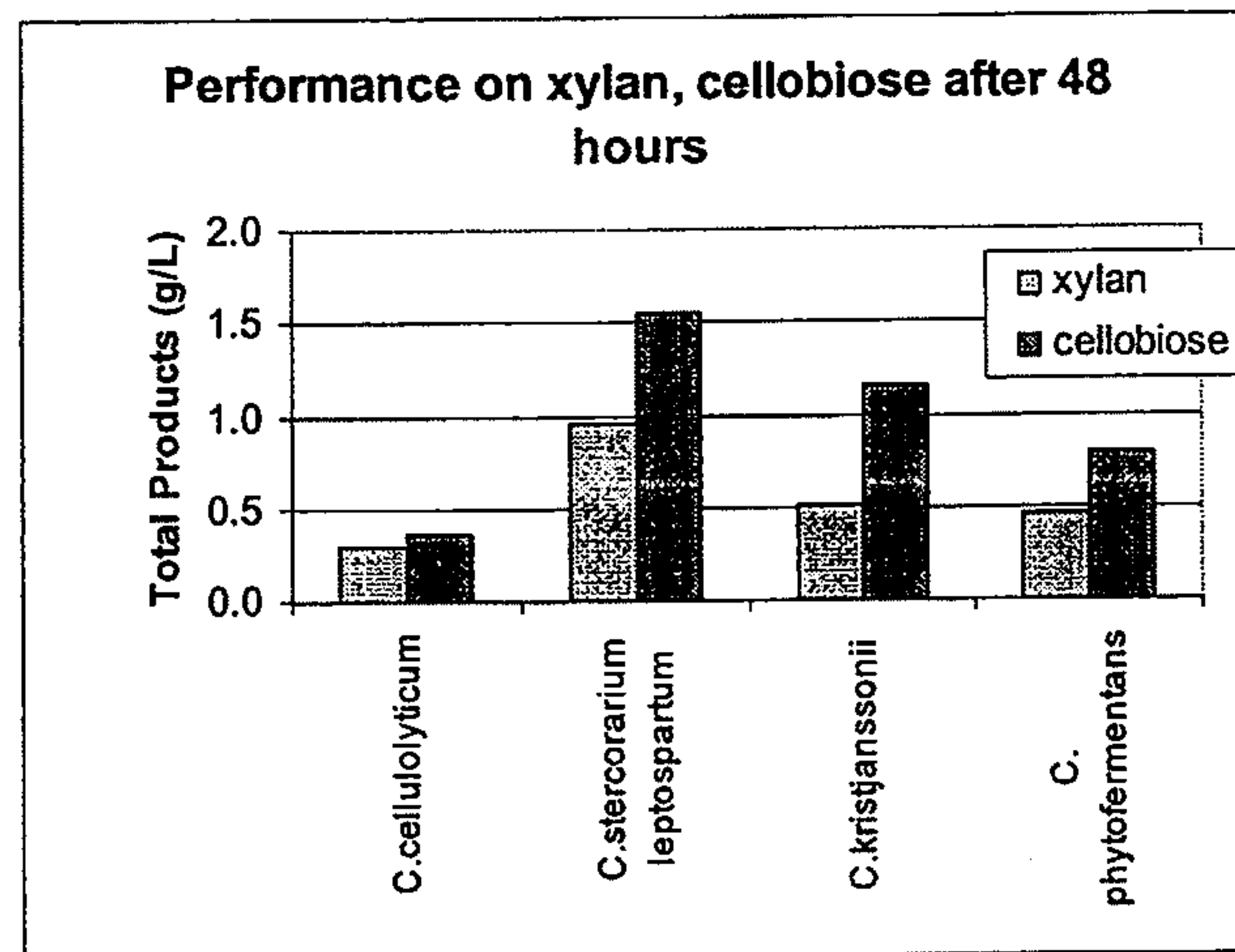


Figure 19

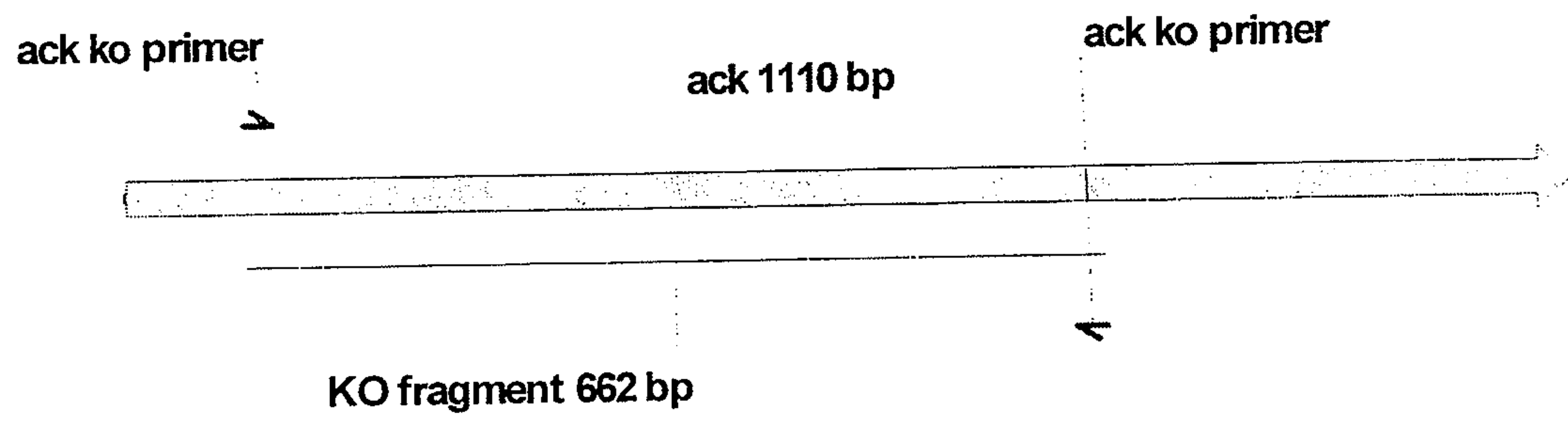


Figure 20

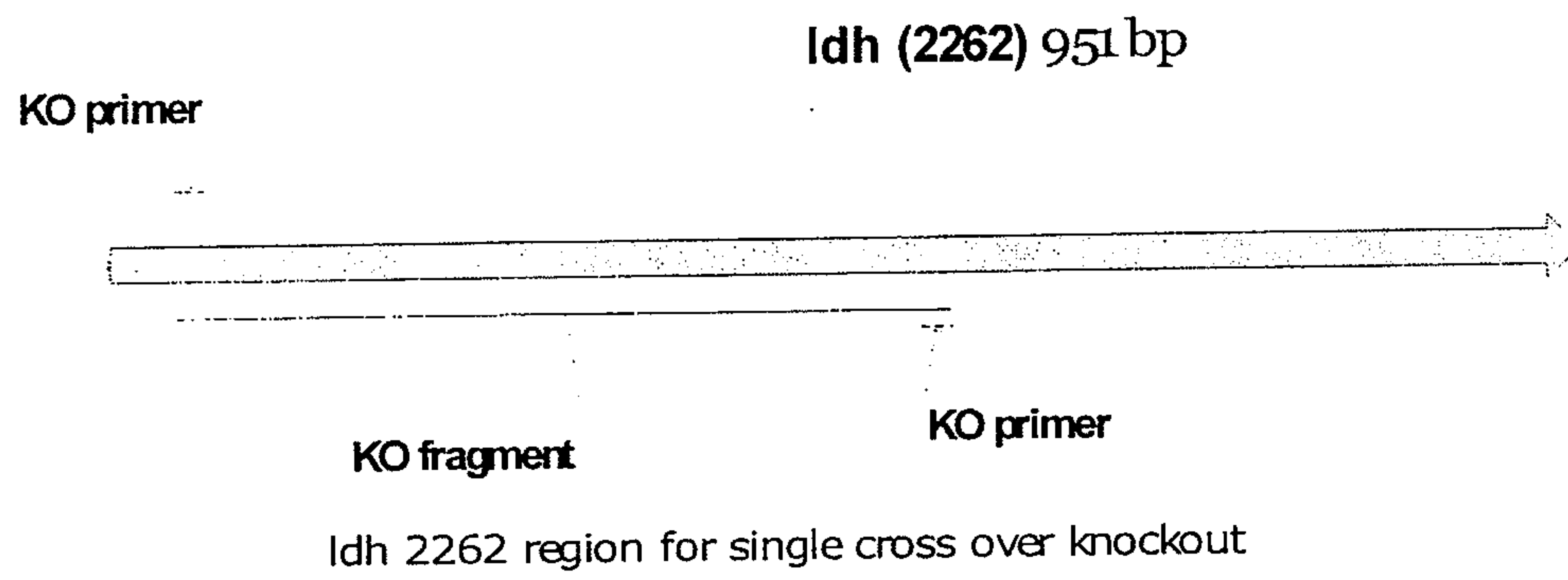
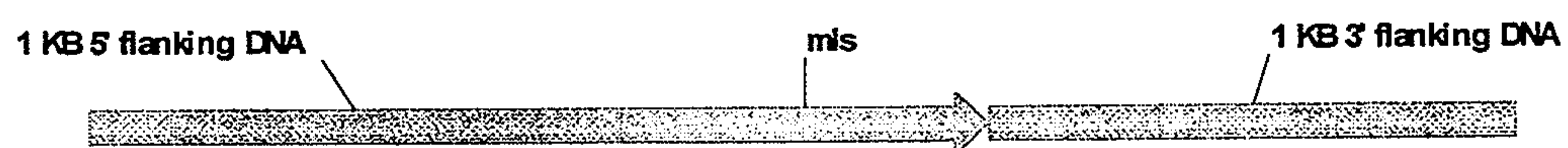


Figure 21



*C. cellulolyticum* ldh 2262 double cross over KO fragment

Figure 22

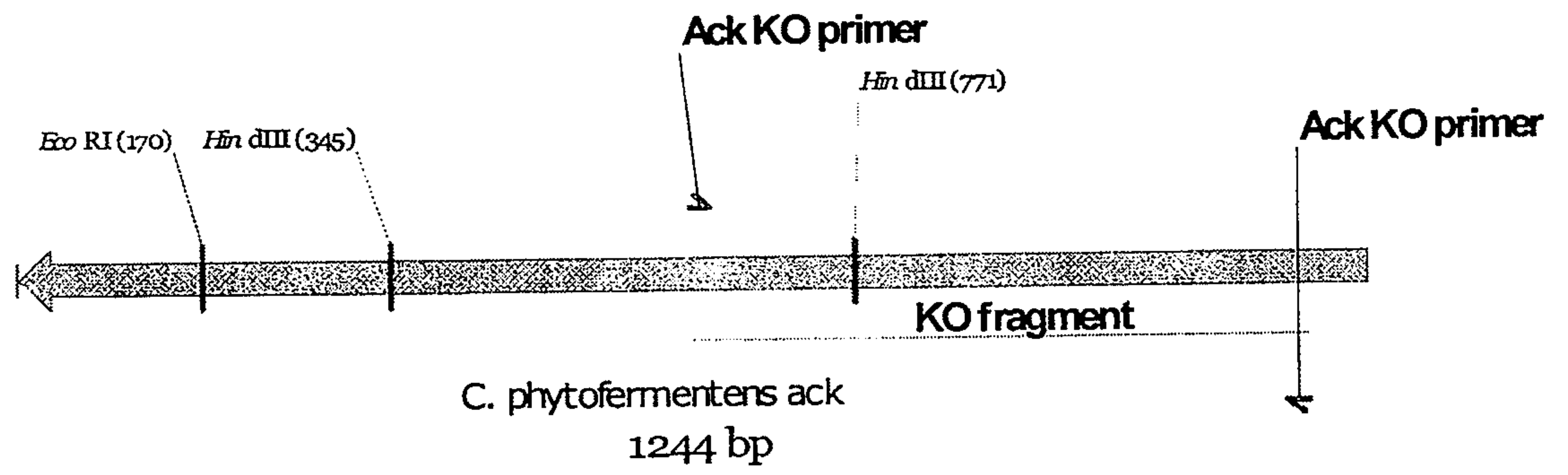


Figure 23

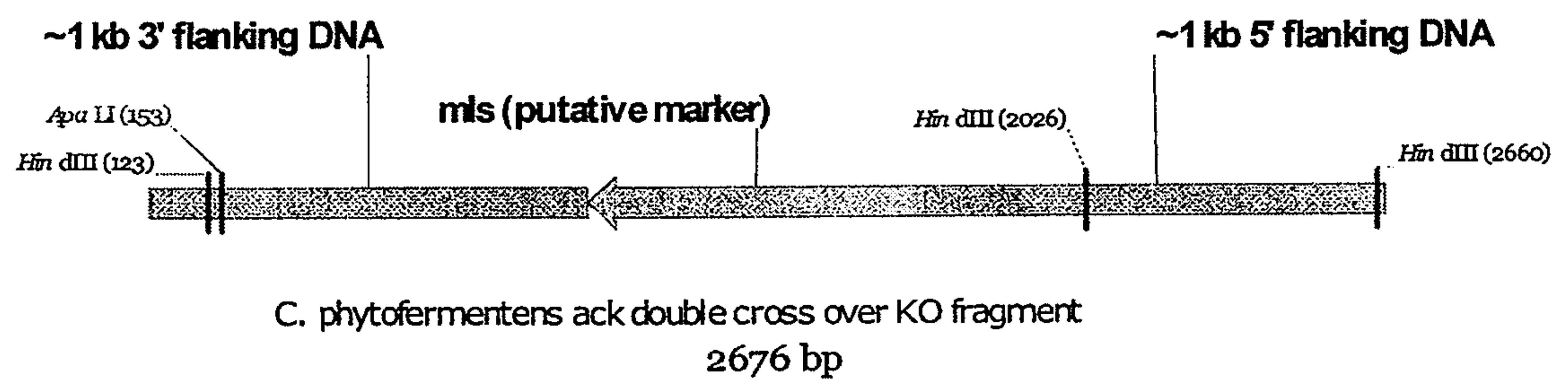
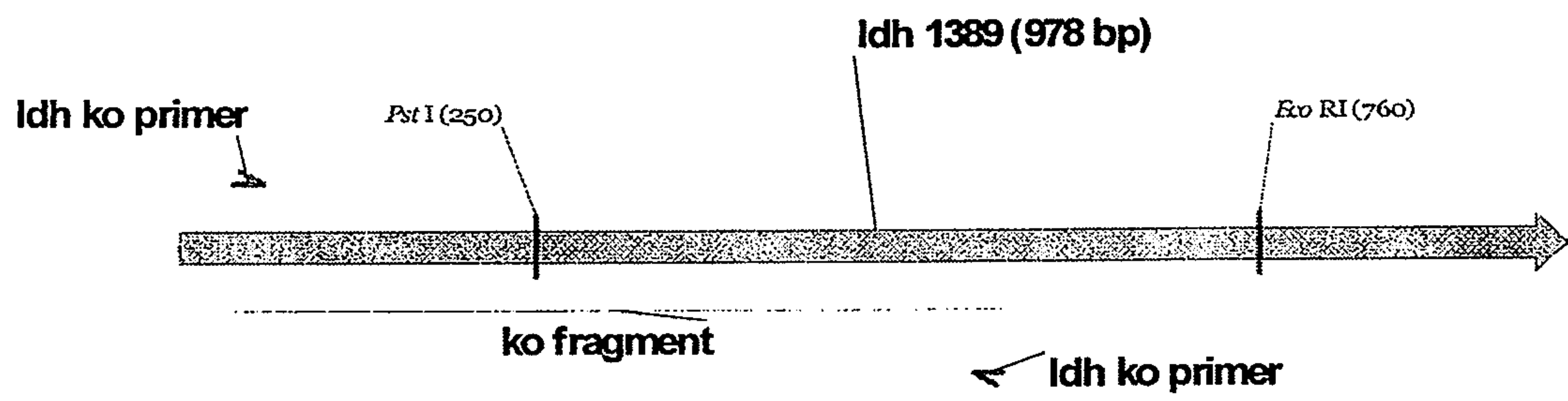
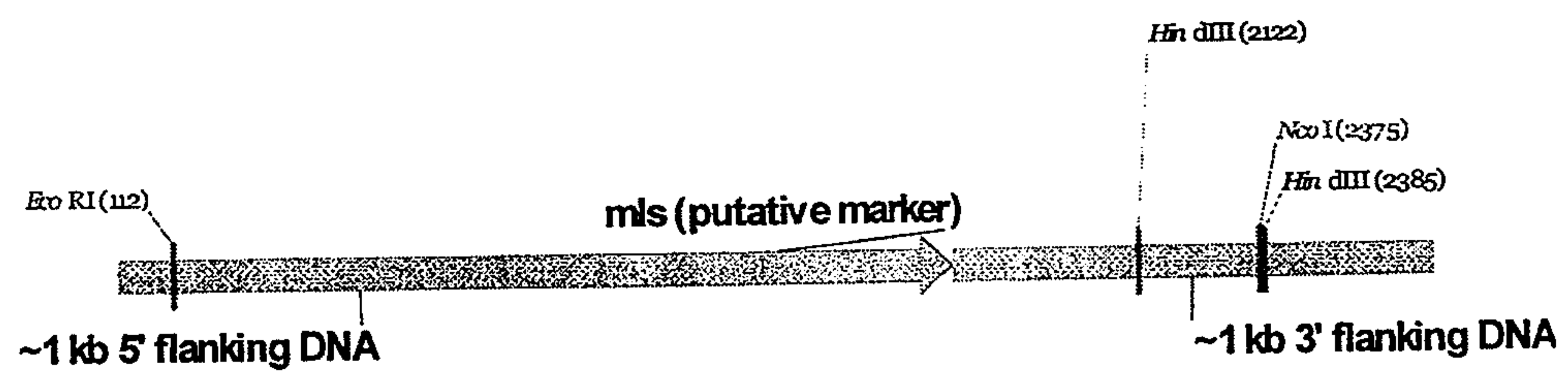


Figure 24



C. phytofermentens Idh 1389 single cross over fragment

Figure 25



C. phytofermentens double crossover ko fragment

2732 bp

Figure 26

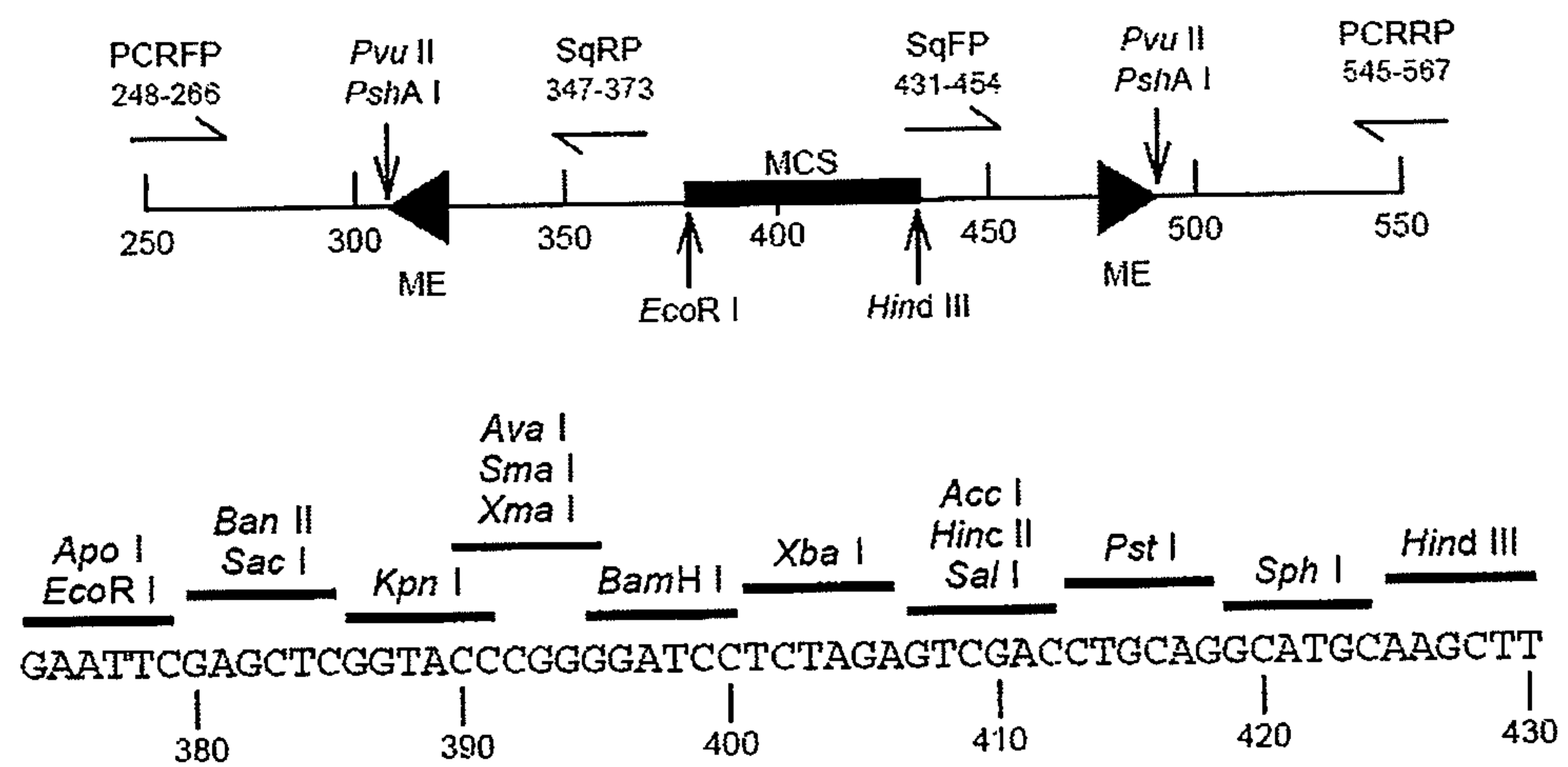


Figure 27

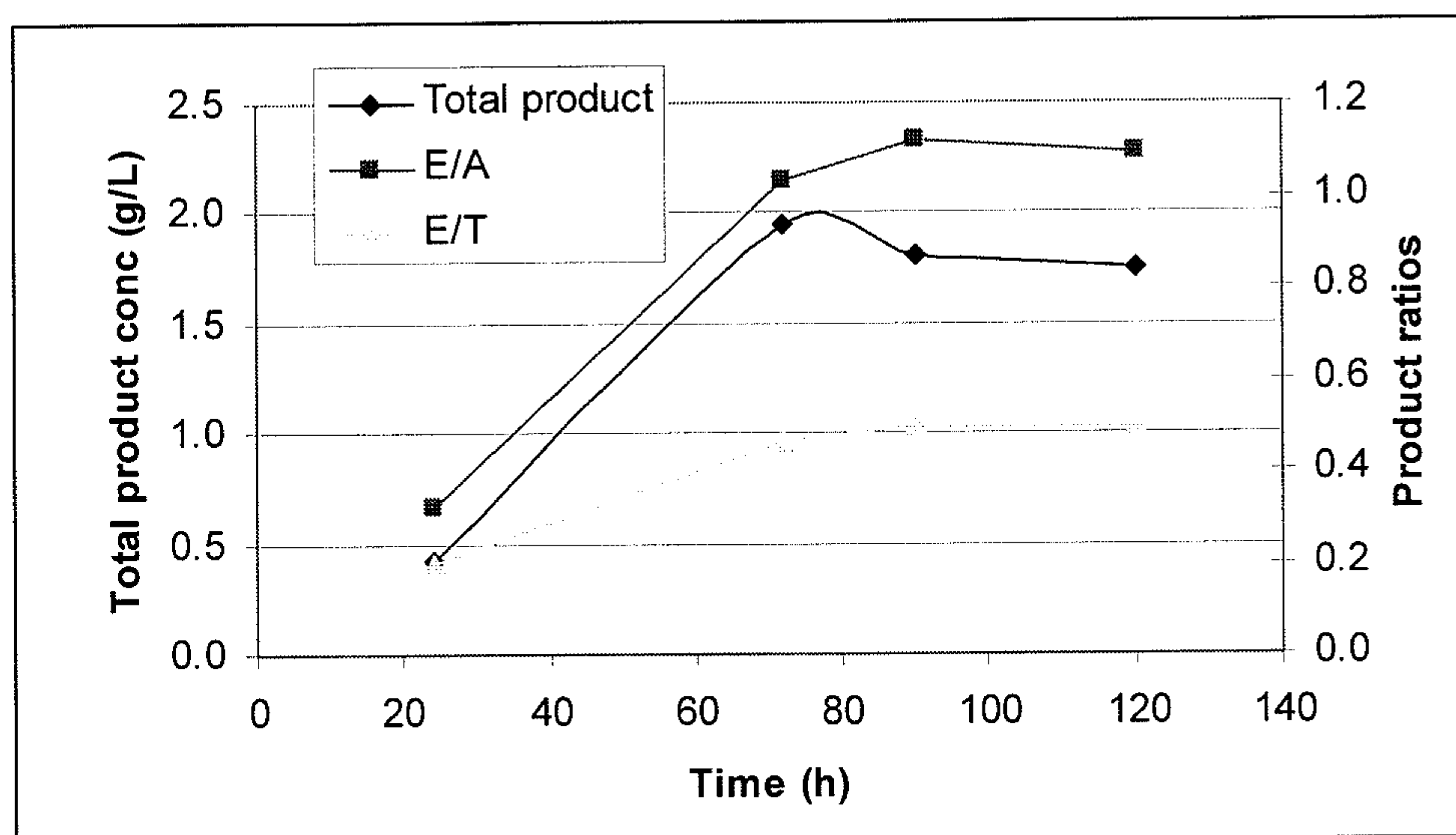


Figure 28

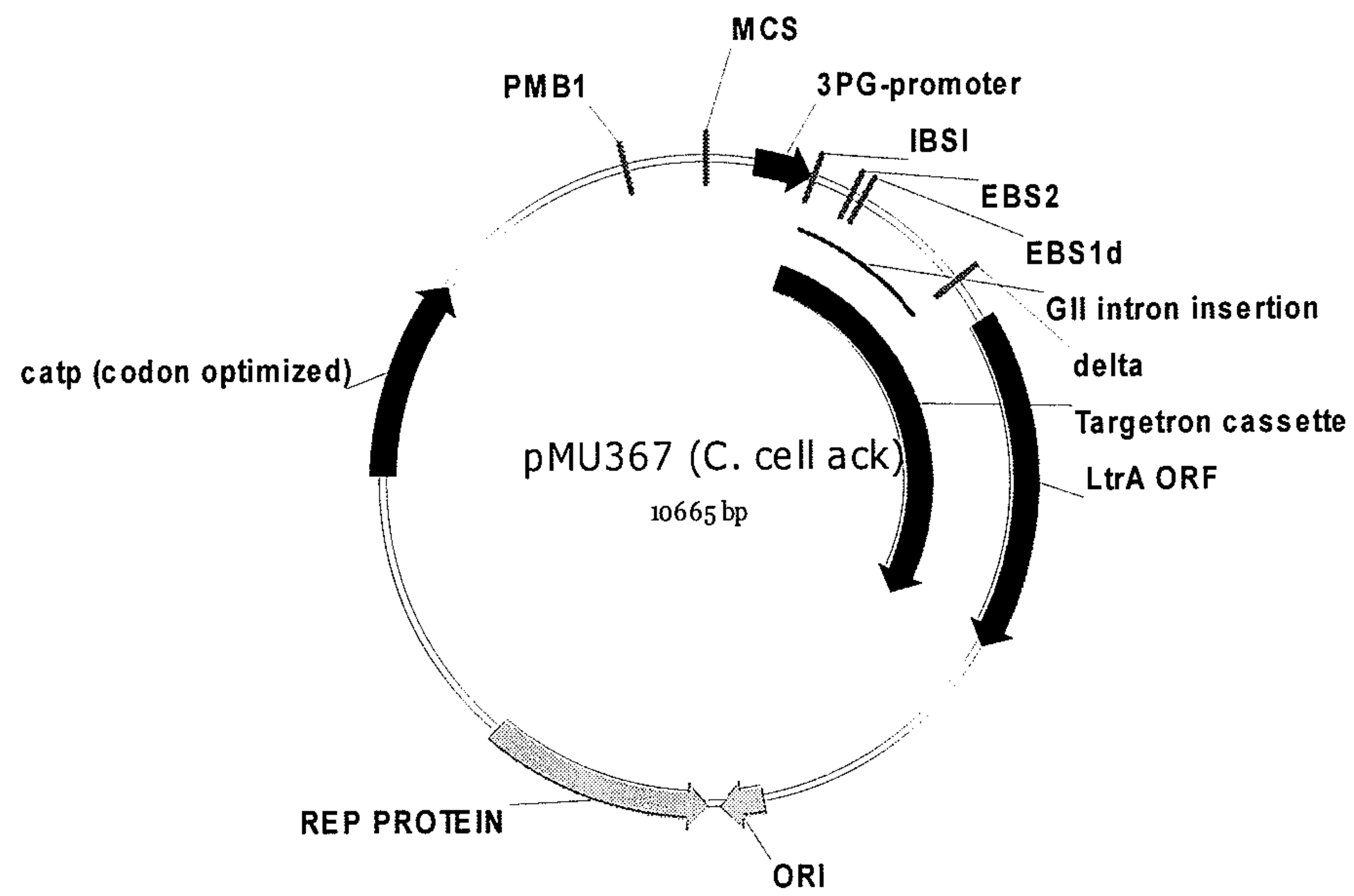
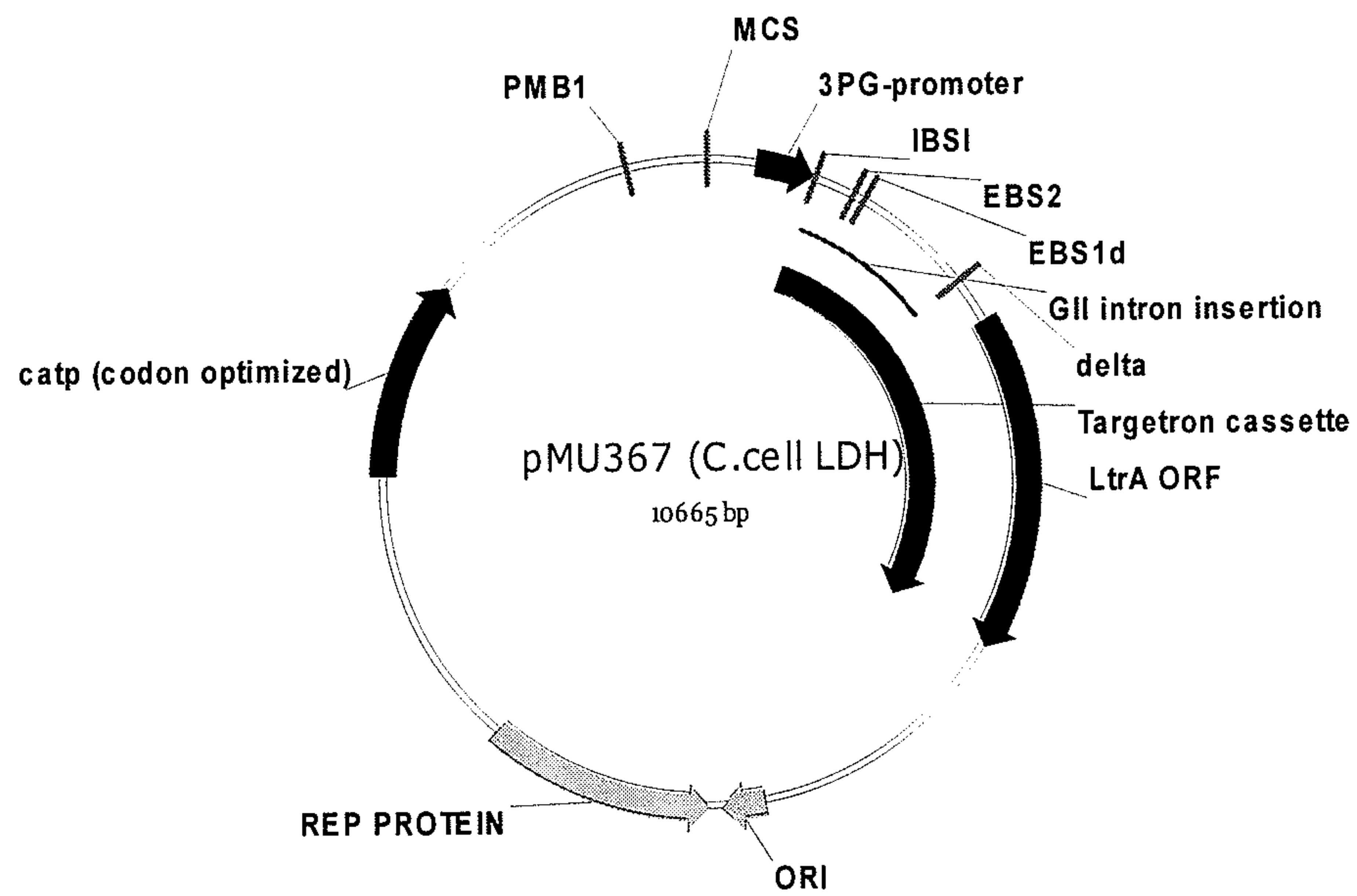


Figure 29



Percent identity of 16S sequence to T. saccharolyticum JW/SL-YS485:  
 Thermoanaerobacterium saccharolyticum B6A-RI - 99.0%  
 Thermoanaerobacter sp. strain 59 - 95.7%  
 Thermoanaerobacter pseudoethanolicus 39E - 83.7%

|                                     |        |                                                     |
|-------------------------------------|--------|-----------------------------------------------------|
| Tethanolicus39E_16S                 | (453)  | GGTAGG CAGCAAG GGG CTA G G G G G T G                |
| Thermoanaerobacter_sp_strain_59_16S | (412)  | GGTAGG TACG GAGGGGGG CTA G G G G G T                |
| TsaccharolyticumB6A-RI_16S          | (492)  | GGTAGG TACGCAAG GGGG CTA G G G G G T                |
| TsaccharolyticumYS485_16S           | (492)  | GGTAGG TACGCAAG GGGGGG GCT G G G G G T              |
| Consensus                           | (501)  | GGTACCATACGAAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATA  |
|                                     |        | 551 600                                             |
| Tethanolicus39E_16S                 | (503)  | GGTAGGGGGGGGAGGGG GGGGGG G G G G G                  |
| Thermoanaerobacter_sp_strain_59_16S | (462)  | GGTAGGGGGG GAGG G G G G G G G G G A A               |
| TsaccharolyticumB6A-RI_16S          | (542)  | GGTAGGGGGGGG GGGG G G G G G G G G G A A             |
| TsaccharolyticumYS485_16S           | (542)  | GGTAGGGGGG GAGG G G G G G G G G G A A               |
| Consensus                           | (551)  | CGTAGGGGGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCACGTAGG |
|                                     |        | 601 650                                             |
| Tethanolicus39E_16S                 | (553)  | GG CG TC AGG G G G A T C G G G A CG ATA             |
| Thermoanaerobacter_sp_strain_59_16S | (512)  | GGG TA AAAG G G A G A A T G G G G G GA TAT          |
| TsaccharolyticumB6A-RI_16S          | (592)  | GGG TG AAAG G G A G A A T G G G G G GA TGT          |
| TsaccharolyticumYS485_16S           | (592)  | GGG TG AAAG G G A G A A T G G G G G GA TGT          |
| Consensus                           | (601)  | CGGCTGTAAAAGTCAGATGTGAAAAACCTGGGCTCAACCGAGGGTGTGCA  |
|                                     |        | 651 700                                             |
| Tethanolicus39E_16S                 | (603)  | CTGAAA GGGGG G A G G C G G G G G T C                |
| Thermoanaerobacter_sp_strain_59_16S | (562)  | TCGAA GAAATA TCA TCA G G G A C C T                  |
| TsaccharolyticumB6A-RI_16S          | (642)  | TCGAAA GAAACA TCA TCA G G G A C C T                 |
| TsaccharolyticumYS485_16S           | (642)  | TCGAAA GAAACA TCA TCA G G G A C C T                 |
| Consensus                           | (651)  | TCTGAAACTAAACAGCTTGAGTCAAGGAGAGGAGAGCGGAATTCCTGGTG  |
|                                     |        | 701 750                                             |
| Tethanolicus39E_16S                 | (653)  | AGGGG G G G G G T G G G G G G G G A                 |
| Thermoanaerobacter_sp_strain_59_16S | (612)  | AGGGG GAAAG G G G G G A A A G G G A G               |
| TsaccharolyticumB6A-RI_16S          | (692)  | AGGGG G G G G G G G G G A A A G G G A G             |
| TsaccharolyticumYS485_16S           | (692)  | AGGGG G G A A G G G G G G A A A G G G A G           |
| Consensus                           | (701)  | TAGCGGTGAAATGCGTAGAGATCAGGAAGAATACCAGTGGCGAAAGCGGC  |
|                                     |        | 751 800                                             |
| Tethanolicus39E_16S                 | (703)  | TCTG G G G G G G G G G G G G G G G G G G            |
| Thermoanaerobacter_sp_strain_59_16S | (662)  | TCTG G G G G G G G G G G G G G G G G G G            |
| TsaccharolyticumB6A-RI_16S          | (742)  | TCTG G G G G G G G G G G G G G G G G G G            |
| TsaccharolyticumYS485_16S           | (742)  | TCTG G G G G G G G G G G G G G G G G G G            |
| Consensus                           | (751)  | TCTCTGGACTTGAAGTACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG   |
|                                     |        | 801 850                                             |
| Tethanolicus39E_16S                 | (753)  | GATTAGATAGCG G G G G G G G G G G G G G              |
| Thermoanaerobacter_sp_strain_59_16S | (712)  | GATTAGATAGCG G G G G G G G G G G G G G              |
| TsaccharolyticumB6A-RI_16S          | (792)  | GATTAGATAGCG G G G G G G G G G G G G G              |
| TsaccharolyticumYS485_16S           | (792)  | GATTAGATAGCG G G G G G G G G G G G G G              |
| Consensus                           | (801)  | GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTGG  |
|                                     |        | 851 900                                             |
| Tethanolicus39E_16S                 | (803)  | CA GCGGA G G T G G G G G T G G G G G G C            |
| Thermoanaerobacter_sp_strain_59_16S | (762)  | G TACAT T A G G G G G G G G G G G G G T             |
| TsaccharolyticumB6A-RI_16S          | (842)  | G TACAT T A G G G G G G G G G G G G G T             |
| TsaccharolyticumYS485_16S           | (842)  | G TACAT T A G G G G G G G G G G G G G T             |
| Consensus                           | (851)  | G TGAGGCATCAT CCGTGCCGGAGTTAACGCAATAAGTATCCCGCCTGG  |
|                                     |        | 901 950                                             |
| Tethanolicus39E_16S                 | (853)  | GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC |
| Thermoanaerobacter_sp_strain_59_16S | (810)  | GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC |
| TsaccharolyticumB6A-RI_16S          | (890)  | GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC |
| TsaccharolyticumYS485_16S           | (890)  | GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC |
| Consensus                           | (901)  | GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC |
|                                     |        | 951 1000                                            |
| Tethanolicus39E_16S                 | (903)  | AAGCGGTGAGGATGCGGTTTAAATTCGAAGCAACGCGAAGAACCTTACCA  |
| Thermoanaerobacter_sp_strain_59_16S | (860)  | AAGCGGTGAGGATGCGGTTTAAATTCGAAGCAACGCGAAGAACCTTACCA  |
| TsaccharolyticumB6A-RI_16S          | (940)  | AAGCGGTGAGGATGCGGTTTAAATTCGAAGCAACGCGAAGAACCTTACCA  |
| TsaccharolyticumYS485_16S           | (940)  | AAGCGGTGAGGATGCGGTTTAAATTCGAAGCAACGCGAAGAACCTTACCA  |
| Consensus                           | (951)  | AAGCAGCGGAGCATGTGGTTTAAATTCGAAGCAACGCGAAGAACCTTACCA |
|                                     |        | 1001 1050                                           |
| Tethanolicus39E_16S                 | (953)  | GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG      |
| Thermoanaerobacter_sp_strain_59_16S | (910)  | GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG      |
| TsaccharolyticumB6A-RI_16S          | (990)  | GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG      |
| TsaccharolyticumYS485_16S           | (990)  | GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG      |
| Consensus                           | (1001) | GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG      |
|                                     |        | 1051 1100                                           |
| Tethanolicus39E_16S                 | (1003) | AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTG  |
| Thermoanaerobacter_sp_strain_59_16S | (955)  | AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTG  |
| TsaccharolyticumB6A-RI_16S          | (1035) | AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTG  |
| TsaccharolyticumYS485_16S           | (1035) | AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTG  |
| Consensus                           | (1051) | AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTG  |
|                                     |        | 1101 1150                                           |
| Tethanolicus39E_16S                 | (1053) | CTGAGATGTTGGGTTAACTGCGGCAAGGAGGAGGAGGAGGAGGAGGAGG   |
| Thermoanaerobacter_sp_strain_59_16S | (1004) | CTGAGATGTTGGGTTAACTGCGGCAAGGAGGAGGAGGAGGAGGAGGAGG   |
| TsaccharolyticumB6A-RI_16S          | (1084) | CTGAGATGTTGGGTTAACTGCGGCAAGGAGGAGGAGGAGGAGGAGGAGG   |
| TsaccharolyticumYS485_16S           | (1084) | CTGAGATGTTGGGTTAACTGCGGCAAGGAGGAGGAGGAGGAGGAGGAGG   |

|                                     |        |                                                      |      |      |
|-------------------------------------|--------|------------------------------------------------------|------|------|
| Consensus                           | (1101) | GTGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCCTGTTGGTAGTT   | 1151 | 1200 |
| Tethanolicus39E_16S                 | (1103) | GACAGCGGAGACCGGCAAGTAGAGGACAGCGGAGCAACAGCGG          |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1054) | ACGAGCGG--TAACAGCGGGGACTCTACCGAGACTGCCGTGGAGAACACCGG |      |      |
| TsaccharolyticumB6A-RI_16S          | (1134) | ACGAGCGG--TAACAGCGGGGACTCTACCGAGACTGCCGTGGAGAACACCGG |      |      |
| TsaccharolyticumYS485_16S           | (1134) | ACGAGCGG--TAACAGCGGGGACTCTACCGAGACTGCCGTGGAGAACACCGG |      |      |
| Consensus                           | (1151) | ACCAGCG TAAAGACGGGGACTCTACCGAGACTGCCGTGGAGAACACCGG   | 1201 | 1250 |
| Tethanolicus39E_16S                 | (1153) | AGGAAGCGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1102) | AGGAAGCGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  |      |      |
| TsaccharolyticumB6A-RI_16S          | (1182) | AGGAAGCGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  |      |      |
| TsaccharolyticumYS485_16S           | (1182) | AGGAAGCGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  |      |      |
| Consensus                           | (1201) | AGGAAGCGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  | 1251 | 1300 |
| Tethanolicus39E_16S                 | (1203) | CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA   |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1152) | CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1232) | CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA   |      |      |
| TsaccharolyticumYS485_16S           | (1232) | CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA   |      |      |
| Consensus                           | (1251) | CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA   | 1301 | 1350 |
| Tethanolicus39E_16S                 | (1253) | GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC   |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1202) | GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1282) | GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC   |      |      |
| TsaccharolyticumYS485_16S           | (1282) | GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC   |      |      |
| Consensus                           | (1301) | GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC   | 1351 | 1400 |
| Tethanolicus39E_16S                 | (1303) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA   |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1252) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1332) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA   |      |      |
| TsaccharolyticumYS485_16S           | (1332) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA   |      |      |
| Consensus                           | (1351) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA   | 1401 | 1450 |
| Tethanolicus39E_16S                 | (1353) | TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC   |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1302) | TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1382) | TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC   |      |      |
| TsaccharolyticumYS485_16S           | (1382) | TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC   |      |      |
| Consensus                           | (1401) | TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC   | 1451 | 1500 |
| Tethanolicus39E_16S                 | (1402) | AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG   |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1351) | AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1431) | AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG   |      |      |
| TsaccharolyticumYS485_16S           | (1432) | AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG   |      |      |
| Consensus                           | (1451) | AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG   | 1501 | 1550 |
| Tethanolicus39E_16S                 | (1451) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1396) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| TsaccharolyticumB6A-RI_16S          | (1481) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| TsaccharolyticumYS485_16S           | (1482) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| Consensus                           | (1501) | GGCTGGATCACCTCCTTTTCCT                               | 1551 | 1572 |
| Tethanolicus39E_16S                 | (1501) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| Thermoanaerobacter_sp_strain_59_16S | (1396) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| TsaccharolyticumB6A-RI_16S          | (1531) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| TsaccharolyticumYS485_16S           | (1532) | GGCTGGATCACCTCCTTTTCCT                               |      |      |
| Consensus                           | (1551) | GGCTGGATCACCTCCTTTTCCT                               |      |      |



|                                      |        |                                                        |      |
|--------------------------------------|--------|--------------------------------------------------------|------|
| Thermoanaerobacter_sp._strain_59_pta | (601)  | TGGGCTAAATGCCAGAGAACTTGGTTCTATACCTTGACAACTGCTAATAA     |      |
| Tpsuedoethanolicus_pta               | (526)  | AATGCTAATGACAGAGCTTACAGCAATGGCAATGCTTCTGCCCCTAC        |      |
| Tsaccharolyticum_B6A-RI_pta          | (601)  | TCACTAAATGCCAGAGAACTTGGTTCTATACCTTGACAACTGCTAATAA      |      |
| Tsaccharolyticum_YS485_pta           | (598)  | TGGGCTAATGCCAGAGAACTTGGTTCTATACCTTGACAACTGCTAATAA      |      |
| Consensus                            | (601)  | TGGGCTAATGCCAGAGAACTTGGTTCTATACCTTGCTGTACAATCTGCTAATAC |      |
|                                      |        | 651                                                    | 700  |
| Thermoanaerobacter_sp._strain_59_pta | (651)  | TGCAAA-CAATTTCTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT     |      |
| Tpsuedoethanolicus_pta               | (576)  | TGCAAA-CAATTTCTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT     |      |
| Tsaccharolyticum_B6A-RI_pta          | (651)  | TGCAAA-CAATTTCTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT     |      |
| Tsaccharolyticum_YS485_pta           | (648)  | TGCAAA-CAATTTCTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT     |      |
| Consensus                            | (651)  | TGCAAA GAATTTCTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT     |      |
|                                      |        | 701                                                    | 750  |
| Thermoanaerobacter_sp._strain_59_pta | (700)  | TCTACAAAAGGTAGTGCATCACATGAATTAGTAGATAAGGTAAGAAAAGC     |      |
| Tpsuedoethanolicus_pta               | (625)  | TCTACAAAAGGTAGTGCATCACATGAATTAGTAGATAAGGTAAGAAAAGC     |      |
| Tsaccharolyticum_B6A-RI_pta          | (700)  | TCTACAAAAGGTAGTGCATCACATGAATTAGTAGATAAGGTAAGAAAAGC     |      |
| Tsaccharolyticum_YS485_pta           | (697)  | TCTACAAAAGGTAGTGCATCACATGAATTAGTAGATAAGGTAAGAAAAGC     |      |
| Consensus                            | (701)  | TCTACAAAAGGTAGTGCATCACATGAATTAGTAGATAAGGTAAGAAAAGC     |      |
|                                      |        | 751                                                    | 800  |
| Thermoanaerobacter_sp._strain_59_pta | (750)  | GACAGAGATAGCAAAAGAATTGATGCCAGATGTTGCTA TCGATGGTGAA     |      |
| Tpsuedoethanolicus_pta               | (675)  | GACAGAGATAGCAAAAGAATTGATGCCAGATGTTGCTA TCGATGGTGAA     |      |
| Tsaccharolyticum_B6A-RI_pta          | (750)  | GACAGAGATAGCAAAAGAATTGATGCCAGATGTTGCTA TCGATGGTGAA     |      |
| Tsaccharolyticum_YS485_pta           | (747)  | GACAGAGATAGCAAAAGAATTGATGCCAGATGTTGCTA TCGATGGTGAA     |      |
| Consensus                            | (751)  | GACAGAGATAGCAAAAGAATTGATGCCAGATGTTGCTA TCGATGGTGAA     |      |
|                                      |        | 801                                                    | 850  |
| Thermoanaerobacter_sp._strain_59_pta | (799)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAGTTGCAGAGCTAAAAGCGCC      |      |
| Tpsuedoethanolicus_pta               | (724)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAGTTGCAGAGCTAAAAGCGCC      |      |
| Tsaccharolyticum_B6A-RI_pta          | (799)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAGTTGCAGAGCTAAAAGCGCC      |      |
| Tsaccharolyticum_YS485_pta           | (796)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAGTTGCAGAGCTAAAAGCGCC      |      |
| Consensus                            | (801)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAGTTGCAGAGCTAAAAGCGCC      |      |
|                                      |        | 851                                                    | 900  |
| Thermoanaerobacter_sp._strain_59_pta | (849)  | AGGAAGCAAAGTTGCGGGATGTGCAAATGTGCTTATATTCCTGATTTAC      |      |
| Tpsuedoethanolicus_pta               | (774)  | AGGAAGCAAAGTTGCGGGATGTGCAAATGTGCTTATATTCCTGATTTAC      |      |
| Tsaccharolyticum_B6A-RI_pta          | (849)  | AGGAAGCAAAGTTGCGGGATGTGCAAATGTGCTTATATTCCTGATTTAC      |      |
| Tsaccharolyticum_YS485_pta           | (846)  | AGGAAGCAAAGTTGCGGGATGTGCAAATGTGCTTATATTCCTGATTTAC      |      |
| Consensus                            | (851)  | AGGAAGCAAAGTTGCGGGATGTGCAAATGTGCTTATATTCCTGATTTAC      |      |
|                                      |        | 901                                                    | 950  |
| Thermoanaerobacter_sp._strain_59_pta | (899)  | AAGCTGGTAATATAGGATATAAGCTTGTACAGAGATTAGCTAAGGCAAT      |      |
| Tpsuedoethanolicus_pta               | (824)  | AAGCTGGTAATATAGGATATAAGCTTGTACAGAGATTAGCTAAGGCAAT      |      |
| Tsaccharolyticum_B6A-RI_pta          | (899)  | AAGCTGGTAATATAGGATATAAGCTTGTACAGAGATTAGCTAAGGCAAT      |      |
| Tsaccharolyticum_YS485_pta           | (896)  | AAGCTGGTAATATAGGATATAAGCTTGTACAGAGATTAGCTAAGGCAAT      |      |
| Consensus                            | (901)  | AAGCTGGTAATATAGGATATAAGCTTGTACAGAGATTAGCTAAGGCAAT      |      |
|                                      |        | 951                                                    | 1000 |
| Thermoanaerobacter_sp._strain_59_pta | (947)  | GCAATTGGACCTATAACACAAGGAATGGGTGCACCGGTTAATGATTTATC     |      |
| Tpsuedoethanolicus_pta               | (874)  | GCAATTGGACCTATAACACAAGGAATGGGTGCACCGGTTAATGATTTATC     |      |
| Tsaccharolyticum_B6A-RI_pta          | (949)  | GCAATTGGACCTATAACACAAGGAATGGGTGCACCGGTTAATGATTTATC     |      |
| Tsaccharolyticum_YS485_pta           | (946)  | GCAATTGGACCTATAACACAAGGAATGGGTGCACCGGTTAATGATTTATC     |      |
| Consensus                            | (951)  | GCAATTGGACCTATAACACAAGGAATGGGTGCACCGGTTAATGATTTATC     |      |
|                                      |        | 1001                                                   | 1050 |
| Thermoanaerobacter_sp._strain_59_pta | (996)  | AAGAGGATGCAGCTATAGAGATATTGTTGACGTAATAGCAACAACAGCTG     |      |
| Tpsuedoethanolicus_pta               | (924)  | AAGAGGATGCAGCTATAGAGATATTGTTGACGTAATAGCAACAACAGCTG     |      |
| Tsaccharolyticum_B6A-RI_pta          | (999)  | AAGAGGATGCAGCTATAGAGATATTGTTGACGTAATAGCAACAACAGCTG     |      |
| Tsaccharolyticum_YS485_pta           | (996)  | AAGAGGATGCAGCTATAGAGATATTGTTGACGTAATAGCAACAACAGCTG     |      |
| Consensus                            | (1001) | AAGAGGATGCAGCTATAGAGATATTGTTGACGTAATAGCAACAACAGCTG     |      |
|                                      |        | 1051                                                   | 1100 |
| Thermoanaerobacter_sp._strain_59_pta | (1044) | TGCAGGCTCAA                                            |      |
| Tpsuedoethanolicus_pta               | (974)  | TGCAGGCTCAA                                            |      |
| Tsaccharolyticum_B6A-RI_pta          | (1049) | TGCAGGCTCAA                                            |      |
| Tsaccharolyticum_YS485_pta           | (1046) | TGCAGGCTCAA                                            |      |
| Consensus                            | (1051) | TGCAGGCTCAA                                            |      |
|                                      |        | 1101                                                   | 1150 |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                                  |      |
| Tpsuedoethanolicus_pta               | (1024) | AATTTTAGTCATGAACTGTGGAAGCTCGTCATTAAAGTATCAATTGTTA      |      |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                                  |      |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                                  |      |
| Consensus                            | (1101) | -----                                                  |      |
|                                      |        | 1151                                                   | 1200 |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                                  |      |
| Tpsuedoethanolicus_pta               | (1074) | GATATGGATAATGGGAAAGTGCTAGCGAAAGGATTGGCGGAAAGGATAGG     |      |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                                  |      |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                                  |      |
| Consensus                            | (1151) | -----                                                  |      |
|                                      |        | 1201                                                   | 1250 |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                                  |      |
| Tpsuedoethanolicus_pta               | (1124) | TATCAATGATTCTCTTTTAACTCATCAAGTAGAGGGCAAAGATAAAATAA     |      |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                                  |      |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                                  |      |

|                                      | Consensus | (1201)                 | 1251 | 1273 |
|--------------------------------------|-----------|------------------------|------|------|
| Thermoanaerobacter_sp._strain_59_pta | (1054)    | -----                  |      |      |
| Tpsuedoethanolicus_pta               | (1174)    | AAATACAAAAGATATGAAAAAT |      |      |
| Tsaccharolyticum_B6A-RI_pta          | (1060)    | -----                  |      |      |
| Tsaccharolyticum_YS485_pta           | (1057)    | -----                  |      |      |
| Consensus                            | (1251)    |                        |      |      |

Figure 32

|                                      |       |                                                        |
|--------------------------------------|-------|--------------------------------------------------------|
|                                      | 1     | 50                                                     |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                  |
| Tpsuedoethanolicus_ack               | (1)   | GCTAATGCTATCGGACCAATTTCTCAAGGTCTTGCAAAACCTATCAATGA     |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----                                                  |
| Tsaccharolyticum_YS485_ack           | (1)   | -----                                                  |
| Consensus                            | (1)   | -----                                                  |
|                                      | 51    | 100                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                  |
| Tpsuedoethanolicus_ack               | (51)  | CTTGTCAAGAGGTTGTAGTGTAGAAGATATTGTTAATGTTATAGCAATAA     |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----                                                  |
| Tsaccharolyticum_YS485_ack           | (1)   | -----                                                  |
| Consensus                            | (51)  | -----                                                  |
|                                      | 101   | 150                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                  |
| Tpsuedoethanolicus_ack               | (101) | CTTGTGTACAAGCTCAAGGGGTGCAAAAATAACTTTGAGGAGGCAGCGAT     |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----ATGAAAAC                                          |
| Tsaccharolyticum_YS485_ack           | (1)   | -----ATGAAAAT                                          |
| Consensus                            | (101) | -----ATGAAAAT                                          |
|                                      | 151   | 200                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                  |
| Tpsuedoethanolicus_ack               | (151) | TATGAAAATTTTAGTCATGAACCTGTGGAAGCTCGTCATTAAGAGTATCAA    |
| Tsaccharolyticum_B6A-RI_ack          | (9)   | TATGAAAATTTCTGGTTATTATTCTGTGGAAGTCTTTCACATAAAA-TATCAA  |
| Tsaccharolyticum_YS485_ack           | (9)   | TATGAAAATTAAGTGGTTATTATTCTGTGGAAGTCTTTCACATAAAA-TATCAA |
| Consensus                            | (151) | TATGAAAATTTCTGGTTATTATTCTGTGGAAGTCTTTCACATAAAA TATCAA  |
|                                      | 201   | 250                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                  |
| Tpsuedoethanolicus_ack               | (201) | TTGTAGATATGGATAATGGGAAAGTGCTAGCGAAAGGATTGGCGGAAAG      |
| Tsaccharolyticum_B6A-RI_ack          | (58)  | TTGATTGAATCAATTGATGGAAATGTGCTGGCAAAAGGCTTGGCTGAAAG     |
| Tsaccharolyticum_YS485_ack           | (58)  | CTGATTGAATCAACTGATGGAAATGTGCTGGCAAAAGGCTTGGCTGAAAG     |
| Consensus                            | (201) | TTGATTGAATCAA TGATGGAAATGTGCTGGCAAAAGGCTTGGCTGAAAG     |
|                                      | 251   | 300                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----TCCTA-CGG-----AGAA                                |
| Tpsuedoethanolicus_ack               | (251) | GATAGGTATCAATGATTCCTTTTAACTCATCA TAG G CCAA G TA       |
| Tsaccharolyticum_B6A-RI_ack          | (108) | AATCGGCATAAATGATTCCTTGTGACGATA T CTA CGC -----AGAA     |
| Tsaccharolyticum_YS485_ack           | (108) | AATCGGCATAAATGATTCCTTGTGACGATA T CTA CGC -----AGAA     |
| Consensus                            | (251) | AATCGGCATAAATGATTCCTTGTGAC CATAATGCTAACGG AGAA         |
|                                      | 301   | 350                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (16)  | AA-TCAGATAAAAAAAGACATGAAAGATCACAAGACGAAATTAATG         |
| Tpsuedoethanolicus_ack               | (301) | AAATCAAGATAAAAAAAGACATGAAAGATCACAAGACGCAATAAAATTG      |
| Tsaccharolyticum_B6A-RI_ack          | (155) | AAATCAAGATAAAAAAAGACATGAAAGATCACAAGACGAAATTAATG        |
| Tsaccharolyticum_YS485_ack           | (155) | AAATCAAGATAAAAAAAGACATGAAAGATCACAAGACGCAATAAAATTG      |
| Consensus                            | (301) | AAATCAAGATAAAAAAAGACATGAAAGATCACAAGACGCAATAAAATTG      |
|                                      | 351   | 400                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (65)  | -----GTTTACATGCTTTTGGTAAGTAGTACTACGGCGTTATAAAGGATATGTC |
| Tpsuedoethanolicus_ack               | (351) | GTTTACATGCTTTTGGTAAGTAGTACTACGGCGTTATAAAGGATATGTC      |
| Tsaccharolyticum_B6A-RI_ack          | (205) | GTTTACATGCTTTTGGTAAGTAGTACTACGGCGTTATAAAGGATATGTC      |
| Tsaccharolyticum_YS485_ack           | (205) | GTTTACATGCTTTTGGTAAGTAGTACTACGGCGTTATAAAGGATATGTC      |
| Consensus                            | (351) | GTTTACATGCTTTTGGTAAGTAGTACTACGGCGTTATAAAGGATATGTC      |
|                                      | 401   | 450                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (114) | TGAGTAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA       |
| Tpsuedoethanolicus_ack               | (401) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tsaccharolyticum_B6A-RI_ack          | (255) | TGAGTAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA       |
| Tsaccharolyticum_YS485_ack           | (255) | TGAGTAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA       |
| Consensus                            | (401) | TGAGTAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA       |
|                                      | 451   | 500                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (164) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tpsuedoethanolicus_ack               | (451) | TGATTCCTATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tsaccharolyticum_B6A-RI_ack          | (305) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tsaccharolyticum_YS485_ack           | (305) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Consensus                            | (451) | CATCATCAGTTCTCATAAATGATGAAGTGTTAAAAGCGATAACAGATTGT     |
|                                      | 501   | 550                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (214) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tpsuedoethanolicus_ack               | (501) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tsaccharolyticum_B6A-RI_ack          | (355) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Tsaccharolyticum_YS485_ack           | (355) | ATCAATAGATGCTGTAGGACATAGAGTTGTTACGGAGGAGAGTCTTTTA      |
| Consensus                            | (501) | ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC     |
|                                      | 551   | 600                                                    |
| Thermoanaerobacter_sp._strain_59_ack | (264) | ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC     |
| Tpsuedoethanolicus_ack               | (551) | ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC     |
| Tsaccharolyticum_B6A-RI_ack          | (405) | ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC     |
| Tsaccharolyticum_YS485_ack           | (405) | ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC     |
| Consensus                            | (551) | TTGCCAGCAAATCATGCCAAACGTTCCAATGGTGGCGGTATTGTATACAG     |
|                                      | 601   | 650                                                    |

|                                      |        |                                                     |      |      |
|--------------------------------------|--------|-----------------------------------------------------|------|------|
| Thermoanaerobacter_sp._strain_59_ack | (314)  | CCCTTACAGGATGTTAAACACAA                             | 651  | 700  |
| Tpsuedoethanolicus_ack               | (601)  | CTTTCAGCAAGATGAGATGATCT                             |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (455)  | CCCTTACAGGATGTTAAACACAA                             |      |      |
| Tsaccharolyticum_YS485_ack           | (455)  | CCCTTACAGGATGTTAAACACAA                             |      |      |
| Consensus                            | (601)  | CCTTTTCATCAGACAATGCCTGATTATGCATATCTTTATCCAATACCTTAT |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (364)  | GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTTCATGGCACATC | 701  | 750  |
| Tpsuedoethanolicus_ack               | (651)  | GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTTCATGGCACATC |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (505)  | GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTTCATGGCACATC |      |      |
| Tsaccharolyticum_YS485_ack           | (505)  | GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTTCATGGCACATC |      |      |
| Consensus                            | (651)  | GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTTCATGGCACATC |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (414)  | GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  | 751  | 800  |
| Tpsuedoethanolicus_ack               | (701)  | GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (555)  | GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |      |      |
| Tsaccharolyticum_YS485_ack           | (555)  | GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |      |      |
| Consensus                            | (701)  | GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (464)  | AAGATTGAAAATCATAACTTGTTCATCTTGAAATGGCTCCAGTATTGCT   | 801  | 850  |
| Tpsuedoethanolicus_ack               | (751)  | AAGATTGAAAATCATAACTTGTTCATCTTGAAATGGCTCCAGTATTGCT   |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (605)  | AAGATTGAAAATCATAACTTGTTCATCTTGAAATGGCTCCAGTATTGCT   |      |      |
| Tsaccharolyticum_YS485_ack           | (605)  | AAGATTGAAAATCATAACTTGTTCATCTTGAAATGGCTCCAGTATTGCT   |      |      |
| Consensus                            | (751)  | AAGATTGAAAATCATAACTTGTTCATCTTGAAATGGCTCCAGTATTGCT   |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (514)  | GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT  | 851  | 900  |
| Tpsuedoethanolicus_ack               | (801)  | GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (655)  | GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT  |      |      |
| Tsaccharolyticum_YS485_ack           | (655)  | GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT  |      |      |
| Consensus                            | (801)  | GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (564)  | AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA  | 901  | 950  |
| Tpsuedoethanolicus_ack               | (851)  | AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (705)  | AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA  |      |      |
| Tsaccharolyticum_YS485_ack           | (705)  | AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA  |      |      |
| Consensus                            | (851)  | AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (614)  | TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT  | 951  | 1000 |
| Tpsuedoethanolicus_ack               | (901)  | TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (755)  | TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT  |      |      |
| Tsaccharolyticum_YS485_ack           | (755)  | TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT  |      |      |
| Consensus                            | (901)  | TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (664)  | ATATTAATAAAAAATCTGGTGTTCACGGTATTTTCAGGAATAAGCAGCGA  | 1001 | 1050 |
| Tpsuedoethanolicus_ack               | (951)  | ATATTAATAAAAAATCTGGTGTTCACGGTATTTTCAGGAATAAGCAGCGA  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (805)  | ATATTAATAAAAAATCTGGTGTTCACGGTATTTTCAGGAATAAGCAGCGA  |      |      |
| Tsaccharolyticum_YS485_ack           | (805)  | ATATTAATAAAAAATCTGGTGTTCACGGTATTTTCAGGAATAAGCAGCGA  |      |      |
| Consensus                            | (951)  | ATATTAATAAAAAATCTGGTGTTCACGGTATTTTCAGGAATAAGCAGCGA  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (714)  | TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC  | 1051 | 1100 |
| Tpsuedoethanolicus_ack               | (1001) | TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (855)  | TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC  |      |      |
| Tsaccharolyticum_YS485_ack           | (855)  | TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC  |      |      |
| Consensus                            | (1001) | TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (764)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT  | 1101 | 1150 |
| Tpsuedoethanolicus_ack               | (1051) | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT  |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (905)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT  |      |      |
| Tsaccharolyticum_YS485_ack           | (905)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT  |      |      |
| Consensus                            | (1051) | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT  |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (814)  | TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTACAGCAGGTGT   | 1151 | 1200 |
| Tpsuedoethanolicus_ack               | (1101) | TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTACAGCAGGTGT   |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (955)  | TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTACAGCAGGTGT   |      |      |
| Tsaccharolyticum_YS485_ack           | (955)  | TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTACAGCAGGTGT   |      |      |
| Consensus                            | (1101) | TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTACAGCAGGTGT   |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (864)  | TGGTGAAAATGGTCTGAGATACGAGAATTTATACTTGATGGATTAGAGT   | 1201 | 1250 |
| Tpsuedoethanolicus_ack               | (1151) | TGGTGAAAATGGTCTGAGATACGAGAATTTATACTTGATGGATTAGAGT   |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (1005) | TGGTGAAAATGGTCTGAGATACGAGAATTTATACTTGATGGATTAGAGT   |      |      |
| Tsaccharolyticum_YS485_ack           | (1005) | TGGTGAAAATGGTCTGAGATACGAGAATTTATACTTGATGGATTAGAGT   |      |      |
| Consensus                            | (1151) | TGGTGAAAATGGTCTGAGATACGAGAATTTATACTTGATGGATTAGAGT   |      |      |
| Thermoanaerobacter_sp._strain_59_ack | (914)  | TCCTACCGTCAGCTTCCATACAAATACACACCAACCAACCAACCAACCA   |      |      |
| Tpsuedoethanolicus_ack               | (1201) | TCCTACCGTCAGCTTCCATACAAATACACACCAACCAACCAACCAACCA   |      |      |
| Tsaccharolyticum_B6A-RI_ack          | (1055) | TCCTACCGTCAGCTTCCATACAAATACACACCAACCAACCAACCAACCA   |      |      |
| Tsaccharolyticum_YS485_ack           | (1055) | TCCTACCGTCAGCTTCCATACAAATACACACCAACCAACCAACCAACCA   |      |      |

|                                      |        |                                                      |      |
|--------------------------------------|--------|------------------------------------------------------|------|
| Consensus                            | (1201) | TTTTAGGGTTCAGCTTGGATAAAGAAAAAATAAAGTCAGAGGAAAAGAA    |      |
|                                      |        | 1251                                                 | 1300 |
| Thermoanaerobacter_sp._strain_59_ack | (964)  | ACTATTATATCTACGCCGAATTGAAAAGTTGGCTGATGCGTTCGCTAC     |      |
| Tpsuedoethanolicus_ack               | (1251) | GAAATTTATATCTACAGAAGTTTAAAAGTTTAAATTTATGCTTATTTCTAC  |      |
| Tsaccharolyticum_B6A-RI_ack          | (1105) | ACTATTATATCTACGCCGAATTGAAAAGTTGGCTGATGCGTTCGCTAC     |      |
| Tsaccharolyticum_YS485_ack           | (1105) | ACTATTATATCTACGCCGAATTGAAAAGTTGGCTGATGCGTTCGCTAC     |      |
| Consensus                            | (1251) | ACTATTATATCTACGCCGAATTGAAAAGTTAGCGTGATGGTTGTGCCTAC   |      |
|                                      |        | 1301                                                 | 1350 |
| Thermoanaerobacter_sp._strain_59_ack | (1014) | TAAATGAGAGAAATATGATTTGCTAAAGATACTGAAAAGATTGTAAAGAGTA |      |
| Tpsuedoethanolicus_ack               | (1301) | AATGAGAGAAATATGATTTGCTAAAGATACTGAAAAGATTGTAAAGAGTA   |      |
| Tsaccharolyticum_B6A-RI_ack          | (1155) | TAAATGAGAGAAATATGATTTGCTAAAGATACTGAAAAGATTGTAAAGAGTA |      |
| Tsaccharolyticum_YS485_ack           | (1155) | TAAATGAGAGAAATATGATTTGCTAAAGATACTGAAAAGATTGTAAAGAGTA |      |
| Consensus                            | (1301) | TAAATGAGAGAAATATGATTTGCTAAAGATACTGAAAAGATTGTAAAGAGTA |      |
|                                      |        | 1351                                                 |      |
| Thermoanaerobacter_sp._strain_59_ack | (1064) | TAAAA---                                             |      |
| Tpsuedoethanolicus_ack               | (1351) | TAAAGTAG                                             |      |
| Tsaccharolyticum_B6A-RI_ack          | (1205) | TAAAA---                                             |      |
| Tsaccharolyticum_YS485_ack           | (1205) | TAAAA---                                             |      |
| Consensus                            | (1351) | TAAAA                                                |      |

|                                 |       |                                                     |              |     |
|---------------------------------|-------|-----------------------------------------------------|--------------|-----|
|                                 |       | 1                                                   |              | 50  |
| Thermoanaerobacter_sp._strain59 | (1)   | ATGAGTAAAGTGGCCATAATAGGTTTCAGGATTGTAGGTGCTACATCTGC  |              |     |
| Tpseudoethanolicus_39E          | (1)   | ATGAACAAAATATCTATAATAGGTTCTCGGATTGTGCGTGCTACTACTGC  |              |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                               |              |     |
| Tsaccharolyticum_Y8485          | (1)   | ATGAGCAAAGTAGCAATAATAGGATCTGGTTTTGTAGGTGCAACATCGGC  |              |     |
| Consensus                       | (1)   | ATGAGCAAAGTAGC ATAATAGGTTCTCGGATTGTAGGTGCTACATCTGC  |              |     |
|                                 |       | 51                                                  |              | 100 |
| Thermoanaerobacter_sp._strain59 | (51)  | ATTTACATTGGCTCTAAGTGGGACTGTGACAGACATTGTTTTAGTAGATT  |              |     |
| Tpseudoethanolicus_39E          | (51)  | ATACACACTGGCTTTGAGTGGGATTGCCAAACTATTGTATTAAATAGATA  |              |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                               |              |     |
| Tsaccharolyticum_Y8485          | (51)  | ATTTACGCTGGCATTAAGTGGGACTGTGACAGATATCGTGCTGGTGGATT  |              |     |
| Consensus                       | (51)  | ATTTACACTGGCTTTAAGTGGGACTGTGACAGATATTGT TTAGTAGATT  |              |     |
|                                 |       | 101                                                 |              | 150 |
| Thermoanaerobacter_sp._strain59 | (101) | TAAACAAGGACAAGGCGAT                                 | A G T T C TA |     |
| Tpseudoethanolicus_39E          | (101) | TTAATAAAGACAAAGCAGA                                 | T T T A C CG |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                               | A G C A T CA |     |
| Tsaccharolyticum_Y8485          | (101) | TAAACAAGGACAAGGCTAT                                 | A G C A T CA |     |
| Consensus                       | (101) | TAAACAAGGACAAGGC ATAGGCGATGCACTGGATATAAGCCATGGCATA  |              |     |
|                                 |       | 151                                                 |              | 200 |
| Thermoanaerobacter_sp._strain59 | (151) | CCGCTAACAGTCAA TGCTTCTCAC CAG A CGA                 |              |     |
| Tpseudoethanolicus_39E          | (151) | CCGT TA TAGT A TG AT C G A T T GT G TTC             |              |     |
| Tsaccharolyticum_B6ARI          | (32)  | CAI A ACAG T AA TG T A T C C AA G TGA               |              |     |
| Tsaccharolyticum_Y8485          | (151) | CCGCA A ACAG T AA TG CT T A T C C AA G GAA          |              |     |
| Consensus                       | (151) | CCGTTTATACAGCCTGTAAATGTGTATGCAGGTGACTACAAAGATGTTGA  |              |     |
|                                 |       | 201                                                 |              | 250 |
| Thermoanaerobacter_sp._strain59 | (201) | GCCGACATGAGTGAGTGAA A A T GCT A A G A A             |              |     |
| Tpseudoethanolicus_39E          | (201) | AGTTTACAATAAACA T AGG A A A A A A G A               |              |     |
| Tsaccharolyticum_B6ARI          | (82)  | ACCAGGCCCAAGTG G A A G T T A A E A T                |              |     |
| Tsaccharolyticum_Y8485          | (201) | AGCCGACATGTAATAGTGAGCA A A T T T A A G C A          |              |     |
| Consensus                       | (201) | AGGCGCAGATGTAATAGTTGTGACAGCAGGTGCTGCTCAAAGCCGGGAG   |              |     |
|                                 |       | 251                                                 |              | 300 |
| Thermoanaerobacter_sp._strain59 | (251) | AGTTAAGTGACCCTCA A A AG A ATCT A C GTC G            |              |     |
| Tpseudoethanolicus_39E          | (251) | AACACAACTCA T A G G GA A GATG T T AGA A             |              |     |
| Tsaccharolyticum_B6ARI          | (132) | AGA GAGG TGACC T G G AA A AGCT A T GTC G            |              |     |
| Tsaccharolyticum_Y8485          | (251) | AGACACGCTCAACC T A A G AA A AGCC A A T GTC G        |              |     |
| Consensus                       | (251) | AGAC AGGCTTGACCTTGTGAAGAAAAATACAGCTATATTTAAGTCCATG  |              |     |
|                                 |       | 301                                                 |              | 350 |
| Thermoanaerobacter_sp._strain59 | (301) | AACCTGAG - AAAA A C A T AA T CC G T A               |              |     |
| Tpseudoethanolicus_39E          | (301) | G GGA A A A A A A A G - A A C CA A A CC T A T       |              |     |
| Tsaccharolyticum_B6ARI          | (182) | AACCTGAG - AAAA G C A C A G T A TT G T C            |              |     |
| Tsaccharolyticum_Y8485          | (301) | AACCTGAG - AAAA G C A A C A G C A A TT G T G        |              |     |
| Consensus                       | (301) | ATACCTGAGCTT TTAAAGTACAATGACAAGGCTATATATTTGATTGT A  |              |     |
|                                 |       | 351                                                 |              | 400 |
| Thermoanaerobacter_sp._strain59 | (350) | CAATCCTGCTATATAT A G T A A A C A AG GAAAC T         |              |     |
| Tpseudoethanolicus_39E          | (350) | CAATCACTAGATATCT A A A C A C A A AT TGGCT G         |              |     |
| Tsaccharolyticum_B6ARI          | (231) | CAATCCTGACACAC G A G C A A A C G AT TGGAC T         |              |     |
| Tsaccharolyticum_Y8485          | (350) | CAATCCC TAGAT A C G G G C A A C A G TT TGGAC T      |              |     |
| Consensus                       | (351) | CAAATCCTGTAGATATACTGACGTACGTTACATACAAGATATCTGGACTT  |              |     |
|                                 |       | 401                                                 |              | 450 |
| Thermoanaerobacter_sp._strain59 | (400) | CCG GGG GC T A C C T A T T C T A A TT CC A          |              |     |
| Tpseudoethanolicus_39E          | (400) | CATAC AAC A A G G T C A T C A A TG GA A             |              |     |
| Tsaccharolyticum_B6ARI          | (281) | CAATGGGCAAC T C C T T C A T T T A A G               |              |     |
| Tsaccharolyticum_Y8485          | (400) | CAATGGGCAAC T T C T T C A C T T A A G               |              |     |
| Consensus                       | (401) | CCATGGGGCAGAGTTTTCGGTTCTGGCACTGTTCTTGACAGTTCAAGGTT  |              |     |
|                                 |       | 451                                                 |              | 500 |
| Thermoanaerobacter_sp._strain59 | (450) | TACGATCTTAACTAA T C T T A - T A A G A T             |              |     |
| Tpseudoethanolicus_39E          | (450) | CACATATCTTAACC AA T T C A A A - G G A A A C         |              |     |
| Tsaccharolyticum_B6ARI          | (331) | TACGACCTTAAAC GG C C C T A A T G A A G C C          |              |     |
| Tsaccharolyticum_Y8485          | (450) | TACATCTCTTAAC AG C C C T A A - G A A G C C          |              |     |
| Consensus                       | (451) | TAGGTATCTTTTAAGCAAGCATTGCAATATAGAT CCGAGAAATGTCCAC  |              |     |
|                                 |       | 501                                                 |              | 550 |
| Thermoanaerobacter_sp._strain59 | (499) | CGAAGCAATTC A A C C TA A AT G G                     |              |     |
| Tpseudoethanolicus_39E          | (499) | CGATATCAAT T G G T C TT T GC A A T                  |              |     |
| Tsaccharolyticum_B6ARI          | (381) | CGAAGCAATTC C G G A T C CA A GT A A A               |              |     |
| Tsaccharolyticum_Y8485          | (499) | CGAAGCAATTC C G G A T C CA A GT A A A               |              |     |
| Consensus                       | (501) | GGAAGGATAATTGGCGAGCATGGTGATACAGAGTTTGACAGCATGGAGCAT |              |     |
|                                 |       | 551                                                 |              | 600 |
| Thermoanaerobacter_sp._strain59 | (549) | AACAAATTTTACCAAT T A A G G A A ATT G C AC           |              |     |
| Tpseudoethanolicus_39E          | (549) | TACGACATAG ACC A C A G A A T A A A ATT A T AA       |              |     |
| Tsaccharolyticum_B6ARI          | (431) | AACAAACMAT T A A A T T A A G G A A C A A C GC       |              |     |
| Tsaccharolyticum_Y8485          | (549) | AACAAACMAATGGCTTAT T A A G G A A C A A C AC         |              |     |
| Consensus                       | (551) | AACAAACATATCAGGAATATCATTTAATGAGTACTGCAGTTTATGCGGAC  |              |     |
|                                 |       | 601                                                 |              | 650 |

|                                 |       |                                                      |
|---------------------------------|-------|------------------------------------------------------|
| Thermoanaerobacter_sp._strain59 | (599) | GAGTTTCTAATACAAATTCAGAAAGGACGGAAGTCAGTTGTCAT         |
| Tpseudoethanolicus_39E          | (599) | AAGCACTGGAAGAGTTTACAGAGCGAATTTTAAATATCTTCAAGA        |
| Tsaccharolyticum_B6ARI          | (481) | CGATCTGCAACACAAATTCAGAAAGGACGGAAGTCAGAAAGACACTCTAAAT |
| Tsaccharolyticum_YS485          | (599) | CGCTCTGCAACACAAATTCAGAAAGGACGGAAGTCAGAAAGACACTCTAAAT |
| Consensus                       | (601) | GCGTCTGTAACACAAATTTTCAGAAAGGAAGTAGAAGATGAAGTTGTAAAT  |
|                                 |       | 651 700                                              |
| Thermoanaerobacter_sp._strain59 | (649) | CGGGCTTACAAAATATATCTTAAAAAGGGTCCCGGTATACCGCTGTCC     |
| Tpseudoethanolicus_39E          | (649) | GCTGCTATACCAATATAGAAAAACGGTGGACATATATATCCGTTC        |
| Tsaccharolyticum_B6ARI          | (531) | CGTGTCTTAAAGAAATAGACAAAAACGGTCTACATCTATGCTGTCCC      |
| Tsaccharolyticum_YS485          | (649) | CGTGTCTTACAGAAATAGACAAAAACGGTCTACATCTATGCTGTCCC      |
| Consensus                       | (651) | GCTGTCTTATAAGATAATAGACAAAAAGGGTGCTACATATTATGCTGTGGC  |
|                                 |       | 701 750                                              |
| Thermoanaerobacter_sp._strain59 | (699) | TCAGGAGTAAGAGAAATCTTAGTCTATCAATACGGAATGAAATTCAA      |
| Tpseudoethanolicus_39E          | (699) | TCCTGGAGTAAGAGAAATCTTAGAGCTATTTTCAGAGATGAATTC        |
| Tsaccharolyticum_B6ARI          | (581) | AGTTGAGTAAGAGAGATCTAGGAGTCCATCTTACAGAGATGAATTC       |
| Tsaccharolyticum_YS485          | (699) | AGTTGAGTAAGAGAGATCTAGGAGTCCATCTTACAGAGATGAATTC       |
| Consensus                       | (701) | TGTTGCAGTAAGAAGGATTGTGGAGTGTATCTTAAGAGATGAAAATTCCA   |
|                                 |       | 751 800                                              |
| Thermoanaerobacter_sp._strain59 | (749) | TTCTTACAGTTTCATCTCCATTAAATGCTCAATACGGTGTAAAGAGATCTA  |
| Tpseudoethanolicus_39E          | (749) | TTTGTAGCTGTGTCATCTCCGCTAACCCGCCAATATGGTGTACAAATCTG   |
| Tsaccharolyticum_B6ARI          | (631) | -----                                                |
| Tsaccharolyticum_YS485          | (749) | CTTCACAGTATCATCTCCATTAAATGCCACGTACGGCGTGAAAGATCTT    |
| Consensus                       | (751) | TTCT ACAGT TCATCTCCATTAAATGG CAATACGGTGT A AGATGT    |
|                                 |       | 801 850                                              |
| Thermoanaerobacter_sp._strain59 | (799) | TCTTTAAGCTTGGCATCAATTTGTGCCCAAAAATGGTETTTGCAAGGGTTCT |
| Tpseudoethanolicus_39E          | (799) | GCTTTGAGGCTTCCCTCCGTTGTTCCGACCAAAATGGAATCGTAAATATAC  |
| Tsaccharolyticum_B6ARI          | (632) | -----                                                |
| Tsaccharolyticum_YS485          | (799) | TCATTAAGCTTGGCATCTATCTTACCCAGGAATGGCCTTCCACAGCATTT   |
| Consensus                       | (801) | TCTTTAAGCTTGGCATC ATTGT GGCAGAAATGG GTTGCAAGGATTCT   |
|                                 |       | 851 900                                              |
| Thermoanaerobacter_sp._strain59 | (849) | GGATTTCGCTTTGGTTCATCACCAAGTTCAGAACTTTAAACATTCGCCAA   |
| Tpseudoethanolicus_39E          | (849) | TGAATTACCACCTTTCACAGCAACAAATTCCTGCTTTTAGAAGATCAGCCG  |
| Tsaccharolyticum_B6ARI          | (632) | -----                                                |
| Tsaccharolyticum_YS485          | (849) | GGACTTCGCTTTTATCTGACCAAGCAAGTTCAGAACTTTAGGCATTCAGCAA |
| Consensus                       | (851) | GGA TTGCCTTT TCTGA GAAGAAGTTGAGAAGTTTAGACATTTCAGCAA  |
|                                 |       | 901 938                                              |
| Thermoanaerobacter_sp._strain59 | (899) | GCCTTATGGCTGATGTTATAAAACAGTTTGACATA---               |
| Tpseudoethanolicus_39E          | (899) | AAGTTATCAAAAGTGTAAATACAAGAGCTTGATATATAA              |
| Tsaccharolyticum_B6ARI          | (632) | -----                                                |
| Tsaccharolyticum_YS485          | (899) | GTGTCATCGGAGATGTCATAAAACAATTACATATA---               |
| Consensus                       | (901) | G GTTATGGCAGATGT ATAAACAGTT GATATA                   |

The diagram illustrates the metabolic pathways of *Clostridium acetobutylicum*, starting from cellobiose and branching into glycolysis and various fermentation products.

**Glycolysis Pathway:**

- cellobiose** is converted to **glucose (Glu)** by **cellobiose phosphorylase** (releasing  $\Delta$  cellobiose) or by  **$\beta$ -glucosidase** (releasing  $1/2$  cellobiose and  $1/2$   $H_2O$ ).
- glucose (Glu)** is converted to **glucose 6-phosphate (G6P)** by **hexokinase** (consuming ATP, producing ADP).
- glucose 6-phosphate (G6P)** is converted to **fructose 6-phosphate (F6P)** by **phosphohexoisomerase**.
- fructose 6-phosphate (F6P)** is converted to **fructose 1,6-diphosphate (FDP)** by **phosphofructokinase (PFK)** (consuming ATP, producing ADP).
- fructose 1,6-diphosphate (FDP)** is cleaved by **aldolase (ALD)** into **glyceraldehyde 3-phosphate (GAP)** and **dihydroxyacetone phosphate (DHP)**.
- glyceraldehyde 3-phosphate (GAP)** is converted to **1,3-diphosphoglycerate (DPG)** by **glyceraldehyde 3-phosphate dehydrogenase (G3PDH)** (consuming NAD, producing NADH and  $P_i$ ).
- 1,3-diphosphoglycerate (DPG)** is converted to **3-phosphoglycerate (3PG)** by **3-phosphoglycerate kinase (PGK)** (consuming ADP, producing ATP).
- 3-phosphoglycerate (3PG)** is converted to **2-phosphoglycerate (2PG)** by **phosphoglycerate mutase (PGM)**.
- 2-phosphoglycerate (2PG)** is converted to **phosphoenolpyruvate (PEP)** by **enolase (ENO)**.
- phosphoenolpyruvate (PEP)** is converted to **pyruvate (Pyr)** by **pyruvate kinase (PK)** (consuming ADP, producing ATP).

**Pyruvate Metabolism and Fermentation Products:**

- pyruvate (Pyr)** can be converted to **L-lactate (Lac)** by **lactate dehydrogenase (LDH)** (consuming NADH, producing NAD).
- pyruvate (Pyr)** can be converted to **acetyl CoA** by **hydrogenase** (consuming  $H_2$  and  $2H^+$ , producing  $NAD(P)H$  and  $NAD(P)$ ) or by **ferredoxin:NAD oxidoreductase** (consuming  $NAD(P)H$ , producing  $NAD(P)$ ).
- acetyl CoA** is converted to **acetyl phosphate** by **phosphotransacetylase (PTA)** (consuming  $CoASH$ , producing  $P_i$ ).
- acetyl phosphate** is converted to **acetate (Ace)** by **acetate kinase (AK)** (consuming ADP, producing ATP).
- acetyl CoA** is converted to **acetaldehyde** by **acetaldehyde dehydrogenase (ACDH)** (consuming NAD, producing NADH and  $CoASH$ ).
- acetaldehyde** is converted to **ethanol (Eth)** by **alcohol dehydrogenase (ADH)** (consuming  $NAD(P)H$ , producing  $NAD(P)$ ).
- acetyl CoA** can also be converted to **formate (For)** by **formate lyase (FL)** (consuming  $CO_2$ , producing  $H_2$ ).

**Other Pathways and Products:**

- fructose 6-phosphate (F6P)** can be converted to **fructose** by **fructokinase** (consuming ATP, producing ADP).
- fructose** can be converted to **xylose** by **xylose isomerase**.
- xylose** is converted to **xylulose** by **xylose isomerase**.
- xylulose** is converted to **xylulose-5-phosphate** by **xylulokinase** (consuming ATP, producing ADP).
- xylulose-5-phosphate** is converted to **glyceraldehyde 3-phosphate (GAP)** by **phosphoketolase**.
- glyceraldehyde 3-phosphate (GAP)** can be converted to **glycerol 1-phosphate (G1P)** by **glycerol 1-phosphate synthase** (consuming NAD, producing NADH).
- glycerol 1-phosphate (G1P)** is converted to **lipid synthesis**.
- dihydroxyacetone phosphate (DHP)** is converted to **methylglyoxal (MGX)** by **methylglyoxal synthase (MGS)** (consuming  $P_i$ ).
- methylglyoxal (MGX)** is converted to **acetol (ATL)** by **glyoxalase** (consuming NADH, producing NAD).
- acetol (ATL)** is converted to **1,2-propanediol (PPD)** by **glyoxalase** (consuming NADH, producing NAD).
- 1,2-propanediol (PPD)** can be converted to **D-lactate (Lac)** by **glyoxalase**.

Figure 35

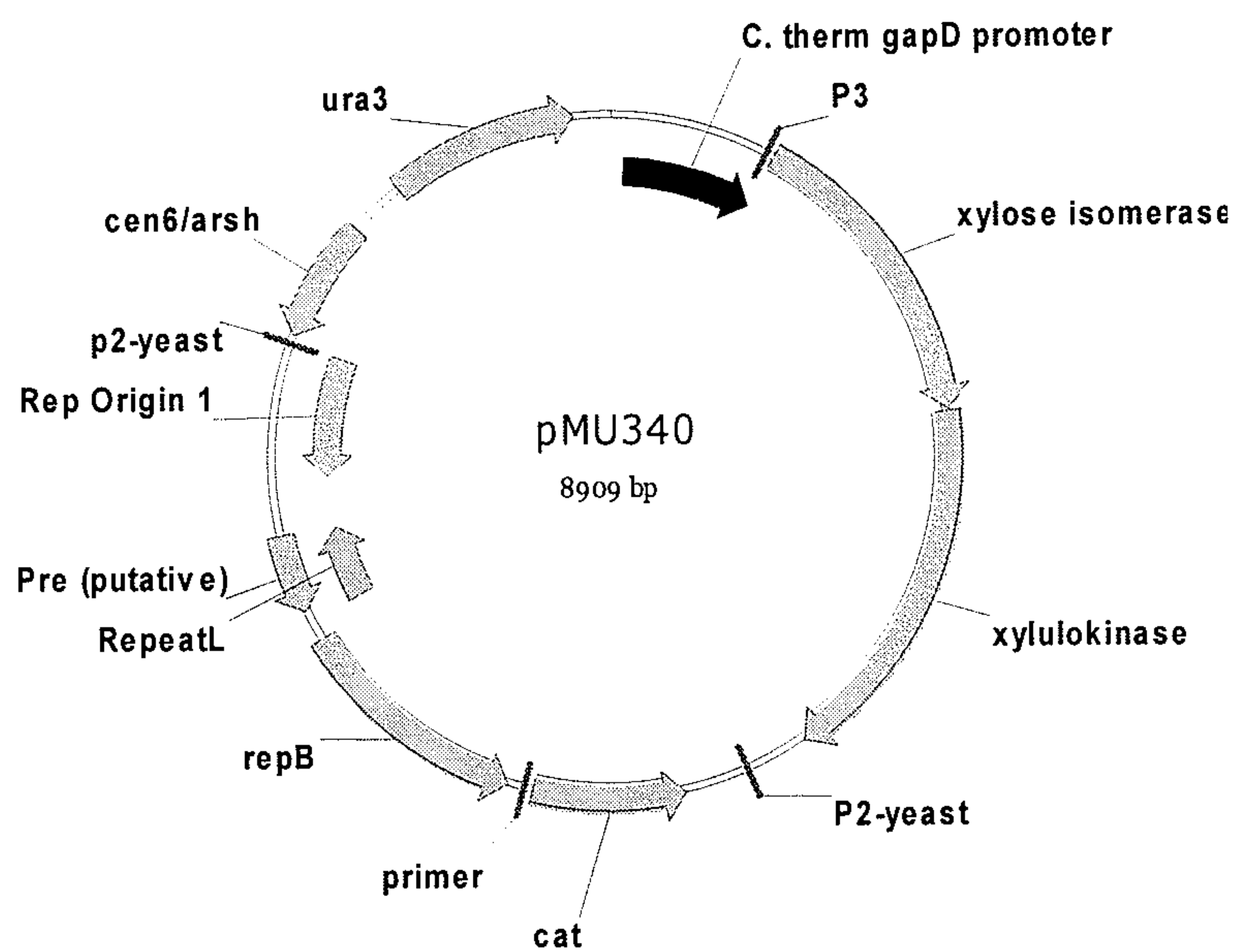


Figure 36

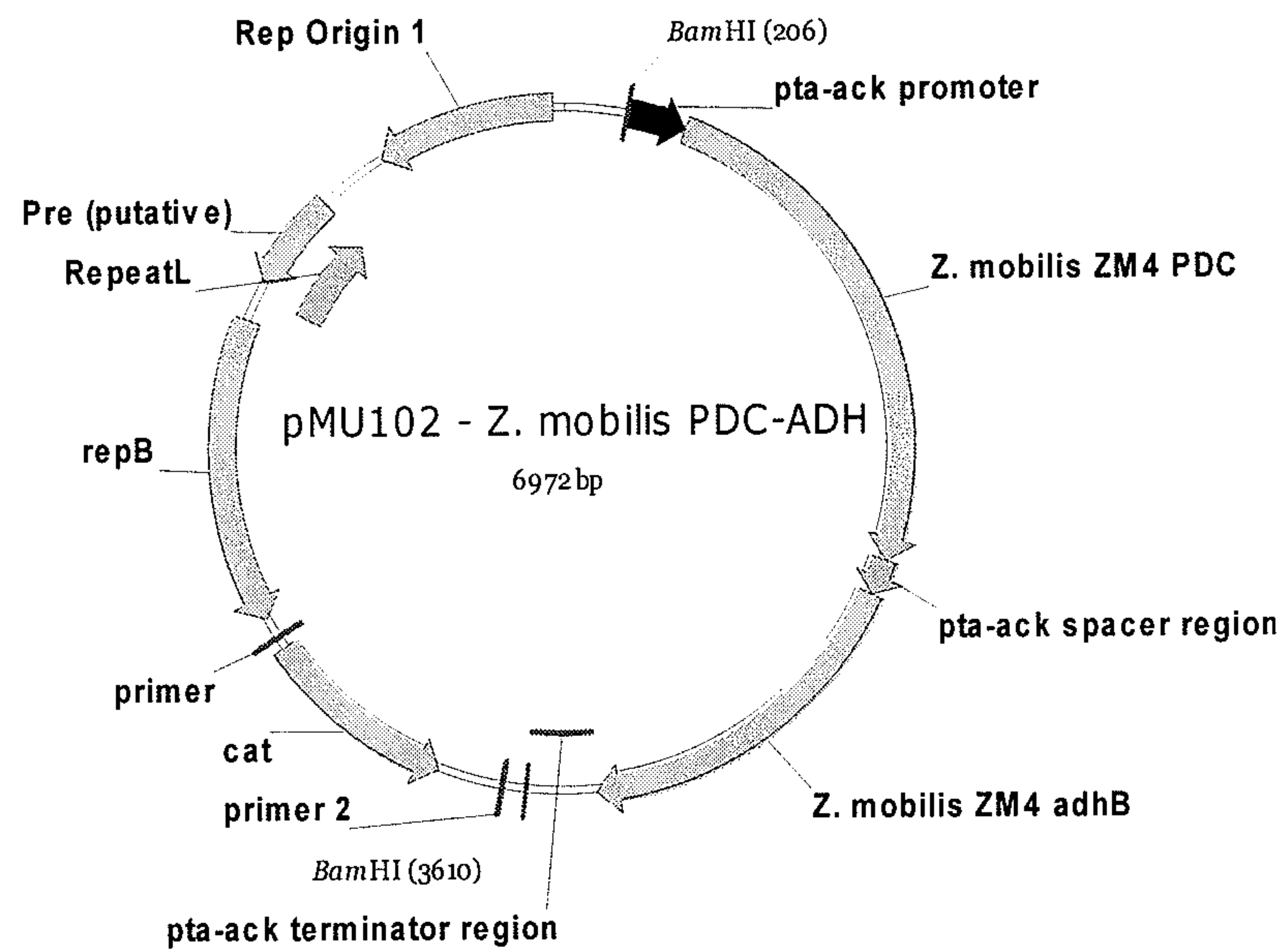


Figure 37

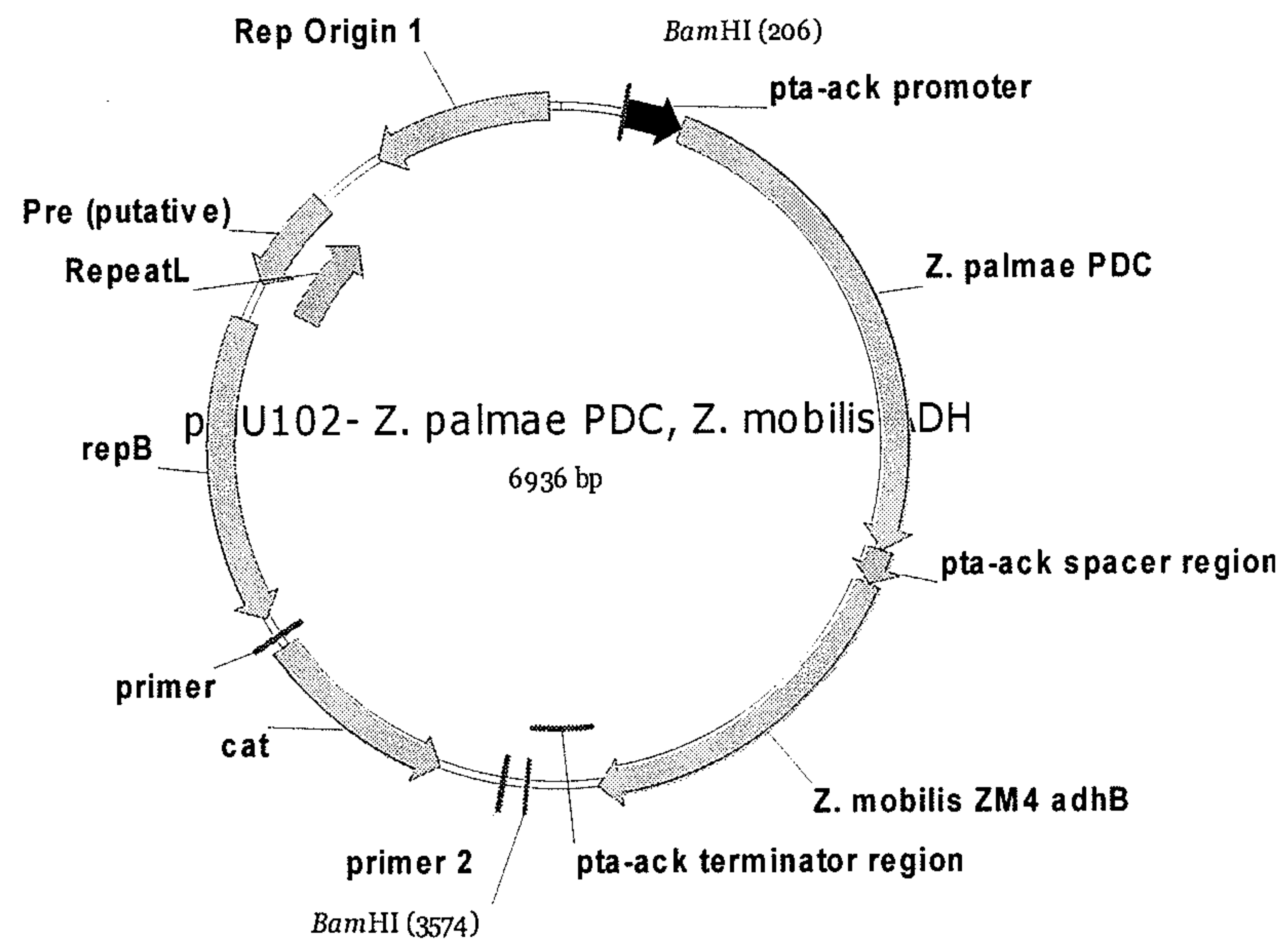


Figure 38

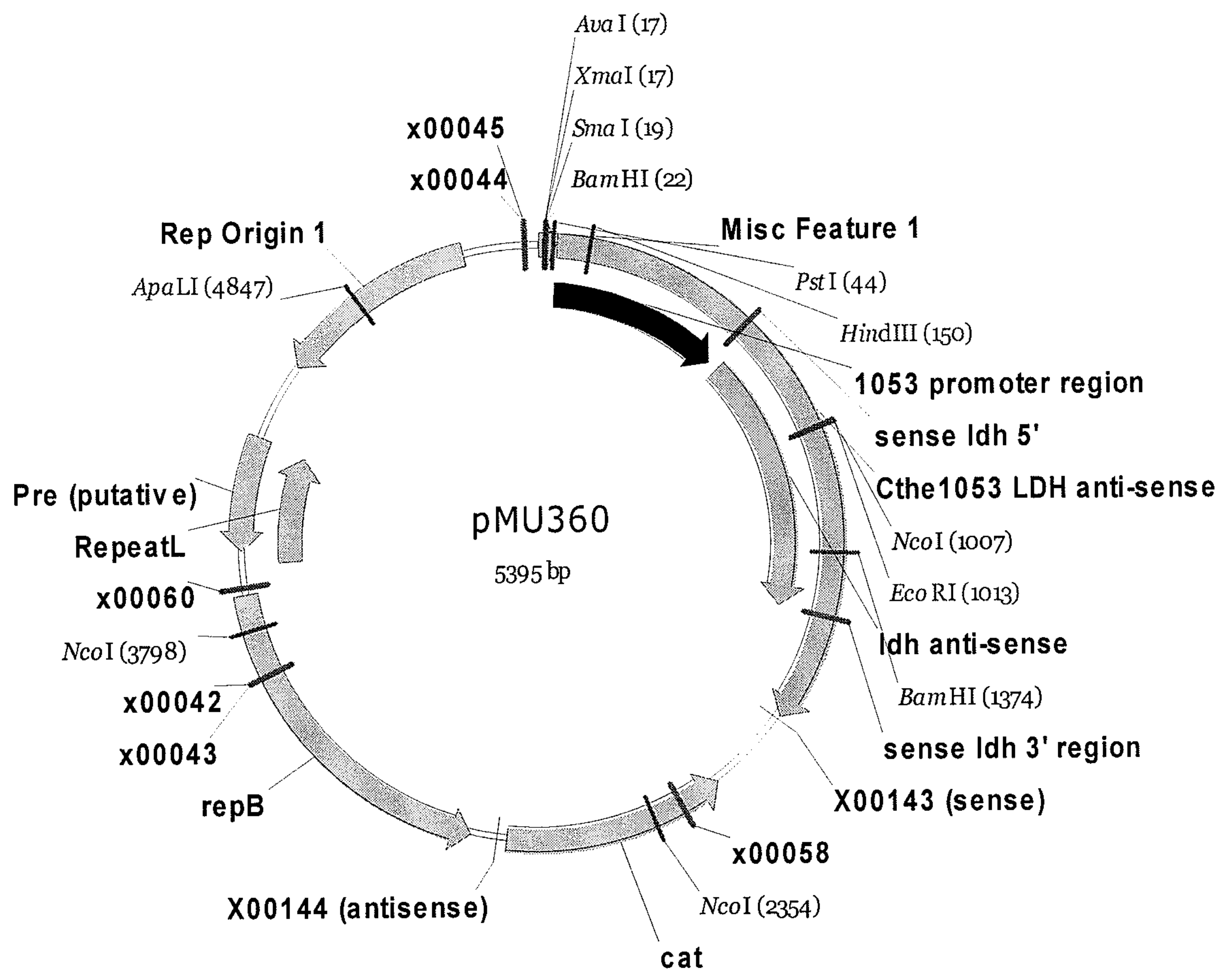


Figure 39

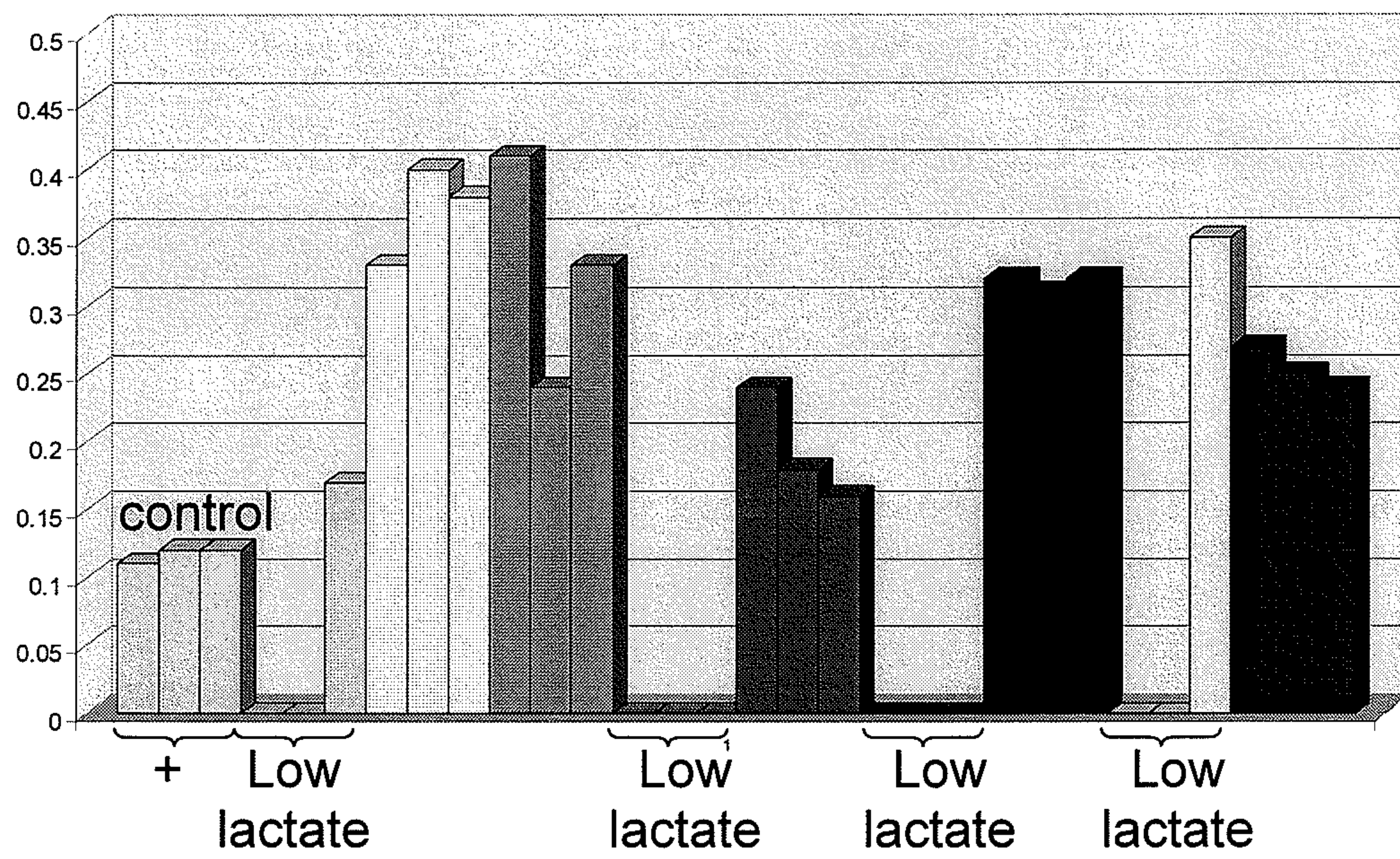


Figure 40

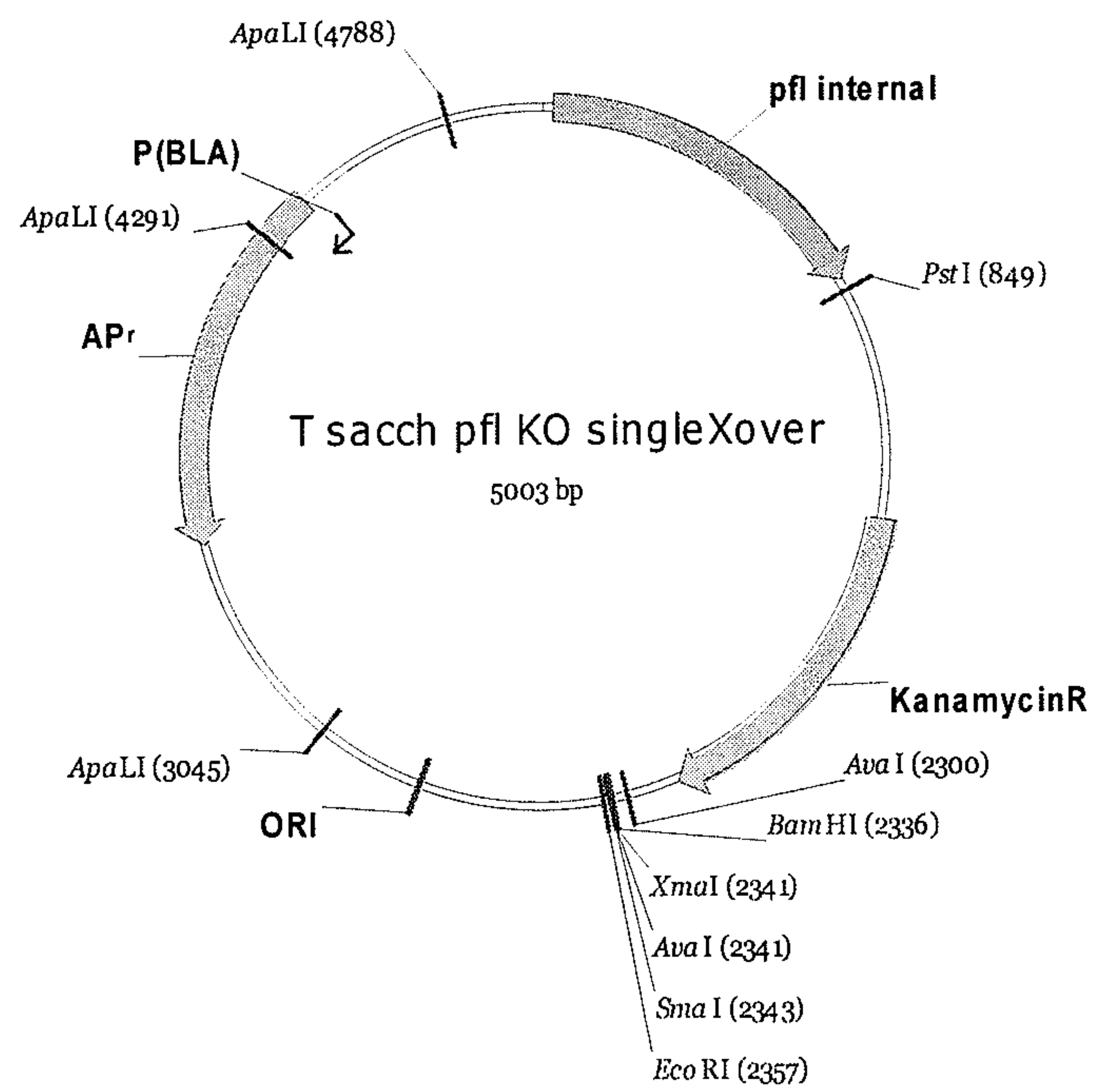


Figure 41

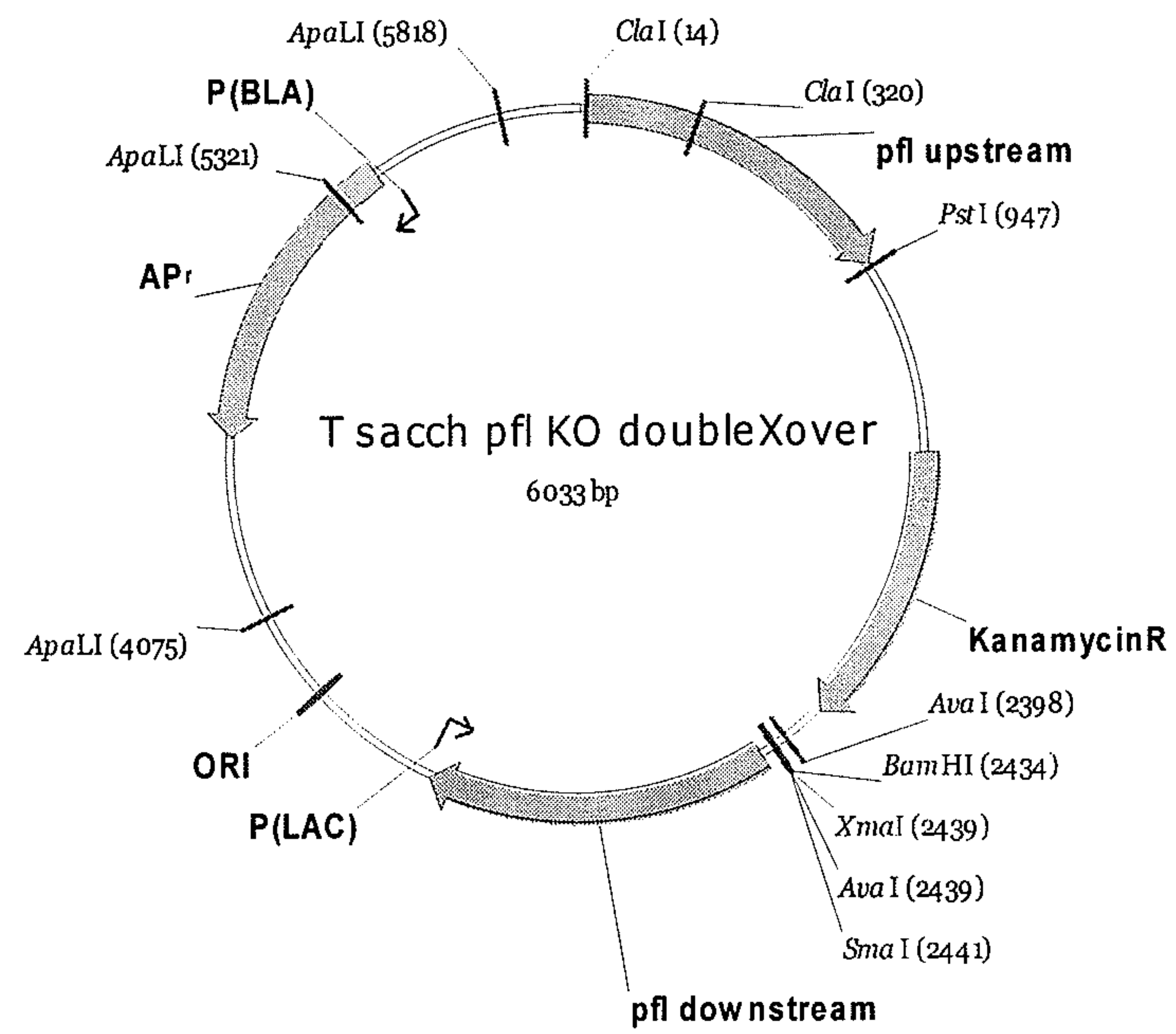


Figure 42

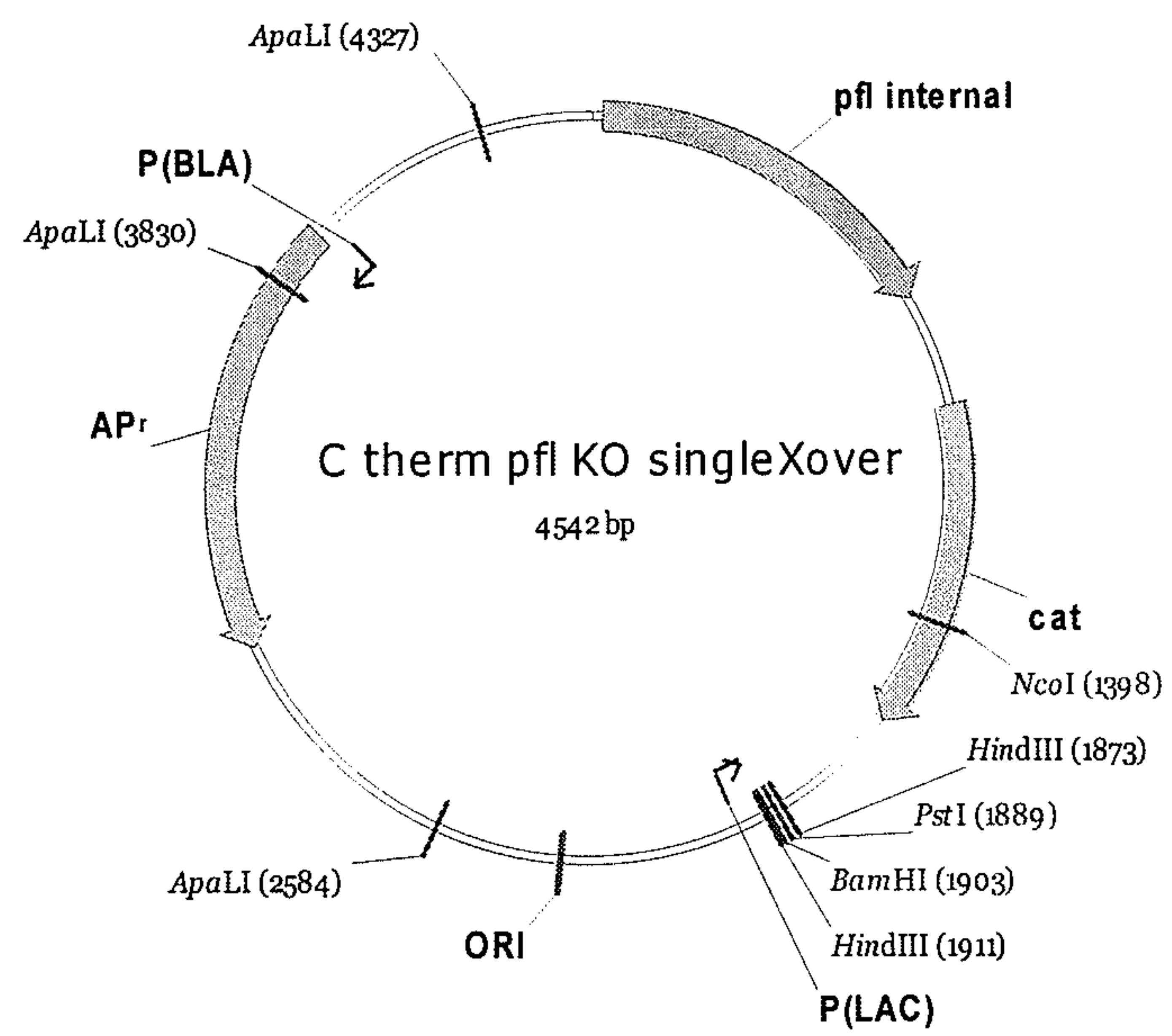


Figure 43

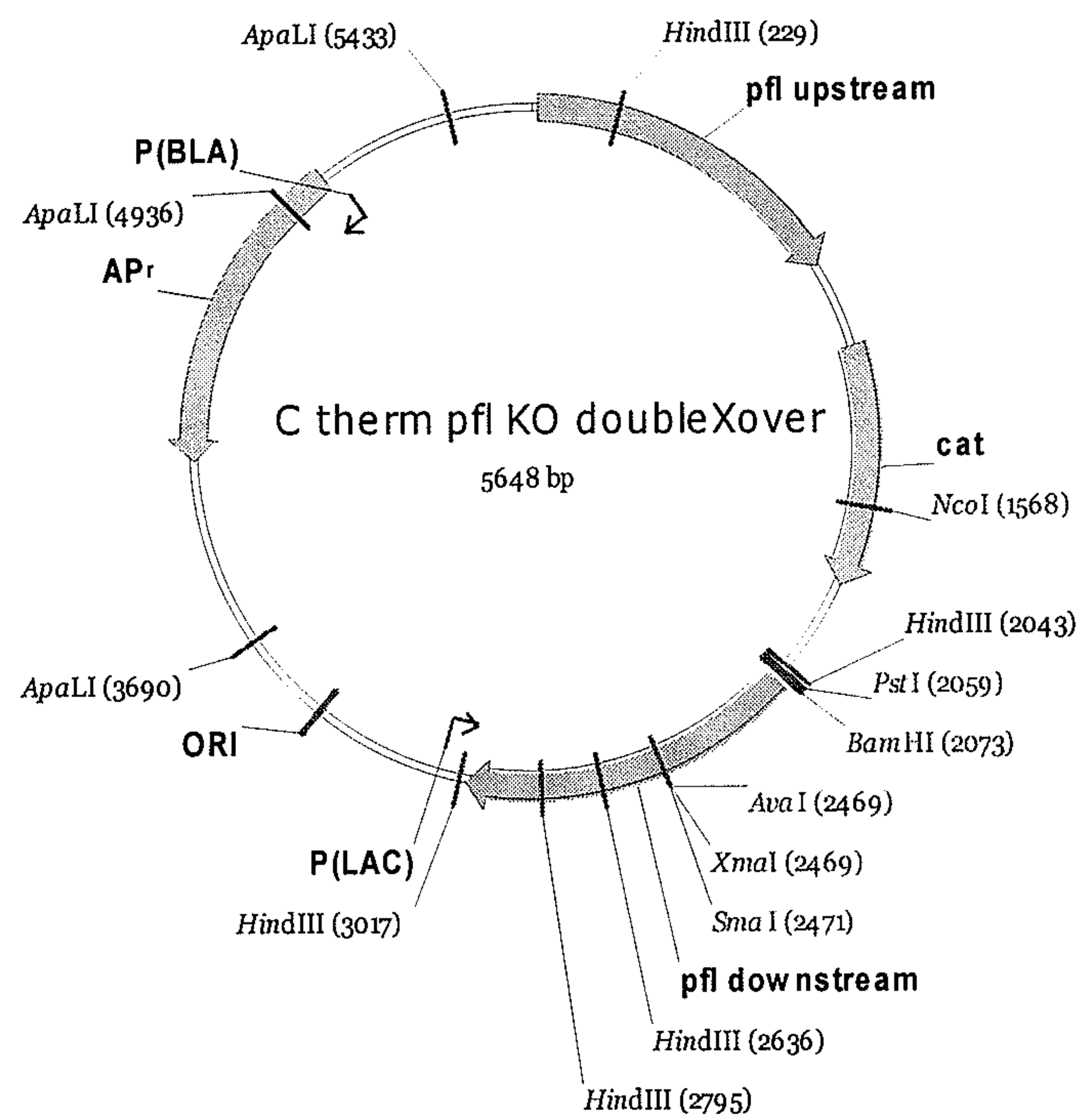


Figure 44

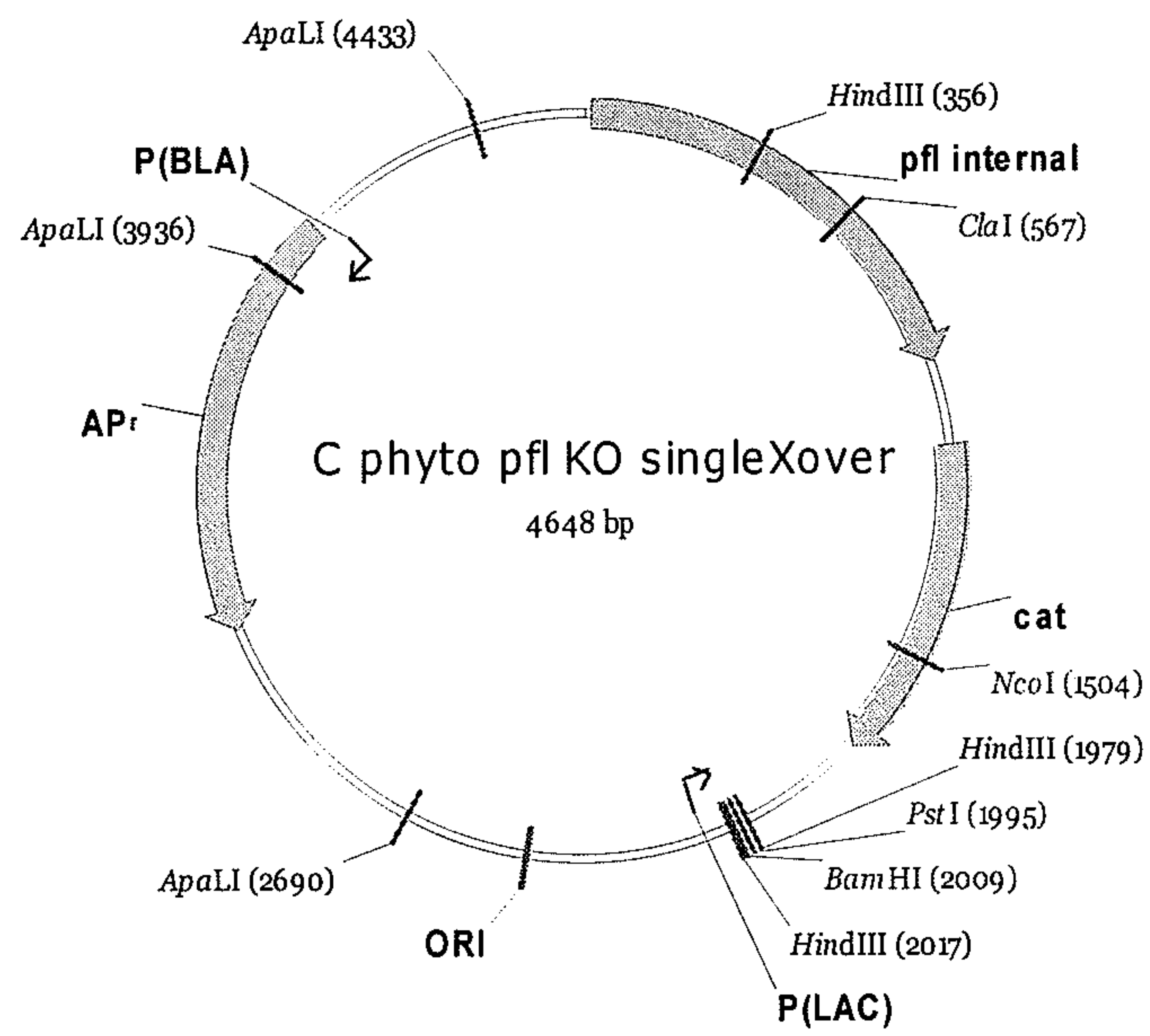


Figure 45

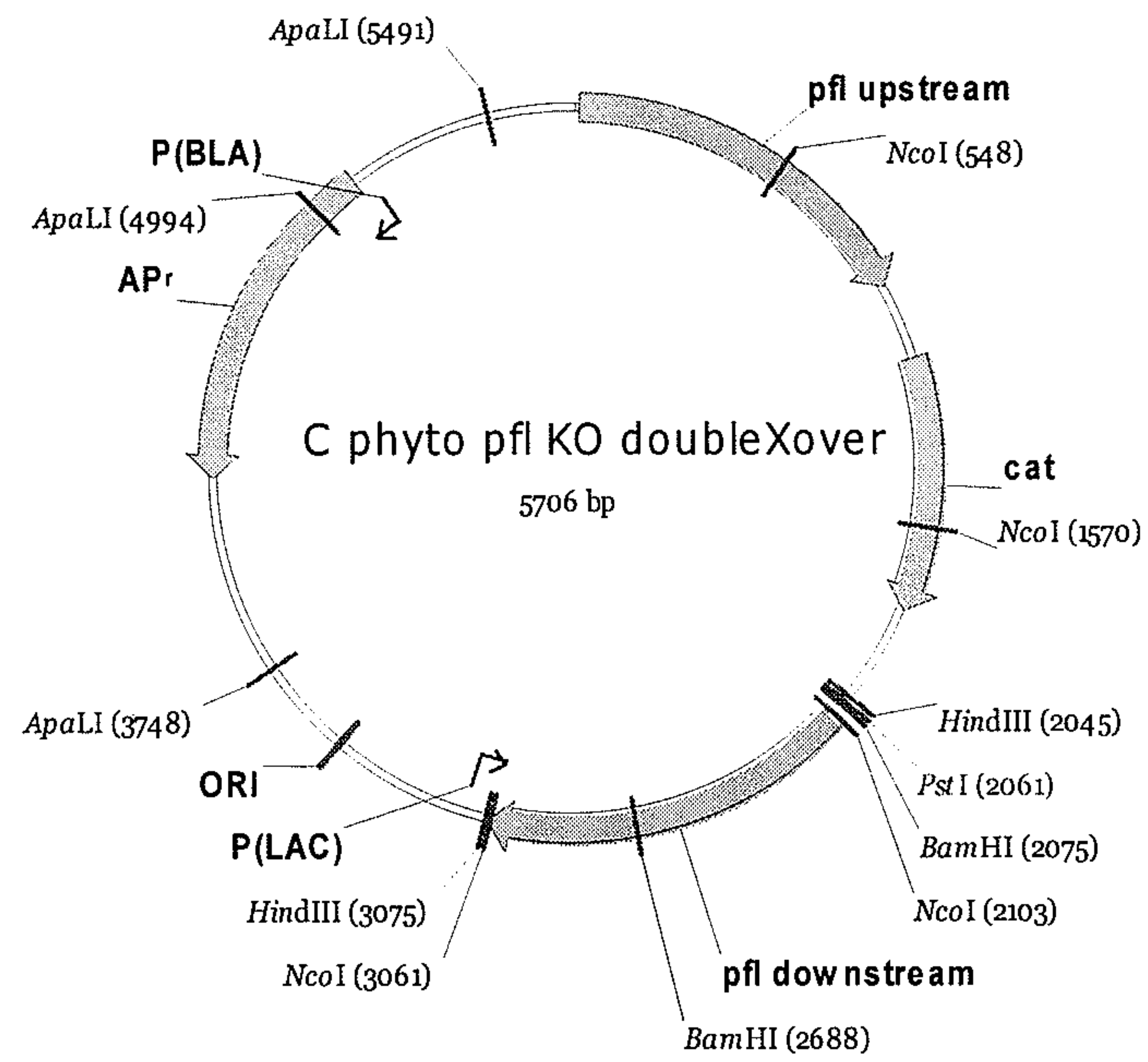


Figure 46

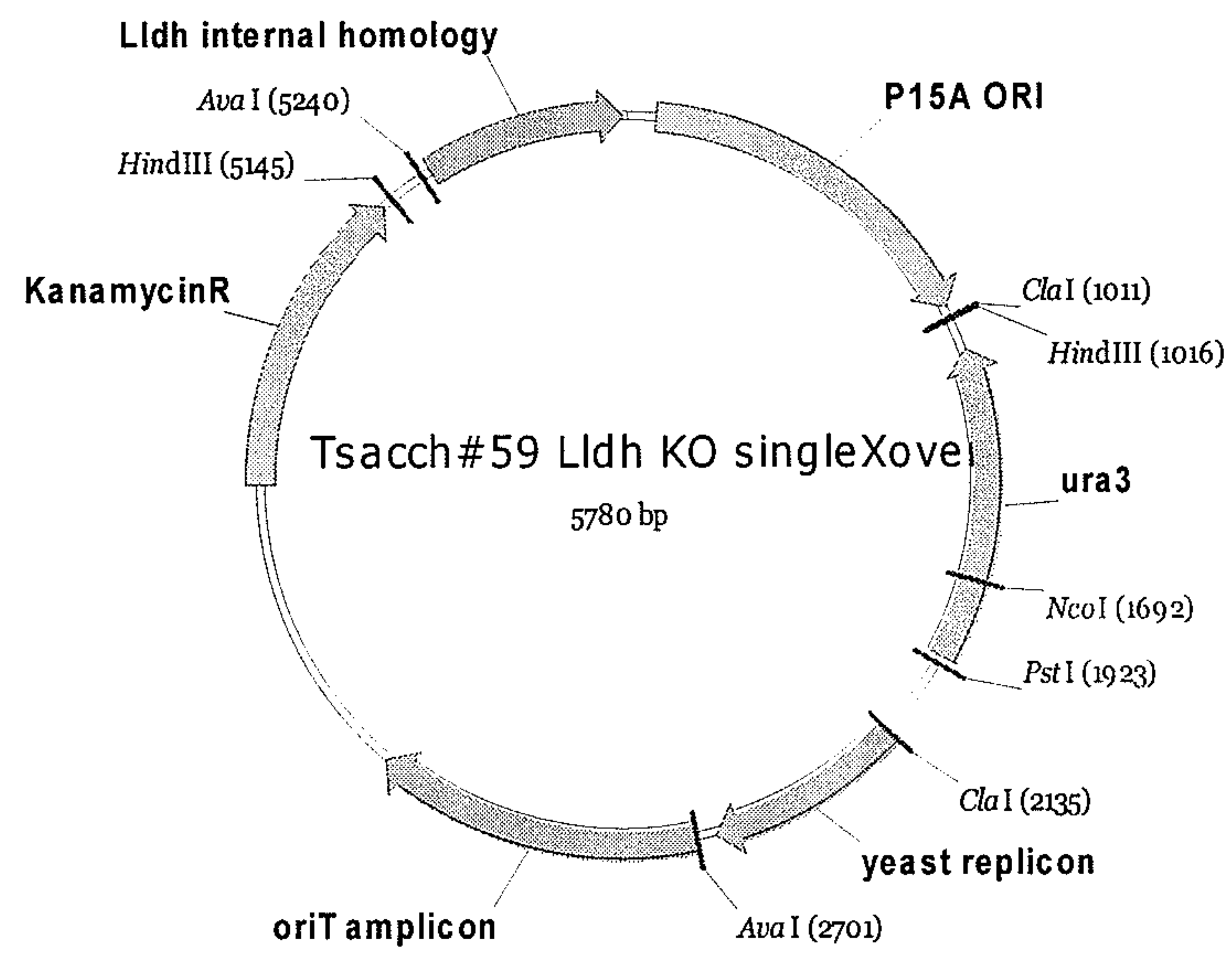


Figure 47

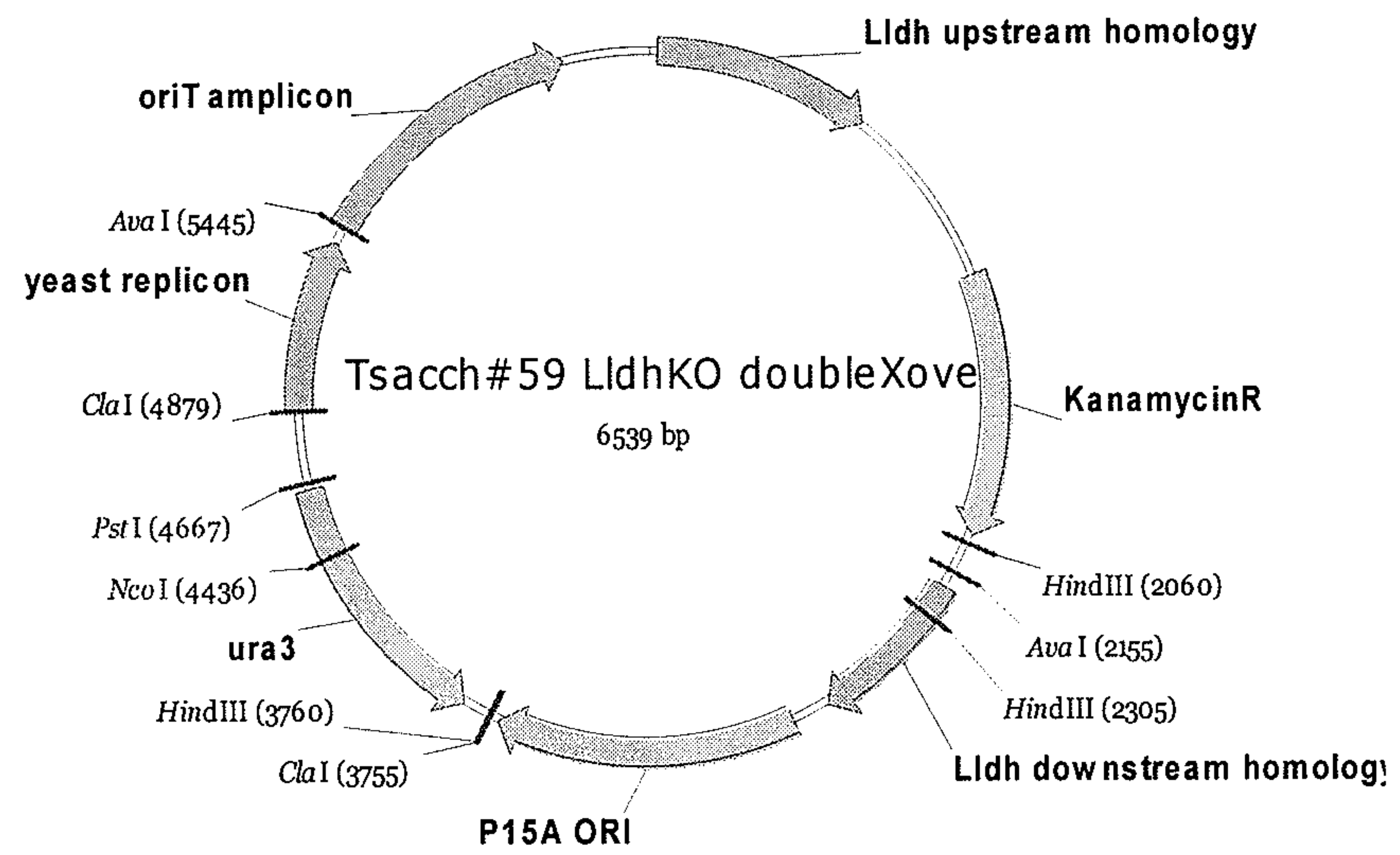


Figure 48

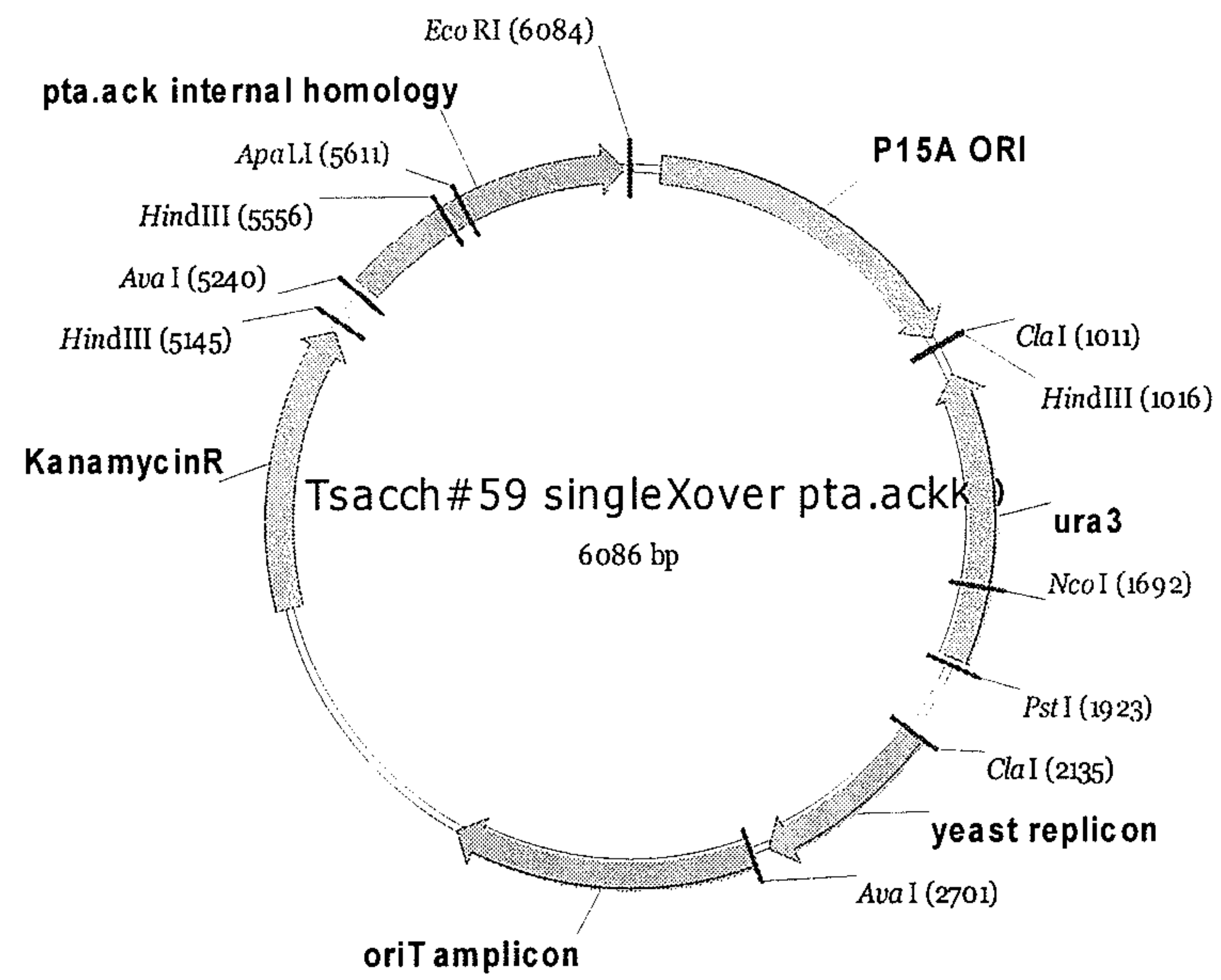


Figure 49

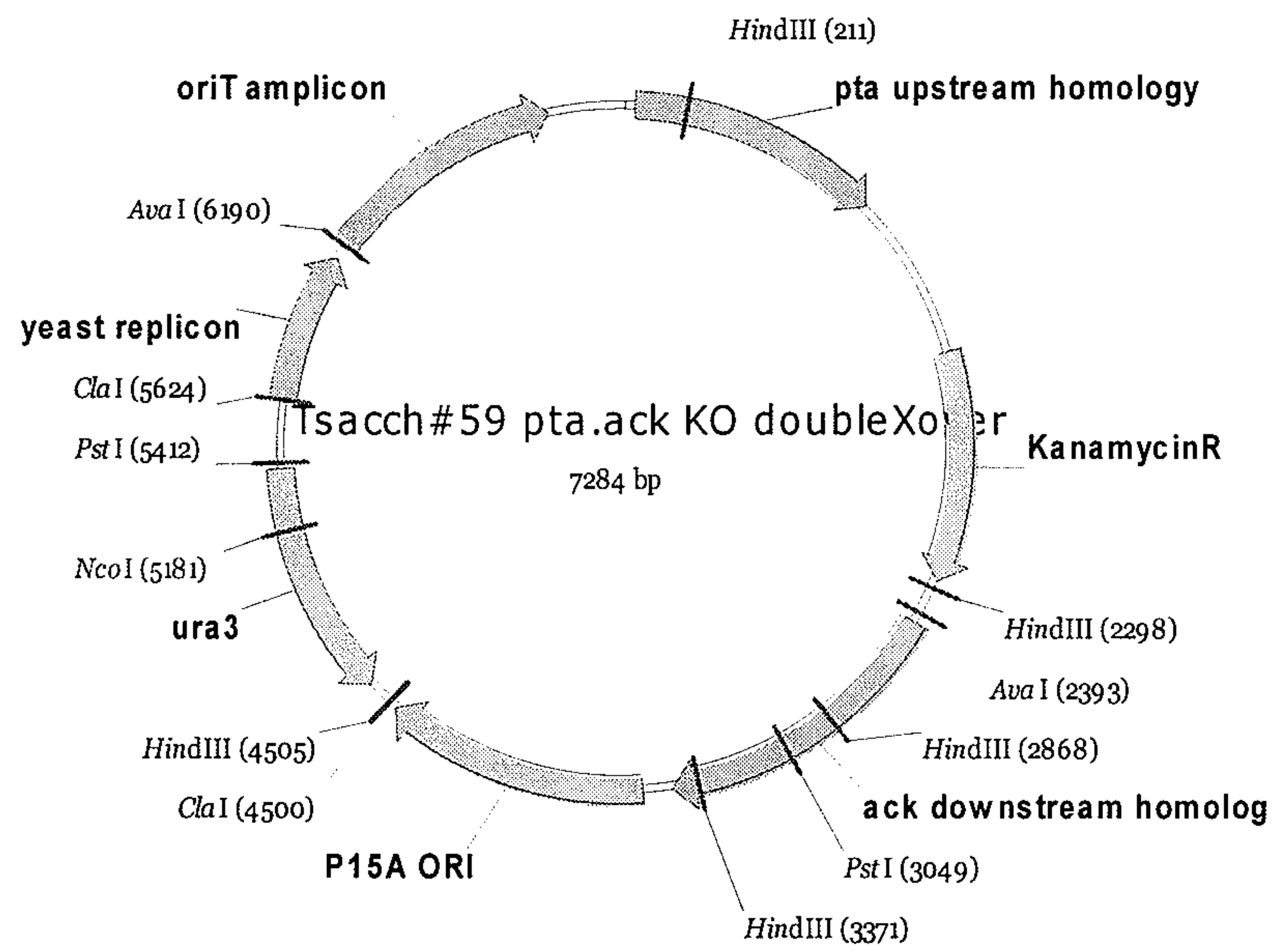


Figure 50

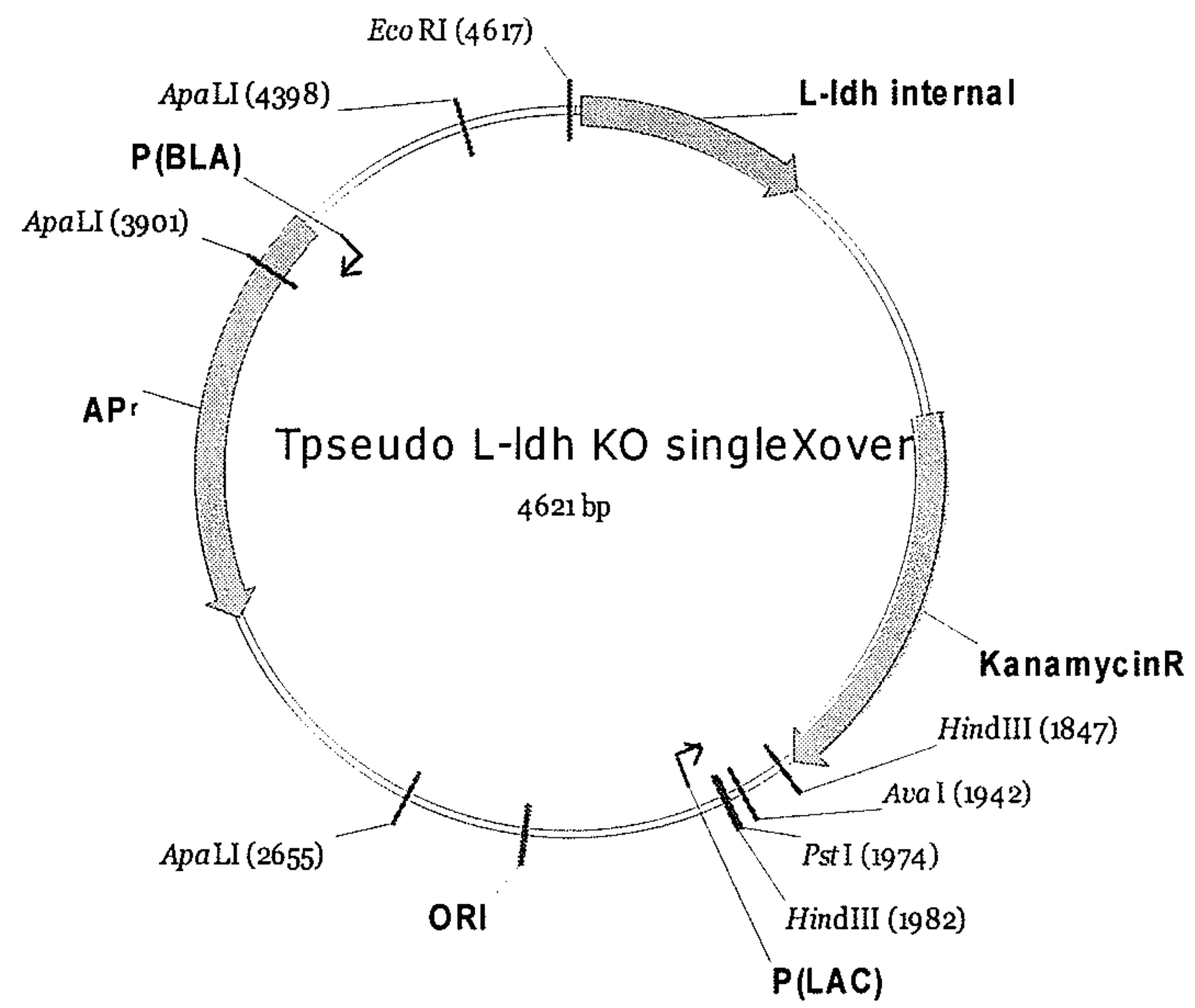


Figure 51

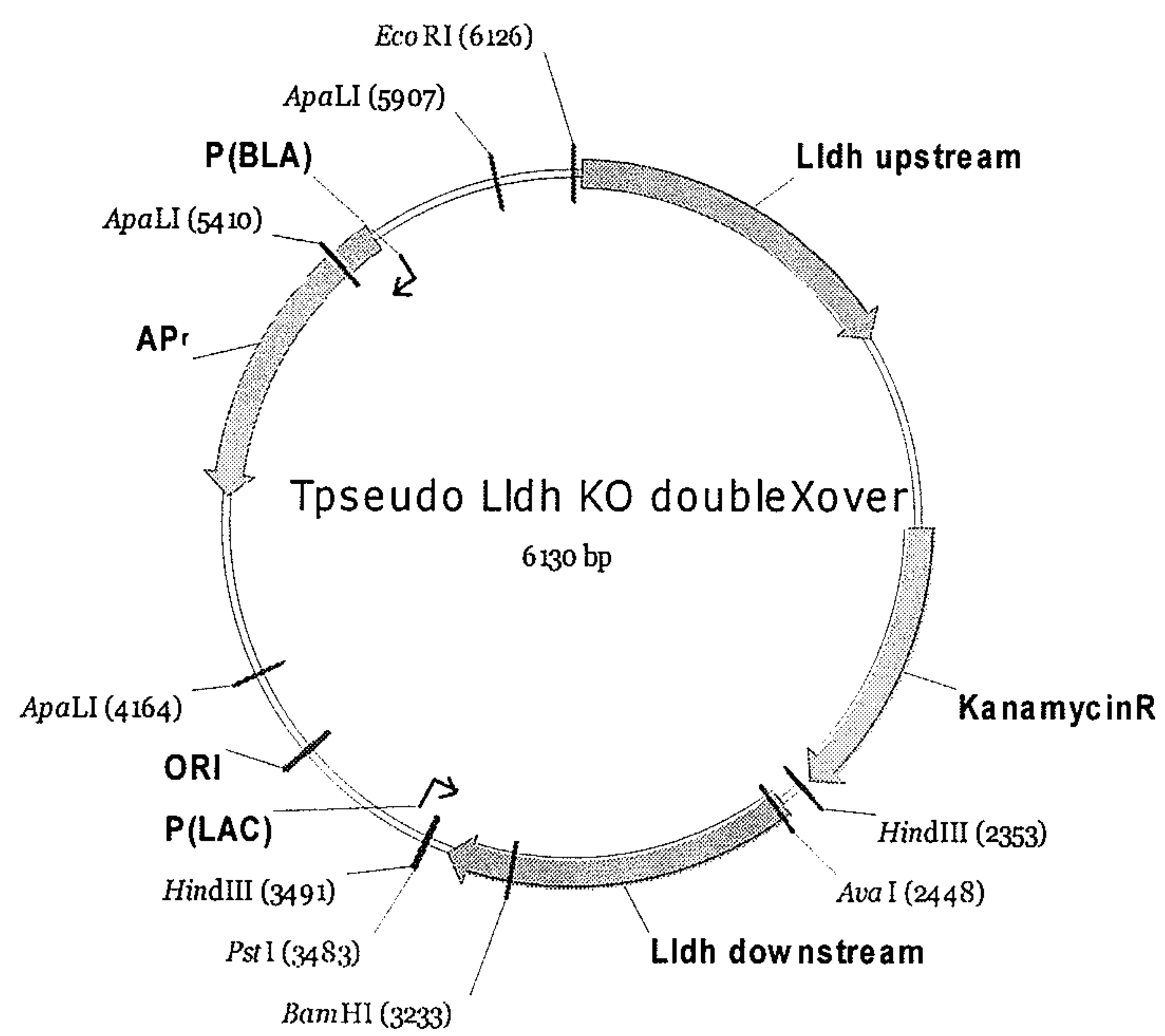


Figure 52

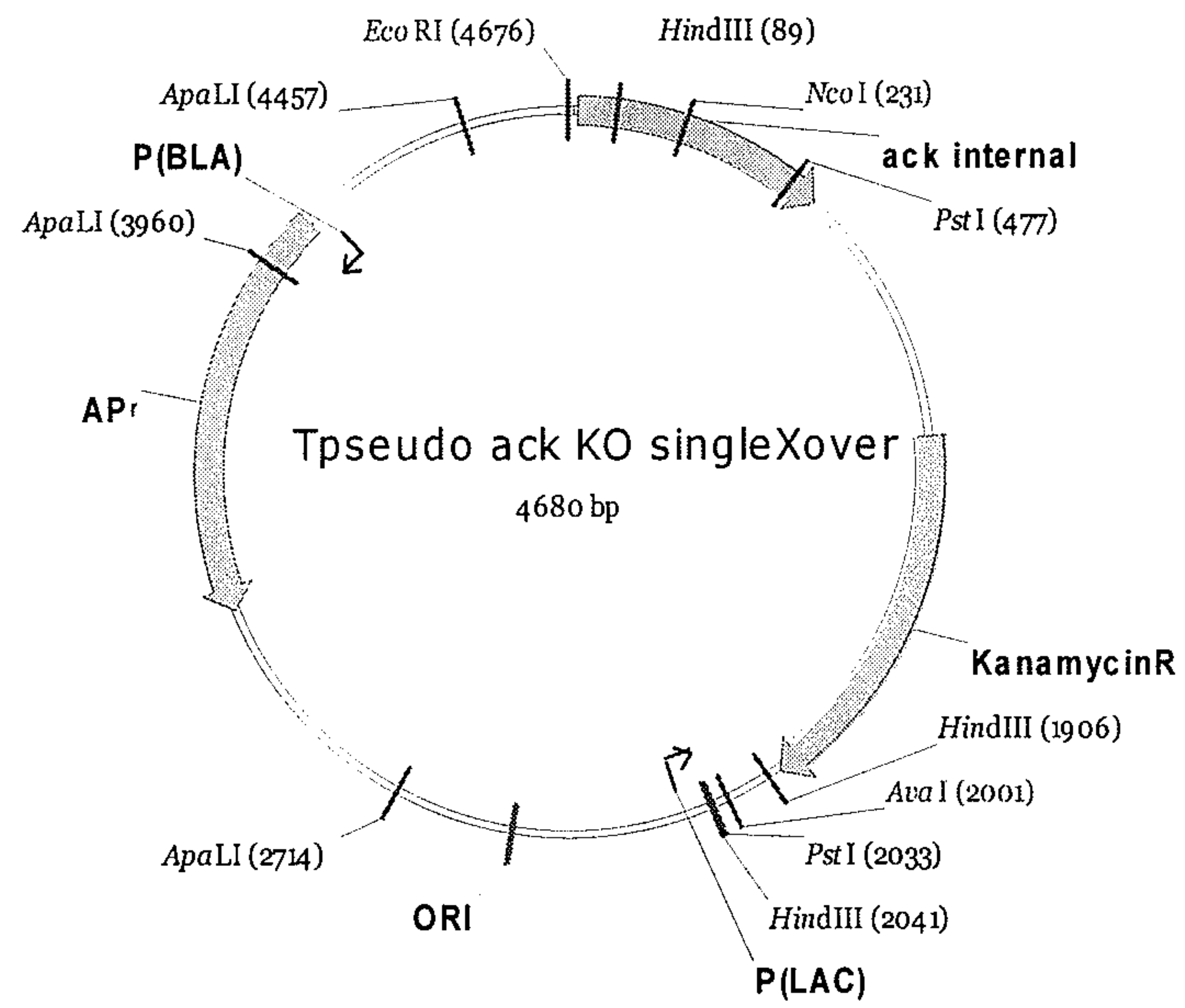


Figure 53

