Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2744659 A1 2010/06/03

(21) 2 744 659

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2009/11/26
(87) Date publication PCT/PCT Publication Date: 2010/06/03
(85) Entrée phase nationale/National Entry: 2011/05/25
(86) N° demande PCT/PCT Application No.: EP 2009/065923
(87) N° publication PCT/PCT Publication No.: 2010/060965
(30) Priorité/Priority: 2008/11/28 (EP08291120.7)

(51) Cl.Int./Int.Cl. *C12N 9/42* (2006.01),
C12P 7/06 (2006.01)

(71) Demandeurs/Applicants:
LE CENTRE NATIONAL DE RECHERCHE SCIENTIFIQUE (CNRS), FR;
TOTAL S.A., FR;
L'UNIVERSITE DE LA MEDITERRANEE, FR;
L'UNIVERSITE DE PROVENCE, FR;
L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES (INSA), FR

(72) Inventeurs/Inventors:
FIEROBE, HENRI-PIERRE, FR;
CHANAL-VIAL, ANGELIQUE, FR

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : POLYPEPTIDES CELLULOLYTIQUES ET LEUR UTILISATION DANS DES MICRO-ORGANISMES POUR LA PRODUCTION DE SOLVANTS ET DE COMBUSTIBLES

(54) Title: CELLULOLYTIC POLYPEPTIDES AND THEIR USE IN MICRO-ORGANISMS FOR THE PRODUCTION OF SOLVENTS AND FUELS

(57) Abrégé/Abstract:
The invention relates to applications of a cellulase of Pseudomonas sp. ND137, functional fragments and/or variants and engineered forms thereof, in the context of industrial bioprocessing, more particularly solvent production.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 June 2010 (03.06.2010)

PCT

(10) International Publication Number
WO 2010/060965 A1

(51) International Patent Classification:
C12N 9/42 (2006.01) *C12P 7/06* (2006.01)

(21) International Application Number:
PCT/EP2009/065923

(22) International Filing Date:
26 November 2009 (26.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
08291120.7 28 November 2008 (28.11.2008) EP

(71) Applicants *(for all designated States except US)*: **TOTAL S.A.** [FR/FR]; 2 Place Jean Millier, La Défense 6, F-92400 Courbevoie (FR). **LE CENTRE NATIONAL DE RECHERCHE SCIENTIFIQUE (CNRS)** [—/FR]; 3 rue Michel-Ange, F-75794 Paris Cedex 16 (FR). **L'UNIVERSITE DE LA MEDITERRANNEE** [FR/FR]; Jardin du Pharo, 58 Boulevard Charles-Livon, F-13284 Marseille Cedex 7 (FR). **L'UNIVERSITE DE PROVENCE** [FR/FR]; 3 Place Victor Hugo, F-13331 Marseille Cedex 3 (FR). **L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES (INSA)** [FR/FR]; 135 Avenue de Rangueil, F-31077 Toulouse Cedex 4 (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FIEROBE, Henri-Pierre** [FR/FR]; 31 chemin Joseph Aiguier, F-13009 Marseille (FR). **CHANAL-VIAL, Angélique** [FR/FR]; 3 place Eugène Bertrand, F-13012 Marseille (FR).

(74) Agents: **PAEMEN, Liesbet Rita Johanna** et al.; De Clercq Brants & Partners, E. Gevaertdreef 10a, B-9830 Sint-Martens-Latem (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

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

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

WO 2010/060965 A1

(54) **Title**: CELLULOLYTIC POLYPEPTIDES AND THEIR USE IN MICRO-ORGANISMS FOR THE PRODUCTION OF SOLVENTS AND FUELS

(57) **Abstract**: The invention relates to applications of a cellulase of *Pseudomonas* sp. ND137, functional fragments and/or variants and engineered forms thereof, in the context of industrial bioprocessing, more particularly solvent production.

CELLULOLYTIC POLYPEPTIDES AND THEIR USE IN MICRO-ORGANISMS FOR THE PRODUCTION OF SOLVENTS AND FUELS

FIELD OF THE INVENTION

The present invention relates to the area of biotechnology and genetic engineering and particularly concerns strategies for utilisation of cellulosic and lignocellulosic materials by means of suitable enzymatic agents and micro-organisms expressing such. More specifically, the invention relates to applications of the cellulolytic polypeptides and their use in solvent production either directly or expressed in solventogenic micro-organisms.

BACKGROUND OF THE INVENTION

Cellulosic and lignocellulosic materials are major constituents of plant biomass, and cellulose polymers found therein can provide a significant source of glucose or other fermentable mono- and oligosaccharides for use in the production of solvents, such as ethanol, acetone or butanol.

Cellulose polymers can be hydrolysed by cellulose-depolymerising enzymes commonly known as cellulolytic enzymes or cellulases. For example, hydrolysis of native cellulose mainly involves four cellulase types: cellobiohydrolase (1,4-β-D-glucan cellobiohydrolase, EC 3.2.1.91), endo-β-1,4-glucanase (endo-1,4-β-D-glucan 4-glucanohydrolase, EC 3.2.1.4), endo-processive cellulases (EC 3.2.1.4./3.2.1.91), and β-glucosidase (EC 3.2.1.21). Cellulases and related enzymes have been widely utilised in various areas of biotechnology including in food, beer, wine, animal feeds, textile production and laundering, pulp and paper industry, agricultural industry and others (for review see Bhat 2000. Biotechnical Advances 18: 355-383).

Cellulases can vary in their characteristics, such as *inter alia* they may act as endoglucanases or as processive exoglucanases, they may generate monomers or oligomers of various lengths, they may have contrasting abilities to hydrolyse distinct cellulose forms such as crystalline cellulose, semi-crystalline cellulose, amorphous cellulose or hemicellulose, they may further display different strength of binding to cellulose substrates, different kinetic parameters, *etc*. Therefore, there is a permanent need for

further characterisation of additional cellulases, so as to identify potent enzymes and optimize enzyme combinations.

The recombinant expression of cellulases in solventogenic micro-organisms allows the production of useful solvents including *inter alia* ethanol by these micro-organisms directly from cellulose-containing materials. To date, primarily enzymes from the cellulolytic Clostridium species have been used for this purpose. However, there remains a need for the identification of enzymes with a stronger enzymatic activity.

SUMMARY OF THE INVENTION

The inventors have identified a cellulolytic enzyme referred to herein as “ACLA”, as an enzyme which is particularly suitable for use in the context of industrial s production of solvents, fuels and chemical intermediates as it has an elevated activity compared to other bacterial cellulase enzymes, such as those previously identified from *Clostridium cellulolyticum* or *thermocellum*.

Therefore, ACLA is advantageous for direct use in in vitro degradation of cellulose and in the context of heterologous production in micro-organism with the aim of producing fermentable sugars, more particularly in solvent-producing (solventogenic) micro-organisms.

In one aspect, the invention provides the use of the ACLA polypeptide, or a functional fragment, homologue or variant thereof in cellulose degradation, particularly in the industrial generation of fermentable sugars, more particularly in solvent production. Particularly, the invention provides methods for the degradation of a substance comprising cellulose, such as lignocellulosic or cellulosic material or biomass, comprising contacting the substance with the ACLA polypeptide or a functional fragment, homologue or variant thereof or with arecombinant micro-organism expressing ACLA or a functional fragment, homologue or variant thereof as taught herein. In particular embodiments, the substance comprises or is enriched for crystalline cellulose.

Accordingly, the invention provides methods for producing a solvent, fuel or chemical intermediate from a substance comprising cellulose, comprising applying the methods for the degradation of a substance comprising cellulose according to the invention, i.e. comprising contacting said substance with a recombinant micro-organism expressing ACLA or a functional fragment, homologues, variants of the ACLA polypeptide and/or with ACLA or functional fragments, homologues, variants of the ACLA polypeptide as taught herein. In particular embodiments, the substance comprises, or is enriched for, crystalline cellulose.

Particular embodiments of the invention provide methods for producing a solvent, fuel or chemical intermediate from a substance comprising cellulose, comprising

a) expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue in a recombinant host cell, and

b) contacting said substance with said ACLA polypeptide.

In further particular embodiments, said methods may further comprise the step of separating the ACLA enzyme or active fragment from the recombinant host cell.

The present inventors have found that ACLA is particularly suitable for the industrial degradation of cellulose-containing material (including lignocellulosic biomass). Accordingly, the present invention provides the use of ACLA in simultaneous saccharification and fermentation (SSF) or sequential (separate) hydrolysis and fermentation (SHF). Such methods involve batch degradation of cellulose-containing material under controlled conditions.

Accordingly, one aspect of the invention provides micro-organisms expressing the ACLA polypeptide, functional fragments, homologues, or variants of the ACLA polypeptide which micro-organisms are capable of degrading cellulose into fermentable sugars. This implies that the enzyme is secreted by the micro-organism such that it can degrade the cellulose-containing material with which it is contacted. Such micro-organisms are of use in industrial methods for producing solvents, fuels and chemical intermediates, as described herein.

However, embodiments where the polypeptides are expressed intracellularly are also contemplated. For example, the invention provides micro-organism expressing the ACLA polypeptide such that it acts on cellulose polymers internalised by the micro-organism. Alternatively, it is envisaged that the ACLA polypeptide is released upon lysis of at least a fraction of the micro-organisms by which it is expressed.

In particular embodiments, the methods for producing a solvent, fuel or chemical intermediate from a substance comprising cellulose according to the invention comprise growing recombinant micro-organisms as described herein on cellulose substrates comprising or enriched in crystalline, semi-crystalline or amorphous cellulose, thus ensuring direct production of fermentable sugars from cellulose-containing substrates which can be used in the production of chemicals, chemical intermediates, fuels or solvents such as ethanol.

A further aspect of the invention provides isolated ACLA polypeptides of *Pseudomonas sp.* Strain ND 137 or homologues thereof, or functional fragments and/or variants of the ACLA polypeptide or of the homologue, and their use in the industrial degradation of cellulose. More particularly the ACLA polypeptide is characterized by the sequence provided herein.

Yet a further aspect of the present invention provides recombinant micro-organisms expressing the ACLA polypeptide, a functional fragment, homologue or variant thereof which can be used in consolidated bioprocessing, i.e. which are capable of converting the degradation products of cellulose material directly into an industrial product such as a solvent, fuel or chemical intermediate. Such micro-organisms may, in addition to the gene encoding the ACLA polypeptide, a functional fragment, homologue or variant, contain other transgenes which ensure production of fuels, solvents or other chemicals or chemical intermediates.

According to particular embodiments of the present invention, the recombinant micro-organisms comprise a recombinant nucleic acid molecule encoding the ACLA polypeptide or homologue thereof, or encoding a functional fragment and/or variant of the ACLA polypeptide or of a homologue thereof, operably linked to to regulatory sequences which allow for expression by the micro-organisms and/or a secretion signal sequence which allows for secretion by the micro-organisms and contain one or more recombinant nucleic acid molecules involved in the production of a product of interest from a fermentable sugar.

In particular embodiments, the invention provides solventogenic micro-organisms expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue.

The recombinant solventogenic micro-organism may comprise a recombinant nucleic acid molecule encoding the ACLA polypeptide or a homologue thereof, or encoding a functional fragment and/or variant of said ACLA polypeptide or of said homologue, operably linked to regulatory sequences which allow for expression in said micro-organism.

In particular embodiments, the recombinant solventogenic micro-organism may produce one or more solvents chosen from ethanol, acetone, butanol, propionic acid, butyric acid, ether and glycerine.

In further embodiments, the recombinant solventogenic micro-organism may produce, or may be engineered to produce, at least or mainly ethanol. The industrial importance of ethanol is rapidly increasing largely due to its utility as an environmentally acceptable fuel. Hence, in an embodiment, the recombinant solventogenic micro-organism may be an ethanologenic micro-organism.

In particular embodiments of the invention, the recombinant solventogenic micro-organism is a bacterium. In more particular embodiments, the recombinant solventogenic bacterium is an ethanologenic bacterium. Alternatively, the recombinant solventogenic micro-organism is a yeast, more particularly an ethanologenic yeast.

In further particular embodiments, the solventogenic bacterium used in the context of the invention is a Gram-negative bacterium, such as for example a bacterium of the *Zymomonas* species. In another embodiment, the bacterium is a Gram-positive bacterium, such as for example a bacterium of the *Clostridium* species. The invention thus also includes introducing a cellulase originating from a Gram-negative bacterium into a Gram-positive bacterium, which has not been previously suggested.

In particular embodiments, the recombinant solventogenic and preferably ethanologenic micro-organism is a *Clostridium* species, more partiuclarly *Clostridium acetobutylicum*. In other embodiments, the recombinant solventogenic and particularly ethanologenic micro-organism is a *Zymomonas* species, more particularly *Zymomonas mobilis*. These bacteria are particularly proficient as solvent- and particularly ethanol-producing bacteria.

According to particular embodiments of the present invention, the recombinant solventogenic micro-organisms comprise a recombinant nucleic acid molecule encoding the ACLA polypeptide or homologue thereof, or encoding a functional fragment and/or variant of the ACLA polypeptide or of a homologue thereof, operably linked to to regulatory

sequences which allow for expression by the micro-organisms and/or a secretion signal sequence which allows for secretion by the micro-organisms.

Accordingly, the invention provide methods for producing a solvent, fuel, chemical or chemical intermediate from a substance comprising cellulose, such as lignocellulosic or cellulosic material or biomass, comprising treating said substance with one or more of the micro-organisms, more particularly solventogenic micro-organisms as taught herein. In particular embodiments, the substance comprises or is enriched for crystalline cellulose. In most particular embodiments, the solvent is ethanol and the micro-organism is an ethanologenic micro-organism.

It shall also be appreciated that while in the foregoing aspects and embodiments the invention has been disclosed primarily in connection with the recombinant micro-organisms, the invention also encompasses aspects relating to recombinant means and reagents useful for obtaining the recombinant micro-organisms, methods for obtaining the recombinant micro-organisms using such recombinant means and reagents, methods for expressing the desired polypeptides in the recombinant micro-organisms, as well as to the expressed polypeptides and combinations thereof *per se* and methods for preparation thereof.

In a furher aspect, the invention teaches homologues of ACLA and functional fragments and/or variants of such homologues, and their uses in enzymatic cellulose degradation and/or for expression in recombinant micro-organisms, such as in solventogenic micro-organisms. More particularly, an ACLA homologue as contemplated herein comprises a domain homologous to the DX domain of ACLA.

Particular embodiments of the invention provide recombinant solventogenic micro-organisms as taught herein, wherein the homologue of the ACLA polypeptide is the Cel5H polypeptide of *Saccharophagus degradans* strain 2-40.

The present inventors have further identified the domain structure of the native ACLA polypeptide, which can be outlined as GH5-DPR-CBM6-PSL-DZ, wherein GH5 stands for its glycoside hydrolase family 5 domain, DPR for the aspartic acid-proline-rich region, CBM6 for carbohydrate-binding module family 6 domain, PSL for the polyserine linker, and, without being limited to this interpretation, DZ represents a C-terminal domain identified by the present inventors as a putative carbohydrate-binding module.

Accordingly, in particular embodiments the invention provides isolated fragments of the ACLA polypeptide or of a homologue or variant thereof, having the structure PSL-DZ, CBM6-PSL-DZ or DPR-CBM6-PSL-DZ, wherein PSL stands for polyserine linker and DPR stands for an aspartic acid-proline-rich region.

Accordingly, a further aspect of the present invention provides functional fragments of the ACLA enzyme and their use in enzymatic cellulose degradation and/or fermentation by micro-organisms. According to particular embodiments, recombinant micro-organisms are provided which express a functional fragment of the ACLA enzyme comprising one or more domains chosen from or corresponding to the GH5 domain, the CBM6 domain and the DZ domain of ACLA. The presence of one or more said domains in the fragment can endow the latter with the functionalities attributed to the respective domains.

A further aspect of the invention provides engineered forms of the ACLA polypeptide, its homologues, fragments and/or variants, fused to one or more heterologous domains which can supply further functions and activities useful in carbohydrate polymer metabolism, and particularly cellulose metabolism, such as cellulose-depolymerising, cellulose-binding, cellulosome-forming or other activities. Hence, particular embodiments provide recombinant micro-organisms as taught herein, wherein the ACLA polypeptide or homologue thereof, or the functional fragment and/or variant of said ACLA polypeptide or of the homologue, is fused to one or more domains heterologous to the ACLA polypeptide or homologue thereof, chosen from a glycoside hydrolase (GH) catalytic domain, a carbohydrate-binding module (CBM) domain, a cohesin-binding domain such as a dockerin domain, and a hydrophilic (X module) domain of a cellulosomal scaffoldin protein. Additionally or alternatively, the Cel5H polypeptide or homologue thereof, or the functional fragment and/or variant of said Cel5H polypeptide or of said homologue is fused to a heterologous or its natural signal peptide.

A further aspect of the invention relates to the combined use of ACLA enzyme and other polypeptides or enzymes, useful in carbohydrate polymer metabolism, and particularly cellulose metabolism, such as cellulose-depolymerising, cellulose-binding or celloluse degrading enzymes. In further particular embodiments the invention provides the combined use of the ACLA enzyme and other enzymes including other cellulolytic enzymes in the degradation of cellulose, such as, but not limited to, in the context of solvent production. In

particular embodiments the invention provides for mixtures (also referred to as “cocktails” of cellulases for use in the enzymatic degradation of cellulose.

Hence, according to particular embodiments isolated compositions are provided comprising the ACLA polypeptide or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, and further comprising one or more polypeptides participating in carbohydrate polymer metabolism, and particularly cellulose metabolism, preferably one or more enzymes capable of degrading lignocellulosic material, more preferably one or more cellulases as taught herein, optionally combined with xylanases. Such mixtures particularly contain 2 or more, more particularly three or more, most particularly four or more, typically five or six or more different enzymes. In particular embodiments of said composition, the recited components are present as discrete or free polypeptides. In other embodiments, the recited polypeptides are comprised in a hybrid and/or covalent cellulosome.

In alternative embodiments, the invention contemplates advantages of co-expressing the ACLA or related polypeptides with other polypeptides or enzymes, preferably recombinant polypeptides or enzymes, useful in carbohydrate polymer metabolism, and particularly cellulose metabolism, such as with cellulose-depolymerising, cellulose-binding, cellulosome-forming (such as, *e.g.*, scaffoldin) or other polypeptides. Such co-expression, and optionally and particularly co-secretion, provides for additive or complementary activities and functions, leading to more efficient depolymerisation and utilisation of cellulose by the micro-organisms envisaged in the context of the present invention. Particular embodiments thus provide recombinant micro-organisms, including but not limited to solventogenic micro-organisms as taught, herein co-expressing and optionally co-secreting the ACLA polypeptide or homologue thereof, or the functional fragment and/or variant of said ACLA polypeptide or of said homologue, with one or more polypeptides (also referred to herein as co-expressed polypeptides), most particularly recombinant polypeptides participating in carbohydrate polymer metabolism, and particularly cellulose metabolism, particularly with one or more enzymes capable of degrading lignocellulosic material, more particularly with one or more glycoside hydrolases, even more particularly with one or more cellulases. In more particular embodiments, the present invention provides recombinant micro-organisms expressing (a) the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, and (b) one or more additional foreign

cellulose-degrading enzymes other than the ACLA polypeptide. In particular embodiments, the enzymes are active in the micro-organism or secreted as discrete or free polypeptides. In other embodiments, the recited polypeptides are comprised in a hybrid and/or covalent cellulosome.

In particular embodiments, the catalytic action of the one or more co-expressed polypeptides, such as co-expressed enzymes, more particularly co-expressed cellulases, is additive or complementary, preferably complementary, to the enzymatic activity of the ACLA polypeptide or homologue thereof, or the functional fragment and/or variant of the ACLA polypeptide or of its homologue.

In further embodiments, the one or more co-expressed cellulases may be chosen from family-5, 6, 8, 9 and 48 cellulases. In further particular embodiments, the one or more co-expressed cellulases may be chosen from *Clostridium cellulolyticum* cellulases Cel48F, Cel9G, Cel9R, Cel9P, Cel9E, Cel9H, Cel9J, Cel9M, Cel8C, Cel5N and Cel5A, and *Saccharophagus degradans* strain 2-40 cellulases Cel5H, Cel9A, Cel9B, Cel5J, Cel5I, Cel5F, Cel5D, Cel5B, Cel9G, Cel5E, Cel5A, Cel5C and Cel6A, and functional fragments and/or variants of any one of said cellulases.

Further embodiments provide solventogenic micro-organisms as taught herein, wherein the ACLA polypeptide or homologue thereof, or the functional fragment and/or variant of said ACLA polypeptide or of said homologue, and optionally the one or more co-expressed polypeptides, more particularly one or more co-expressed cellulases as taught above, are comprised in a hybrid and/or covalent cellulosome or minicellulosome (both encompassed throughout this specification by the generic reference to cellulosome). Cellulosomes can provide for supra-molecular organisation of *inter alia* cellulose-binding and cellulose-depolymerising activities, thereby achieving greater efficiency of carbohydrate polymer metabolism, particularly cellulose metabolism.

Another aspect provides recombinant nucleic acid sequences encoding the ACLA polypeptide or homologue thereof, or encoding the functional fragment and/or variant of said ACLA polypeptide or of said homologue.

Further embodiments of this aspect of the invention provide recombinant nucleic acid sequences encoding the ACLA polypeptide or homologue thereof, or encoding the functional fragment and/or variant of said ACLA polypeptide or of said homologue, operably linked to regulatory sequences which allow for expression in a host micro-organism of interest. In particular embodiments the ACLA polypeptide is secreted by the host micro-organism. In particular embodiments, the host is a micro-organism which is capable of converting the degradation product(s) of cellulose into a product of interest. In further particular embodiments, the host is a solventogenic micro-organism. In more particular embodiments, said host is an ethanologenic micro-organism as taught elsewhere in this specification. In particular embodiments of the recombinant nucleic acid molecules provided herein, the nucleic acid sequence encoding the ACLA polypeptide or homologue thereof, or encoding the functional fragment and/or variant of said ACLA polypeptide or of said homologue, is adapted to a host micro-organism codon bias.

Further embodiments provide recombinant nucleic acid sequences encoding the ACLA polypeptide or homologue thereof, or encoding the functional fragment and/or variant of said ACLA polypeptide or of said homologue, operably linked to a secretion signal sequence which allows for secretion in a host micro-organism of interest, particularly in a solventogenic and more preferably ethanologenic micro-organism as taught elsewhere in this specification.

Further embodiments provide recombinant nucleic acid molecules encoding the ACLA polypeptide or homologue thereof, or encoding the functional fragment and/or variant of said ACLA polypeptide or of said homologue, operably linked to a secretion signal sequence which allows for secretion in a host micro-organism of interest, and operably linked to regulatory sequences which allow for expression in the host micro-organism of interest, particularly in a solventogenic and more particularly in an ethanologenic micro-organism as taught elsewhere in this specification.

A further aspect of the invention relates to vectors, such as for instance cloning vectors, shuttle vectors and/or expression vectors, comprising one or more of the recombinant nucleic acids as disclosed above.

A further aspect provides host micro-organisms of interest transformed with one or more of the recombinant nucleic acids or with a vector as described above. In an embodiment, the micro-organism may be a bacterium, more particularly a bacterium chosen from

Escherichia coli, *Salmonella tymphimurium*, *Serratia marcescens*, *Salmonella typhimurium* and *Bacillus subtilis*. Such bacteria are particularly suitable for overproduction of recombinant polypeptides, for example for purposes of purification. In particular embodiments, the host micro-organism is a solventogenic and preferably an ethanologenic micro-organism as taught elsewhere in this specification. A further aspect provides a cell lysate or cell extract directly obtained or obtainable from so-transformed micro-organisms.

Hence, also provided are methods for obtaining recombinant micro-organisms according to the invention, including but not limited to solventogenic micro-organisms. In particular embodiments methods are provided for obtaining ethanologenic micro-organisms as taught elsewhere in this specification, comprising transforming a host micro-organism with a recombinant nucleic acid or vector as taught herein.

Further provided are methods for production, and optionally and most particularly secretion, of one or more ACLA polypeptides or homologues thereof, or functional fragments and/or variants of said ACLA polypeptides or of said homologues, comprising transforming a host micro-organism with one or more of the respective recombinant nucleic acids or vectors as taught herein, and culturing or maintaining the so-transformed micro-organisms under conditions effective to cause expression of the respective polypeptides. Advantageously, the respective polypeptides are produced in a biologically active form. Optionally, the methods may comprise a step of isolating the respective polypeptides.

Hence, also contemplated is the use of the recombinant nucleic acids, vectors or host cells as disclosed herein for the production, and optionally and particularly secretion of ACLA polypeptide or homologue thereof, or functional fragment and/or variant of said ACLA polypeptide or of said homologue. Preferably, the respective polypeptides are produced in a biologically active form.

Accordingly, in a further aspect, the present invention provides the use of

(i) the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,

(ii) a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,

(iii) an expression vector comprising a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, wherein said nucleic acid sequence is operably linked to regulatory sequences which allow for expression in a host cell;

(iv) a host cell expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,

(v) a host cell transformed with a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, or

(vi) the recombinant solventogenic micro-organism according to any of claims 1 to 6,

for the production of a solvent, fuel or chemical intermediate from a substance comprising cellulose.

In particular embodiments of the above polypeptides, recombinant nucleic acids, vectors, transformed host micro-organisms, methods and uses are characterized by one or more of the following features:

- the ACLA homologue comprises a domain homologous to the DZ domain of ACLA; and/or
- the ACLA homologue is the Cel5H polypeptide of *Saccharophagus degradans* strain 1-40; and/or
- the functional fragment comprises one or more domains chosen from or corresponding to the GH5 domain, the CBM6 domain and the DZ domain of ACLA; and/or
- the one or more ACLA polypeptides or homologues thereof, or the functional fragments and/or variants of the one or more ACLA polypeptides or said homologue, is fused to one or more domains heterologous to the ACLA polypeptide or homologue thereof, chosen from a GH catalytic domain, a CBM domain, a cohesin-binding domain such as a dockerin domain, and an X module domain of a cellulosomal scaffoldin protein; and/or

- the one or more ACLA polypeptides or homologues thereof, or the functional fragment and/or variant of said one or more ACLA polypeptides or of said homologues, may be co-expressed and optionally co-secreted with one or more polypeptides, preferably recombinant polypeptides or enzymes, participating in carbohydrate polymer metabolism, particularly cellulose metabolism, preferably with one or more enzymes capable of degrading lignocellulosic material, more preferably with one or more heterologous cellulases as taught herein; and/or

- the one or more ACLA polypeptides or homologues thereof, or the functional fragment and/or variant of said ACLA polypeptides or of said homologues, and optionally the one or more co-expressed polypeptides, more particularly co-expressed cellulases, as taught above, are comprised in a hybrid and/or covalent cellulosome.

The inventors have further performed a thorough analysis of the structure and organisation of the ACLA polypeptide and have identified the presence of a new domain not previously identified as such, near the C-terminal end of the ACLA polypeptide (termed domain DZ hereafter). Without limitation, the inventor's data suggest a putative carbohydrate-binding module (and/or oligomerisation module function) for the DZ domain and indicate that its presence may be at least partly responsible for the good ability of ACLA to depolymerise crystalline cellulose. Therefore, it is a further realisation of the invention that the DZ domain or portions of the ACLA polypeptide comprising the DZ domain can be employed in a variety of enzymatic systems to improve their ability to digest and degrade crystalline cellulose.

Accordingly, a further aspect is an isolated domain (DZ domain) of the ACLA polypeptide of *Pseudomonas* sp. ND137 from amino acid 465 to amino acid 558 of said ACLA polypeptide (such as, *e.g.*, in the exemplary, mature ACLA polypeptide SEQ ID NO: 3 shown in Fig. 1C) or an isolated domain homologous thereto, or a functional fragment and/or variant of said DZ domain or of said homologous domain.

In particular embodiments, a domain homologous to the ACLA DZ is a domain stretching from amino acid 496 to amino acid 596 of the Cel5H polypeptide of *Saccharophagus degradans* 2-40 (such as, *e.g.*, in the exemplary, mature Cel5H polypeptide SEQ ID NO: 5 shown in Fig. 1E).

Particular embodiments provide for an isolated fragment of the ACLA polypeptide or of a homologue or variant thereof, having the structure PSL-DZ, CBM6-PSL-DZ or DPR-CBM6-

PSL-DZ, wherein DPR, CBM6 and PSL have meanings as explained elsewhere in this specification. Such fragments comprise the DZ domain which provides for its advantageous functions, and optionally, such fragment may include the ACLA carbohydrate-binding module CBM6 which may further stimulate the ability to degrade crystalline cellulose or hemicellulosic substrates. Moreover, endogenous linker sequences can provide advantageous means for fusing the DZ and optionally CBM6 domains to other polypeptides while retaining their structure and function.

Hence, particular embodiments provide for an isolated fragment of the Cel5H polypeptide comprising a domain homologous to the DZ domain of ACLA.

Further contemplated are uses of the isolated DZ domain or the functional fragment and/or variant thereof, or of the isolated fragment as taught above, as a carbohydrate-binding module. Such use can endow novel carbohydrate- or cellulose-binding properties on various enzyme systems.

Another aspect relates to chimeric polypeptides comprising an isolated DZ domain most particularly of ACLA or a functional fragment and/or variant thereof, or comprising an isolated fragment of ACLA comprising the DZ domain, and further comprising one or more domains heterologous to the ACLA polypeptide or homologue thereof, chosen from a GH catalytic domain, a CBM domain, a cohesin-binding domain such as a dockerin domain, and an X module domain of a cellulosomal scaffoldin protein. Such modular combinations allow the generation of novel cellulose-degrading enzymes with useful activity towards crystalline cellulose or hemicellulosic substrates.

Other aspects provide chimeric polypeptides comprising the isolated DZ domain or the functional fragment and/or variant thereof, or an isolated fragment of ACLA comprising the DZ domain, fused to one or more polypeptides participating in carbohydrate polymer metabolism, particularly cellulose metabolism, preferably to one or more enzymes capable of degrading lignocellulosic material, more preferably to one or more glycoside hydrolases, even more preferably to one or more cellulases, which is (are) heterologous to the ACLA polypeptide or homologue thereof. Nevertheless, fusions of additional DZ domain(s) with a ACLA polypeptide or a homologue thereof, or fragments and/or variants thereof as described herein, are also contemplated.

A further aspect provides recombinant nucleic acid molecules encoding the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant

thereof, or encoding an isolated fragment of ACLA comprising the DZ domain or homologue thereof, or encoding the chimeric polypeptide as described above.

A particular embodiment of this aspect provides a recombinant nucleic acid molecule encoding the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or encoding the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or encoding the chimeric polypeptide as described above, operably linked to regulatory sequences which allow for expression in a host micro-organism of interest, preferably in a solventogenic and more preferably in an ethanologenic micro-organism as taught elsewhere in this specification.

A further embodiment provides a recombinant nucleic acid molecule encoding the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or encoding the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or encoding the chimeric polypeptide as described above, operably linked to a secretion signal sequence which allows for secretion in a host micro-organism of interest, preferably in a solventogenic and more preferably in an ethanologenic micro-organism as taught elsewhere in this specification.

Yet a further embodiment of this aspect provides a recombinant nucleic acid molecule encoding the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or encoding the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or encoding the chimeric polypeptide as described above, operably linked to a secretion signal sequence which allows for secretion in a host micro-organism of interest, and operably linked to regulatory sequences which allow for expression in the host micro-organism of interest, preferably in a solventogenic and more preferably in an ethanologenic micro-organism as taught elsewhere in this specification.

A further aspect provides vectors, such as for instance a cloning vector, a shuttle vector and/or an expression vector, comprising any of the above recombinant nucleic acid molecules.

A further aspect provides a host micro-organism of interest transformed with the above described recombinant nucleic acid molecule or vector. In particular embodiments, the micro-organism is a bacterium, more preferably a bacterium chosen from *Escherichia coli*, *Salmonella tymphimurium*, *Serratia marcescens*, *Salmonella typhimurium* and *Bacillus*

subtilis. In particular embodiments, the micro-organism may be a solventogenic and is preferably an ethanologenic micro-organism as taught elsewhere in this specification. A further aspect provides a cell lysate or cell extract directly obtained or obtainable from so-transformed micro-organisms.

In particular embodiments, the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or the chimeric polypeptide as described above, and optionally the one or more co-expressed polypeptides involved in carbohydrate polymer metabolism, particularly cellulose metabolism, may be comprised in a hybrid and/or covalent cellulosome.

In a further aspect, the invention provides methods for the degradation of a substance comprising cellulose, such as lignocellulosic or cellulosic material or biomass, comprising contacting said substance with the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or with the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or with the chimeric polypeptide as described above, or with a recombinant micro-organism expressing such. In an embodiment, the substance may comprise or be enriched for crystalline cellulose, semi-crystalline cellulose or amorphous cellulose.

A related aspect provides methods for producing a solvent from a substance comprising cellulose, such as lignocellulosic or cellulosic material or biomass, comprising treating said substance with the isolated DZ domain or the isolated domain homologous thereto or the functional fragment and/or variant thereof, or with the isolated fragment of ACLA comprising the DZ domain or a homologue thereof, or with the chimeric polypeptide as described above, or with a recombinant solventogenic micro-organism expressing such. In an embodiment, the substance may comprise or be enriched for crystalline cellulose. Preferably, the solvent can be ethanol and the micro-organism may be an ethanologenic micro-organism.

A further aspect provides compositions comprising cellulose degradation products and/or solvents, more particularly ethanol, obtained or obtainable by the methods described herein. Such compositions may be crude cultivation media or (at least) partially enriched and/or purified for the product of the ACLA polypeptide (or homologue or fragment thereof) or partially enriched and/or purified for the solvent. In particular embodiments, such

compositions are characterized by the increased glucose, cellobiose and cellotriose content. In further particular embodiments compositions are provided wherein the content of glucose, cellobiose and/or cellotriose is greater than 10 wt%, greater than 20 wt%, greater than 30 wt%, greater than 40 wt%, greater than 50 wt%, greater than 60 wt%, greater than 70 wt%, greater than 80 wt%, greater than 90 wt% or more.

In further particular embodiments, such compositions are characterized by the high solvent content. In further particular embodiments compositions are provided wherein the content of solvent is greater than 10 wt%, greater than 20 wt%, greater than 30 wt%, greater than 40 wt%, greater than 50 wt%, greater than 60 wt%, greater than 70 wt%, greater than 80 wt%, greater than 90 wt%, greater than 90 wt%, or between 96% and 99,9 wt%. In particular embodiments the solvent concentration of the composition is greater than about 10 g/L, and preferably greater than about 15 g/L.

These and other aspects and embodiments of the invention are further explained here below and in the appended claims, as well as illustrated by non-limiting examples.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates exemplary amino acid sequence of native ACLA protein of *Pseudomonas* sp. ND137 containing **(A)** or not containing **(B)** the N-terminal secretion signal sequence; exemplary coding nucleic acid sequence **(C)** and amino acid sequence **(D-E)** of native Cel5H polypeptide of *Saccharophagus degradans* strain 2-40 containing **(D)** or not containing **(E)** the N-terminal secretion signal sequence; and exemplary coding nucleic acid sequence **(F)** of native ACLA protein of *Pseudomonas* sp. ND137.

Figure 2 schematically illustrates comparison of domains of ACLA and Cel5H.

Figure 3 compares the general organisation of ACLA with that of Cel5H. ACLA has the following organisation (from N-terminal to C-terminal ends of the enzyme): signal sequence/catalytic module (GH-family 5)/aspartic acid- and proline linker/carbohydrate binding module belonging to CBM-family 6/serine-rich-linker, in other denotation signal sequence-GH5-DPR-CBM6-PSL.

Figure 4 compares the activities of ACLA and Cel5H on Phosphoric Acid Swollen Cellulose at 37°.

Figure 5 compares the activities of ACLA and Cel5H on Avicel PH101 at 37°.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps.

The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

The term "about" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, is meant to encompass variations of +/-10% or less, preferably +/-5% or less, more preferably +/-1% or less, and still more preferably +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" refers is itself also specifically, and preferably, disclosed.

Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. By means of further guidance, term definitions are included to better appreciate the teaching of the present invention.

As used herein, the terms “ACLA”, “ACLA polypeptide” or “ACLA protein” refer to the polypeptide or protein of *Pseudomonas sp.* Strain ND 137. The above designations particularly refer to such polypeptides with a native sequence, *i.e.*, polypeptides of which the primary sequence is the same as that of the ACLA protein of *Pseudomonas sp.* Strain ND 137 as found in or derived from nature. A skilled person understands that native sequences of ACLA might differ between different sub-strains or sub-cultures of *Pseudomonas sp.* Strain ND 137 due to genetic divergence or spontaneous mutation(s). The native sequences of ACLA may further diverge due to post-transcriptional or post-

translational modifications. Accordingly, all ACLA sequences found in or derived from nature are considered “native”.

The above designations encompass the ACLA polypeptide as present in a living organism, micro-organism or cells, as well as at least partly isolated from such sources. The terms also encompass ACLA polypeptides when produced by recombinant or synthetic means.

The ACLA polypeptide is normally secretory and may exist in a precursor form comprising a cleavable N-terminal secretion signal sequence, or in a mature form lacking said signal sequence. Depending on context readily understood by a skilled person, a reference herein to the ACLA polypeptide may encompass said mature form and/or said precursor forms.

Exemplary ACLA polypeptides include without limitation the sequence as annotated under the Uniprot/Swissprot accession number Q8VUT3 (sequence version 1 entered on March 1, 2002) also reproduced in Fig. 1A (SEQ ID NO: 1). This sequence corresponds to the ACLA precursor comprising the signal sequence. An exemplary mature ACLA polypeptide, *i.e.*, with the signal sequence processed away, is represented by amino acid positions 28 to 585 of the respective precursor, as exemplarily reproduced in Fig. 1B (SEQ ID NO: 2).

Depending on context readily understood by a skilled person, a reference herein to the ACLA polypeptide may encompass said mature form and/or said precursor forms. Further, unless indicated otherwise expressly or by context, whenever domains or other motifs of ACLA are identified herein by reference to amino acid positions, such reference is to the mature form of ACLA, such as in particular that shown in SEQ ID NO: 5.

Further, unless indicated otherwise expressly or by context, whenever domains or other motifs of ACLA are identified herein by reference to amino acid positions, such reference is to the mature form of ACLA, such as in particular that shown in SEQ ID NO: 2.

As used herein, the term “homology” denotes structural similarity between two macromolecules, particularly between two polypeptides or polynucleotides, from same or different taxons, wherein said similarity is due to shared ancestry. Hence, the term “homologues” denotes so-related macromolecules having said structural similarity. Preferably, homologues of the ACLA polypeptide as intended herein encompass said homologues with native sequence.

A polypeptide homologue of the ACLA polypeptide may preferably show at least about 30%, more preferably at least 40%, even more preferably at least 50%, still more preferably

at least 60%, yet more preferably at least 70%, such as very preferably at least 80% or at least 90% or at least 95% sequence identity as defined elsewhere in this specification to the ACLA polypeptide. Alternatively or preferably in addition, a polypeptide homologue of the ACLA polypeptide may show at least about 50%, more preferably at least 60%, even more preferably at least 70%, still more preferably at least 80%, yet more preferably at least 90%, such as very preferably at least 95% sequence similarity as defined elsewhere in this specification to the ACLA polypeptide. Further preferably, a polypeptide homologue of the ACLA polypeptide may comprise one or more functional domains corresponding to domains of the ACLA polypeptide, and preferably one or more of GH family 5 catalytic domain, CBM family 6 domain and DZ domain. Preferably, said domains may have organisation analogous to ACLA, in particular GH5-CBM6-DZ, usually with interposing linkers.

As noted, preferred homologues of the ACLA polypeptide are the Cel5H polypeptides of *Saccharophagus degradans* 2-40. Exemplary Cel5H polypeptides include without limitation the polypeptide encoded by the nucleic acid coding sequence as annotated under the NCBI Entrez Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene) locus tag Sde_3237 (GeneID: 3965710), and under the NCBI Genbank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) accession number NC 007912 (sequence version 1, database entry updated 20July 2008), also reproduced in Fig. 1C (SEQ ID NO: 3); and the Cel5H polypeptide as annotated under the NCBI Genbank accession number YP 528706 (sequence version 1, database entry update 20July 2008) and the Uniprot/Swissprot (http://www.expasy.org/) accession number Q21FN5 (sequence version 1 entered on April 18, 2006), as exemplarily reproduced in Fig. 1D (SEQ ID NO: 4).

This sequence corresponds to a Cel5H precursor comprising the signal sequence. An exemplary mature Cel5H polypeptide, *i.e.*, with the signal sequence processed away, is represented by amino acid positions 35 to 630 of said precursor sequence, as exemplarily reproduced in Fig. 1E (SEQ ID NO: 5).

One shall appreciate that the actual cleavage by signal peptidase may potentially occur at a peptide bond nearby the predicted cleavage site between amino acids 34 and 35 of the precursor Cel5H such as shown in Fig. 1D. For example, the actual cleavage might occur at any of the peptide bonds within the amino acid region from position 30 to 40, preferably 31 to 39, more preferably 32 to 38, even more preferably 32 to 37, still more preferably 33 to 36, of the precursor Cel5H such as shown in Fig. 1B.

The term “fragment” with reference to a given polypeptide such as ACLA or a homologue thereof, or with reference to a given domain such as the DZ domain of the ACLA or homologue thereof, generally refers to a truncated form of said polypeptide that has an N-terminal and/or C-terminal deletion of one or more amino acid residues as compared to said polypeptide such as said ACLA or homologue or said domain, but where the remaining primary sequence of the fragment is identical to the corresponding positions in the amino acid sequence of said polypeptide such as said ACLA or homologue or said domain.

For example, a fragment of a given polypeptide such as ACLA or a homologue thereof, may include a sequence of about ≥5 consecutive amino acids, preferably ≥10 consecutive amino acids, more preferably ≥20 consecutive amino acids, even more preferably ≥30 consecutive amino acids, *e.g.*, ≥40 consecutive amino acids, and most preferably ≥50 consecutive amino acids, *e.g.*, ≥60, ≥70, ≥80, ≥90, ≥100, ≥200 or ≥500 consecutive amino acids of said polypeptide such as said ACLA or homologue.

For example, a fragment of the DZ domain of ACLA or a homologue thereof, may include a sequence of about ≥5 consecutive amino acids, preferably ≥10 consecutive amino acids, more preferably ≥20 consecutive amino acids, even more preferably ≥30 consecutive amino acids, *e.g.*, ≥40 consecutive amino acids, and most preferably ≥50 consecutive amino acids, *e.g.*, ≥60, ≥70, ≥80, or ≥90, consecutive amino acids of said DZ domain or homologue thereof.

Further, a fragment of a given polypeptide such as ACLA or a homologue thereof or DZ domain may represent at least about 30%, *e.g.*, at least 50% or at least 70%, preferably at least 80%, *e.g.*, at least 85%, more preferably at least 90%, and yet more preferably at least 95% or even about 99% of the amino acid sequence of said polypeptide such as said ACLA or homologue or DZ domain.

The term “variant” of a given polypeptide such as ACLA or a homologue thereof or of a DZ domain refers to polypeptides the amino acid sequence of which is substantially identical (*i.e.*, largely but not wholly identical) to a native sequence of said polypeptide such as said ACLA or a homologue thereof or a DZ domain. “Substantially identical” refers to at least about 85% identical, *e.g.*, preferably at least 90% identical, *e.g.*, at least 91% identical, 92% identical, more preferably at least 93% identical, *e.g.*, 94% identical, even more preferably

at least 95% identical, *e.g.*, at least 96% identical, yet more preferably at least 97% identical, *e.g.*, at least 98% identical, and most preferably at least 99% identical.

Sequence identity between two polypeptides can be determined by optimally aligning (optimal alignment of two protein sequences is the alignment that maximises the sum of pair-scores less any penalty for introduced gaps; and may be preferably conducted by computerised implementations of algorithms, such as “Gap”, using the algorithm of Needleman and Wunsch 1970 (J Mol Biol 48: 443-453), or “Bestfit”, using the algorithm of Smith and Waterman 1981 (J Mol Biol 147: 195—197), as available in, e.g., the GCG™ v. 11.1.2 package from Accelrys) the amino acid sequences of the polypeptides and scoring, on one hand, the number of positions in the alignment at which the polypeptides contain the same amino acid residue and, on the other hand, the number of positions in the alignment at which the two polypeptides differ in their sequence. The two polypeptides differ in their sequence at a given position in the alignment when the polypeptides contain different amino acid residues at that position (amino acid substitution), or when one of the polypeptides contains an amino acid residue at that position while the other one does not or vice versa (amino acid insertion or deletion). Sequence identity is calculated as the proportion (percentage) of positions in the alignment at which the polypeptides contain the same amino acid residue versus the total number of positions in the alignment. Further suitable algorithms for performing sequence alignments and determination of sequence identity include those based on the Basic Local Alignment Search Tool (BLAST) originally described by Altschul et al. 1990 (J Mol Biol 215: 403-10), such as the "Blast 2 sequences" algorithm described by Tatusova and Madden 1999 (FEMS Microbiol Lett 174: 247-250), such as using defaults settings thereof.

The variant of a given polypeptide such as ACLA or a homologue thereof or a DZ domain, as used herein, also specifically encompasses polypeptides having a certain degree of similarity to said polypeptide such as said ACLA or a homologue thereof or a DZ domain. Preferably, such variants can be at least about 90% similar, *e.g.*, preferably at least 91% similar, *e.g.*, at least 92% similar, 93% similar, more preferably at least 94% similar, *e.g.*, 95% similar, even more preferably at least 96% similar, *e.g.*, at least 97% similar, yet more preferably at least 98% similar, *e.g.*, at least 99% similar.

Sequence similarity between two polypeptides can be determined by optimally aligning (see above) the amino acid sequences of the polypeptides and scoring, on one hand, the

number of positions in the alignment at which the polypeptides contain the same or similar (*i.e.*, conservatively substituted) amino acid residue and, on the other hand, the number of positions in the alignment at which the two polypeptides otherwise differ in their sequence. The two polypeptides otherwise differ in their sequence at a given position in the alignment when the polypeptides contain non-conservative amino acid residues at that position, or when one of the polypeptides contains an amino acid residue at that position while the other one does not or vice versa (amino acid insertion or deletion). Sequence similarity is calculated as the proportion (percentage) of positions in the alignment at which the polypeptides contain the same or similar amino acid residue versus the total number of positions in the alignment.

It shall be understood that under reference to fragments and/or variants the specification also broadly contemplates fragments of variants of a given polypeptide such as ACLA or a homologue thereof or a DZ domain, as well as variants of fragments of such polypeptide.

The term “functional” with reference to the fragments and/or variants of a given polypeptide such as ACLA or a homologue thereof or a DZ domain denotes that such fragments and variants at least partly retain the biological activity or functionality of the corresponding polypeptide such as said ACLA or homologue thereof or domain such as DZ domain. Preferably, such functional fragments and/or variants may retain at least about 20%, *e.g.*, at least 30%, or at least 40%, or at least 50%, *e.g.*, at least 60%, more preferably at least 70%, *e.g.*, at least 80%, yet more preferably at least 85%, still more preferably at least 90%, and most preferably at least 95% or even 100% of the activity of the corresponding polypeptide such as said ACLA or homologue thereof or DZ domain. Potentially, such functional fragments and/or variants may even have higher activity than the corresponding polypeptide such as said ACLA or homologue or DZ domain.

Functional fragments and/or variants of a given polypeptide such as ACLA or homologue thereof or DZ domain may be functionally equivalent to said polypeptide in at least one and preferably more or all aspects of its biological activity or function. Relevant aspects of biological function of said ACLA or homologue thereof include in particular but without limitation cellulose-depolymerising or cellulase activity, activity profile with respect to different cellulose substrates (*e.g.*, crystalline, semi-crystalline or amorphous cellulose or hemicellulose), ability or preference to degrade crystalline cellulose substrates, and ability to interact with or bind to cellulose substrates. Relevant aspects of biological function of

said DZ domain of ACLA or homologue thereof include in particular but without limitation cellulose-binding and particularly crystalline cellulose-binding, and/or oligomerisation activities. The biological activity of a given fragment and/or variant of ACLA or of homologue thereof or of DZ domain, including the above aspects of such activity, can be determined by standard tests, such as for example the enzyme activity tests presented in the examples section.

The terms “solventogenic” or “solvent-producing” have their art-established meaning and in particular denote the ability of micro-organisms such as bacteria (*i.e.*, solventogenic bacteria) to produce one or more non-gaseous organic liquids or solvents, such as *inter alia* ethanol, acetone, butanol, propionic acid, butyric acid, ether or glycerine, from a carbohydrate source such as for example hexoses, pentoses or oligosaccharides. In particular, the term encompasses naturally occurring solventogenic organisms, solventogenic organisms with naturally occurring or induced mutations, and solventogenic organisms which have been genetically modified. Accordingly, in the context of the present invention, the term “non-solventogenic” refers to micro-organisms such as bacteria which are not capable of producing solvents from a carbohydrate source. Examples of micro-organisms which are naturally non-solventogenic micro-organisms include most micro-organisms used for recombinant production such as, but not limited to, *E. coli, Pichia pastoris, Saccharomyces cerevisiae*, etc.

The terms “ethanologenic” or “ethanol-producing” is intended to denote the ability of micro-organisms such as bacteria (*i.e.*, ethanologenic bacteria) to produce at least or preferably mainly ethanol from a carbohydrate source such as for example hexose, pentose or oligosaccharides, more preferably to produce ethanol from carbohydrates as the most abundant non-gaseous fermentation product. In particular, the term encompasses naturally occurring ethanologenic organisms, ethanologenic organisms with naturally occurring or induced mutations, and ethanologenic organisms which have been genetically modified. Examples of suitable solventogenic micro-organism include but are not limited to strains of Clostridium and Zymomonas, such as, but not limited to those exemplified herein.

The term “isolating” with reference to a particular component generally denotes separating that component from at least one other component of a composition from which the former component is thereby “isolated”. In particular, the terms “isolating” and “isolated” may refer herein to increasing amount of a desired protein(s) or polypeptide(s) in a sample. The

relative amount may be expressed as the ratio between the concentration or amount of the desired protein(s) or polypeptide(s) to be isolated and the concentration or amount of total proteins in the sample. In addition, the term may encompass separating desired protein(s) or polypeptide(s) from non-protein cellular components (*e.g.*, cell walls, lipids, nucleic acids, *etc.*).

The terms “recombinant nucleic acid” or “recombinant nucleic acid molecule” as used herein generally refer to nucleic acid molecules (such as, *e.g.*, DNA, cDNA or RNA molecules) comprising segments generated and/or joined together using recombinant DNA technology, such as for example molecular cloning and nucleic acid amplification. Usually, a recombinant nucleic acid molecule may comprise one or more non-naturally occurring sequences, and/or may comprise segments corresponding to naturally occurring sequences that are not positioned as they would be positioned in a source genome which has not been modified. When a recombinant nucleic acid molecule replicates in the host organism into which it has been introduced, the progeny nucleic acid molecule(s) are also encompassed within the term “recombinant nucleic acid molecule”.

In particular embodiments, a recombinant nucleic acid molecule can be stably integrated into the genome of a host organism, such as for example integrated at one or more random positions or integrated in a targeted manner, such as, *e.g.*, by means of homologous recombination, or the recombinant nucleic acid molecule can be present as or comprised within an extra-chromosomal element, wherein the latter may be auto-replicating.

Standard reference works setting forth the general principles of recombinant DNA technology include Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook *el al.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Current Protocols in Molecular Biology, ed. Ausubel *et al.*, Greene Publishing and Wiley-Interscience, New York, 1992 (with periodic updates); Innis *et al.*, PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990. General principles of microbiology are set forth, for example, in Davis, B. D. et al., Microbiology, 3d edition, Harper & Row, publishers, Philadelphia, Pa. (1980).

The term “recombinant polypeptide” as used herein refers to a polypeptide or protein produced by a host organism through the expression of a recombinant nucleic acid molecule, which has been introduced into said host organism or an ancestor thereof, and which comprises a sequence encoding said polypeptide or protein.

Further, the terms “recombinant” or “transformed” as used herein with reference to host organisms, micro-organisms or cells, encompass such host organisms, micro-organisms or cells into which a recombinant nucleic acid molecule has been introduced, as well as the recombinant progeny of such host organisms, micro-organism or cells.

Hence, the term “transformation” encompasses the introduction or transfer of a foreign nucleic acid such as a recombinant nucleic acid into a host organism, micro-organism or cell. The so-introduced nucleic acid may be preferably maintained throughout the further growth and cell division of said host organism, micro-organism or cell. Any conventional gene transfer methods may be used to achieve transformation, such as without limitation electroporation, electropermeation, chemical transformation, lipofection, virus- or bacteriophage-mediated transfection, *etc*.

By “encoding” is meant that a nucleic acid sequence or part(s) thereof corresponds, by virtue of the genetic code of an organism in question to a particular amino acid sequence, *e.g.*, the amino acid sequence of a desired polypeptide or protein. By means of example, nucleic acids “encoding” a particular polypeptide or protein may encompass genomic, hnRNA, pre-mRNA, mRNA, cDNA, recombinant or synthetic nucleic acids.

Preferably, a nucleic acid encoding a particular polypeptide or protein may comprise an open reading frame (ORF) encoding said polypeptide or protein. An “open reading frame” or “ORF” refers to a succession of coding nucleotide triplets (codons) starting with a translation initiation codon and closing with a translation termination codon known *per se*, and not containing any internal in-frame translation termination codon, and potentially capable of encoding a polypeptide. Hence, the term may be synonymous with “coding sequence” as used in the art.

In an embodiment, the nucleic acid sequence or ORF encoding the present polypeptide(s) may be codon optimised as known *per se* for expression in a particular organism, *e.g.*, micro-organism, more particularly a bacterium of interest. Codon usage bias and codon frequencies from various organisms are available, for example via the Codon Usage Database (http://www.kazusa.or.jp/codon/) described by Nakamura et al. 2000 (Nucl Acids Res 28: 292).

Expression of polypeptides of interest in recombinant micro-organisms as taught herein can be achieved through operably linking the nucleic acid sequences or ORF(s) which encode said polypeptides with regulatory sequences allowing for expression in said micro-

organisms. In this regard, the term “expressing” a polypeptide when referring to a recombinant organism encompasses recombinant organisms capable of expressing the polypeptide, e.g. under suitable culture conditions or upon addition of inducers (e.g. where inducible regulatory sequences are used).

An “operable linkage” is a linkage in which regulatory nucleic acid sequences and sequences sought to be expressed are connected in such a way as to permit said expression. For example, sequences, such as, *e.g.*, a promoter and an ORF, may be said to be operably linked if the nature of the linkage between said sequences does not: (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter to direct the transcription of the ORF, (3) interfere with the ability of the ORF to be transcribed from the promoter sequence.

The precise nature of regulatory sequences or elements required for expression may vary between organisms, but typically include a promoter and a transcription terminator, and optionally an enhancer.

Reference to a “promoter” or “enhancer” is to be taken in its broadest context and includes transcriptional regulatory sequences required for accurate transcription initiation and where applicable accurate spatial and/or temporal control of gene expression or its response to, *e.g.*, internal or external (*e.g.*, exogenous) stimuli. More particularly, “promoter” may depict a region on a nucleic acid molecule, preferably DNA molecule, to which an RNA polymerase binds and initiates transcription. A promoter is preferably, but not necessarily, positioned upstream, *i.e.*, 5’, of the sequence the transcription of which it controls. Typically, in prokaryotes a promoter region may contain both the promoter *per se* and sequences which, when transcribed into RNA, will signal the initiation of protein synthesis (*e.g.*, Shine-Dalgarno sequence).

In embodiments, promoters contemplated herein may be constitutive or inducible.

By means of example, constitutive promoters suitable for expression in Zymomonas species, such as *Zymomonas mobilis* include without limitation the promoter of the pyruvate decarboxylase gene of *Z. mobilis* (P_{pdc} promoter) and the artificial P_{tac} promoter that is derived from *trp* and *lac* promoters.

By means of example, constitutive promoters suitable for expression in Clostridium species such as *Clostridium acetobutylicum* include without limitation the promoter of the thiolase

gene of *C. acetobutylicum* (P_{thl} promoter), the promoter of the acetoacetate decarboxylase gene (P_{adc} promoter), the promoter of the phosphotransbutyrylase gene (P_{ptb} promoter), the xylose-inducible promoter of the *Staphylococcus xylosus* xylose operon (P_{xyl} promoter)

Hence, in particular embodiments a promoter is provided which is a promoter controlled by an operator. As used herein, the term "operator" refers to a nucleotide sequence, preferably DNA sequence, which controls the initiation and/or maintenance of transcription of a sequence from a promoter. Typically, an operator may be generally placed between a promoter and a downstream sequence the transcription of which the promoters controls. Usually, an operator is capable of binding a repressor polypeptide, whereby it reduces the transcription from the said promoter. A useful repressor can alternate between a state in which it binds the operator and a state in which it does not and such alternation may be advantageously controlled by an external condition, *e.g.*, an external substance or a particular metabolite. Accordingly, in host cells comprising a compatible repressor, the inclusion of an operator in the nucleic acid of the invention may allow to control the activity of the promoter and expression there from. Operator sequences may be generally derived from bacterial chromosomes.

It shall also be appreciated that the nucleic acids of the invention may encode one or more than one polypeptides of interest. Moreover, where expression of two or more polypeptides is intended, said expression may be from independent expression units, or may be from a single expression unit (*i.e.*, multi-cistronic expression) comprising two or more sequential ORFs controlled by common regulatory elements. By means of example, an expression unit encoding two or more polypeptides may be generated by associating translation initiation codons of the respective ORFs with sequences controlling translation initiation (*e.g.*, ribosome entry sequences).

The terms "terminator" or "transcription terminator" refer generally to a sequence element at the end of a transcriptional unit which signals termination of transcription. For example, a terminator is usually positioned downstream of, *i.e.*, 3' of ORF(s) encoding a polypeptide of interest. For instance, where a recombinant nucleic acid contains two or more ORFs, *e.g.*, successively ordered and forming together a multi-cistronic transcription unit, a transcription terminator may be advantageously positioned 3' to the most downstream ORF.

In a preferred embodiment, regulatory sequences controlling the expression of polypeptide(s) as taught herein may be advantageously provided on the recombinant

nucleic acid to be used for transforming the present micro-organisms. However, also contemplated are embodiments in which the recombinant nucleic acid provides for coding sequence(s) lacking one or more regulatory sequences, whereas the required regulatory sequences are provided by the transformed micro-organism (such as, *e.g.*, wherein the recombinant nucleic acid inserts into a chromosomal or episomal region comprising suitable regulatory sequences).

Clostridium and *Clostridium acetobutylicum* as used herein refer to bacterial taxons known as such in the art, and particularly encompass bacteria, bacterial species, strains, subspecies and cultures belonging to said taxons, as well as modified or engineered, such as for example mutated or genetically engineered, derivatives of naturally-occurring or wild-type specimens. By means of guidance and not limitation, isolated strains of *Clostridium* available from the American Type Culture Collection (ATCC) can be ordered *inter alia* under accession numbers ATCC 824, ATCC 4259, ATCC 39236 or ATCC 43084 for *Clostridium acetobutylicum.* Isolated strains of *Clostridium beijerinckii* can, for example be ordered inter alia under accession numbers ATCC 858 and ATCC 25752.

Zymomonas and *Zymomonas mobilis* as used herein refer to bacterial taxons known as such in the art, and particularly encompass bacteria, bacterial species, strains, subspecies and cultures belonging to said taxons, as well as modified or engineered, such as for example mutated or genetically engineered, derivatives of naturally-occurring or wild-type specimens. By means of guidance and not limitation, isolated strains of *Zymomonas mobilis* available from the American Type Culture Collection (ATCC) can be ordered *inter alia* under accession numbers ATCC 29191, ATCC 10988, ATCC 39676 or ATCC 31821.

As noted, the polypeptides expressed by the present solventogenic micro-organisms may be targeted for secretion. To achieve secretion, a nucleic acid sequence encoding a polypeptide may be operably linked to a sequence encoding a secretion signal sequence. In this connection, “operably linked” denotes that the sequence encoding the signal sequence and the sequence encoding the polypeptide to be secreted are connected in frame or in phase, such that upon expression the signal peptide facilitates the secretion of the polypeptide so-linked thereto.

It shall be appreciated that suitable signal sequences may depend on the type of micro-organism in which secretion is desired. For example, distinct signal sequences may be required in Gram-positive bacteria vs. Gram-negative bacteria.

By means of example and not limitation, secretion in Gram-positive bacteria, and in particular in *Clostridium* such as *C. acetobutylicum*, may be achieved using the signal sequence of the Cel5A precursor polypeptide of *C. cellulolyticum* (exemplary sequence: Genbank acc. no. AAA51444, seq version 1 revised on October 31, 1994), or of the CipC precursor scaffolding protein of *C. cellulolyticum* (exemplary sequence: Genbank acc. no. AAC28899, seq. version 2 revised on December 5, 2005), or of the CipA precursor scaffolding protein of *C. acetobutylicum* (exemplary sequence: Genbank acc. no. AAK78886, seq. version 1 revised on January 19, 2006).

Further without limitation, secretion in Gram-negative bacteria, and in particular in *Zymomonas* such as *Z. mobilis*, may be achieved via the *Sec* pathway, *e.g.*, using the signal sequence of the gluconolactonase precursor polypeptide of Z. mobilis (exemplary sequence: Genbank acc. no. CAA47637, seq. version 1 revised on October 19, 2006) or of the carbohydrate selective porine OprB (exemplary sequence: Genbank acc. no. AAV88688, seq. version 1 revised on January 25, 2005), or via the twin-arginine translocation (Tat) pathway, *e.g.*, using the signal sequence of the gluco fructose oxidoreductase precursor polypeptide of Z. mobilis (exemplary sequence: Genbank acc. no. CAB02496, seq. version 1 revised on October 19, 2006).

It shall also be appreciated that native (or homologous, endogenous) signal peptides of polypeptides to be expressed by the micro-organisms as taught herein may be employed; insofar they are functional in said micro-organisms. Hence, by means of example, secretion of ACLA or related polypeptides may be achieved using the endogenous or homologous secretion signal sequence of the ACLA precursor polypeptide. In particular, since *Pseudomonas sp.* Strain ND 137 is a Gram-negative bacterium, the endogenous or homologous signal sequence of ACLA may be particularly suited for secretion thereof in other Gram-negative bacteria, such as for example in *Zymomonas*, more particularly *Zymomonas mobilis*. In other embodiments, polypeptides such as ACLA and related polypeptides may be secreted using heterologous signal sequences.

The domain structure of the native ACLA polypeptide can be outlined as GH5-DPR-CBM6-PSL-DZ, with schematic indication of the domain boundaries in mature ACLA shown in Fig. 2. Functional fragments of ACLA or homologues thereof preferably contain one or more of the GH5, CBM6 and DZ domains, or domains corresponding thereto (*e.g.*, analogous domains within a homologue and/or variant of ACLA). Preferably, domains comprised in

such fragments may be functional as intended herein. Hence, in particular embodiments the invention contemplates fragments such as those corresponding to of the following domain organisations: GH5, GH5-DPR, GH5-DPR-CBM6, GH5-DPR-CBM6-PSL, DPR-CBM6, DPR-CBM6-PSL, DPR-CBM6-PSL-DZ, CBM6, CBM6-PSL, CBM6-PSL-DZ, and PSL-DZ, DZ. The invention further envisages fragments corresponding to combinations of the domains of mature ACLA, such as, but not limited to PSL-DZ-PSL-DZ.

As noted, the invention also relates to various fusions of ACLA and related polypeptides with useful heterologous domains. In this context, the term "heterologous" denotes that said domain(s) are derived from polypeptides other than the ACLA polypeptide or the homologue thereof. Accordingly, "heterologous" when referring to combinations of domains refers to the fact that the different domains do not originate from (or do not naturally occur in) the same protein. Nevertheless, fusions of additional ACLA-derived domain(s) with ACLA polypeptide or homologue thereof, or fragments and/or variants thereof, can also be contemplated.

The term "fusion" is broadly contemplated herein as encompassing *inter alia* genetic fusions (*i.e.*, fusing distinct parts into a single open reading frame) as well as fusions induced by chemical and/or physical means (*e.g.*, chemical cross-linking or photo-coupling). Furthermore, fusion includes direct fusion or fusion via a linker. Suitable linkers may include without limitation peptide sequences of at least 3, preferably at least 4 or 5, more preferably at least 7, and yet more preferably at least 12 amino acids, such as between 3 and 15 amino acids, and preferably comprising, being enriched for or consisting of non-charged amino acids, preferably chosen from glycine, serine, cysteine, asparagine, tyrosine, glutamine, alanine, valine, proline, threonine, and more preferably glycine or serine.

Exemplary glycoside hydrolase (GH) catalytic domains may be derived from glycoside hydrolases (glycosidases) known *per se* or future ones identified and isolated *inter alia* using established procedures. Exemplary glycosidases include those grouped under the EC 3.2.1. "Glycosidases" classification of NC-IUBMB, more preferably glycosidases capable of depolymerising cellulosic and lingo-cellulosic substrates, such as, *e.g.*, cellulases (EC 3.2.1.4.), cellulose 1,4-β-cellobiosidases (EC 3.2.1.91) and processive endocellulases (EC 3.2.1.4./EC 3.2.1.91.).

The term carbohydrate-binding module(s) (CBM) is known in the art and commonly broadly covers all non-catalytic sugar- or carbohydrate-binding modules found in or derived from

glycoside hydrolases. CBM modules preferred herein may include those binding specifically to cellulose or cellulosic substrates and alternatively referred to as cellulose-binding domains (CBD). CMB and their structural and functional properties are reviewed *inter alia* in Tomme *et al*. 1995 ("Cellulose-binding domains: classification and properties", in Enzymatic Degradation of Insoluble Polysaccharides, Saddler JN & Penner M, eds., pp. 142-163, American Chemical Society, Washington) and Boraston *et al*. 2004 ("Carbohydrate-binding modules: fine tuning polysaccharide recognition". Biochem J 382: 769-81).

Cohesin-binding domains as used herein refer generally to all polypeptide domains capable of specifically binding to one or more receptor (cohesin) domains of cellulosomal scaffoldin subunits. Exemplary cohesin-binding domains include dockerin domains found in or derived from glycoside hydrolases that assemble to cellulosomes. Both type I and type II dockerin domains are contemplated herein. Dockerin domains are explained *inter alia* in Lytle *et al*. 2001 ("Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain". J Mol Biol 307: 745-753), Adams *et al*. 2005 ("Structural characterization of type II dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition". Biochemistry 44: 2173-82) and in Bayer *et al*. 1998 ("Cellulosomes-structure and ultrastructure". J Struct Biol 124: 221-234). A suitable assay for determining dockerin-cohesin interactions, such as to guide the choice of a suitable dockerin domain herein, is described *inter alia* in Haimovitz *et al*. 2008 ("Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules". Proteomics 8: 968-979).

Further contemplated is co-expression and optionally co-secretion of polypeptides taught herein with other polypeptides or enzymes, preferably recombinant polypeptides or enzymes, useful in carbohydrate polymer metabolism, particularly in cellulose metabolism. The term "co-expression" generally relates to expression of two or more proteins or polypeptides (said to be "co-expressed") by the same micro-organism, particularly by a same cell of said micro-organism. Suitable systems for achieving co-expression of two or more polypeptides are known *per se*. Without limitation, co-expressed polypeptides may be encoded by separate cistrons controlled by the same or different regulatory elements, or may be encoded by a single multi-cistronic element; such cistrons may be on a chromosome or on same or different epigenetic elements, or combinations thereof *etc*.

Also contemplated are mixtures of enzymes comprising the ACLA polypeptide, fragment, variant or homologue thereof, for use in cellulose degradation. More particularly, the mixtures may comprise, in addition to the ACLA polypeptide, other cellulase enzymes. In the mixtures according to the invention, the cellulase enzymes may have complementary activity or additive activity.

The mixtures according to the present invention can contain the ACLA polypeptide, fragment, variant or homologue thereof in different amounts. In particular embodiments the ACLA polypeptide, fragment, variant or homologue thereof is present in the enzyme mixture at an amount relative to all enzymes of from about 15 to about 99%. More particularly, the amount of ACLA, fragment, variant or homologue thereof in the enzyme mixture is from about 50 to about 99%. More particularly the amount of ACLA polypeptide, fragment, variant or homologue thereof in the enzyme mixture is from about 80 to about 99%.

The terms “glycoside hydrolase” or “glycosidase” generally encompass enzymes and enzyme complexes capable of hydrolysing glycosidic linkages, more particularly O-, N- or S- glycosidic linkages, even more preferably such linkages in oligosaccharide and/or polysaccharide substrates. The terms encompass exoglycosidases as well as endoglycosidases. The terms particularly encompass glycosidases grouped under the EC 3.2.1. “Glycosidases” classification of NC-IUBMB.

The term “cellulase” generally encompasses enzymes and enzyme complexes capable of hydrolysing cellulose and cellulose-containing substrates. The term encompasses without limitation the following cellulase types: cellobiohydrolase(1,4-β-D-glucan cellobiohydrolase, EC 3.2.1.91), endo-β-1,4-glucanase (endo-1,4-β-D-glucan 4-glucanohydrolase, EC 3.2.1.4), processive endocellulase (EC 3.2.1.4./EC 3.2.1.91.) and β-glucosidase (EC 3.2.1.21). Cellulases further encompass endoglucanases as well as (processive) exoglucanases, and cover enzymes which produce monomers and/or oligomers of various lengths. The term may also encompass cellobiases, oxidative cellulases, glucosidases, cellulose phosphorylases, and other enzymes commonly denoted thereby. The term encompasses enzymes capable of degrading various forms of cellulose, such as without limitation crystalline cellulose, semi-crystalline cellulose, amorphous cellulose, hemicellulose and/or chemically or biologically modified cellulose forms.

By means of example and not limitation, a first cellulase may be considered as having activity complementary to a second cellulase if the first and second cellulases act

preferentially on distinct substrates (such as, *e.g.*, crystalline cellulose, semi-crystalline cellulose, amorphous cellulose or hemicellulose), or if the first and second cellulases produce distinct products (*e.g.*, distinct populations of sugar monomers and/or oligomers), or if the first cellulase acts preferentially on reaction products of the second cellulase or *vice versa*, *etc*.

By means of example and not limitation, the cellulolytic system of *Clostridium cellulolyticum* has been reviewed by Bélaich *et al*. 1997 ("The cellulolytic system of Clostridium cellulolyticum". J Biotechnol 57: 3-14); and of Saccharophagus degradans in Taylor *et al*. 2006 (*supra*).

The invention further contemplates engineered cellulosomes comprising polypeptides, particularly ACLA and related polypeptides (*e.g.*, DZ-domain and polypeptides comprising such), as taught herein. As known in the art, native cellulosomes represent extracellular, multienzyme cellulolytic complexes of certain bacterial taxons, such as *inter alia Clostridium* and *Bacteroides*. The basic organisation of native cellulosomes involves an organising polypeptide (scaffoldin), commonly comprising one or more same or distinct carbohydrate-binding modules (CBM) and one or more same or distinct receptor (cohesin) domains that facilitate integration of glycosidases or cellulolytic enzymes into the cellulosome complex through interaction with cognate dockerin domains in said enzymes.

These structural features of native cellulosomes may also be employed in engineered cellulosomes, whereby desired carbohydrate-binding and carbohydrate-depolymerising activities may be integrated into engineered cellulosomes. Techniques for producing engineered cellulosomes are established in the art (see, *e.g.*, Fierobe *et al*. 2002 ("Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components". J Biol Chem 277: 49621-30); Fierobe *et al*. 2005 ("Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin". J Biol Chem 280: 16325-34); and Perret *et al*. 2004 ("Production of heterologous and chimeric scaffoldins by Clostridium acetobutylicum ATCC 824". J Bacteriol 186: 253-7)).

Commonly, an engineered cellulosome may contain a hybrid or chimeric scaffoldin polypeptide comprising one or more CBM modules of the same or divergent carbohydrate substrate specificity, and one or more cohesin domains capable of binding the same or

divergent cognate dockerin domains. Enzymes, such as particularly glycosidases or cellulolytic enzymes, may be incorporated into the cellulosome by means of cognate dockerin modules (normally present in and/or engineered into said enzymes) for the cohesin domains found in the scaffoldin polypeptide. Said enzymes may also comprise CBM or other non-catalytic domains. Such engineered cellulosomes are commonly known in the art as “chimeric” or “hybrid” cellulosomes or “minicellulosomes”.

Alternatively or additionally, the one or more enzymes, such as glycosidases or cellulolytic enzymes, may be covalently linked with the scaffoldin backbone, such as, *e.g.*, by being expressed as a part of the same polypeptide chain (i.e., genetic fusion). Such cellulosomes are commonly known as “covalent” (Mingardon et al. 2007, “Exploration of new geometries in cellulosome-like chimeras” Appl. Environ. Microbiol. 73: 7138-7149).

Hence, engineered cellulosomes, such as hybrid, chimeric, mini-, or covalent cellulosomes, or any combination of these modalities, or any other engineered cellulosome forms, are contemplated for use with the polypeptides taught herein.

The term “cellulose” is known *per se* in the art. By means of further explanation and not limitation, the term may encompass all forms of cellulose, such as without limitation naturally and non-naturally occurring cellulose forms, such as, *e.g.*, crystalline cellulose, semi-crystalline cellulose, microcrystalline cellulose, amorphous cellulose, hemicellulose, regenerated cellulose, and/or chemically or biologically modified or derivatised cellulose forms, such as, *e.g.*, ethers or esters thereof, *etc*.

A substance enriched in crystalline cellulose may comprise, without limitation, at least about 1% (w/w), *e.g.*, ≥ 5% (w/w), preferably at least about 10% (w/w), *e.g.*, ≥ 20%, ≥ 30% or ≥ 40% (w/w), more preferably at least about 50% (w/w), *e.g.*, ≥ 60% or ≥ 70% (w/w), yet more preferably at least about 80% (w/w), *e.g.*, ≥ 90% (w/w) of crystalline cellulose, semi-crystalline and/or microcrystalline cellulose as understood in the art, more preferably of crystalline cellulose.Methods and processes for treating or contacting cellulose-containing substances with the ACLA enzymes, fragments, variants and homologues thereof described herein or recombinant micro-organisms such as taught herein are generally known in the art, and may without limitation comprise industrial enzymatic degradation or fermentation processes of various scale. The material to be so degraded and/or fermented may be provided, *e.g.*, in solid, semi-solid, liquid or soluble form, or as a homogenate or extract, or as a (aqueous) dispersion or suspension, or as (partly) isolated or purified and/or

pre-treated (*e.g.*, by dilute-acid pre-treatment, sodium hydroxide pretreatment, lime pre-treatment, pre-treatment with organic solvents with water, hydrothermal processes, or AFEX) lignocellulosic substrates or cellulosic substrates or cellulose, or in other suitable forms used *per se* for depolymerisation and/or fermentation of cellulosic materials and biomass. Accordingly, the invention also contemplates starter cultures comprising the polypeptides or micro-organisms of the invention, optionally in combination with one or more other enzymes or micro-organisms, respectively.

The pretreatment of cellulose is performed by methods known to the person skilled in the art and inclused thermal, mechanical, chemical, or combinations of these methods such as to disrupt the fiber structure of cellulosic substrates and to enhance subsequent enzymatic hydrolysis of the cellulosic substrate. Any pretreatment process may be employed in combination with the method of the present invention including, but not limited to pulping, such as kraft, sulfite, mechanical, thermal, and chemi-thermal-mechanical pulping. Alternatively, pretreatment processes may also involve the addition of organic solvents to aid in the fractionation of lignin, cellulose and hemicellulose from lignocellulosic raw materials,

Biotechnological approaches for cellulose utilisation are reviewed *inter alia* in Lynd *et al*. 2002 (“Microbial cellulose utilization: fundamentals and biotechnology”. Microbiol Mol Biol Rev 66: 506-77), and in Himmel et al. 2007 (“Biomass recalcitrance: engineering plants and enzymes for biofuels production” Science 315: 804-807).

The term “chemical intermediate” as used herein refers to any chemical substance produced during the conversion of some reactant to a product. In the context of the present invention, the processing involves transformation of a degradation product of biomass (such as glucose or another fermentable sugar) to some desired product through a succession of steps. Accordingly, all the substances generated by the degradation of biomass by the ACLA polypeptide according to the invention and which can be used in succeeding steps in bioprocessing are considered intermediates.

Methods for the production of chemical intermediates may include both methods whereby the chemical intermediate is generated in bulk (for use in further, physically separated bioprocessing methods) and methods of consolidated bioprocessing whereby the intermediate is not isolated but immediately converted into an end-product.

The methods for the production of solvents, fuels or chemical intermediates according to the invention which comprise degrading cellulose by contacting cellulose-containing substances with the ACLA enzymes, fragments, variants and homologues thereof or with a (non-solventogenic) micro-organism according to the invention, typically further include steps which involve the conversion of the degradation products of cellulose or cellulose-containing material, more particularly glucose to a solvent, fuel or chemical intermediate. In particular embodiments this is ensured by contacting the degradation products obtained by the cellulose degradation methods according to the invention with a solventogenic micro-organism, capable of converting glucose to the desired solvent. Such solventogenic micro-organisms include micro-organisms which are naturally solventogenic and organisms which have been genetically modified in order to increase their ability to produce solvents from a carbohydrate source. Solventogenic micro-organisms are known to the skilled person and include but are not limited to those as described herein. In further particular embodiments the degradation products are converted to other products either by micro-organisms or by chemical processing.

As described herein, the invention further provides the use of the ACLA polypeptide, or a functional fragment, homologue or variant thereof in industrial cellulose degradation, more particularly in the production of solvents, fuels or chemical intermediates. Indeed, it has been found that the isolated ACLA enzyme has a particularly high cellulolytic activity compared to other bacterial cellulose degrading enzymes. Accordingly, this enzyme is particularly suitable for application in the bulk degradation of cellulose, such as large-scale degradation of cellulose-containing biomass in bioreactors.

More particularly, the invention provides methods for the degradation of a substance comprising cellulose, such as lignocellulosic or cellulosic material or biomass, comprising contacting the substance with the ACLA polypeptide or a functional fragment, homologue or variant thereof or with a recombinant micro-organism expressing ACLA or a functional fragment, homologue or variant thereof as taught herein. It has been observed that ACLA has endo-processing activity, which makes it of particular interest for use in the industrial degradation of biomass. It has also been found that the ACLA enzyme has the further advantage of being able to directly degrade cellulose in its crystalline form, i.e. in its naturally occurring form. Thus, in particular embodiments, the invention provides methods for the degradation of a substance comprising crystalline cellulose, comprising contacting the substance with the ACLA polypeptide or a functional fragment, homologue or variant

thereof or with the recombinant micro-organism expressing ACLA or a functional fragment, homologue or variant thereof as taught herein.

The above aspects and embodiments are further supported by the following non-limiting examples.

EXAMPLES

Example 1: Identification of the DZ domain of ACLA

ACLA is encoded by the gene *aclA* from *Pseudomonas* sp ND137. Fig. 3 compares the general organisation of ACLA with that of Cel5H. ACLA has the following organisation (from N-terminal to C-terminal ends of the enzyme): signal sequence/catalytic module (GH-family 5)/aspartic acid- and proline linker/carbohydrate binding module belonging to CBM-family 6/serine-rich-linker, in other denotation signal sequence-GH5-DPR-CBM6-PSL.

The gene encoding ACLA has already been cloned and over-expressed in *E. coli*, but the corresponding enzyme was not fully characterized (Aoki & Kamei 2006. European Journal of Phycology 41: 321-328). The coding sequence of ACLA is provided in Fig. 1F (SEQ ID NO: 6).

ACLA displays an important overall homology with Cel5H from *S. degradans*, and also possesses a DZ domain.

Example 2: Production and purification of ACLA in *E. coli*

Using molecular biology techniques a synthetic DNA encoding the mature form of ACLA (without signal sequence) was cloned in an *E. coli* expression vector (pET22b(+), Novagen). The resulting vector is used to transform the *E. coli* strain BL21 (DE3) (Novagen). Similarly, modified genes encoding truncated or engineered forms of ACLA are obtained by PCR and cloned in pET22b(+), prior to transformation of *E. coli* strain BL21 (DE3).

In order to facilitate purification, four His codons were grafted at the C-terminus extremity of the recombinant protein to facilitate purification on Nickel resin (Ni-NTA, Qiagen).

The recombinant strains were grown in Luria Bertani medium and the expression of the cloned genes was triggered using IPTG as the inducer. The cultures were centrifuged and the harvested cells are broken in a French press.

To obtain the produced ACLA protein in a soluble form that could be purified on Nickel resin a denaturation (using urea) and a subsequent renaturation step were necessary.

The synthesis of the recombinant ACLA was verified by denaturing polyacrylamide gel electrophoresis (SDS-PAGE).

It was observed that purification on Nickel resin and anion-exchange FPLC yielded a 70% pure protein batch (some degradation material was also observed.

Example 3: Evaluation of the activity of ACLA on various substrates including para-nitrophenyl-β-D-cellobioside (pNPC), carboxymethyl cellulose (CMC), phosphoric acid-swollen cellulose (PASC), Avicel and a raw substrate, the hatched straw

.Activity of the purified ACLA enzyme was tested on a variety of substrates using standard conditions (37°C) and compared to the activity of purified recombinant Cel5H on the same substrates.

The enzymatic activity of ACLA was tested on the soluble chromogenic substrate, para-nitrophenyl-β-D-cellobioside (pNPC) and compared to that of Cel5H (at 95% purity). Upon cleavage by the enzyme of the link between the para-nitrophenyl group and the cellobiose, para-nitrophenol and cellobiose is liberated and its concentration can be directly determined by means of spectrophotometric analysis. The enzymatic activity of ACLA on para-nitrophenyl-β-D-cellobioside was 17,9 ui/µmol, whereas the enzymatic activity of Cel5H on the same substrate was 18,2 ui/µmol. Accordingly, ACLA showed 98% of enzymatic activity compared to Cel5H on para-nitrophenyl-β-D-cellobioside substrate.

For testing of activity on PASC (amorphous cellulose), the enzyme concentration used was 5 nM and the substrate concentration was at 3.5 g/L (pH 6.0). The enzymatic activity of ACLA on PASC was 780 ui/µmol, whereas the enzymatic activity of Cel5H on the same

substrate was 1600 ui/µmol. Accordingly, ACLA showed 49% of enzymatic activity compared to Cel5H on PASC. The liberation of sugars as a function of time using PASC as a substrate is shown in Fig. 4 for both ACLA and Cel5H.

For testing of activity on Avicel (crystalline cellulose), the enzyme concentration is 100 nM and substrate is at 3.5 g/L (pH 6.0). The enzymatic activity of ACLA on Avicel was 56 µM/24h, whereas the enzymatic activity of Cel5H on the same substrate was 140 µM/24h. Accordingly, ACLA showed 40% of enzymatic activity compared to Cel5H on Avicel. The liberation of sugars as a function of time using Avicel as a substrate is shown in Fig. 5 for both ACLA and Cel5H.

For testing of activity on CMC, the enzyme concentration is 1 nM and substrate is at 8 g/L CMC (pH 6.0).

Example 4: Comparison of ACLA avicelase activity with wild type and engineered cellulases from *C. cellulolyticum*

Using an assay design as described in example 3, the ACLA avicelase activity is compared with wild type and engineered cellulases from *C. cellulolyticum*.

The enzyme concentration is 100 nM and substrate is at 3.5 g/L (pH 6.0).

The activity of ACLA (ACLAwt) is also tested on hatched straw (enzyme and substrate at 100 nM and 3.5 g/L, respectively). All kinetic experiments described above are performed at 37°C in 20 mM Tris-maleate pH 6.0, 1 mM $CaCl_2$ (azide 0.01% w/v) buffer. The avicelase activity of ACLA is also measured in presence of 1 % sodium chloride (171 mM NaCl) in the same buffer. It is observed that the ACLA enzyme is significantly more active that the wild-type and engineered cellulases from C. cellulolyticum.

Example 5: ACLA activity combined with the activity of other cellulases on crystalline cellulose

Kinetic experiments are performed to investigate whether ACLA activity is complementary or additive to that of other cellulase enzymes in the degradation of pure crystalline cellulose.

In these experiments, the following enzymes are tested:

- wild type enzymes from *C. cellulolyticum* : Cel5A (endo), Cel9G (endo), Cel9M (endo), Cel9P (endo), Cel48F (processive) et Cel9E (processive).
- Modified enzymes from *C. cellulolyticum* : Cip0-G (endo)
- Enzymes isolated from organisms other than *C. cellulolyticum* : Cel6A (exo) de *Neocallimastix patriciarum* (fungus).

It is observed that the combined activity of ACLA with other cellulases is at least additive. This is an indication of the potential use of the combination of ACLA with other enzymes for the degradation of cellulose.

Example 6: Production and secretion of ACLA by *C. acetobutylicum*

The wild type gene encoding ACLA is amplified by PCR from the strain *Pseudomonas sp.*, and cloned in the *E. coli-C. acetobutylicum* pSOS952 shuttle vector. This plasmid is an expression vector for *C. acetobutylicum* (Perret *et al*. 2004. J Bacteriol 186: 253-7). The expression of the gene of interest is under the control of the strong and constitutive promoter (p*thl*) of the gene encoding the thiolase enzyme. Two *lac* operators are introduced upstream and downstream of the p*thl* to prevent any expression of the gene in *E. coli*. The vector is subsequently methylated (*in vivo* using the strain ER2275[pAN1] and *in vitro* using the methylase CpG) and used to electro-transform *C. acetobutylicum* strain as formerly described. Recombinant clones are obtained and PCR tests on colonies indicated that the vector remains intact in the solventogenic *Clostridium*. Several clones are grown in 2YTC medium, but no additional activity on para-nitro-phenyl-cellobioside is detected in the culture supernatant compared to control strain harbouring the “empty” pSOS952 (CMCase test on plates).

Transformation of *C. acetobutylicum* with this new vector (after methylation) generates recombinant colonies. *Pseudomonas sp.* is a Gram-negative bacterium which displays two membranes (outer membrane and cytoplasmic membrane) whereas *C. acetobutylicum* (Gram-positive bacterium) only possesses a cytoplasmic membrane. In one construct, the native signal sequence of ACLA is replaced by that of Cel5A from *C. cellulolyticum*, since the latter is well secreted by *C. acetobutylicum* (up to 5 mg/L). *C. acetobutylicum* is transformed with the vector containing the hybrid gene. The secretion yield of ACLA using the Cel5A signal sequence is investigated.

A culture of the recombinant bacterial strain is grown in a fermentor in an environment of 3 g/L 2YT-PASC with a pH fixed at 5,5. The degradation of PASC, the consummation of residual cellobiose and bacterial growth are measured.

A second culture of the recombinant bacterial strain is grown in a fermentor in an environment of 2YT (however not containing any PASC) with a pH fixed at 5,5. The goal of this experiment is to check the correlation between bacterial growth and the consummation of PASC. Furthermore, a third culture is set up, to check whether the observed bacterial growth is due to the activity of ACLA and not to the activity of the endogenous enzymatic system of *C. acetobutylicum*. Therefore, a culture of a strain containing an empty vector (pSOS) is grown in a fermentor in an environment of 3 g/L 2YT-PASC with a pH fixed at 5,5.

Finally, a fourth culture is grown in a fermentor in an environment of 5 g/L 2YT-PASC with a pH fixed at 5,5 to check whether the PASC concentration is the limiting factor for bacterial growth. Thus, a higher concentration of PASC may potentially result in higher OD values (i.e. higher than the maximal OD value of 0.45 observed in the previous experiment).

CLAIMS

1. A method for producing a solvent, fuel or chemical intermediate from a substance comprising cellulose, comprising:
a) expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue in a recombinant host cell, and
b) contacting said substance with said ACLA polypeptide.

2. The method of claim 1, which further comprises the step of separating said ACLA enzyme or active fragment from said recombinant host cell.

3. A recombinant micro-organism expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, wherein said micro-organism is capable of degrading cellulose.

4. The recombinant micro-organism according to claim 3 which is a solventogenic micro-organism.

5. The recombinant solventogenic micro-organism according to claim 4, which is an ethanologenic micro-organism.

6. A recombinant solventogenic micro-organism according to claim 4 or 5, which is a solventogenic bacterium.

7. The recombinant solventogenic bacterium according to claim 6, which is *Clostridium acetobutylicum*.

8. The method according to claim 1 or 2, comprising contacting said substance with the recombinant micro-organism according to any one of claims 3 to 7.

9. A method for the degradation of a substance comprising cellulose, comprising contacting said substance with the recombinant micro-organism according to any one of claims 3 to 7.

10. The method according to any one of claims 1, 2, 8 or 9, wherein said substance comprises or is enriched in crystalline cellulose.

11. A recombinant micro-organism expressing
(a) the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, and
(b) one or more additional foreign cellulose-degrading enzymes other than said ACLA polypeptide.

12. A mixture of enzymes comprising the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue and one or more cellulose-degrading enzymes other than said ACLA polypeptide.

13. Use of
(i) the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,
(ii) a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,
(iii) an expression vector comprising a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue, wherein said nucleic acid sequence is operably linked to regulatory sequences which allow for expression in a host cell;
(iv) a host cell expressing the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a functional fragment and/or variant of said ACLA polypeptide or of said homologue,
(v) a host cell transformed with a nucleotide sequence encoding the ACLA polypeptide of Pseudomonas sp. ND137 or a homologue thereof, or a

functional fragment and/or variant of said ACLA polypeptide or of said homologue, or

(vi) the recombinant micro-organism according to any of claims 3 to 7 and 11,

for the production of a solvent, fuel or chemical intermediate from a substance comprising cellulose.

Figure 1

FIG 1A

MKHSLHQRFLLPVTLAALSLSASMTNADVAPISTNGNQLLFGGAVDSVAGPSLFWSNN
GWGGEKFYNAGAVASAQQDWNAEIIRAAMGVDEPGGYLEDASANLNRVRAVVDAAI
ANDMYVIIDWHSHHAESYTQAAVSFFQQMASEYGQHDNVIYEIYNEPLSVSWSNTIKP
YAEQVIGAIRAVDPDNLIVVGTPTWSQDVDAAANDPITNYNNIAYTLHFYAGTHTQYL
RDKAQYALDMGIPLFVTEWGTVNANGDGGVAYNETNTWMDFLKANNISHANWALN
DKAEGSSALVTGTNPSGNWADNQYTASGTFVRDIVRDWSDGDPVDPDPTCTRINMPGT
IEAESFCDMDGIQTESTTDTGGGLNIGWTDAGDWTSYEVNVPAAGRYKVSYRVAAAQ
NSGMLQLEAAGGFPTYGSITTPVTGGWQSWQTISHEVDLPAGDQDLAIAVVSGGWNL
NWIKVEPAGGSSSSSSSSSSSSSSSSSSSTSGCDTANATSITGNTITVSEGQCIRYEHTW
GSLQLGSWSAAAGTTYDVINCNGQVIADVAQVQNGFSTVATGTNHCNLYVYVKQAPT
SFDLQFGSW (SEQ ID NO: 1)

FIG 1B

DVAPISTNGNQLLFGGAVDSVAGPSLFWSNNGWGGEKFYNAGAVASAQQDWNAEIIR
AAMGVDEPGGYLEDASANLNRVRAVVDAAIANDMYVIIDWHSHHAESYTQAAVSFFQ
QMASEYGQHDNVIYEIYNEPLSVSWSNTIKPYAEQVIGAIRAVDPDNLIVVGTPTWSQD
VDAAANDPITNYNNIAYTLHFYAGTHTQYLRDKAQYALDMGIPLFVTEWGTVNANGD
GGVAYNETNTWMDFLKANNISHANWALNDKAEGSSALVTGTNPSGNWADNQYTASG
TFVRDIVRDWSDGDPVDPDPTCTRINMPGTIEAESFCDMDGIQTESTTDTGGGLNIGWT
DAGDWTSYEVNVPAAGRYKVSYRVAAAQNSGMLQLEAAGGFPTYGSITTPVTGGWQ
SWQTISHEVDLPAGDQDLAIAVVSGGWNLNWIKVEPAGGSSSSSSSSSSSSSSSSSST
SGCDTANATSITGNTITVSEGQCIRYEHTWGSLQLGSWSAAAGTTYDVINCNGQVIADV
AQVQNGFSTVATGTNHCNLYVYVKQAPTSFDLQFGSW (SEQ ID NO: 2)

FIG 1C

ATGAAATCAGCAACCACAAATCAATCGAGGGCACGCAGTAGCGCCTTTAAAAATAT
GTTGGCGGCATCGCTCGCAGGTTTAGGGCTACTATCAGCTTCTGCATTTGCCGATGT
AGCCCCGCTAACCGTAGACGGCAATAAAATTCTTAGCGGTGGCCAGCAAGCCAGTT
TTGCCGGTAATAGCTTATTTTGGTCTAACAATGGCTGGGGCGGTGAGAAGTATTAC
ACGGCCGGTACCGTTGAATGGCTAAAGCAAGACTGGGGCAGTAATTTAGTTCGCGC
CGCAATGGGTGTCGATGAAAACGGCGGCTACTTAGAAGACCCAGCAGGAAACAAA
GCGAAAGTAACAACCGTTGTAGATGCAGCCATCGCTAACGATATGTATGTAATTAT
CGATTGGCACAGCCACCACGCCGAAGACTACCAAAACCAAGCCATTAGCTTTTTCC
AAGATATGGCTCGCACCTACGGTAACAACAACAACGTTATATACGAAATTTATAAC
GAGCCATTACAGGTTTCTTGGAGCGGCACCATCAAGCCTTACGCAGAAGCGGTAAT
TGGCGCAATTCGCGCAATCGACCCAGATAACCTTATTATTGTGGGCACGCCTACTTG
GTCGCAGGATGTAGACGTAGCCTCGCGCGACCCCATCACGCAGTACAGCAACATTG
CCTACACTATTCACTTTTATGCGGGCACCCACAAACAATCCCTACGCGATAAAGCA
CAAACCGCATTAAATAATGGTATTGCTTTGTTTGCTACCGAATGGGGTACAGTAAAT
GCCAACGGTGACGGCGGTGTAGACGCAGCCGAAACTGATCGTTGGATGCAGTTTTT
TAAAGCGAATCATATAAGCCATGCCAACTGGGCCTTAAACGATAAAGCCGAAGGCT
CTTCTGCATTAAAGCCTGGCTCTAACGCAAACGGCGGCTGGAGCAATTCCGACTTA
ACCGCCTCTGGTACCTATGTTAAAAACTTAATTAAAACATGGAACGACGGCTCACC
GAGCAGCAGCTCATCTAGCAGCACCAGTTCTTCTTCAAGCAGCTCCTCGTCTAGTAG
CTCATCATCTAGCAGCTCTTCATCTAGTAGTTCTGGCGGTACCAATTTACCCGCGCG

CATTGAAGCAGAAAACTACGATAGCGCACCGGTAGAAACCACTGCAGGTAATAGC
GGCTCACCCACCAATTGTTCGTATAAAGGTATGGGCGTAGATGTAGAAAACTCTAC
TGAAGGTGCTTGTAATATTGGCTGGACTGCGGCAGGCGAAAAAGTAACTTACAACA
TTGGCAATGCCGATGGCACTTACGATATTGCATTGCGCGTAGCCTCTATGGATGCGG
GCAAACGTATCTCTGTGCATGTAAACAACAGCCTAGCAGATACCGTAACCACACAA
GGTGGCGGCTGGCAGGCATGGACTACCGAAACCATTTCTAACGTGTATATCCCATC
AAACTCGGTAATTACCGTTGAGTTTTACGATAGTGGCTCTAACCTAAACTTTTTAAA
CATTACCGAAAGCTCGGGTACCGAACCACCTGTAGAACCACCCGTTGAGCCGCCAG
TAGAACCACCCGTAGACAACGGTAACTTCCCATGTAACGACGGTAACTCTACGCTT
GCCAACAACGGCGCCTCCATTAACCTTAACCAAGGAGCGTGTGTTAAATACAATCA
CGGCTGGGGCGATATTCGTTTAGGCACCTGGAGCGGCAACGGTACCATTCGATACG
ACGTACTAGACTGCAATAACAACGTAATGAGTGATATTGCACAAAAACTTAATGAC
TTTACTGCTGTAGACACCGCAACAATGAACTGCGCACACTACATTTATGTAAAACA
AGCCCCTAGCAGCTACACCCTGCAATTTGGTAGCTGGTAG (SEQ ID NO: 3)

FIG 1D

MKSATTNQSRARSSAFKNMLAASLAGLGLLSASAFADVAPLTVDGNKILSGGQQASFA
GNSLFWSNNGWGGEKYYTAGTVEWLKQDWGSNLVRAAMGVDENGGYLEDPAGNK
AKVTTVVDAAIANDMYVIIDWHSHHAEDYQNQAISFFQDMARTYGNNNNVIYEIYNEP
LQVSWSGTIKPYAEAVIGAIRAIDPDNLIIVGTPTWSQDVDVASRDPITQYSNIAYTIHFY
AGTHKQSLRDKAQTALNNGIALFATEWGTVNANGDGGVDAAETDRWMQFFKANHIS
HANWALNDKAEGSSALKPGSNANGGWSNSDLTASGTYVKNLIKTWNDGSPSSSSSSST
SSSSSSSSSSSSSSSSSSSSSGGTNLPARIEAENYDSAPVETTAGNSGSPTNCSYKGMGV
DVENSTEGACNIGWTAAGEKVTYNIGNADGTYDIALRVASMDAGKRISVHVNNSLAD
TVTTQGGGWQAWTTETISNVYIPSNSVITVEFYDSGSNLNFLNITESSGTEPPVEPPVEPP
VEPPVDNGNFPCNDGNSTLANNGASINLNQGACVKYNHGWGDIRLGTWSGNGTIRYD
VLDCNNNVMSDIAQKLNDFTAVDTATMNCAHYIYVKQAPSSYTLQFGSW (SEQ ID
NO: 4)

FIG 1E

FADVAPLTVDGNKILSGGQQASFAGNSLFWSNNGWGGEKYYTAGTVEWLKQDWGSN
LVRAAMGVDENGGYLEDPAGNKAKVTTVVDAAIANDMYVIIDWHSHHAEDYQNQAI
SFFQDMARTYGNNNNVIYEIYNEPLQVSWSGTIKPYAEAVIGAIRAIDPDNLIIVGTPTW
SQDVDVASRDPITQYSNIAYTIHFYAGTHKQSLRDKAQTALNNGIALFATEWGTVNAN
GDGGVDAAETDRWMQFFKANHISHANWALNDKAEGSSALKPGSNANGGWSNSDLTA
SGTYVKNLIKTWNDGSPSSSSSSTSSSSSSSSSSSSSSSSSSSSSGGTNLPARIEAENYDS
APVETTAGNSGSPTNCSYKGMGVDVENSTEGACNIGWTAAGEKVTYNIGNADGTYDI
ALRVASMDAGKRISVHVNNSLADTVTTQGGGWQAWTTETISNVYIPSNSVITVEFYDS
GSNLNFLNITESSGTEPPVEPPVEPPVEPPVDNGNFPCNDGNSTLANNGASINLNQGACV
KYNHGWGDIRLGTWSGNGTIRYDVLDCNNNVMSDIAQKLNDFTAVDTATMNCAHYI
YVKQAPSSYTLQFGSW (SEQ ID NO: 5)

FIG 1F

ATGAAACATTCACTACATCAGCGGTTTTTGTTACCCGTAACCCTGGCCGCCCTATCG
CTATCGGCTAGTATGACTAATGCCGACGTTGCGCCTATCTCCACCAACGGTAACCA
ATTACTGTTTGGCGGCGCGGTCGACAGCGTGGCCGGGCCCAGCCTGTTTTGGAGCA
ATAACGGCTGGGGCGGCGAAAAGTTTTACAACGCCGGTGCCGTTGCCAGCGCCCAG
CAAGATTGGAACGCCGAGATTATCCGAGCCGCTATGGGGGTCGATGAACCAGGCG
GCTATCTAGAAGACGCCTCCGCCAACCTCAACCGGGTGCGCGCCGTGGTCGATGCC
GCGATTGCCAACGATATGTATGTCATCATCGATTGGCACAGCCACCATGCCGAGAG
CTATACCCAGGCAGCGGTAAGCTTCTTTCAACAAATGGCGTCCGAATACGGTCAGC
ACGACAACGTTATTTACGAAATTTACAACGAGCCGCTCAGCGTATCCTGGAGCAAC
ACCATTAAGCCTTACGCCGAGCAAGTCATCGGCGCTATTCGCGCCGTCGACCCGGA
TAACCTTATTGTGGTGGGTACCCCTACTTGGTCTCAAGACGTAGACGCCGCCGCTAA
TGACCCTATCACCAACTACAACAACATTGCCTATACCCTGCACTTTTACGCCGGCAC
CCACACCCAATACCTACGCGATAAAGCCCAATATGCGCTGGACATGGGTATTCCGC
TATTTGTTACCGAGTGGGGCACCGTCAACGCCAACGGGGACGGCGGCGTGGCCTAC
AACGAAACCAATACTTGGATGGACTTTTTAAAGGCCAACAACATCAGCCATGCCAA
CTGGGCGTTAAACGACAAGGCTGAGGGGTCATCTGCACTGGTCACTGGCACCAACC
CCAGCGGCAATTGGGCCGACAACCAATACACCGCATCGGGTACGTTTGTGCGCGAT
ATTGTCCGCGATTGGTCTGACGGCGACCCGGTTGATCCGGACCCAACCTGCACCCG
CATTAACATGCCCGGCACTATTGAGGCGGAATCGTTCTGTGATATGGACGGCATTC
AAACCGAATCCACCACCGATACCGGTGGCGGCTTAAACATCGGCTGGACTGACGCG
GGCGATTGGACGAGCTACGAGGTTAATGTCCCGGCAGCGGGCCGCTACAAAGTCAG
CTATCGCGTCGCAGCCGCACAAAATAGCGGCATGCTACAGCTTGAAGCCGCTGGCG
GTTTTCCCACTTATGGCAGTATTACCACACCGGTTACCGGAGGCTGGCAATCGTGGC
AAACCATTTCGCACGAGGTTGACCTACCCGCCGGCGACCAAGATCTAGCCATCGCT
GTTGTCAGTGGCGGCTGGAACCTTAACTGGATCAAAGTCGAACCTGCCGGCGGCTC
TAGCTCTTCGAGCAGCTCCTCCTCTAGTAGTTCTTCTTCCAGCAGTTCGTCGTCCAGC
AGCACGTCTGGCTGTGACACCGCCAACGCTACGTCCATCACAGGCAATACCATCAC
GGTGTCCGAGGGACAGTGCATTCGCTATGAACACACCTGGGGCAGCTTGCAGCTGG
GCAGCTGGAGCGCCGCGGCGGGCACAACCTACGATGTCATCAACTGCAACGGTCAA
GTGATTGCCGATGTCGCCCAAGTGCAAAATGGTTTTAGCACTGTCGCCACCGGCAC
TAACCACTGCAACCTCTATGTTTACGTTAAACAAGCACCCACCAGTTTTGACCTGCA
ATTTGGCAGCTGGTAA (SEQ ID NO: 6)

Figure 2

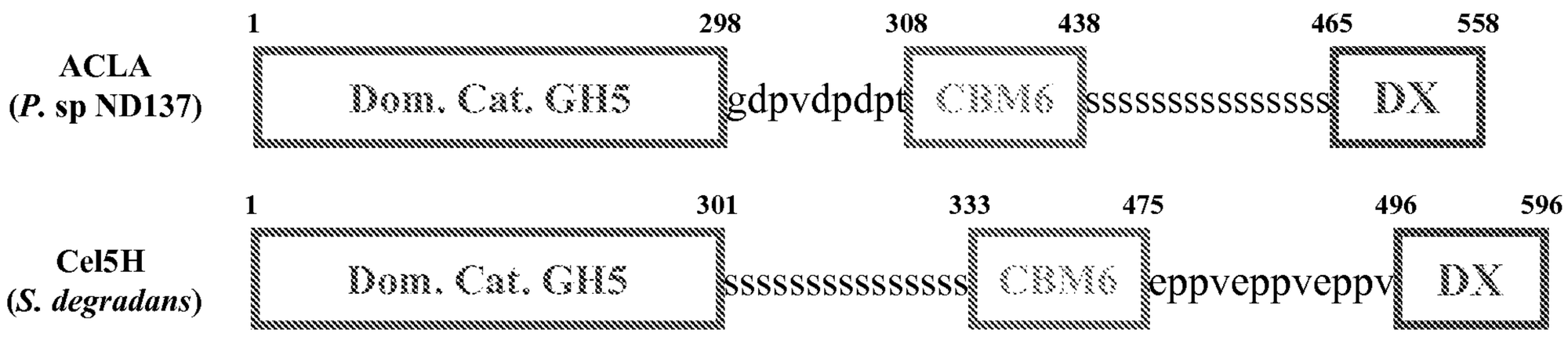

Figure 3

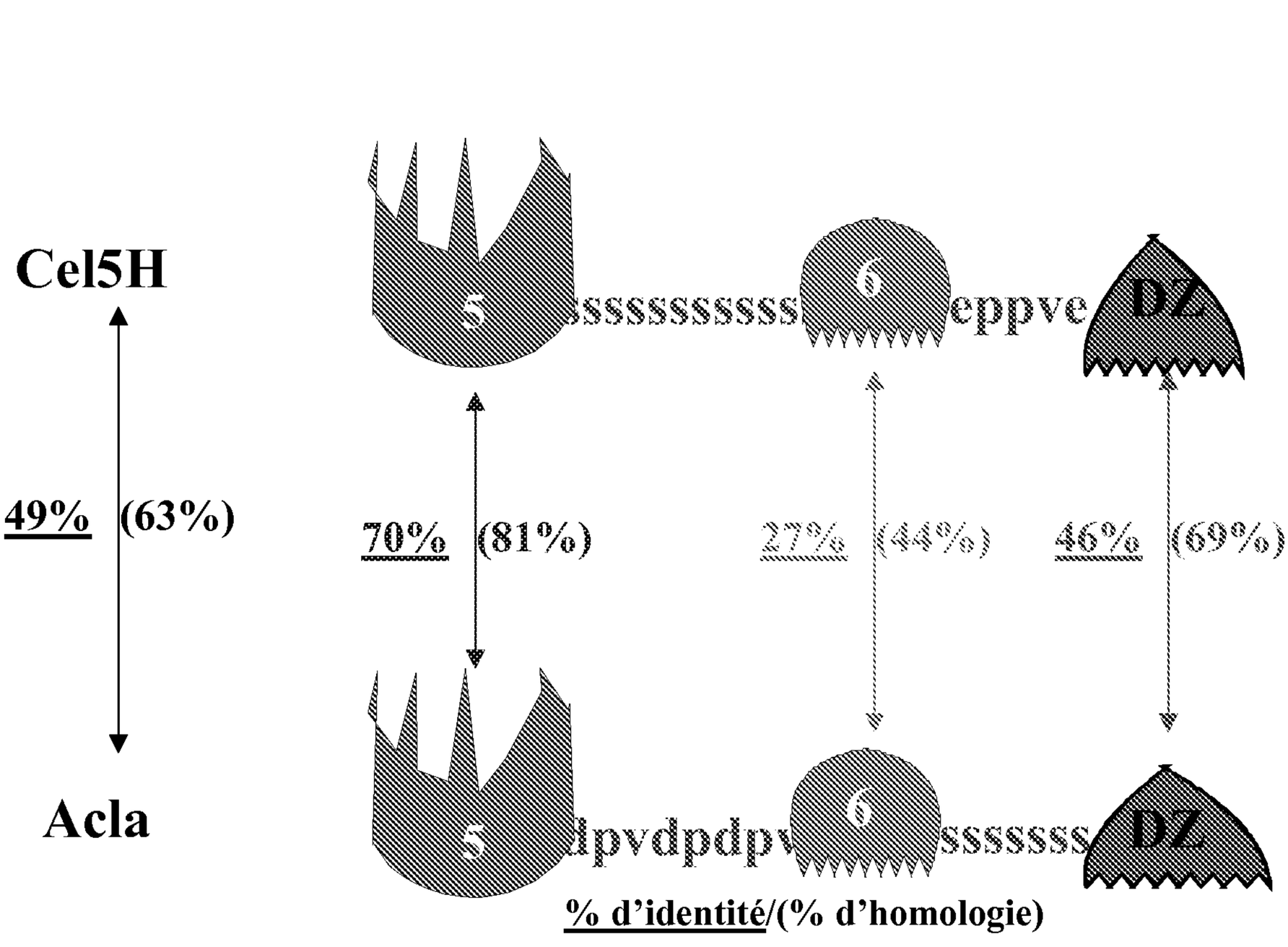

Figure 4

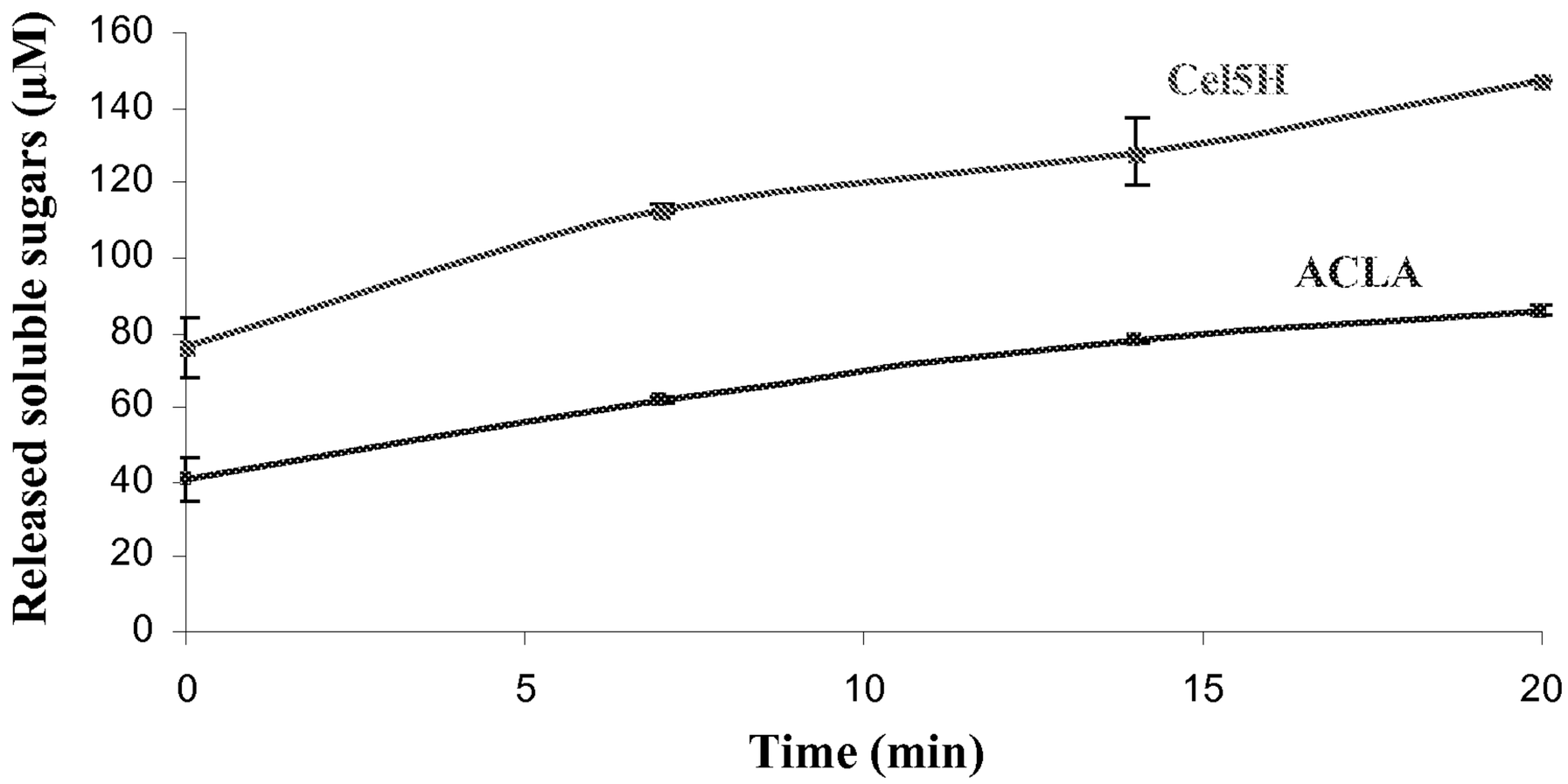

Figure 5