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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2017 (26.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/013438 A1

(51) International Patent Classification:
C12P 19/02 (2006.01) *C07G 1/00* (2011.01)

(21) International Application Number:
PCT/GB2016/052219

(22) International Filing Date:
21 July 2016 (21.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1512831.7 21 July 2015 (21.07.2015) GB

(71) Applicant: **UNIVERSITY OF WARWICK** [GB/GB]; University House, Kirby Corner Road, Coventry CV4 8UW (GB).

(72) Inventors: **BARKER, Guy**; School of Life Sciences, Gibbet Hill Campus, The University of Warwick, Coventry CV4 7AL (GB). **EASTWOOD, Daniel**; 30 Lundy Drive, West Cross, Swansea SA3 5QL (GB). **NURIKA, Irnia**; Agricultural Technology Facility, University of Brawijaya, Beteran Street 1, Malang, East Java, S5145 (ID).

(74) Agents: **HAMER, Christopher** et al.; Mathys & Squire LLP, The Shard, 32 London Bridge Street, London SE1 9SG (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

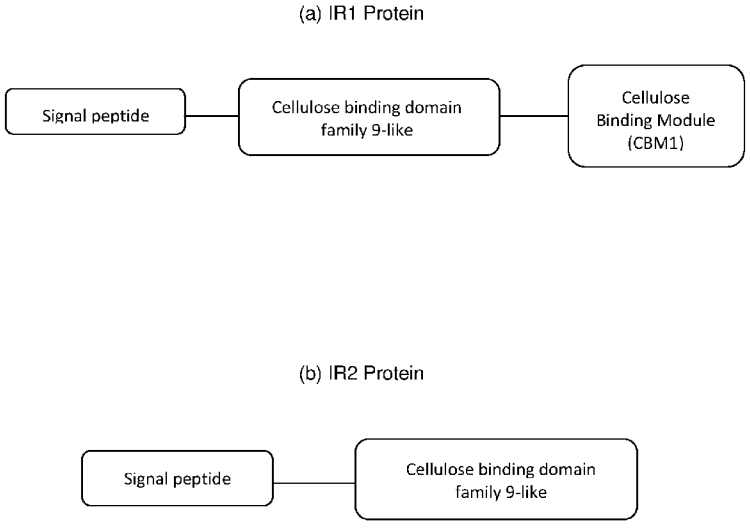
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: METHODS FOR EXTRACTING SUGARS AND LIGNIN DERIVED PRODUCTS FROM LIGNOCELLULOSIC BIOMASS MATERIAL

Figure 1



(57) Abstract: A method for extracting sugars and/or lignin-derived products from lignocellulosic biomass material, comprising: contacting the lignocellulosic biomass material with a composition comprising a polypeptide in the presence of a reducible substrate and an oxidising agent, wherein said polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars and/or lignin-derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

WO 2017/013438 A1

METHODS FOR EXTRACTING SUGARS AND LIGNIN DERIVED PRODUCTS FROM LIGNOCELLULOSIC BIOMASS MATERIAL

FIELD

The present invention relates to polypeptide-based methods for extracting sugars and /or lignin derived products from lignocellulosic biomass materials, uses of a polypeptide for extracting sugars and/or lignin derived products from lignocellulosic biomass materials, and recombinant micro-organisms that express said polypeptide.

BACKGROUND

Lignocellulose is the major structural component of woody plants, and non-woody plants such as grass, and consists of three main components: lignin, cellulose and hemicellulose. These components provide essential resources for the production of components for use in various industries including agriculture, food production, waste management, textiles and paper, as well as renewable energy and high-value chemical products. Much focus has therefore been placed on identifying effective methods for isolating and/ or further processing lignin, cellulose and hemicellulose from waste lignocellulosic biomass material.

Biofuels, such as biogasoline, biodiesel and bioethanol, offer an alternative renewable energy source to petroleum-based fuels, and have the potential to provide a sustainable source of energy with reduced greenhouse emissions. Lignocellulosic biomass material provides a key source of biological material for production of biofuels. The cellulose and hemicellulose components can be broken down into simple sugars using enzymatic or chemical hydrolysis reactions. Hydrolysis of cellulose results in the release of hexose sugars, such as glucose, mannose and galactose; whilst hemicellulose can be hydrolysed into pentose sugars, such as xylose and arabinose. The extracted sugars can be fermented to alcohol, such as ethanol, and used to produce biogasoline. Alternatively, the extracted sugars may be processed into esters for use in the production of biodiesel. Lignin may also be used to produce biofuels, for example through the generation of syngas (carbon monoxide/ hydrogen), which in turn can be used to generate chemical alcohols and produce biogasoline.

High value chemical products can also be produced from the lignin component of lignocellulosic biomass material. For example, lignin can be depolymerised and used to produce a variety of aromatic and polyaromatic compounds, such as phenol, BTX

chemicals (such as benzene, toluene and xylene) and terephthalic acid. Lignin can also be used to produce a wide variety of other products, including carbon fibre materials, plastic materials, polymer modifiers, resins, adhesives, binders, dispersants, emulsifiers, wetting agents, fertilisers, dyestuffs, agglomerants and chelants.

Lignocellulosic biomass material is a relatively inexpensive potential source of cellulose, hemicellulose and lignin because it can be obtained from widely available waste materials, such as wood waste, grass and agricultural waste. However, lignocellulose has a complex structure, in which the cellulose is in the form of fibres that are locked within a rigid structure of hemicellulose and lignin. In order to access the cellulose, hemicellulose and lignin components, lignocellulosic biomass material is therefore typically subjected to an initial 'pre-treatment' process that damages or alters the lignocellulose structure to separate and release the individual lignocellulose components. The requirement for this pre-treatment process increase the costs associated with using lignocellulosic biomass material. There is therefore an increasing emphasis on improving the methods of extraction from lignocellulosic biomass materials.

Various pre-treatment methods are known in the art. Examples include thermochemical methods such as steam explosion, in which the lignocellulosic material is heated to 130°C - 230°C by steam injection (optionally including the addition of a catalyst or caustic agent to facilitate degradation); or wet oxidation, which involves exposure of the biomass material to oxygen at 150-185°C. Another conventional pre-treatment method is acid hydrolysis, in which the lignocellulosic material is subjected to an acid (such as sulphuric acid) that hydrolyses the cellulose and hemicellulose components to their monomeric sugars. Acid hydrolysis is also often performed at high temperatures to promote further hydrolysis of the biomass material. Organosolve pre-treatment methods also exist and involve the use of solvents together with acids or bases, and may be performed at high temperatures. These organosolve methods act by disrupting the interactions between the lignin, hemicellulose and cellulose and permit the lignin and hemicellulose components to be separated from cellulose in conjunction with the solvent.

Although these pre-treatment methods allow for efficient breakdown of lignocellulosic biomass material, they require expensive reaction vessels, and are energy intensive. The thermochemical and acid hydrolysis methods occur at high temperatures and extreme pH conditions which are often incompatible with downstream processes, such as the enzymatic hydrolysis reaction used to release sugars for production of biofuels.

The pre-treatment processes also result in the production of compounds or by-products that inhibit downstream processes. For example, these pre-treatment processes have been shown to result in the production of furfural from xylose and phenolic fragments from lignin, both of which inhibit the downstream fermentation process used to produce biogasoline. The organosolve pre-treatment methods are also hindered by drawbacks - in particular, because they require large volumes of solvent. The solvents used are not easily recyclable, and it has been shown that any solvent remaining in the extracted sugars can inhibit the downstream fermentation process. As a result, these pre-treatment methods are often performed separately to the downstream processes that utilise the cellulose, hemicellulose and lignin components of the lignocellulosic biomass material.

As an alternative to these harsh physical and chemical pre-treatment methods, biological based approaches have also been proposed, with the aim of providing a more efficient, eco-friendly and cost-effective way of pre-treating lignocellulosic biomass material. These biological pre-treatment methods include the use of intact micro-organisms that naturally degrade lignocellulosic biomass material. By way of example, filamentous fungi (such as the white rot fungi basidiomycete *P. chrysosporium)* have been identified as potentially suitable candidates for use in pre-treatment processes, because they produce enzymes that permit degradation of lignin, hemicellulose and cellulose, via oxidative and hydrolytic mechanisms (Dashtban *et al.* Fungal Bioconversion of Lignocellulosic Residues; Opportunities and Perspectives; Int. J. Biol. Sci; 2009; 5; p578-595; Liu *et al.* Development of Highly Efficient, Low-Cost Lignocellulolytic Enzyme Systems in the Post-Genomic Era; Biotechnology Advances; 2013; 31; 962-975). White rot fungi degrade lignin via an oxidative mechanism, in which the lignin is attacked by reactive oxygen species, such as hydroxyl free radicals. White rot fungi are also known to attack lignin using enzymes such as laccases, which catalyse free radical mediated lignin degradation. White rot fungi also produce hydrolytic enzymes (including those with endocellulase, exocellulase and cellobiohydrolase activities), which degrade glycosidic linkages in cellulose and hemicellulose to release monomeric sugars.

In light of these observations, several studies have been performed to demonstrate that intact filamentous fungi can be used to extract sugars from lignocellulosic material (Potumarthi *et al*. Simultaneous Pretreatment and Saccharification of Rice Husk By *Phanerochete Chrysosporium* for Improved Production of Reducing Sugars; Bioresour. Technol.; 2013; 128; 113-117; Dias *et al*. Enzymatic Saccharification of Biologically Pre-

treated Wheat Straw with White-Rot Fungi; Bioresourc. Technol.; 2010; 101; 6045-6050). Despite these findings, biological pre-treatment methods are yet to be applied in practice because numerous technical difficulties have been encountered. By way of example, it can be difficult to ensure that the lignocellulose degrading micro-organisms produce sufficiently high levels of the lignocellulosic targeting enzymes to permit efficient lignocellulose degradation. In addition, the micro-organisms used consume high levels of the cellulose and hemicellulose material leading to low output and increased costs. Operational costs are also further increased by the need to perform extraction processes under sterile conditions.

In summary, in order to extract sugars and/ or lignin derived products from lignocellulosic biomass materials, conventional methods typically require extensive pre-treatment using harsh physical and chemical processes. These methods are costly and environmentally damaging. Due to the harsh conditions used for pre-treatment methods, the downstream processing of cellulose, hemicellulose and/ or lignin components must be performed separately to the initial pre-treatment process. The need to perform multiple steps makes the overall method more costly, and reduces the economic feasibility of using lignocellulosic biomass material for production of biofuels and/ or high value chemical products.

Alternative biological delignification approaches that utilise whole micro-organisms have been contemplated, but are hampered by low efficiency and high costs.

There is therefore a need in the art for improved methods for extracting sugars and lignin derived products from lignocellulosic biomass material. In particular, there is a need to develop alternative low cost and environmentally friendly pre-treatment methods. There is also a need to reduce the number of processing steps required to extract sugars and lignin derived products from lignocellulosic biomass material.

SUMMARY OF INVENTION

The present invention presents a solution to the above problems by providing a method for extracting sugars and/or lignin-derived products from lignocellulosic biomass material, comprising: contacting the lignocellulosic biomass material with a composition comprising a polypeptide in the presence of a reducible substrate and an oxidising agent, wherein said polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof;

wherein said polypeptide is capable of extracting sugars and/or lignin-derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

The present invention also provides a method for extracting sugars from lignocellulosic biomass material, wherein the lignocellulosic biomass material has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure; the method comprising: contacting the pre-treated lignocellulosic biomass material with a composition comprising a polypeptide, wherein said polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars from the pre-treated lignocellulosic biomass material.

The 'lignocellulosic biomass material' used in the present invention includes plant and/or wood material, such as herbaceous material, softwood, hardwood, wood waste, sawdust, agricultural crop, plant residue, forestry residue, municipal solid waste, pulp or paper mill residue, waste paper, recycling paper, or construction debris. Examples of suitable wood material include, but are not limited to, spruce, pine, hemlock, fir, birch, aspen, maple, poplar, alder, salix, cottonwood, rubber tree, marantii, eucalyptus, sugi, and acase. Examples of suitable plant material include, but are not limited to aquatic plants such as kelp, algae, lily, and hyacinth; and fibrous plants such as grass (including switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus and Napier grass), straw (including rice straw, barley straw, cereal straw, sugarcane straw, wheat straw, canola straw, and oat straw), sugarcane bagasse, agricultural wastes, rice hulls, corn cobs, oat hulls, corn fiber, stover, soybean stover and corn stover.

The lignocellulosic biomass material obtained from these plant and/ or wood materials comprises lignin, hemicellulose, and cellulose. The term "lignin" used throughout the present specification broadly refers to a biopolymer that may be part of secondary cell walls in plants, such as complex highly cross-linked aromatic polymer that covalently links to hemicellulose. The term "hemicellulose" used throughout the present specification broadly refers to a branched sugar polymer composed mostly of pentoses, such as with a generally random amorphous structure and up to hundreds of thousands of pentose units. The term "cellulose" used throughout the present specification broadly refers to an organic compound with a formula $(C_6H_{10}O_5)_Z$ where z includes any suitable

integer. Cellulose may include a polysaccharide with a linear chain of several hundred to over ten thousand hexose units and a high degree of crystalline structure.

The lignocellulosic biomass material used in the present invention may have been subjected to a pre-treatment step such that lignin and/ or hemicellulose present in the lignocellulosic structure is at least partially depolymerised. Subjecting lignocellulosic biomass material to a pre-treatment step may result in the separation of the cellulose, hemicellulose and lignin components of the lignocellulosic biomass material so as to increase the surface area and/ or accessibility of the material.

A skilled person will be familiar with conventional 'pre-treatment' processes known in the art. For example, the lignocellulosic biomass material may be exposed to a hydrolysis process, an acidic process (pH 7 and below), a basic or alkali process (pH above 7), an enzymatic process, a solvent-based process, a thermo-mechanical process, a thermo-chemical process, a steam based process, including but not limited to steam explosion and hot water based treatment, and/ or a supercritical process. Acid processes may include concentrated and/ or dilute acid steps, such as with sulphuric acid, sulphurous acid, hydrochloric acid, phosphoric acid, and/ or organic acids. Basic processes may include caustic materials, such as ammonia, calcium hydroxide, calcium oxide, magnesium hydroxide, sodium hydroxide, potassium hydroxide, sodium carbonate, potassium carbonate, sodium bicarbonate, and/or potassium bicarbonate. Pre-treatment processes may also involve subjecting lignocellulosic biomass material to a mechanical de-sizing process which includes any method for reducing the particle size of biomass such as, but not limited to, processes involving grinding, milling or crushing. Other pre-treatment processes known to a skilled person in the art may also be used to prepare pre-treated lignocellulosic material for use in the present invention.

The present invention provides polypeptide based methods for extracting sugars and/or lignin derived products from lignocellulosic biomass material.

The term 'sugars' used throughout the present specification means any carbohydrate compound having the general formula $C_XH_{2X}O_X$, where x includes any suitable integer, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, and/ or the like. Sugars may be building blocks or components of more complex molecules, such as cellulose, which comprises hexose sugars; and hemicellulose, which comprises pentose sugars.

The term "lignin derived products" used throughout the present specification means any product obtained from lignin that can be used to produce other useful compounds or products, including high value chemicals and/ or biofuels, such as biogasoline, bioethanol or biodiesel. A skilled person will be familiar with the various types of lignin derived products that can be obtained from lignin, and used to produce other useful compounds or products.

Non-limiting examples of lignin derived products that can be used to produce other useful compounds or products include guaiacols (such as methylguaiacol, ethylguaiacol, vinylguaiacol and guaiacylacetone, eugenol and isoeugenol); syringols; phenols and phenolic aldehydes (such as vanillin, acetovanillone, veratraldehyde, acetoveratron, trimethoxy benzaldehyde, 4-hydroxy-benzaldehyde, acetanisole, acetosyringone and syringaldehyde). A skilled person will be familiar with other types of lignin derived products that can also be used to produce other useful compounds or products.

The polypeptides used in the methods of the present invention are derived from two iron reductase enzymes, IR1 (defined in Genbank accession number EGN95518.1) and IR2 (defined in Genbank accession number EGN95519.1). These enzymes are encoded by the genome of the brown rot fungus, *S. lacrymans*, which is a natural decomposer of lignocellulosic material.

The identification of IR1 and IR2 was previously reported following sequencing and analysis of the *S. lacrymans* genome (Eastwood *et al.* The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi; Science; 2011; 333; 762-765). These studies revealed that the IR1 enzyme (protein ID 452187) contained sequences corresponding to an iron reductase domain (also known as a 'heme' domain) and a cellulose binding module, whilst the IR2 enzyme (protein ID 417465) contained sequences corresponding to an iron reductase domain, but did not contain sequences corresponding to a cellulose binding module. Figure 1 shows a comparison of the structures of the IR1 and IR2 enzymes.

In light of the domains proposed to exist in IR1, IR1 was proposed to play a role in the process of lignocellulose breakdown by *S. lacrymans*. It has been suggested that brown rot fungi mediate lignocellulose breakdown using a non-enzymatic reaction in which hydroxyl free radicals are generated via the Fenton reaction ($Fe^{2+} + H_2O_2 + H^+ > Fe^{3+} + OH^- + H_20$) and used to depolymerise the lignocellulosic structure. Hydrogen peroxide (H_2O_2) for use in the Fenton reaction has been shown to be metabolically generated by

oxidase enzymes produced by brown rot fungi. Brown rot fungi have also been shown to produce metabolites (such as variegatic acid) that reduce Fe^{3+} to Fe^{2+} for use in the Fenton reaction. It was therefore proposed that *S. lacrymans* may initiate lignocellulose depolymerisation by producing metabolites such as variegatic acid. Following this initial lignocellulose depolymerisation, it was proposed that IR1 may localise to cellulose in the partially depolymerised lignocellulosic structure via its cellulose binding module, and promote further generation of Fe^{2+} using its iron reductase domain, leading to localised hydroxyl free radical production (via the Fenton reaction) and further depolymerisation of the lignocellulosic structure.

Following the initial genomic analysis, further research into the mechanism of lignocellulose breakdown by *S. lacrymans* demonstrated that, instead of variegatic acid, the fungal metabolite, 2,5-dimethoxyhydroquinone, may be required to initiate lignocellulose depolymerisation by the brown rot fungus *S. lacrymans (*Korripoly *et al.* Evidence from *Serpula Lacrymans* that 2,5 dimethoxyhydroquinone is a Lignocellulytic Agent of Divergent Brown Rot Basidiomycetes; Applied and environmental microbiology; 2013; 79; 7; 2377-2383). These findings therefore suggest that *S. lacrymans* may utilise multiple mechanisms for initiating breakdown of the lignocellulosic structure.

It has now been surprisingly found that both IR1 and IR2 can induce depolymerisation of the lignocellulosic structure without requiring any other fungal metabolites or fungal derived compounds (such as variegatic acid and 2,5 dimethoxyhydroquinone) to initiate depolymerisation of the lignocellulosic structure. In particular, a simplified method has been developed for targeting depolymerisation of the lignocellulosic structure using IR1 or IR2 polypeptides together with a reducible substrate (such as Fe^{3+} ions) and an oxidising agent (such as H_2O_2). Using this method, sugars and/or lignin derived products can be extracted from lignocellulosic biomass material, without the need to perform any pre-treatment process, and without the need to utilise intact *S. lacrymans* or any other *S. lacrymans* derived metabolites or compounds.

It has also been unexpectedly found that both IR1 and IR2 have the ability to directly depolymerise polysaccharides present in lignocellulosic biomass material (such as cellulose) to produce sugars. This depolymerisation activity has been shown to occur independently of the iron reductase activity shared by both IR1 and IR2, and does not require the presence of a cellulose binding domain, such as that present in IR1. The enzymatic activity of these enzymes also does not require the presence of any other fungal metabolites or fungal derived compounds. A simplified method has therefore

been developed for extracting sugars from lignocellulosic biomass material that avoids the need to add other cellulase enzymes, or utilise intact *S. lacrymans.*

In light of these findings, polypeptide based methods for extraction of sugars and/ or lignin derived products from lignocellulosic biomass material have been developed using polypeptides derived from IR1 or IR2.

In one embodiment, the polypeptide used in the methods of the present invention may comprise (or consist of) an amino acid sequence having at least 70% identity to the sequence of SEQ ID NO: 1 or 2, or a fragment of said sequence. In accordance with this embodiment, the amino acid sequence defined in SEQ ID NO: 1 corresponds to the polypeptide sequence of IR1 (Genbank accession number EGN95518.1) but does not include the IR1 signal peptide sequence. Likewise, the amino acid sequence defined in SEQ ID NO: 2 corresponds to the polypeptide sequence of IR2 (Genbank accession number EGN95519.1) but does not include the IR2 signal peptide sequence.

The term "polypeptide" used throughout the present specification is synonymous with the terms "oligopeptide", "peptide" and "protein". These terms are used interchangeably and do not refer to a specific length of the product. These terms embrace post-translational modifications such as glycosylation, acetylation and phosphorylation.

In one embodiment, the polypeptide may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence of SEQ ID NO: 1, or a fragment of said sequence. In one embodiment, the polypeptide may comprise (or consist of) the amino acid of SEQ ID NO: 1, or a fragment of said sequence.

In one embodiment, the polypeptide may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence of SEQ ID NO: 2, or a fragment of said sequence. In one embodiment, the polypeptide may comprise (or consist of) the amino acid sequence of SEQ ID NO: 2, or a fragment of said sequence.

In all embodiments of the present invention, a fragment of a polypeptide comprises (or consists of) a truncated form of the polypeptide. For example, a fragment of a polypeptide may comprise a series of consecutive amino acid residues from the sequence of the polypeptide. A fragment of a polypeptide may have an N-terminal truncation or a C-terminal truncation (as compared with the polypeptide).

In one embodiment, the polypeptide may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence of SEQ ID NO: 1, or a fragment of said sequence having at least 10 consecutive amino acids thereof (eg. least 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225 or 230 consecutive amino acids thereof).

In one embodiment, the polypeptide may comprise (or consist of) the amino acid sequence of SEQ ID NO: 1, or a fragment of said sequence having at least 10 consecutive amino acids thereof (eg. least 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225 or 230 consecutive amino acids thereof).

In accordance with these embodiments, the polypeptide fragment may comprise (or consist of) an amino acid sequence having at least 70% identity to the sequence of the IR1 iron reductase domain (defined as amino acid residues 4-175 of SEQ ID NO: 1), or a fragment thereof; and/ or an amino acid sequence having at least 70% identity to the sequence of the IR1 cellulose binding domain (defined as amino acid residues 202-229 of SEQ ID NO: 1), or a fragment thereof.

For example, the polypeptide fragment may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning amino acid residues 4 to 175 of SEQ ID NO: 1, or a fragment thereof having at least 50 consecutive amino acids thereof (eg. least 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, or 170 consecutive amino acids thereof); and/ or an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning amino acid residues 202 to 229 of SEQ ID NO: 1, or a fragment thereof having at least 15 consecutive amino acids thereof (eg. least 20 or 25 consecutive amino acids thereof).

In one embodiment, the polypeptide may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence of SEQ ID NO: 2, or

a fragment of said sequence having at least 10 consecutive amino acids thereof (eg. at least 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180 or 185 consecutive amino acids thereof).

In one embodiment, the polypeptide may comprise (or consist of) the amino acid sequence of SEQ ID NO: 2, or a fragment of said sequence having at least 10 consecutive amino acids thereof (eg. at least 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180 or 185 consecutive amino acids thereof).

In accordance with these embodiments, the polypeptide fragment may comprise (or consist of) an amino acid sequence having at least 70% identity to the sequence of the IR2 iron reductase domain (defined as amino acid residues 4-175 of SEQ ID NO: 2), or a fragment thereof.

For example, the polypeptide fragment may comprise (or consist of) an amino acid sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning amino acid residues 4 to 175 of SEQ ID NO: 2, or a fragment thereof having at least 50 consecutive amino acids thereof (eg. least 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, or 170 consecutive amino acids thereof).

In one embodiment, the polypeptides, polypeptide variants and polypeptide fragments for use in the methods of the present invention are capable of extracting sugars from lignocellulosic biomass material that has been previously treated so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulosic structure.

In one embodiment, the polypeptides, polypeptide variants and polypeptide fragments for use in the methods of the present invention are capable of extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

In one embodiment, the polypeptides, polypeptide variants and polypeptide fragments for use in the methods of the present invention have reductase activity and/ or polysaccharide cleavage activity. For example, the polypeptides, polypeptide variants

and polypeptide fragments for use in the methods of the present invention may have iron reductase activity and/ or cellulase and/ or hemicellulase activity.

Conventional methods can be used to determine amino acid sequence identity. There are many established algorithms available to align two amino acid sequences. Typically, one sequence acts as a reference sequence, to which test sequences may be compared. The sequence comparison algorithm calculates the percentage sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alignment of amino acid sequences for comparison may be conducted, for example, by computer implemented algorithms (eg. GAP, BESTFIT, FASTA or TFASTA), or BLAST and BLAST 2.0 algorithms. The BLOSUM62 table shown below is an amino acid substitution matrix derived from about 2,000 local multiple alignments of protein sequence segments, representing highly conserved regions of more than 500 groups of related proteins (Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-10919, 1992). Amino acids are indicated by the standard one-letter codes. The percent identity is calculated as:

$$\frac{\text{Total number of identical matches}}{\text{[length of the longer sequence plus the number of gaps Introduced into the longer sequence in order to align the two sequences]}} \times 100$$

BLOSUM62 table

```
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V
A  4
R -1  5
N -2  0  6
D -2 -2  1  6
C  0 -3 -3 -3  9
Q -1  1  0  0 -3  5
E -1  0  0  2 -4  2  5
G  0 -2  0 -1 -3 -2 -2  6
H -2  0  1 -1 -3  0  0 -2  8
I -1 -3 -3 -3 -1 -3 -3 -4 -3  4
L -1 -2 -3 -4 -1 -2 -3 -4 -3  2  4
K -1  2  0 -1 -3  1  1 -2 -1 -3 -2  5
M -1 -1 -2 -3 -1  0 -2 -3 -2  1  2 -1  5
F -2 -3 -3 -3 -2 -3 -3 -3 -1  0  0 -3  0  6
P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4  7
S  1 -1  1  0 -1  0  0  0 -1 -2 -2  0 -1 -2 -1  4
T  0 -1  0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1  1  5
W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1  1 -4 -3 -2 11
Y -2 -2 -2 -3 -2 -1 -2 -3  2 -1 -1 -2 -1  3 -3 -2 -2  2  7
V  0 -3 -3 -3 -1 -2 -2 -3 -3  3  1 -2  1 -1 -2 -2  0 -3 -1  4
```

In a sequence identity comparison, the identity may exist over a region of the sequences that is at least 10 amino acid residues in length (eg. at least 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225 or 230 amino acid residues in length – eg. up to the entire length of the reference sequence).

Polypeptides used in the methods of the present invention may have one or more amino acid substitutions, deletions, or additions. In many embodiments, those changes are of a minor nature, for example, involving only conservative amino acid substitutions. Conservative substitutions are those made by replacing one amino acid with another amino acid within the following groups: Basic: arginine, lysine, histidine; Acidic: glutamic

acid, aspartic acid; Polar: glutamine, asparagine; Hydrophobic: leucine, isoleucine, valine; Aromatic: phenylalanine, tryptophan, tyrosine; Small: glycine, alanine, serine, threonine, methionine. Polypeptides used in the methods of the present invention may also encompass those comprising other substitutions that do not significantly affect the folding or activity of the polypeptide; small deletions, typically of 1 to about 10 amino acids (such as 1-5 amino acids); and small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues, or an affinity tag. The polypeptide may also comprise non-naturally occurring amino acid residues.

Essential amino acids in the polypeptide can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis. Sites of biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. The identities of essential amino acids can also be inferred from analysis of homologies with related family members of the polypeptide of interest.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening. Methods are known for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display.

Routine deletion analyses of nucleic acid molecules can be performed to obtain functional fragments of a nucleic acid molecule that encodes a polypeptide. As an illustration, DNA molecules can be digested with Bal31 nuclease to obtain a series of nested deletions. These DNA fragments are then inserted into expression vectors in proper reading frame, and the expressed polypeptides are isolated and tested for the desired activity. An alternative to exonuclease digestion is to use oligonucleotide-directed mutagenesis to introduce deletions, or stop codons to specify production of a desired fragment. Alternatively, particular polynucleotide fragments can be synthesized using the polymerase chain reaction.

The polynucleotide sequence of SEQ ID NO: 3 encodes the IR1 derived polypeptide defined in SEQ ID NO: 1; and the polynucleotide sequence of SEQ ID NO: 4 encodes the IR2 derived polypeptide defined in SEQ ID NO: 2.

In one embodiment of the present invention, the polypeptide is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity to the sequence of SEQ ID NO: 3 or 4, or a fragment of said sequence.

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity) to the sequence of SEQ ID NO: 3, or a fragment of said sequence. In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) the sequence of SEQ ID NO: 3, or a fragment of said sequence.

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity) to the sequence of SEQ ID NO: 4, or a fragment of said sequence. In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) the sequence of SEQ ID NO: 4, or a fragment of said sequence.

In all embodiments of the present invention, a fragment of a polynucleotide comprises (or consists of) a truncated form of the polynucleotide. For example, a “fragment” of a polynucleotide may comprise a series of consecutive nucleotides from the sequence of the polynucleotide. For example, a fragment of a polynucleotide may encode a polypeptide having an N-terminal truncation or a C-terminal truncation (as compared with the reference polypeptide).

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity) to the sequence of SEQ ID NO: 3, or a fragment of said sequence having at least 30 consecutive nucleotides thereof (eg. at least 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, 510, 525, 540, 555, 570, 585, 600, 615, 630, 645, 660, 675 or 690 consecutive nucleotides thereof).

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) the sequence of SEQ ID NO: 3, or a fragment of said sequence having at least 30 consecutive nucleotides thereof (eg. at least 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, 510, 525, 540, 555, 570, 585, 600, 615, 630, 645, 660, 675 or 690 consecutive nucleotides thereof).

In accordance with these embodiments, the polynucleotide fragment may comprise (or consist of) a nucleotide sequence having at least 70% identity to the nucleotide sequence encoding the IR1 iron reductase domain (corresponding to nucleotides 10-525 of SEQ ID NO: 3), or a fragment thereof; and/ or a nucleotide sequence having at least 70% identity to the sequence encoding the IR1 cellulose binding domain (corresponding to nucleotides 604-687 of SEQ ID NO: 3), or a fragment thereof.

For example, the polynucleotide fragment may comprise (or consist of) a nucleotide sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning nucleotides 10-525 of SEQ ID NO: 3, or a fragment thereof having at least 150 consecutive nucleotides thereof (eg. at least 165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, or 510 consecutive nucleotides thereof); and/ or a nucleotide sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning nucleotides 604-687 of SEQ ID NO: 3, or a fragment thereof having at least 45 consecutive nucleotides thereof (eg. at least 60 or 75 consecutive nucleotides thereof).

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity) to the sequence of SEQ ID NO: 4, or a fragment of said sequence having at least 30 consecutive nucleotides thereof (eg. at least 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, 510, 525, 540 or 555 consecutive nucleotides thereof).

In one embodiment, the polypeptide is encoded by a polynucleotide that comprises (or consists of) the sequence of SEQ ID NO: 4, or a fragment of said sequence having at least 30 consecutive nucleotides thereof (eg. at least 45, 60, 75, 90, 105, 120, 135, 150,

165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, 510, 525, 540 or 555 consecutive nucleotides thereof).

In accordance with these embodiments, the polynucleotide fragment may comprise (or consist of) a nucleotide sequence having at least 70% identity to the nucleotide sequence encoding the IR2 iron reductase domain (corresponding to nucleotides 10-525 of SEQ ID NO: 4), or a fragment thereof.

For example, the polynucleotide fragment may comprise (or consist of) a nucleotide sequence having at least 70% identity (eg. at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity) to the sequence spanning nucleotides 10-525 of SEQ ID NO: 4, or a fragment thereof having at least 150 consecutive nucleotides thereof (eg. at least 165, 180, 195, 210, 225, 240, 255, 270, 285, 300, 315, 330, 345, 360, 375, 390, 405, 420, 435, 450, 465, 480, 495, or 510 consecutive nucleotides thereof).

In one embodiment, the polynucleotides, polynucleotide variants and polynucleotide fragments for use in the methods of the present invention encode polypeptides that are capable of extracting sugars from lignocellulosic biomass material that has been previously pre-treated so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulosic structure.

In one embodiment, the polynucleotides, polynucleotide variants and polynucleotide fragments for use in the methods of the present invention encode polypeptides that are capable of extracting sugars and lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

In one embodiment, the polynucleotides, polynucleotide variants and polynucleotide fragments for use in the methods of the present invention encode polypeptides that have reductase activity and/ or polysaccharide cleavage activity. For example, the polynucleotides, polynucleotide variants and polynucleotide fragments for use in the methods of the present invention encode polypeptides that have iron reductase activity and/ or cellulase and/ or hemicellulase activity.

The term “polynucleotide sequence” includes nucleic acid sequences that have been removed from their naturally occurring environment, recombinant or cloned DNA isolates, and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

The polynucleotides may be prepared by any means known in the art. For example, large amounts of the polynucleotides may be produced by replication in a suitable host cell. The natural or synthetic DNA fragments coding for a desired fragment will be incorporated into recombinant nucleic acid constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the DNA constructs will be suitable for autonomous replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to and integration within the genome of a cultured insect, mammalian, plant or other eukaryotic cell lines.

The polynucleotides may also be produced by chemical synthesis, eg. by the phosphoramidite method or the triester method, and may be performed on commercial automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

The term “recombinant” as used herein intends a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; or (2) is linked to a polynucleotide other than that to which it is linked in nature; and (3) does not occur in nature. This artificial combination is often accomplished by via conventional chemical synthesis techniques, or by the artificial manipulation of isolated segments of nucleic acids – e.g., by conventional genetic engineering techniques.

Conventional methods may be used to determine nucleotide sequence identity in the present invention. In view of the degeneracy of the genetic code, considerable sequence variation is possible among the polynucleotides encoding the at least polypeptide. Degenerate codons encompassing all possible codons for a given amino acid are set forth below:

Amino Acid	Codons	Degenerate Codon
Cys	TGC TGT	TGY
Ser	AGC AGT TCA TCC TCG TCT	WSN
Thr	ACA ACC ACG ACT	ACN
Pro	CCA CCC CCG CCT	CCN
Ala	GCA GCC GCG GCT	GCN
Gly	GGA GGC GGG GGT	GGN
Asn	AAC AAT	AAY
Asp	GAC GAT	GAY
Glu	GAA GAG	GAR
Gln	CAA CAG	CAR
His	CAC CAT	CAY
Arg	AGA AGG CGA CGC CGG CGT	MGN
Lys	AAA AAG	AAR
Met	ATG	ATG
Ile	ATA ATC ATT	ATH
Leu	CTA CTC CTG CTT TTA TTG	YTN
Val	GTA GTC GTG GTT	GTN
Phe	TTC TTT	TTY
Tyr	TAC TAT	TAY
Trp	TGG	TGG
Ter	TAA TAG TGA	TRR
Asn/ Asp		RAY
Glu/ Gln		SAR
Any		NNN

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequences of the polypeptide described herein.

Those of skill in the art will also recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the disclosure. The native DNA

sequence encoding the polypeptides described are referenced herein merely to illustrate an embodiment of the disclosure, and the disclosure includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides utilised in the methods of the disclosure. The invention contemplates and provides each and every possible variation of nucleic acid sequence encoding the polypeptides could be made by selecting combinations based on possible codon choices.

A nucleic acid sequence or fragment thereof is "identical" to a reference sequence if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 70%, 75%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 97%, 98%, or 99% of the nucleotide bases. Identity determination is performed as described *supra* for polypeptides.

Alternatively, a nucleic acid sequence or fragment thereof is "identical" to a reference sequence if it is capable of hybridizing under stringent (eg. highly stringent) hybridization conditions. Nucleic acid sequence hybridization will be affected by such conditions as salt concentration (e.g. NaCl), temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions are preferably employed, and generally include temperatures in excess of 30°C, typically in excess of 37°C and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. The pH is typically between 7.0 and 8.3. The combination of parameters is much more important than any single parameter.

One of ordinary skill in the art appreciates that different species exhibit "preferential codon usage". As used herein, the term "preferential codon usage" refers to codons that are most frequently used in cells of a certain species, thus favouring one or a few representatives of the possible codons encoding each amino acid. For example, the amino acid threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian host cells ACC is the most commonly used codon; in other species, different Thr codons may be preferential. Preferential codons for a particular host cell species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into

recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species.

Thus, in one embodiment of the invention, the nucleic acid sequence is codon optimized for expression in a host cell.

The polypeptide used in the present invention may be produced using a recombinant micro-organism expressing the polypeptide, or using other recombinant cell-types that express the polypeptide described herein (eg. because they are transformed with recombinant nucleic acids encoding the polypeptide).

Expression of the polypeptide from the recombinant micro-organism or cell-type is via a chimeric gene that is operably linked to regulatory sequences and encodes the polypeptide defined herein. Polynucleotides encoding the polypeptide can be prepared using methods that are well known in the art, and described herein. Expression vectors containing the chimeric gene may be used to transform an appropriate host cell. A skilled person will be familiar with methods for recombinant expression of proteins in host cells, and any number of expression vectors are available or can be constructed using routine methods.

Any vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used. For example, vectors may include a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), or a yeast artificial chromosome (YAC). For example, suitable vectors may include derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses and many others.

Vectors may further comprise regulatory sequences, including, for example, a promoter, operably linked to the protein encoding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art. The construct may optionally include nucleotide sequences to facilitate integration into a host genome and/or results in amplification of construct copy number in vivo.

To obtain high levels of expression in a particular host it is often useful to express polypeptides under control of a heterologous promoter. Typically a promoter sequence may be operably linked to the 5' region of the polypeptide coding sequence. It will be

recognized that in making such a construct it is not necessary to define the bounds of a minimal promoter. Instead, the DNA sequence 5' to the polypeptide gene start codon can be replaced with DNA sequence that is 5 ' to the start codon of a given heterologous gene. This 5' "heterologous" sequence thus includes, in addition to the promoter elements per se, a transcription start signal and the sequence of the 5' untranslated portion of the transcribed chimeric mRNA. Thus, the promoter-gene construct and resulting mRNA may comprise a sequence encoding the polypeptide described herein and a heterologous 5' sequence upstream to the start codon of the sequence encoding the polypeptide. In some, but not all, cases the heterologous 5' sequence will immediately abut the start codon of the polynucleotide sequence encoding the polypeptide. In some embodiments, gene constructs may be employed in which a polynucleotide encoding the polypeptide is present in multiple copies. Such embodiments, may employ the endogenous promoter for the IR1 or IR2 genes or may employ a heterologous promoter.

In one embodiment, the polypeptide may be expressed without a signal peptide. For example, the polypeptide may be derived from the sequence defined in SEQ ID NO: 1, which corresponds to the amino acid sequence of the IR1 polypeptide without its signal peptide; or the polypeptide may be derived from the sequence defined in SEQ ID NO: 2, which corresponds to the amino acid sequence of the IR2 polypeptide without its signal peptide.

In one embodiment, the polypeptide may be expressed as a pre-protein including the naturally occurring signal peptide of the IR1 or IR2 polypeptides. In accordance with this embodiment, the polypeptide may be expressed as a pre-protein with the signal peptide of the IR1 polypeptide having the amino acid sequence 'MFSHLLTTIILSIGFRAVTVVAQS' (defined in SEQ ID NO: 5). In accordance with this embodiment, the polypeptide may be expressed as a pre-protein with the signal peptide of the IR2 polypeptide having the amino acid sequence 'MFQKLLVASLLFLGIQFVNAVPN' (defined in SEQ ID NO: 6).

In one embodiment, the polypeptide may be expressed as a pre-protein with a heterologous signal peptide.

In one embodiment, the polypeptide may be produced using recombinant micro-organisms using techniques, such as using yeast cells and filamentous fungal cells;

algal cells; and prokaryotic cells, including gram positive, gram negative and gram-variable bacterial cells.

In one embodiment, the recombinant micro-organism may be a fungal host cell including, but are not limited to, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. In one embodiment, fungal host cells may be yeast cells and filamentous fungal cells. Filamentous fungal host cells may include all filamentous forms of the subdivision Eumycotina and Oomycota. For example, the filamentous fungal host cell may be a cell of a species of, but not limited to Achlya, Acremonium, Aspergillus, Aureobasidium, Bjerkandera, Ceriporiopsis, Cephalosporium, Chrysosporium, Cochliobolus, Corynascus, Cryphonectria, Cryptococcus, Coprinus, Coriolus, Diplodia, Endothia, Fusarium, Gibberella, Gliocladium, Humicola, Hypocrea, Myceliophthora, Mucor, Neurospora, Penicillium, Podospora, Phlebia, Piromyces, Pyricularia, Rhizomucor, Rhizopus, Schizophyllum, Scytalidium, Sporotrichum, Talaromyces, Thermoascus, Thielavia, Trametes, Tolypocladium, Trichoderma, Verticillium, Volvariella, or teleomorphs, or anamorphs, and synonyms or taxonomic equivalents thereof.

In one embodiment, the recombinant micro-organism may be a yeast host cell, including a cell of a species of, but not limited to Candida, Hansenula, Saccharomyces, Schizosaccharomyces, Pichia, Kluyveromyces, and Yarrowia. For example, the yeast cell may be Hansenula polymorpha, Saccharomyces cerevisiae, Saccaromyces carlsbergensis, Saccharomyces diastaticus, Saccharomyces norbensis, Saccharomyces kluyveri, Schizosaccharomyces pombe, Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia kodamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia quercuum, Pichia pijperi, Pichia stipitis, Pichia methanolica, Pichia angusta, Kluyveromyces lactis, Candida albicans, and Yarrowia lipolytica.

In one embodiment, the recombinant micro-organism may be an algal host cell such as, Chlamydomonas (e.g., C. reinhardtii) and Phormidium (P. sp. ATCC29409).

In one embodiment, the recombinant micro-organism may be a prokaryotic cell. Suitable prokaryotic cells include gram positive, gram negative and gram-variable bacterial cells. The host cell may be a species of, but not limited to, Agro bacterium, Alicyclobacillus, Anabaena, Anacystis, Acinetobacter, Acidothermus, Arthrobacter, Azobacter, Bacillus,

Bifidobacterium, Brevibacterium, Butyrivibrio, Buchnera, Campestris, Camplyobacter, Clostridium, Corynebacterium, Chromatium, Coprococcus, Escherichia, Enterococcus, Enterobacter, Erwinia, Fusobacterium, Faecalibacterium, Francisella, Flavobacterium, Geobacillus, Haemophilus, Helicobacter, Klebsiella, Lactobacillus, Lactococcus, Ilyobacter, Micrococcus, Microbacterium, Mesorhizobium, Methylobacterium, Mycobacterium, Neisseria, Pantoea, Pseudomonas, Prochlorococcus, Rhodobacter, Rhodopseudomonas, Rhodopseudomonas, Roseburia, Rhodospirillum, Rhodococcus, Scenedesmus, Streptomyces, Streptococcus, Synechococcus, Saccharomonospora, Staphylococcus, Serratia, Salmonella, Shigella, Thermoanaerobacterium, Tropheryma, Tularensis, Temecula, Thermosynechococcus, Thermococcus, Ureaplasma, Xanthomonas, Xylella, Yersinia and Zymomonas.

In another embodiment, the bacterial host strain may be non-pathogenic to humans. In some embodiments the bacterial host strain may be an industrial strain. Numerous bacterial industrial strains are known and suitable in the present invention.

Strains that may be used in the practice of the invention including both prokaryotic and eukaryotic strains, are readily accessible to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

The methods of the present invention utilise a composition that comprises the polypeptide, as defined herein.

The term “composition” used throughout the present specification is intended to designate a reagent for use in the methods of the present invention which comprises the polypeptide isolated from its natural wild type *S. lacrymans* host (ie. when the polypeptide is not present in its natural wild type *S. lacrymans* host). In this way, the present invention encompasses only compositions that comprise the polypeptide when it is isolated from its natural wild type host (ie. when the polypeptide is not present in its natural wild type *S. lacrymans* host).

In one embodiment, the composition used in the methods of the present invention may comprise (or consist of) the polypeptide in the form of an isolated polypeptide.

In one embodiment, the composition used in the methods of the present invention may comprise a crude, semi-purified, or purified preparation of polypeptide produced using a recombinant micro-organism or other recombinant cell-type, as described herein.

For example, the polypeptide produced using a recombinant micro-organism may be provided in a crude cell mass fermentation broth, which may be treated to prevent further microbial growth (for example, by heating or addition of antimicrobial agents).

For example, the polypeptide produced using other recombinant cell-types may be provided in cell culture media or in a cell lysate obtained from the recombinant cells expressing the polypeptide. For example, after producing the polypeptide by culturing a host cell transformed with a polynucleotide or vector encoding the polypeptide, the polypeptide need not be isolated from the culture medium (i.e., if the polypeptide is secreted into the culture medium) or cell lysate (i.e., if the polypeptide is not secreted into the culture medium) or used in a purified form to be useful. Any composition, cell culture medium, or cell lysate containing the polypeptide may be suitable for use in the methods of the present invention.

For example, the expressed polypeptide may be purified by, for instance, a combination of hydrophobic interaction chromatography, ion exchange chromatography and ceramic hydroxyl apatite chromatography. Other chromatographic techniques well known to the art of protein purification, such size exclusion chromatography, may be used. Polypeptide purity or homogeneity may be indicated by, for example, polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel, or using HPLC.

In one embodiment, the isolated polypeptide is substantially free from other proteins with which it is co-produced as well as from other contaminants. For instance, an isolated polypeptide is substantially free of material or other proteins from the cell, bacterial, or tissue source from which it was derived.

As used herein, a “purified” molecule is substantially free of its original environment and is sufficiently pure for use in pharmaceutical compositions. A substantially pure polypeptide, as used herein, refers to a polypeptide that is at least about 50% (w/w) pure; or at least about 60%, 70%, 80%, 85%, 90% or 95% (w/w) pure; or at least about 95%, 96%, 97%, 98%, 99%, or 100% (w/w) pure.

In an alternative embodiment, the composition used in the methods of the present invention may comprise (or consist of) a recombinant micro-organism that expresses the polypeptide.

For example, the polypeptide may be recombinantly expressed using one or more of the recombinant micro-organisms described herein, including one or more of the specific yeast cells and filamentous fungal cells; algal cells; and prokaryotic cells, including gram positive, gram negative and gram-variable bacterial cells described herein.

In one embodiment, the composition used in the methods of the present invention may comprise (or consist of) a recombinant micro-organism that expresses the polypeptide in combination with one or more additional polypeptides that facilitate extraction of sugars and lignin derived products from the lignocellulosic biomass material.

For example, the recombinant micro-organism may be modified so as to express one or more enzymes that degrade the cellulose, lignin and hemicellulose components of the lignocellulosic structure. For example, the one or more enzymes may be polysaccharases including cellulases, beta-glucanases, xylanases, pectinases, alpha glucuronidases, alpha-L-arabinefuranosidases, alpha amylases, beta-amylases, glucoamylases, pullulanases, beta-glucanases, hemicellulases, arabinosidases, mannanases, pectin hydrolyases, polygalacturonases, exopolygalaturonases and/ or pectate lyases. For example, the one or more enzymes may facilitate depolymerisation of lignin, including oxidases, peroxidases, and laccases. The one or more enzymes may also encode esterases for degradation of hemicellulose.

In one embodiment, the composition used in the methods of the present invention may further comprise one or more chelating agents. A skilled person will be aware of suitable chelating agents available in the art. For example, the composition may further comprise 2,3 dihydroxybenzoic acid (2,3-DHBA).

In one embodiment, the composition used in the methods of the present invention may further comprise 2,3 dihydroxybenzoic acid (2,3-DHBA) at a concentration from about 10μM to about 100μM. For example, the composition used in the methods of the present invention may further comprise 2,3 dihydroxybenzoic acid (2,3-DHBA) at a concentration of at least 10μM, at least 20μM, at least 30μM, at least 40μM, at least 50μM, at least 60μM, at least 70μM , at least 80μM, at least 90μM or at least 100μM. For example, the composition used in the methods of the present invention may further comprise 2,3

dihydroxybenzoic acid (2,3-DHBA) at a concentration of up to 10μM, up to 20μM, up to 30μM, up to 40μM, up to 50μM, up to 60μM, up to 70μM , up to 80μM, up to 90μM or up to 100μM. In one embodiment, the composition used in the methods of the present invention may further comprise 2,3 dihydroxybenzoic acid (2,3-DHBA) at a concentration of about 50μM.

The addition of one or more chelating agents to the compositions used in the methods of the present invention further improves the efficiency of the sugar and/ or lignin derived product extraction methods.

The present invention further provides a recombinant micro-organism that expresses a polypeptide comprising an amino acid sequence that has at least 70% identity (eg. up to 100% identity) to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent; and wherein said polypeptide is capable of extracting sugars from lignocellulosic biomass material that has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure.

In one embodiment, the recombinant micro-organism expresses a polypeptide that is encoded by a polynucleotide comprising a nucleic acid sequence having at least 70% identity (eg. up to 100% identity) to the sequence of SEQ ID NO: 3 or 4, or a fragment thereof, wherein said polypeptide is capable of extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent; and wherein said polypeptide is capable of extracting sugars from lignocellulosic biomass material that has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure.

All embodiments of the polypeptides, polypeptide variants and polypeptide fragments, and polynucleotides, polynucleotide variants and polynucleotide fragments, as described herein with respect to the methods of the present invention, apply equally to said recombinant micro-organism of the present invention.

All embodiments of the recombinant micro-organisms as described herein with respect to the methods of the present invention apply equally to said recombinant micro-organism of the present invention.

The present invention provides a method for extracting lignin derived products from lignocellulosic biomass material, comprising contacting the lignocellulosic biomass material with a composition comprising a polypeptide, as defined herein, in the presence of a reducible substrate and an oxidising agent, wherein said polypeptide is capable of extracting lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

The phrase “extracting lignin derived products” refers to the process by which the polypeptide induces the depolymerisation of lignocellulosic biomass material to produce lignin derived products.

The term “reducible substrate” used throughout the present specification means any substrate that can be reduced by the iron reductase domain of the IR1 or IR2 polypeptide. By ‘reduction’, we mean the process whereby electrons are donated to the reducible substrate.

For example, the reducible substrate may comprise Fe^{3+} ions, such as those present in $FeCl_3$. Fe^{3+} ions may be reduced using the iron reductase domain of IR1 or IR2 to generate Fe^{2+} ions.

The term “oxidising agent” used throughout the present specification means any compound that is capable of accepting electrons. For example, the oxidising agent may comprise hydrogen peroxide.

In accordance with this embodiment, the polypeptide has reductase activity and permits electrons to be donated to a reducible substrate, which in turn promotes the production of free radicals in the presence of an oxidising agent. By promoting the production of free radicals, the polypeptide is able to induce depolymerisation of the lignocellulosic structure to produce lignin derived products.

For example, the polypeptide may have iron reductase activity and may act by reducing Fe^{3+} ions to generate Fe^{2+} ions. In the presence of an oxidising agent, Fe^{2+} ions promote the generation of hydroxyl free radicals (OH^-) via the Fenton Reaction: $Fe^{2+} + H_2O_2 + H^+ \rightarrow Fe^{3+} + OH^- + H_20$.

In accordance with this embodiment, the lignocellulosic biomass material for use in these methods may or may not have been subjected to a pre-treatment process so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulosic structure. This is because the polypeptide used in the method has reductase activity and is able to promote the production of free radicals that induce depolymerisation of the lignocellulosic biomass material. There is therefore no need to subject the lignocellulosic biomass material to a pre-treatment process (such as the conventional thermo-chemical methods described herein) prior to performing the method of the present invention. The lignocellulosic biomass material for use in these methods may therefore comprise pre-treated or untreated lignocellulosic biomass material.

In accordance with this embodiment, the present method for extracting lignin derived products from lignocellulosic biomass material is performed using a polypeptide that is capable of extracting lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent. In accordance with this embodiment, the method does not require the presence of any other *S. lacrymans* metabolites or *S. lacrymans* derived compounds. In accordance with this embodiment, the method does not require the presence of intact *S. lacrymans.*

The present invention also provides methods for extracting sugars from lignocellulosic biomass material.

The phrase “extracting sugars” refers to the process by which the polypeptide induces the depolymerisation of lignocellulosic biomass material to produce sugars.

In one embodiment, the present invention provides a method for extracting sugars from lignocellulosic biomass material, wherein the lignocellulosic biomass material has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure; the method comprising: contacting the pre-treated lignocellulosic biomass material with a composition comprising a polypeptide, as defined herein, that is capable of extracting sugars from the pre-treated lignocellulosic biomass material.

For example, the polypeptide used in the method of the present invention may have ‘cellulase’ activity that cleaves the glycosidic bonds present in cellulose to produce sugars, such as hexose sugars. Hexose sugars include six carbon member sugars or saccharides (monomers), corresponding dissacharides (dimers), corresponding

trisaccharides (trimmers), corresponding tetrasaccharides (tetramers) and/or the like. Hexose includes glucose, galactose, sucrose, fructose, allose, altrose, gulose, idose, mannose, sorbose, talose, tagatose, any other isomer of six carbon sugars, and/ or the like.

For example, the polypeptide used in the method of the present invention may have 'hemicellulase' activity that cleaves the glycosidic bonds present in hemicellulose to produce sugars, such as pentose sugars. Pentose sugars include five carbon member sugars or saccharides (monomers), corresponding dissacharides (dimers), corresponding trisaccharides (trimers), corresponding tetrasaccharides (tetramers) and/or the like. Pentose includes xylose, ribose, arabinose, ribulose, xylulose, lyxose, any other isomer of five carbon sugars, and/ or the like.

In accordance with this embodiment, the polypeptide may be capable of extracting sugars from the cellulose and/or hemicellulose components from the pre-treated lignocellulosic biomass material without requiring the presence of a reducible substrate and/or an oxidising agent.

For example, the method of the present invention for extracting sugars from pre-treated lignocellulosic biomass may be performed (and operates) substantially in the absence of a reducible substrate and/ or an oxidising agent. The lignocellulosic biomass material for use in this embodiment must have been subjected to a pre-treatment process so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulosic structure.

The phrase "operates substantially in the absence of a reducible substrate and/ or an oxidising agent" means that the reaction is performed in the presence of less than about 1% of a reducible substrate and/ or an oxidising agent (eg. less than 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1% or 0.01% of a reducible substrate and/ or an oxidising agent). In one embodiment, the reaction may be performed in the presence of up to about 1% of a reducible substrate and/ or an oxidising agent (eg. up to 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1% or 0.01% of a reducible substrate and/ or an oxidising agent).

The method of the present invention for extracting sugars from pre-treated lignocellulosic biomass may also be performed (and operates) entirely in the absence of a reducible substrate and/ or an oxidising agent. The lignocellulosic biomass material for use in this

embodiment must have been subjected to a pre-treatment process so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulosic structure.

In accordance with this embodiment, the present method for extracting sugars from pre-treated lignocellulosic biomass material is performed using a polypeptide that is capable of extracting sugars from pre-treated lignocellulosic biomass material. In accordance with this embodiment, the method does not require the presence of any other *S. lacrymans* metabolites or *S. lacrymans* derived compounds. In accordance with this embodiment, the method does not require the presence of intact *S. lacrymans.*

In an alternative embodiment, the present invention provides a method for extracting sugars from lignocellulosic biomass material, comprising contacting the lignocellulosic biomass material with a composition comprising a polypeptide, as defined herein, in the presence of a reducible substrate and an oxidising agent, wherein said polypeptide is capable of extracting sugars from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

In accordance with this embodiment, the polypeptide used in the methods of the present invention acts as a reductase to add electrons to a reducible substrate, which in turn promotes the production of free radicals in the presence of an oxidising agent. By promoting the production of free radicals, the polypeptide induces depolymerisation of the lignocellulosic biomass material to produce sugars.

For example, the polypeptide may have iron reductase activity and may act by reducing Fe^{3+} ions to generate Fe^{2+} ions. In the presence of an oxidising agent, Fe^{2+} ions promote the generation of hydroxyl free radicals (OH^-) via the Fenton Reaction: $Fe^{2+} + H_2O_2 + H^+ \rightarrow Fe^{3+} + OH^- + H_2O$.

In one embodiment, the polypeptide not only has reductase activity, but may additionally have 'cellulase' and/ or 'hemicellulase' activity to produce sugars from the cellulose and/ or hemicellulose components of lignocellulosic biomass material. For example, the polypeptide may additionally have 'cellulase' activity that cleaves the glycosidic bonds present in cellulose to produce sugars, such as hexose sugars. For example, the polypeptide may additionally have 'hemicellulase' activity that cleaves the glycosidic bonds present in hemicellulose to produce sugars, such as pentose sugars.

The lignocellulosic biomass material for use in this embodiment may or may not have been subjected to a pre-treatment process so as to at least partially depolymerise lignin

and/ or hemicellulose present in the lignocellulosic structure. This is because the polypeptide used in the method has reductase activity and is able to promote the production of free radicals that induce depolymerisation of the lignocellulosic biomass material. There is therefore no need to subject the lignocellulosic biomass material to a pre-treatment process (such as the conventional thermo-chemical methods described herein) prior to performing the method of the present invention. The lignocellulosic biomass material for use in these methods may therefore comprise pre-treated or untreated lignocellulosic biomass material.

In accordance with this embodiment, the present method for extracting sugars from lignocellulosic biomass material is performed using a polypeptide that is capable of extracting sugars from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent. In accordance with this embodiment, the method does not require the presence of any other *S. lacrymans* metabolites or *S. lacrymans* derived compounds. In accordance with this embodiment, the method does not require the presence of intact *S. lacrymans.*

An advantage of the methods of the present invention which relate to the extraction of sugars and/ or lignin derived products in the presence of a reducible substrate and an oxidising agent is that it is not necessary to "pre-treat" the lignocellulosic biomass material so as to at least partially depolymerise lignin and/ or hemicellulose present in the lignocellulose structure. By avoiding the need to subject the lignocellulosic biomass material to a pre-treatment process, the methods of the present invention overcome the disadvantages associated with performing conventional pre-treatment processes. For example, conventional pre-treatment processes often result in the accumulation of 'inhibitory' compounds that must be removed from the reaction mix before downstream processes can take place (such as the hydrolysis and fermentation processes required to produce biofuels). By comparison, the methods of the present invention permit lignocellulosic biomass material to be depolymerised without accumulating these 'inhibitory' by-products or breakdown products. Hence, in one embodiment, the method of the present invention may operate as a single step method.

The term "single step method" means that the entire method for extracting sugars and/ or lignin derived products from lignocellulosic biomass material takes place without requiring any intermediate purification or "clean-up" steps to be performed. This has significant advantages in improving the economy of the methods for extracting sugars and/ or lignin derived products from lignocellulosic biomass materials.

The methods of the present invention may operate in the presence of a reducible substrate that can be reduced by the iron reductase domain of IR1 or IR2. For example, the reducible substrate may comprise Fe^{3+} ions. In one embodiment, the reducible substrate may comprise $FeCl_3$. Alternative reducible substrates will also work in the methods of the present invention and will be familiar to a person of skill in the art. These can be identified using routine methods known in the art, such as those described herein.

The methods of the present invention may take place in the presence of $FeCl_3$ at a concentration of at least 0.06mM, 0.07mM, 0.08mM, 0.09mM, 0.1mM, 0.2mM, 0.3mM, 0.4mM or 0.5mM $FeCl_3$. For example, the methods of the present invention may take place in the presence of $FeCl_3$ at a concentration of up to 0.06mM, 0.07mM, 0.08mM, 0.09mM, 0.1mM, 0.2mM, 0.3mM, 0.4mM or 0.5mM $FeCl_3$. For example, the methods of the present invention may take place in the presence of $FeCl_3$ at a concentration of from about 0.05mM to about 0.5mM. For example, the methods of the present invention may take place in the presence of 0.1mM $FeCl_3$.

The methods of the present invention may operate in the presence of an oxidising agent. In one embodiment, the oxidising agent used in the methods of the present invention is hydrogen peroxide. Alternative oxidising agents may include any other inorganic peroxide. A person of skill in the art will also be familiar with other alternative oxidising agents that will work in the methods of the present invention. These can be identified using routine methods known in the art, such as those described herein.

The oxidising agent will be used at a concentration sufficient to induce free radical production in the presence of a reducible substrate and the polypeptide described herein.

For example, the methods of the present invention may take place in the presence of hydrogen peroxide at a concentration of at least 0.5mM, 1mM, 1.5mM, 2mM, 2.5mM, 3mM, 3.5mM, 4mM, 4.5mM, 5mM, 5.5mM, 6mM, 6.5mM, 7mM, 7.5mM, 8mM, 8.5mM, 9mM, 9.5 M or 10mM H_2O_2. For example, the methods of the present invention may take place in the presence of hydrogen peroxide at a concentration of up to 0.5mM,1mM, 1.5mM, 2mM, 2.5mM, 3mM, 3.5mM, 4mM, 4.5mM, 5mM, 5.5mM, 6mM, 6.5mM, 7mM, 7.5mM, 8mM, 8.5mM, 9mM, 9.5 M or 10mM H_2O_2. For example, the methods of the present invention may take place in the presence of hydrogen peroxide at a concentration of from about 0.5mM to about 10mM H_2O_2 (eg. at a concentration of from

about 1mM to about 9mM H_2O_2, about 2mM to about 8mM H_2O_2, about 3mM to about 7mM H_2O_2, about 4mM to about 6mM H_2O_2, about 4mM to about 5mM H_2O_2, about 3mM to about 5mM H_2O_2, about 3.5 mM to about 4.5mM H_2O_2. In one embodiment, the methods of the present invention may take place in the presence of hydrogen peroxide at a concentration of at least 4mM H_2O_2.

All methods of the present invention comprise the step of 'contacting' lignocellulosic biomass material with compositions that comprise a polypeptide, as defined herein.

The term 'contacting' used throughout the present specification refers to the placing of the polypeptide in sufficiently close proximity to the components of the lignocellulosic biomass material to enable extraction of sugars and/or lignin derived products, in accordance with the methods of the present invention.

It will be apparent to those skilled in the art that 'contacting' may comprise a step of mixing or combining the lignocellulosic biomass material with the composition comprising the polypeptide. In one embodiment, the 'contacting' step takes place in a solution.

It will also apparent to a skilled person that the 'contacting' step is not limited only to a mixing or combining step in which the polypeptide directly contacts or interacts with components of the lignocellulosic biomass material, but also includes a mixing or combining step in which the polypeptide is in solution with but spatially separated from components of the lignocellulosic biomass material (ie. the polypeptide does not necessarily need to come into direct contact or interact directly with components of the lignocellulosic biomass material in order to effect the 'contacting' step).

In one embodiment, the 'contacting' step of the method involves mixing or combining lignocellulosic biomass material with a culture of the recombinant micro-organism expressing the polypeptide.

In one embodiment, the 'contacting' step of the method involves mixing or combining lignocellulosic biomass material with the isolated polypeptide. For example, by mixing or combining untreated or pre-treated lignocellulosic biomass material with the isolated polypeptide in solution.

The methods of the present invention may be performed using conditions suitable for maintaining the enzymatic activities of the polypeptide.

For example, the reaction conditions used in the methods of the present invention may be suitable for maintaining the polysaccharide cleavage activity of the polypeptide.

For example, the reaction conditions used in the methods of the present invention may be suitable for maintaining the 'cellulase' and/ or 'hemicellulase' activity of the polypeptide.

For example, the reaction conditions used in the methods of the present invention may be suitable for maintaining the reductase activity of the polypeptide, such as the iron reductase activity described herein.

In one embodiment, the methods may take place under mild conditions that do not include extreme heat or acid treatment. In one embodiment, the methods may take place at a temperature of from about 20°C to about 70°C, and at a pH range from about pH 4.5 to about pH 9.

For example, the methods of the present invention may take place at a temperature of at least 25°C, 30°C, 35°C, 37°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C or 70°C. For example, the methods of the present invention may take place at a temperature of up to 25°C, 30°C, 35°C, 37°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C or 70°C. For example, the methods of the present invention may take place at a temperature of from about 20°C to about 70°C (eg. about 25°C to about 65°C, about 30°C to about 60°C, about 35°C to about 55°C, about 40°C to about 55°C, about 45°C to about 55°C or about 45°C to about 50°C). For example, the methods of the present invention take place at a temperature of at least 50°C. For example, the methods of the present invention may take place at a temperature of up to 50°C.

For example, the methods of the present invention may take place at a pH of at least pH 4.5, pH 5.0, pH 5.5, pH 6.0, pH 6.5, pH 7.0, pH 7.5, pH 8.0, pH 8.5 or pH 9.0. For example, the methods of the present invention may take place at a pH of up to pH 4.5, pH 5.0, pH 5.5, pH 6.0, pH 6.5, pH 7.0, pH 7.5, pH 8.0, pH 8.5 or pH 9.0. For example, the methods of the present invention may take place at a pH of from about pH 4.5 to about pH 9.0 (eg. about pH 5.0 to about pH 8.5, pH 5.5 to about pH 8.0, pH 6.0 to about pH 7.5, pH 6.5 to about pH 7.5, pH 7.0 to about pH 7.5, pH 6.5 to about pH 8.0, pH 7.0 to about pH 8.0). For example, the methods of the present invention may be performed at a least pH 7.5. For example, the methods of the present invention may be performed at up to pH 7.5.

The methods of the present invention may take place from several minutes to several hours. In one embodiment, the methods may take place from about 6 hours to about 120 hours, such as from about 6 hours to about 48 hours, from about 6 to about 24 hours, or for about 6 hours. In one embodiment, the methods may take place for at least 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours or 120 hours. In one embodiment, the methods of the present invention may take place for at least 24 hours.

In one embodiment, the methods of the present invention may be performed using conditions suitable for maintaining a culture of the recombinant micro-organism expressing the polypeptide, such as under suitable temperature, pH, and/ or culture media formulations. Reaction conditions will vary depending on the micro-organism in question, and will be familiar to a person skilled in the art.

In one embodiment, the methods of the present invention take place under sterile reaction conditions, such as those routinely used in the manufacture of biofuels.

In one embodiment, the methods of the present invention are performed under reaction conditions that result in release of substantial amounts of sugar and/ or lignin derived products from the lignocellulosic biomass material. By substantial amount is intended at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or more of available sugars and/ or lignin derived products.

The present invention further provides the use of a polypeptide for extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent, wherein the polypeptide comprises an amino acid sequence that has at least 70% identity (eg. up to 100% identity) to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; and wherein said polypeptide is capable of extracting sugars and/or lignin derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

In one embodiment, the polypeptide that is capable of extracting sugars and lignin derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. up to 100% identity) to the sequence of SEQ ID NO: 3 or 4, or a fragment of said sequence.

All embodiments of the polypeptides, polypeptide variants and polypeptide fragments, and polynucleotides, polynucleotide variants and polynucleotide fragments described

herein, with respect to the methods of the present invention for extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent, apply equally to said uses of the present invention.

The present invention further provides the use of a polypeptide for extracting sugars from lignocellulosic biomass material that has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure; wherein the polypeptide comprises an amino acid sequence that has at least 70% identity (eg. up to 100% identity) to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; and wherein said polypeptide is capable of extracting sugars from the pre-treated lignocellulosic biomass material.

In one embodiment, the polypeptide that is capable of releasing sugars from pre-treated lignocellulosic biomass material is encoded by a polynucleotide that comprises (or consists of) a sequence having at least 70% identity (eg. up to 100% identity) to the sequence of SEQ ID NO: 3 or 4, or a fragment of said sequence.

All embodiments of the polypeptides, polypeptide variants and polypeptide fragments, and polynucleotides, polynucleotide variants and polynucleotide fragments described herein, with respect to the method of the present invention for extracting sugars from pre-treated lignocellulosic biomass material, apply equally to said uses of the present invention.

The invention further provides a method for the production of high-value chemicals comprising carrying out any one of the methods of the present invention for extracting sugars and/ or lignin derived products from lignocellulosic biomass material, and subsequently using the extracted sugars and/ or lignin derived products to produce high-value chemicals.

In accordance with this embodiment, the term 'high-value chemicals' refers to compounds or products that have a high value relative to the value of the starting materials used to produce the compounds or products. A skilled person will be familiar with the types of high-value chemicals that can be obtained from sugars and lignin derived products that have been extracted from lignocellulosic biomass material. Non-limiting examples include high value chemicals such as eugenol, syringols, coniferols, guaiacols, wood preservatives and nutraceuticals/ drugs.

The invention further provides methods for the production of biofuels, such as biogasoline, bioethanol and biodiesel from lignocellulosic biomass material.

The invention provides a method for the production of biogasoline comprising carrying out any one of the methods of the present invention for extracting sugars and/ or lignin derived products from lignocellulosic biomass material, and subsequently producing biogasoline using the extracted sugars and/ or lignin derived products.

The term “biogasoline” used throughout the present specification refers to a gasoline product containing non-oxygenated hydrocarbons that is produced from biomass material.

For example, the invention may provide a method for producing biogasoline that comprises carrying out any one of the methods of the present invention for extracting sugars from lignocellulosic biomass material, subsequently fermenting the extracted sugars to produce fermentation end products, and using the fermentation end products to produce biogasoline.

The terms “fermentation” or “fermenting” used throughout the present specification refer to an enzyme controlled aerobic or anaerobic breakdown of an energy-rich compound, such as a carbohydrate to carbon dioxide and an alcohol or an organic acid. Fermentation may also include an enzyme controlled transformation of an organic compound.

Fermentation processes may be performed using yeast, bacteria, cyanobacteria, algae, and/ or enzymes to produce fermentation end products, such as alcohol or oxygen containing compounds. For example, fermentation may be performed using naturally occurring hexose or pentose consumers and/ or genetically modified hexose or pentose consumers. Naturally occurring organisms may produce alcohols or other oxygen containing compounds, such as may be used directly or may be converted to an ether.

The fermentation end products are then further processed using conventional techniques to produce biogasoline. For example, the fermentation end products may be converted into non-oxygenated hydrocarbons using conventional chemical processing techniques. The biogasoline may be optionally blended with a gasoline supply.

Alternatively, the invention may provide a method for producing biogasoline that comprises carrying out any one of the methods of the present invention for extracting

lignin derived products from lignocellulosic biomass material, and subsequently producing biogasoline from the lignin derived products.

Conventional methods may be used to produce biogasoline using lignin derived products. For example, gasification may be used to convert lignin into syngas (carbon monoxide/ hydrogen), which in turn is used to prepare methanol and/ or dimethyl ether, from which biogasoline can be prepared. Pyrolysis and hydroliquefaction processes may also be used to produce biogasoline from lignin derived products.

The invention further provides a method for the production of bioethanol comprising carrying out any one of the methods of the present invention for extracting sugars from lignocellulosic biomass material, subsequently fermenting the extracted sugars to produce ethanol, and processing the ethanol into bioethanol.

The term “bioethanol” used throughout the present specification refers to the biofuel product containing ethanol that is produced from biomass material.

All embodiments of the fermentation process described above with respect to the production of biogasoline apply equally to this embodiment of the invention, with the additional limitation that the fermentation process specifically produces ethanol.

Conventional techniques may be used to process ethanol into bioethanol. For example, the ethanol obtained following fermentation may be subjected to fractional distillation and/ or dehydration processes to remove excess water. The ethanol may be optionally blended with a gasoline supply. A skilled person will be aware of the conventional techniques used to process ethanol into bioethanol.

The invention further provides a method for the production of biodiesel comprising carrying out any one of the methods of the present invention for extracting sugars and/ or lignin derived products from lignocellulosic biomass material, and subsequently producing biodiesel using the extracted sugars and/ or lignin derived products.

The term “biodiesel” used throughout the present specification refers to the diesel product containing long-chain alkyl esters that is produced from biomass material.

In accordance with this embodiment, conventional processes may be used to convert the extracted sugars and/ or lignin derived products into biodiesel. The extracted sugars may be converted into one or more suitable biodiesel materials, such as triglycerides, fatty acids, alkanes, alkenes, and/or pure hydrocarbons, from which long-chain alkyl

esters can be produced. For example, fermentation may be used to convert the extracted sugars into one or more suitable biodiesel materials. All embodiments of the fermentation process described above with respect to the production of biogasoline apply equally to this embodiment of the invention. For example, pentose sugars obtained from hydrolysis of hemicellulose may be processed into fatty acids using naturally occurring pentose consumers or genetically modified pentose consumers. Naturally occurring organisms may produce fatty acids, which may be esterified with an alcohol and/ or hydrogenated with hydrogen to produce long-chain alkyl esters suitable for forming a biodiesel product. Lignin derived products may be converted into biodiesel using processes such as pyrolysis. A skilled person will be aware of alternative conventional processes for converting extracted sugars and/ or lignin derived products into biodiesel. The biodiesel may be optionally blended with a diesel supply.

Sequences

SEQ ID NO: 1 = Amino acid sequence of IR1 derived polypeptide (without signal peptide)

SEQ ID NO: 2 = Amino acid sequence of IR2 derived polypeptide (without signal peptide)

SEQ ID NO: 3 = Nucleotide sequence encoding IR1 derived polypeptide (without signal peptide)

SEQ ID NO: 4 = Nucleotide sequence encoding IR2 derived polypeptide (without signal peptide)

SEQ ID NO: 5 = signal peptide of IR1

SEQ ID NO: 6 = signal peptide of IR2

SEQ ID NO: 7 = AttB1-IR1plusSP-F primer

SEQ IS NO: 8 = AttB2-IR1minSTP-R

SEQ ID NO: 9 = AttB1-IR2minSP-F

SEQ ID NO: 10 = AttB2-IR2plusSTP-R

SEQ ID NO: 11 = AttB1-adapter-F

SEQ ID NO: 12 = AttB2-adapter-R

SEQ ID NO: 13 = M13 primer

SEQ ID NO: 14 = M13 primer

SEQ ID NO: 1 (IR1 Polypeptide without signal peptide)

MATAYCDSTSGICYAGYTDPTLNITVGLVLPPLSTGTTPNATEFVAELIAPASYGWTGLS
VGGTMADSLLFTMWPYNGEVIFGPRWTSGYVQPLPYSGPIITMLPDTMVNSTHIKASFR
CQNCTTWNGGALGSGDLSGFQLIAYVASDDTPVDDPSNVASNFTEHDQMNFFGLDLSI
SHSTNYSTYIGTTGTSPSEPTQTE***YGQCGGTGWTGPTVCVTGTDCTEVSPPY***YSQCL

Heme domain (underlined) - corresponds to residues 4-175

Cellulose binding domain (bold and italics) – corresponds to residues 202-229

SEQ ID NO: 2 (IR2 Polypeptide without signal sequence)

MATAYCDSTSGICYQGYTDPTLDITVGLTFPPVSTDGGANSTAFIAEIVAPVTYGWTGIT
VGGSMAYSLLFVLWPYEGEVILSPRWTTGYTQPTPYYGPIITMLPGTSVNSTSITASFLC
ENCTYWEGGAIGGGDLEGTQLIEYVANNDTEVFDPANYASNFTIHDVYGSFELNLADA
HFSDFFSLL

Heme domain (underlined) – corresponds to residues 4-175

SEQ ID NO: 3 (Polynucleotide encoding the IR1 polypeptide without signal sequence)

ATGGCTACAGCTTACTGCGATTCCACGAGTGGTATCTGTTATGCAGGCTATACAGA
CCCAACTTTGAATATCACAGTTGGCCTTGTCTTACCCCCTTTGTCTACCGGAACCAC
GCCTAACGCGACTGAATTCGTTGCGGAATTGATTGCGCCTGCCAGCTATGGCTGGA
CCGGTCTCAGCGTTGGGGGTACTATGGCAGACAGCCTTTTGTTCACGATGTGGCCT
TACAATGGCGAAGTCATATTTGGCCCTAGGTGGACTTCCGGTTATGTGCAACCTCT
TCCCTACTCTGGTCCCATTATCACCATGCTTCCCGACACGATGGTAAACAGTACTCA
CATCAAGGCATCTTTCCGCTGCCAGAATTGCACAACATGGAATGGCGGCGCACTTG
GAAGTGGTGATCTGAGCGGATTCCAGCTTATCGCCTATGTCGCGTCGGATGACACA
CCCGTTGATGATCCTTCCAACGTCGCATCCAACTTCACGGAGCATGACCAGATGAA
CTTCTTCGGTCTTGATTTATCGATCTCCCACTCTACGAATTACTCCACATATATCGG
GACCACCGGCACATCTCCAAGTGAGCCTACTCAGACCGAG***TACGGTCAGTGCGGT***
GGGACTGGTTGGACGGGACCAACAGTGTGTGTAACTGGGACTGATTGCACTGAA
GTCAGCCCCCCGTATTATAGCCAGTGCCTGTGA

Heme domain (underlined) – corresponds to nucleotides 10-525

Cellulose binding domain (bold and italics) – corresponds to nucleotides 604-687

SEQ ID NO: 4 (Polynucleotide encoding the IR2 polypeptide without signal sequence)

ATGGCTACTGCATACTGCGACTCGACTAGTGGCATCTGCTATCAGGGTTACACGGA
TCCCACCCTCGATATCACCGTCGGCCTCACATTCCCTCCCGTATCGACAGACGGTG
GTGCCAACTCGACCGCCTTCATCGCAGAAATAGTAGCACCCGTTACCTATGGTTGG
ACTGGTATCACTGTCGGAGGCTCCATGGCTTACAGCTTACTGTTCGTTCTCTGGCC

ATATGAGGGCGAAGTCATCTTGTCACCGAGGTGGACAACTGGTTATACACAGCCTA
CTCCTTACTACGGACCAATAATCACCATGCTCCCTGGTACCTCGGTCAACTCTACGT
CCATCACAGCTTCGTTCTTATGCGAGAACTGCACTTACTGGGAGGGAGGAGCCATT
GGAGGCGGTGACTTGGAGGGAACACAGCTCATCGAATACGTCGCCAATAACGATA
CAGAAGTCTTTGACCCAGCCAACTACGCGTCAAACTTCACCATCCACGACGTCTAT
GGCTCTTTCGAGCTCAACCTCGCCGATGCGCACTTTTCCGACTTTTTCAGTCTCTTG
TGA

Heme domain (underlined) – corresponds to nucleotides 10-525

SEQ ID NO: 5 (IR1 signal peptide)

MFSHLLTTIILSIGFRAVTVVAQS

SEQ ID NO: 6 (IR2 signal peptide)

MFQKLLVASLLFLGIQFVNAVPN

SEQ ID NO: 7

AAAAAGCAGGCTTCatgGCTACAGCTTACTGCGATTC

SEQ ID NO: 8

AGAAAGCTGGGTTCACAGGCACTGGCTATAATAC

SEQ ID NO: 9

AAAAAGCAGGCTTCatgGCTACTGCATACTGCGACTC

SEQ ID NO: 10

AGAAAGCTGGGTTCACAAGAGACTGAAAAAGTC

SEQ ID NO: 11

GGGGACAAGTTTGTACAAAAAAGCAGGCT

SEQ ID NO: 12

GGGGACCACTTTGTACAAGAAAGCTGGGT

SEQ ID NO: 13

GTAAAACGACGGCCAG

SEQ ID NO: 14

CAGGAAACAGCTATGAC

Figures

Figure 1

Figure 1 shows the domain structure of the (a) IR1 and (b) IR2 enzymes encoded by *Serpula lacrymans.* The IR1 and IR2 proteins both contain a cellulose binding domain family 9-like (IPR008960) / (iron reductase domain (IPR015920); IR1 also has a cellulose binding module-1 (CBM1- PDOC00486) domain.

Figure 2

Figure 2 shows the results of a western blot analysis performed on purified recombinant protein GST-IR1 and GST-IR2 by SDS-PAGE gels, as described in Example 2.

Figure 3

Figure 3 shows the results obtained from the Ferrozine spot assay performed using extracts of IR1 and IR2 protein, as described in Example 3.

Figure 4

Figure 4 shows absorbance at 540nm observed for the dichlorophenol indophenol (DCPIP) assay performed using extracts of IR1 and IR2 protein over a 15 minute period, as described in Example 4.

Figure 5

Figure 5 shows the mean change of absorbance at 540nm observed for the dichlorophenol indophenol (DCPIP) assay performed using extracts of IR1 and IR2 protein over a 15 minute period, as described in Example 4.

Figure 6

Figure 6 shows absorbance at 430nm observed for the nitrated lignin assay performed using extracts of IR1 and IR2 protein over a 20 minute period, as described in Example 5.

Figure 7

Figure 7 shows the mean change of absorbance at 430nm observed at 1 and 20 minutes for the nitrated lignin assay performed using extracts of IR1 and IR2 protein, as described in Example 5.

Figure 8

Figure 8 shows the amount of total reducing sugars extracted from Avicel and wheat straw following 24 hours incubation with recombinant IR1 and IR2.

Examples

Example 1: Domain characterization

IR1 and IR2 sequences were amplified from RNA collected from *Serpula lacrymans.* The RNA was collected after 28 days of culture showing that these polypeptides are expressed in the later stage of *Serpula lacrymans* growth. Amplified products were cloned and the nucleotide sequences and predicted translated amino acid sequence of IR1 and IR2 were aligned with the sequences of other fungal genes.

BLASTN alignment with the NCBI databases found no significant alignments with previously described nucleotide sequences. However, comparison of the iron reductase amino acid sequence by BLASTP showed that IR1 shares high identity (74.2%) with the amino acid sequence of carbohydrate binding module (CBM1) of different fungi e.g. *Coniophora puteana* (accession number EIW84939), 65.9% identity with that from carbohydrate binding cytochrome (CBcyt b562) from *Stereum*

hirsutum (accession number EIM89944.1), 61.4% identity with cellulose binding cytochrome b562 from *Phanerochaete chrysosporium* (accession number BAD95668), and 59.8% identity with IR2 (protein ID: 417465). BLASTP analysis of the IR2 amino acid sequence showed that IR2 has 57.5% identity with CBM1 from *Coniophora puteana* (accession number EIW84939), 54.8% identity with carbohydrate binding cytochrome b562 from *Stereum hirsutum* (accession number EIM89944.1) and 46.9% identity with CBcyt b562 from *Phanerochaete chrysosporium* (accession number BAD95668).

Using MrBayes version 3.2 software (Fredrik et al. 2010), a phylogenetic tree was constructed based on the amino acid sequences for IR1, IR2 and 14 other fungal genes encoding either cellobiose dehydrogenases (CDH) or cellulose/ carbohydrate binding modules (CBM). Phylogenetic analysis showed that IR1 and IR2 genes are more closely related to the CBM type of genes than to the CDH gene family. For example, 74% identity was observed between IR1 and the sequence for a CBM from *Coniphora puteana.* Blast comparison of the IR1 and IR2 genes sequences indicated that these genes have a low level of sequence similarity to other fungal genes.

Based on sequence homology analysis, both the IR1 and IR2 genes were found to contain a cellulose binding (CBD) 9 family cytochrome domain (also referred to as a heme domain). The IR1 gene was also found to contain a C-terminal cellulose binding module (CBM1) (see Figure 1).

The alignments also showed that both IR1 and IR2 contain certain conserved amino acids. These included: a methionine at amino acid residue '65' of the IR1 sequence (as defined in SEQ ID NO: 1) and amino acid residue '65' of the IR2 sequence (as defined in SEQ ID NO: 2) and a histidine at amino acid residue '165' of the IR1 sequence (as defined in SEQ ID NO: 1) and amino acid residue '165' of the IR1 sequence (as defined in SEQ ID NO: 2). Cysteine residues are also present at amino acid residues '6', '13', '120' and '123' of the IR1 sequence (as defined in SEQ ID NO: 1) and at amino acid residues '6', '13', '120' and '123' of the IR2 sequence (as defined in SEQ ID NO: 2).

It has been reported that cellulose binding domains normally only contain one aromatic residue, such as tyrosine (Tomme et al., 1998). The recombinant protein IR1 contains three aromatic residues including a tryptophan at amino acid residue

'210' of the IR1 sequence (as defined in SEQ ID NO: 1), and a tyrosine at amino acid residues '229' and '230' of the IR1 sequence (as defined in SEQ ID NO: 1).

Strains, culture condition and plasmid

The brown rot basidiomycete *Serpula lacrymans* S7.3 was obtained from the culture collection of Warwick HRI (School of Life Sciences) and grown in the dark on malt extract agar (MEA) plate at 20ºC for 3-4 weeks. The MEA medium containing malt extract 20g/l and agar 12g/l.

cDNA synthesis and Polymerase Chain Reaction (PCR)

IR1 and IR2 were amplified using primers, designed from the sequence from the *Serpula lacrymans* genome (Table 1). All primers were ordered from INVITROGEN. 5µg RNA and 2µM Oligo dT18 (Invitrogen) were denatured at 65˚C for 5min and cooled on ice for 2 minutes. The 5x cDNA synthesis mix, (0.1M DTT (Invitrogen), 1xSuperScript Buffer (Invitrogen), 10mM dNTPs (Invitrogen), x1 RNaseOUT (Invitrogen), x1 SuperScript RT (Invitrogen) and DEPC- H_2O) were added and the following cycles was completed: 1 cycle of 96ºC (5min); 30cycles of 95ºC (20sec), 59ºC(20sec), 73ºC(40sec) and 1 cycle of 73ºC(10min). The cDNA was then stored at -20ºC.

Table 1. Primers for amplifying IR1 and IR2

Primer	**Primer Sequence 5'-3'**	**SEQ ID NO**
AttB1-IR1plusSP-F	AAAAAGCAGGCTTCatgGCTACAGCTTACTGCGATTC	7
AttB2-IR1minSTP-R	AGAAAGCTGGGTTCACAGGCACTGGCTATAATAC	8
AttB1-IR2minSP-F	AAAAAGCAGGCTTCatgGCTACTGCATACTGCGACTC	9
AttB2-IR2plusSTP-R	AGAAAGCTGGGTTCACAAGAGACTGAAAAAGTC	10
AttB1-adapter-F	GGGGACAAGTTTGTACAAAAAAGCAGGCT	11
AttB2-adapter-R	GGGGACCACTTTGTACAAGAAAGCTGGGT	12

PCR amplification and Cloning of iron reductase from the brown rot fungus Serpula lacrymans

cDNA encoding iron reductases (IR1 and IR2) were cloned using a binary vector/plasmid and following a protocol from TA cloning kit INVITROGEN Cat no K2020-20. All PCR products were ligated into pCR.2.1. The ligation was performed overnight at 4ºC by incubating 2µl PCR product, 1µl 10x ligation buffer, 2µl pCR 2.1 vector and 1µl t4 DNA ligase.

The constructs then were transformed into *E. coli* (DH5α cells) and the transformants were cultured overnight at 37ºC in LB agar plate containing 50µg/ml kanamycin (KAN) selective media. Plasmid DNA from positive colonies was isolated using the alkaline lysis miniprep method (QIAGEN Plasmid Mini Purification protocol). The presence of a fragment of the correct size was confirmed by PCR using gene specific primers (see Table 1) or plasmid specific M13 Forward and M13 Reverse primers (see Table 2). The yield of DNA was determined using a UV spectrophotometer (Nano-drop) and by quantitative analysis on a 1.2% agarose gel.

Table 2. M13 Primers used for sequencing

Primer	Primer Sequence 5'-3'	Primer sequence 5'-3'
M13	GTAAAACGACGGCCAG (SEQ ID NO: 13)	CAGGAAACAGCTATGAC (SEQ ID NO: 14)

cDNA sequencing analysis

Sequencing reactions were performed using the ABI BigDye terminator V.1.1/3.1 seq Kit. Each reaction contained a vector specific primer (3.2pmol), 2µl ready reaction mix (Big dye V3.1), 1µl big dye sequencing buffer and 1µl of plasmid cDNA (100ng) samples. Each reaction was made up to 10µl with pure distilled water. PCR cycles (25) were as followed: 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. The product were then analysed using a ABI3130xl sequencer at the School of Life Sciences-Wellesbourne campus. Searches were performed using BLAST algorithms against various databases in the GenBank (http://www.ncbi.nlm.nih.gov/BLAST) and also from the *S. lacrymans* genome databases.

Example 2: Expression and purification of the recombinant protein in *E. coli* (BL21)

The coding regions of both IR1 (SEQ ID NO: 3) and IR2 genes (SEQ ID NO: 4) without a signal peptide were cloned and expressed in *E. coli* BL21 as GST fusion proteins. The soluble and insoluble fractions were analysed on SDS-PAGE with a 4-12% gel gradient. Crude extracts derived from the soluble fraction of pelleted cells following 5 hours induction showed the highest production of the recombinant proteins of the expected size. The total amount of protein obtained in the recombinant IR1 cell extracts was 4.84 mg ml^{-1} and for IR2 was 4.98 mg ml^{-1}. Crude protein extracts were purified using glutathione sepharose beads to obtain IR1 protein at a concentration of 0.066 mg ml^{-1} and IR2 protein at a concentration of 0.043 mg ml^{-1}. Western blotting was performed using an anti-GST antibody to confirm that GST-IR1 and GST-IR2 were correctly purified. GST-IR1 was correctly observed to have a molecular weight of 55kDa; and GST-IR2 was correctly observed to have a molecular weight of 49kDa (Figure 2).

Figure 2 shows the western blotting results obtained for purified GST-IR1 Figure 2(a) and purified GST-IR2 Figure 2(b). In Figure 2(a), the arrow indicates a band corresponding to recombinant IR1 having a molecular weight of 55kDa. Lanes 1 and 2 correspond to the flow through fraction; lane 3 and 4 correspond to the wash fractions; lanes 5-7 correspond to the elute fractions; lane 8 is empty; and lane 9 corresponds to further elute fractions. In Figure 2(b), the arrow indicates a band corresponding to recombinant IR2 having a molecular weight of 49kDa IR2. Lanes 1 and 2 correspond to the flow through fraction; lanes 3-7 correspond to the wash fractions; and lanes 8-9 correspond to the elute fractions.

Gateway cloning strategy

The Invitrogen Gateway Cloning system (www.invitrogen.com) was used to clone the coding regions of IR1 (SEQ ID NO: 3) and IR2 (SEQ ID NO: 4) genes. This system uses site-specific recombination *att*B x *att*P→*att*L x *att*R of α phage, which is schematically presented below:

*att*B1-gene-*att*B2 x *att*P1-*ccd*B-*att*P2 ↔ *att*L1-gene-*att*L2 x *att*R1-*ccd*B-*att*R2

(Expression clone) (pDONR) (Entry clone) (Destination vector)

The major steps of the Gateway cloning system are the BP and LR reactions. The *att*B x *att*P reaction is mediated by Gateway BP clonase II enzyme mix, while the *att*L x *att*R reaction is mediated by Gateway LR clonase II enzyme mix. The BP reactions utilize the recombination between attB of the DNA segment of interest and attP of the donor to create entry clones.

Primers design

IR1 and IR2 fragments with a plus stop codon and without the signal peptide were PCR amplified from plasmids containing the IR1 and IR2 genes. Primers used for PCR amplification are as defined in Table 1.

PCR

In order to amplify the target sequence of both genes (IR1 and IR2), the first and second steps of the gateway system (50µl) were carried out as follows. 2µl (10mM) primers, 2µl plasmid cDNA (100ng), 25µl taq DNA polymerase were mixed with 21µl pure water. The PCR reaction was performed using 1 cycle of 94ºC (3 min), followed by 5 cycles of denaturation (30sec at 94ºC), annealing (30sec at 55ºC) and extension (1.5min at 72ºC) and then 25 cycles: 94ºC for 30sec, 65ºC for 30sec and 72ºC for 1.5min, and a final extension at 72ºC for 7 min. 5µl of product was electrophoresed on a 1.2%(w/v) agarose gel, followed by gel purification (Qiaquick gel extraction protocol). The first PCR product was then used for the second plasmid step using the same kit and conditions as the first PCR, but the primers were adapter primers (see Table 1). The purified amplicon from the second PCR reactions was used for the Gateway BP reaction.

Recombination attB primers with pDONR/Zeo vector to create entry clone

To introduce the DNA fragment into the entry clone (pDONR/Zeo, Invitrogen), *in vitro* BP clonase recombination reactions were carried out according to the manufacturer's instructions (Invitrogen). The product of the recombination reactions (BP reactions) was used to transform competent DH5α *E. coli* using heat shock. Positive transformants were cultured overnight at 37˚C in 5ml LB containing antibiotic selection 30 µg ml^{-1} Zeocin. The plasmid was extracted using the Qiagen plasmid mini-prep kit.

Performing the LR Recombination Reaction

The entry clone from the BP reaction and destination vector pDEST15 was mixed with the LR clonase II enzyme mix. pDEST15 is N-terminal fusion vectors which contain an ATG initiation codon upstream of GST tag. The product of recombination of LR reaction was transformed into the DH5α *E. coli* strain. Positive transformants were cultured overnight at 37˚C in 5ml LB containing antibiotic selection 30µg/ml Zeocin. The plasmid was then purified using the Qiagen plasmid mini-prep kit.

Verification of Gateway Product

Using the ABI BigDye terminator V.1.1/3.1 seq Kit, the BP and LR products and appropriate primers were verified by sequencing. Each sequencing reaction contains 3.2pmol primers, 2µl ready raction mix (Big dye V3.1), 1µl big dye sequencing buffer and 1µl of BP product. Each reaction was made up to 10µl with pure distilled water, and the sequencing condition were as followed: 1 cycles 96ºC for 2min; 35cycles: 96°C for 10 sec, 50°C for 10sec and 60°C for 3 min. These were sequenced using an ABI3130XL.

Expressing the recombinant protein

The transformant colonies were inoculated into 10ml of LB medium containing the selective antibiotics 50µg/ml carbenicellin and 34µg/ml chloramphenicol and grown overnight at 37ºC with shaking 220rpm. 2.5ml of overnight culture was inoculated into 50ml of prewarmed LB media (with antibiotics) on the shaking incubator (220 rpm for approximately 1.5 hours), until the OD_{600} is 0.5—0.7. The transformants were induced using 0.4mM of isopropyl-β-D-thiogalactopiranoside (IPTG) and the culture incubated at 30ºC for an additional 5-6 hours. 1ml induced samples were collected in different time points; 0, 3, 5, 12 and 20 hours. The optical density (OD) significantly increased after 5 hours induction. The cells were harvested by centrifugation at 5000rpm (20min) for the SDS-PAGE analysis and resuspended in an appropriate volume lysis buffer prior to sonication and purification over the Gluthatione Sepharose.

The cell pellet was resuspended in 1 ml of lysis buffer containing 50mM Tris-HCl pH 8; 1mM EDTA pH 8,0; 1mM tris2 carboxyethyl-phosphine (TCEP); 1mM phenyl methylsulfonyl- fluoride (PMSF); 200mM NaCl, and deionized water (dH_2O). The cell pellet was frozen using liquid nitrogen and thawed in cold water. The cells were then

sonicated for 6 x 10sec with 10 sec pauses at 200-300W and the lysate was centrifuged at 5000 x g at 4ºC for 20min. The supernatant was obtained (ie. 'the crude extract') and used for protein analysis.

SDS PAGE and Western blotting analysis

The soluble and insoluble fractions were tested for the presence of recombinant protein using SDS-PAGE with a 12% SDS-PAGE gel. 15µl samples were added to 5µl 2x SDS-PAGE sample buffer and heated at 95ºC for 5 min. Prior to the samples being separated by 12% SDS-PAGE gel for approximately 1 hour, samples were centrifuged at 13,000 x rpm for 1 min. The gel was stained with Coomasie instant blue from Expedeon. All the procedures were done according the manufacturer's recommendations (Biolab color plus prestained 7-175kDa).

Western blotting was carried out using standard protocols. The protein was transferred onto nitrocellulose membrane for 1.5 hour and treated for 2-3 hours at room temperature using 5% skimmed milk as the blocking agent. The membrane was then incubated overnight at 4ºC with primary antibody (monoclonal anti-GST antibody (SIGMA G-1160)) at a dilution of 1:2000. The membrane was washed three times using PBST (Phosphate Buffer Saline with Tween 20) for 5-10min. The membrane was then incubated with secondary antibody (anti-GST antibody-peroxidase conjugate produced in mouse (SIGMA-A4416)) at a dilution of 1:10,000). Secondary antibody was incubated for 2 hours at room temperature. The blot was washed three to five times for 15 minutes using buffer PBST. The blot was then incubated with ECL (Enhanced chemiluminesence) Western blotting detection reagents (according to manufacture instructions from Amersham) for 5 minutes at room temperature. Analysis of the blot was performed using a hyperprocessor machine.

Purification of recombinant iron reductases

500ml LB medium was prepared for the purification of recombinant protein (IR1 and IR2) as described above. All of the protein purification was undertaken at 4°C. The supernatant was centrifuged at 5000 x g for 20 minutes, 4°C using a SORVALL RC 5B and resuspended in lysis buffer (containing 50mM Tris-HCl pH 8; 1mM EDTA pH 8,0; 1mM tris2 carboxyethyl-phosphine (TCEP); 1mM phenyl methylsulfonyl- fluoride (PMSF); 200mM NaCl, and deionized water (dH_2O)). The cells were lysed using a

combination of freeze thaw and sonication. Lysed cells were then centrifuged at 13,000 x g for 10 minutes at 4°C, and the supernatant was collected for purification. The soluble fractions of recombinant protein (IR1 and IR2) were purified using the Glutathione Sepharose 4B beads (GE Healthcare, UK) according to the manufacturer's instructions. The crude cell extract was passed through a column pre-equilibrated with binding buffer PBS pH 7.5 (140mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, and 1.8mM KH_2PO_4). The columns were prepared according to the manual (Glutathione Separose 4B, 52-2303-00 AK). After extensive washing using binding buffer, the GST fusion proteins were eluted with elution buffer (50mM Tris-HCl, 20mM reduced glutathione, pH 8.0).

Protein analysis

Concentration of the recombinant protein was determined using the Bradford RC-DC protein assay (from BIO-RAD). 1mg ml^{-1} BSA was used as the standard and absorption was measured at 750nm.

Example 3: Determination of the function of recombinant iron reductase (IR1 and IR2)

The IR proteins were both predicted to have iron reducing activity due to the presence of the CBD9 (iron reductase) domain. The Ferrozine spot assay was used to detect the release of Fe^{2+} following reduction of Fe^{3+}. 50μM of 2,3 dihydroxybenzoic acid (2,3-DHBA) was used as a positive control for the assay, and shown to significantly increase absorbance.

Crude protein extracts containing recombinant IR1 and IR2 were used in the assay, and shown to increase absorbance significantly ($P \leq 0.05$) within 30 minutes of incubation (Figure 3). No 2,3 dihydroxybenzoic acid (2,3-DHBA) was used in the reaction mixes containing IR1 or IR2. In contrast, the negative control for the assay (ie. buffer without 2,3-DHBA) showed no significant change within 30 minutes of incubation. These results indicate that IR1 and IR2 have the capacity to generate Fe^{2+} from Fe^{3+}.

Figure 3 illustrates the results of the Ferrozine spot assay. Changes in absorbance at 550nm were observed following the addition of recombinant protein IR1 or IR2 or in the presence of the positive control 2,3 dihydroxybenzoic acid (2,3-DHBA) after 30 minutes of incubation. No significant change in absorbance was observed for the negative

control buffer without 2,3-DHBA. The error bars represent the least significant different (LSD 5%).

Iron reductase assay (Ferrozine assay)

Reduction of iron was detected using the Ferrozine reagent [3-(2-pyridyl)-5,6-bis-(4-phenylsulfonic acid)-1,2,triazine] (Sigma) using a modified method developed by Arantes et al., (2009) and Kerem et al., (1999) (Arantes, V., *et al.* (2009); Effect of pH and oxalic acid on the reduction of $Fe3^+$ by a biomimetic chelator and on $Fe3^+$ desorption/adsorption onto wood: Implications for brown-rot decay; International Biodeterioration and Biodegradation; 63; 478-483; and Kerem, Z., *et al.* (1999); Biodegradative mechanism of the brown rot basidiomycete Gloeophyllum trabeum: evidence for an extracellular hydroquinone-driven fenton reaction; FEBS Letters; 446; 49-54).

The experiment was conducted in 96-well micro titer plates. 50µl of crude extract/supernatant from soluble fusion protein of IR1 and IR2 cultures were combined with 0.1 mM $FeCl_3$, 1M acetate buffer pH4.6, in the presence and absence of 50µM 2,3 dihydroxybenzoic-acid (DHBA). After 10 minutes incubation 10µM Ferrozine reagent was added to the reaction. The absorbance was measured at 550nm using a spectrophotometer TECAN-Genious plate reader for 30 minutes kinetically.

Example 4: The ability of iron reductases to act as an electron acceptor as measured by the effect of 2,3-DHBA in the presence of H_2O_2

In order to further demonstrate that IR1 and IR2 can reduce electron acceptors, the dichlorophenol indophenol (DCPIP) based assay was used (as set out in Baminger, U., *et al.;* A simple assay for measuring cellobiose dehydrogenase activity in the presence of laccase; Journal of Microbiological Methods; 1999; 35; 253-259). This assay measures the reduction of the electron acceptor DCPIP over 30 minutes by measuring a change in absorbance.

To examine the significance of the reduction in DCPIP absorbance due to the different treatment, at least 10 minutes observation is required by which time all DCPIP is assumed to be completely degraded and no remaining DCPIP left in the solution. After 10 minutes, a significant reduction ($P<0.05$) in the absorbance of DCPIP was apparent

in the presence of IR1 and IR2 supplemented with Fe^{3+}, 2,3-DHBA and hydrogen peroxide (H_2O_2) (Figures 4 and 5).

Meanwhile, there was no reduction in DCPIP absorbance observed with the negative control treatment (including buffer alone and *E. coli*). Although, the negative controls (either *E. coli* or buffer alone) gave an initial increase in absorbance at the beginning of the reaction, this leveled off after 2 minutes. This phenomenon was hypothesized to be due to a lack of iron reductase enzyme raising the absorbance of DCPIP (Figures 4 and 5).

In the absence of 2,3-DHBA, both IR1 and IR2 continued to decrease absorbance in the assay, indicating that a reduction of DCPIP occurs despite the absence of 2,3-DHBA (Figure 4). This demonstrates that IR1 and IR2 can act as reducing agents and have a similar effect as the positive control 2,3-DHBA.

Figure 4 shows the results obtained from the DCPIP assay, measuring absorbance at 540nm over a period of 15 minutes. The DCPIP assay was performed in the presence of IR1 (with and without 2,3 DHBA) and IR2 (with and without 2,3 DHBA). Negative controls used in the assay include E.Coli (with and without 2,3 DHBA) and buffer (with and without 2,3 DHBA).

Figure 5 shows the mean change in absorbance observed at 540nm for the DCPIP assay following 15 minutes of incubation. The DCPIP assay was performed in the presence of IR1 (with and without 2,3 DHBA) and IR2 (with and without 2,3 DHBA). Negative controls used in the assay include E.Coli (with and without 2,3 DHBA) and buffer (with and without 2,3 DHBA). The error bars represent the least significant different (LSD 5%).

DCPIP assay

The DCPIP assay performed was a modified version of the methods described in Baminger *et al.*, (1999) and Nakagame *et al.,* (2006) (Baminger *et al.* A simple assay for measuring cellobiose dehydrogenase activity in the presence of laccase; Journal of Microbiological Methods; 1999; 35; 253-259; and Nakagame, S. *et al;* Purification and characterization of cellobiose dehydrogenase from white-rot basidiomycete Trametes hirsata; Bioscience Biotechnology and Biochemistry; 2006; 70; 1629-1635).

Recombinant enzyme activity was determined at room temperature using 0.1M 2,6-dichlorophenol-indophenol (DCPIP; Sigma-Aldrich) as an electron acceptor in two different buffers 50mM sodium acetate buffer (pH 5) and 50mM tris-HCL (pH 7.5) with cellobiose as the substrate. The reaction mixture was prepared in a total volume 200µl and contained 100µl of recombinant IR1 or IR2 protein, 40µl of 0.6mM cellobiose, 10µl of Fe^{3+} (Ferric chloride), 10µl 2,3 dihydroxyl-benzoic acid (2,3 DHBA), 10µl of 4mM H_2O_2 and 10µl 0.5mM DPCIP. Reducing activity was measured by following decrease in absorbance of the electron acceptor DCPIP. The decrease in absorbance of DPCIP was monitored using kinetic spectrophotometry at 540nm every minute from the first 60s until 30 minutes. Absorbance was measured using a spectrophotometer TECAN GENious plate reader. The assay was performed also in the absence of cellobiose, 2,3-DHBA and recombinant proteins. All readings were taken in quadruplicate.

Example 5: The ability of iron reductases to degrade nitrated lignin

In order to demonstrate the role of IR1 and IR2 in the depolymerisation of lignin, the nitrated lignin assay was performed, and the release of phenolic compounds was measured.

The activities of IR1 and IR2 were tested by incubating IR1 or IR2 (with and without 2,3-DBHA) in the presence of iron (Fe^{3+}) and hydrogen peroxide (H_2O_2). An increase in the absorbance at 430nm was observed in all cases after 20 minutes of incubation. However, IR1 showed greater activity than IR2 in the presence and absence of 2,3-DHBA, which indicates IR1 under these conditions has a greater potential to degrade nitrated lignin compared to IR2. After 20 minutes of incubation, IR2 only showed a significant difference in absorbance at 430nm when it was added in the presence of both Fe^{3+} and 2,3-DHBA. The presence of 2,3-DHBA appeared to be additive increasing the potency of both IR1 and IR2.

As a negative control for the assay, IR-1 and IR-2 (with and without the addition of 2,3-DBHA) were also incubated in the absence of hydrogen peroxide (H_2O_2). No change in absorbance at 430nm was observed (data not shown). This demonstrates that hydrogen peroxide is required for the reducing activity of IR1 and IR2, and provides evidence to support the theory that these proteins are involved in facilitating the Fenton reaction in conjunction with hydroxyl radical generated when lignocellulosic depolymerisation occurs.

Figure 6 shows the results obtained for the nitrated lignin assay, measuring absorbance at 430nm over a period of 20 minutes. The nitrated lignin assay was performed in the presence of IR1 (with and without 2,3 DHBA) and IR2 (with and without 2,3 DHBA).

Figure 7 shows the mean change in absorbance observed at 430nm for the nitrated lignin assay following 1 and 20 minutes of incubation. The nitrated lignin assay was performed in the presence of IR1 (with and without 2,3 DHBA) and IR2 (with and without 2,3 DHBA). The error bars represent the least significant different (LSD 5%).

Preparation of nitrated organosolve lignin solution

A stock solution of nitrated organosolve lignin was prepared from the mixture of 25mg organosolve lignin with 5ml of glacial acetic acid (80mM of organosolv lignin in glacial acteic acid). The solution was filtered to remove insoluble material. The solution was then added to 750µl of concentrated nitric acid (HNO_3) and stirred on ice for 1 hour. The reaction was neutralized with 1M NaOH at pH 7.0 and added to 10ml of H_2O. The nitrated lignin stock solution was stored at 4-5ºC and diluted 25-fold in deionized H_2O.

Nitrated lignin assay

The nitrated lignin assay was performed using a method modified from Ahmad *et al.*, (2010), using Fe^{3+} as a substrate (Ahmad, M. *et al.;* Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders; 2010; Molecular Biosystems; 6; 815-821).

110 µl of diluted nitrated organosolve lignin was added to each well of a 96 well plate, followed by 40µl 0.1mM $FeCl_3$, 10µl of 50uM 2,3dihydroxybenzoic-acid (2,3-DHBA), 30µl recombinant protein of IR1 or IR2 and 10µl 4mM H_2O_2. The assay was monitored at 430nm every minute for 20 minutes and carried out in quadruplicate. The whole plate was repeated as above but with 2,3-DHBA and/or H_2O_2 being replaced by deionized H_20. The bacteria lignin degrading enzyme (dypB) and an *E. coli* GFP construct was used as positive and negative controls respectively (data not shown).

Example 6: The ability of iron reductases to degrade cellulose

The DNS (Dinitrosalicylic Acid) assay was used to demonstrate that IR1 and IR2 proteins have the capacity to degrade cellulose into its component sugars.

The assay was performed in the presence of either IR1 or IR2 and using cellulose in the form of purified cellulose (Avicel) or in the form of lignocellulosic cellulose (wheat straw powder). Both IR1 and IR2 enzymes were observed to degrade both forms of cellulose into their component sugars (see increase in absorbance in Figure 8).

Initial results observed following 1 hour of incubation demonstrated little difference between depolymerisation of the different cellulose types. However, after 24 hours, significant differences were seen, with powdered wheat straw giving higher readings when compare to the Avicel (Figure 8). The average sugar concentration released by recombinant IR1 and IR2 following incubation for 1 and 24 hours using Avicel and wheat straw ranged between 3.89 to 5.58 µg mg^{-1} samples. Recombinant IR2 showed a marginally greater ability to convert straw compared to IR1 despite the lack of a CBM module in the IR2 protein.

These results demonstrate that both IR1 and IR2 have the capacity to degrade cellulose. Based on the similar findings for IR1 (which contains a CBM domain) and IR2 (which lacks a CBM domain), these results indicate that the cellulose binding module present in IR1 is not involved in the cellulose depolymerisation activity.

In addition, these results were obtained in the absence of iron and/ or hydrogen peroxide, indicating that these reagents are not required for cellulose depolymerisation.

Figure 8 shows the results observed for the DNS (Dinitrosalicylic Acid) assay. The total amount of reducing sugars (µg/ml) released following 1 hour and 24 hours incubation of Avicel or straw with IR1 or IR2 is shown. A negative control containing buffer was included in the assay.

Total reducing sugars assay by the recombinant proteins

The CBM assay performed was modified from Hall *et al.*, (2010) and Yoon *et al.*, (2005) (Hall, M. *et al.;* Cellulose crystallinity - a key predictor of the enzymatic hydrolysis rate; FEBS Journal; 2010; 277; 1571-1582; and Yoon, J.J. *et al.;* Degradation of

crystalline cellulose by the brown-rot basidiomycete Fomitopsis palustris; 2005; Journal of Microbiology; 43; 487-492).

The ability of recombinant iron reductase to degrade microcrystalline cellulose was tested using Avicel-PH 101 (Sigma Aldrich) while the depolymerisation of lignocellulosic was detected using wheat straw powder as the substrate. Avicel (30mg/ml) and wheat straw powder (30mg/ml) were incubated in the presence of crude extracts of IR1 or IR2 protein. Incubation was performed for 24 hours at 50°C, pH 7.5. An aliquot of the crude extract (250μl) was taken and centrifuged for 1 min at 13.000 rpm. Total reducing sugars were measured using the DNS (Dinitrosalicylic Acid) method, with glucose as the standard. The absorbance was measured at 540nm using a TECAN GENious spectrophotometer plate reader. The assay was performed in quadruplicate, and negative controls corresponding to buffer alone (ie. no IR1 or IR2 extract) were included.

CLAIMS

1. A method for extracting sugars and/or lignin-derived products from lignocellulosic biomass material, comprising:

contacting the lignocellulosic biomass material with a composition comprising a polypeptide in the presence of a reducible substrate and an oxidising agent, wherein said polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars and/or lignin-derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

2. A method according to Claim 1, wherein the reducible substrate comprises Fe^{3+} ions.

3. A method according to Claim 1, wherein the reducible substrate comprises $FeCl_3$.

4. A method according to any of Claims 1-3, wherein the oxidising agent comprises hydrogen peroxide.

5. A method for extracting sugars from lignocellulosic biomass material, wherein the lignocellulosic biomass material has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure; the method comprising:

contacting the pre-treated lignocellulosic biomass material with a composition comprising a polypeptide, wherein said polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars from the pre-treated lignocellulosic biomass material.

6. A method according to any of Claims 1-5, wherein the polypeptide comprises an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or a fragment thereof.

7. A method according to any of Claims 1-5, wherein the polypeptide comprises an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 2, or a fragment thereof.

8. A method according to any of Claims 1-5, wherein the polypeptide is encoded by a polynucleotide that comprises a nucleic acid sequence having at least 70% identity to the nucleic acid sequence of SEQ ID NO: 3 or 4, or a fragment thereof.

9. A method according to any preceding claim, wherein the composition comprises a recombinant micro-organism that expresses the polypeptide.

10. Use of a polypeptide for extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent, wherein the polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; and wherein said polypeptide is capable of extracting sugars and/or lignin derived products from the lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent.

11. Use according to Claim 10, wherein the reducible substrate comprises Fe^{3+} ions.

12. Use according to Claim 10, wherein the reducible substrate comprises $FeCl_3$.

13. Use according to any of Claims 10-12, wherein the oxidising agent comprises hydrogen peroxide.

14. Use of a polypeptide for extracting sugars from lignocellulosic biomass material that has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure;

wherein the polypeptide comprises an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; and wherein said polypeptide is capable of extracting sugars from the pre-treated lignocellulosic biomass material.

15. Use according to any of Claims 10 to 14, wherein the polypeptide is encoded by a polynucleotide that comprises a nucleic acid sequence having at least 70% identity to the sequence of SEQ ID NO: 3 or 4, or a fragment thereof.

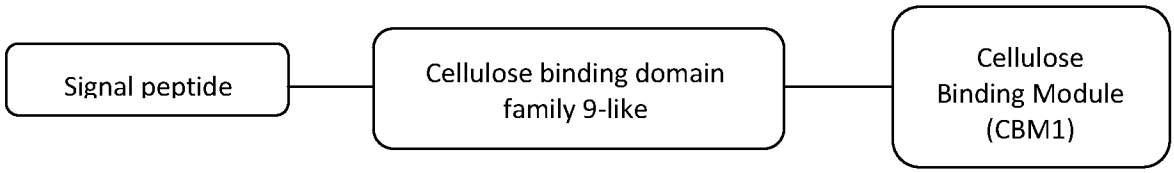
16. A recombinant micro-organism that expresses a polypeptide comprising an amino acid sequence that has at least 70% identity to the amino acid sequence of SEQ ID NO: 1 or 2, or a fragment thereof; wherein said polypeptide is capable of extracting sugars and/or lignin derived products from lignocellulosic biomass material in the presence of a reducible substrate and an oxidising agent; and wherein said polypeptide is capable of extracting sugars from lignocellulosic biomass material that has been subjected to a pre-treatment process that at least partially depolymerises lignin and/ or hemicellulose present in the lignocellulosic structure.

17. A recombinant micro-organism according to Claim 16, wherein the polypeptide is encoded by a polynucleotide that comprises a nucleic acid sequence having at least 70% identity to the sequence of SEQ ID NO: 3 or 4, or a fragment thereof.

18. A method, use or recombinant micro-organism substantially as hereinbefore described, and/ or as illustrated in the Figures and/ or Examples.

Figure 1

(a) IR1 Protein



(b) IR2 Protein

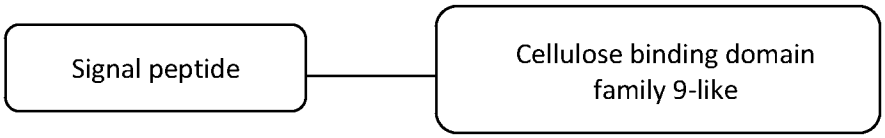
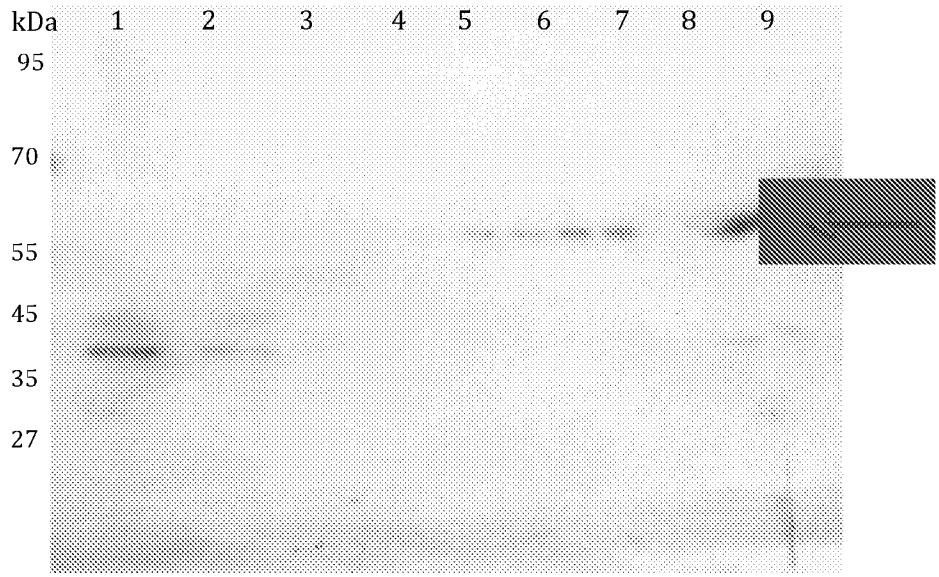


Figure 2

(a) IR1 Protein



(b) IR2 Protein

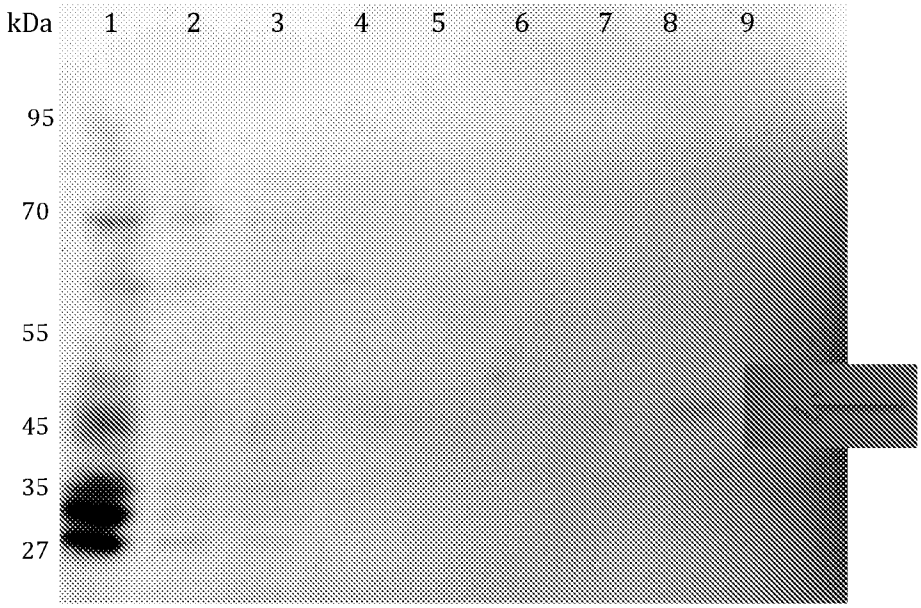


Figure 3

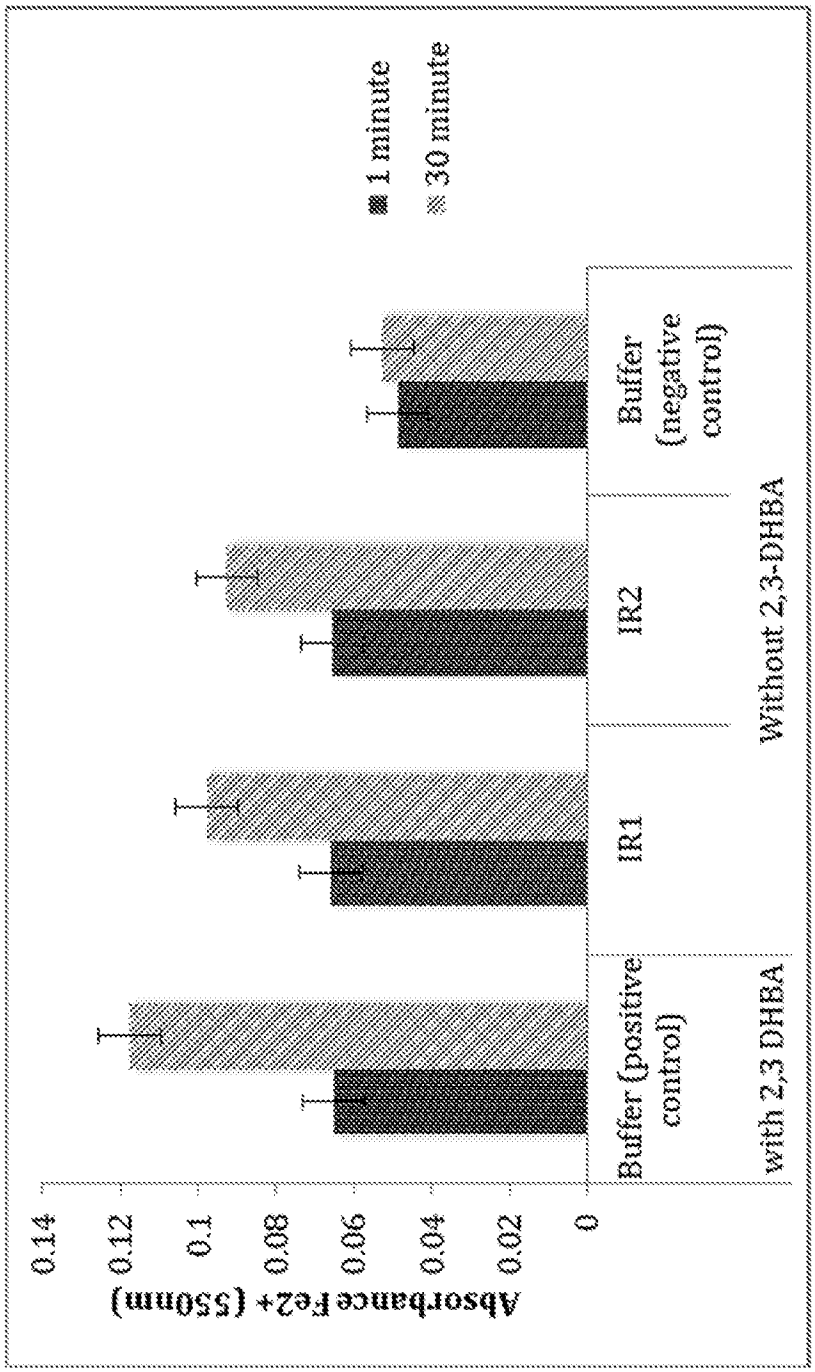

Absorbance Fe2+ (550nm)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 minute
30 minute
Buffer (positive control)
IR1
IR2
Buffer (negative control)
with 2,3 DHBA
Without 2,3-DHBA

Figure 4

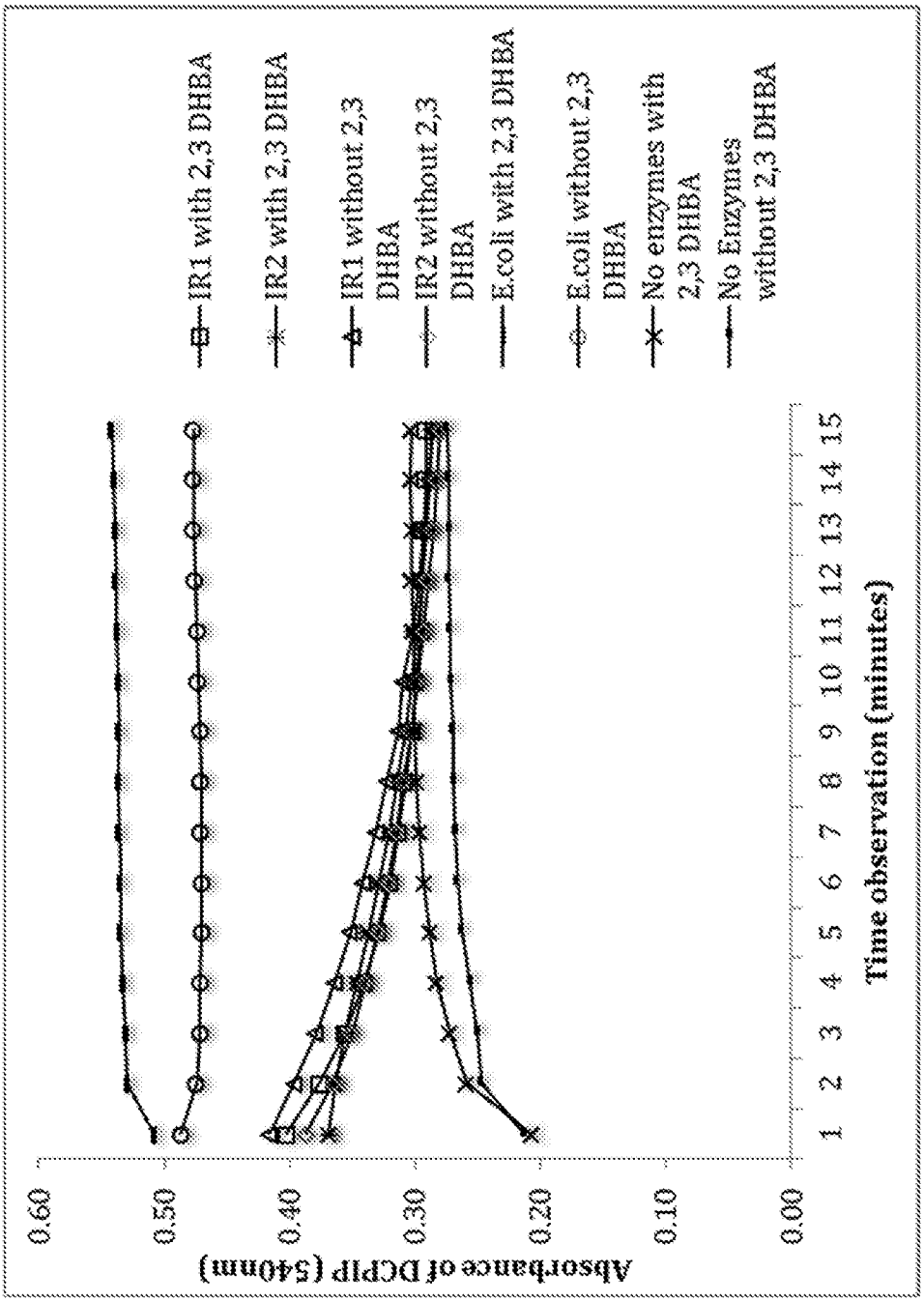

Absorbance of DCPIP (540nm)
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Time observation (minutes)
IR1 with 2,3 DHBA
IR2 with 2,3 DHBA
IR1 without 2,3 DHBA
IR2 without 2,3 DHBA
E.coli with 2,3 DHBA
E.coli without 2,3 DHBA
No enzymes with 2,3 DHBA
No Enzymes without 2,3 DHBA

Figure 5

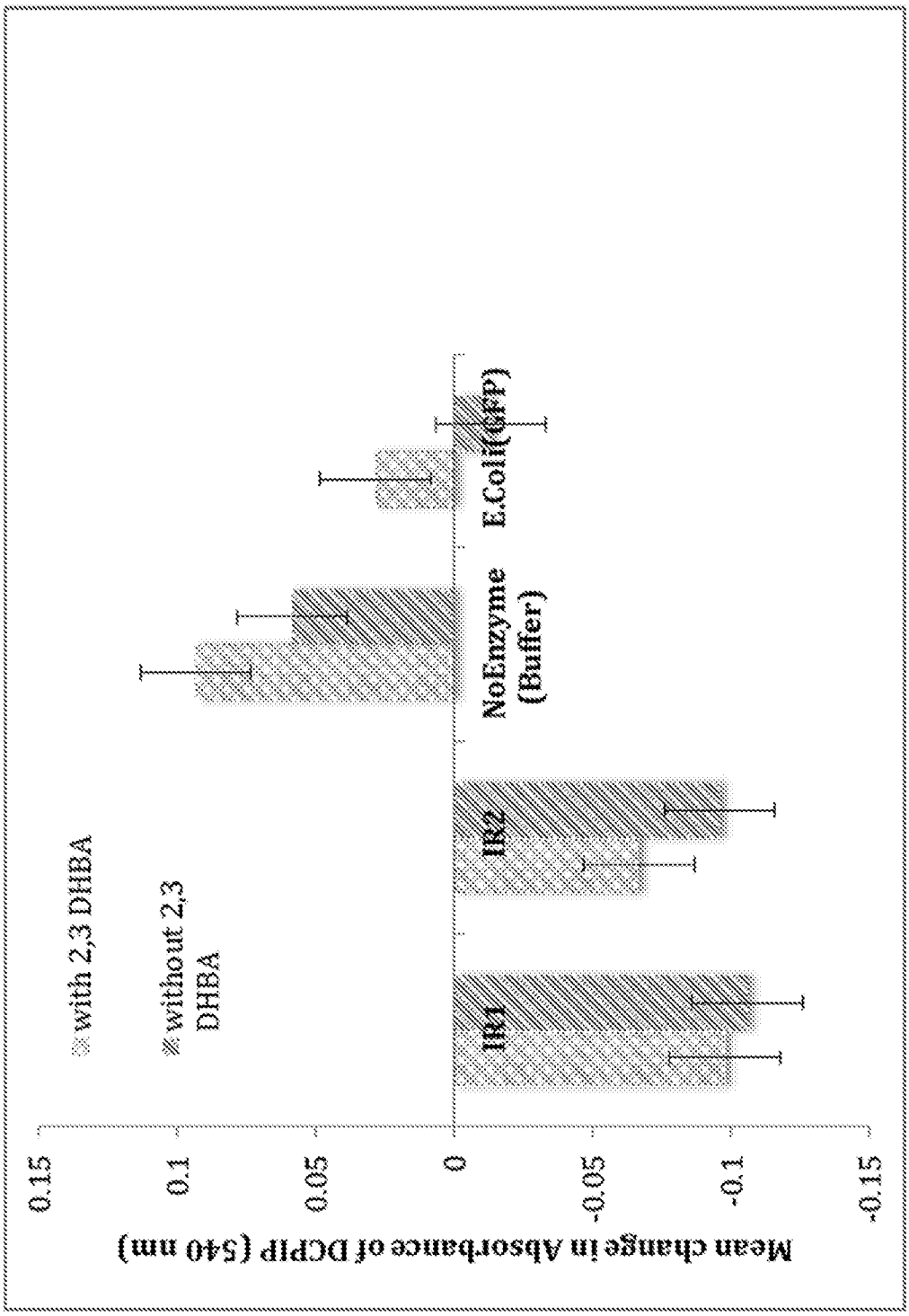

Mean change in Absorbance of DCPIP (540 nm)
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
IR1
IR2
NoEnzyme (Buffer)
E.Coli(GFP)
with 2,3 DHBA
without 2,3 DHBA

Figure 6

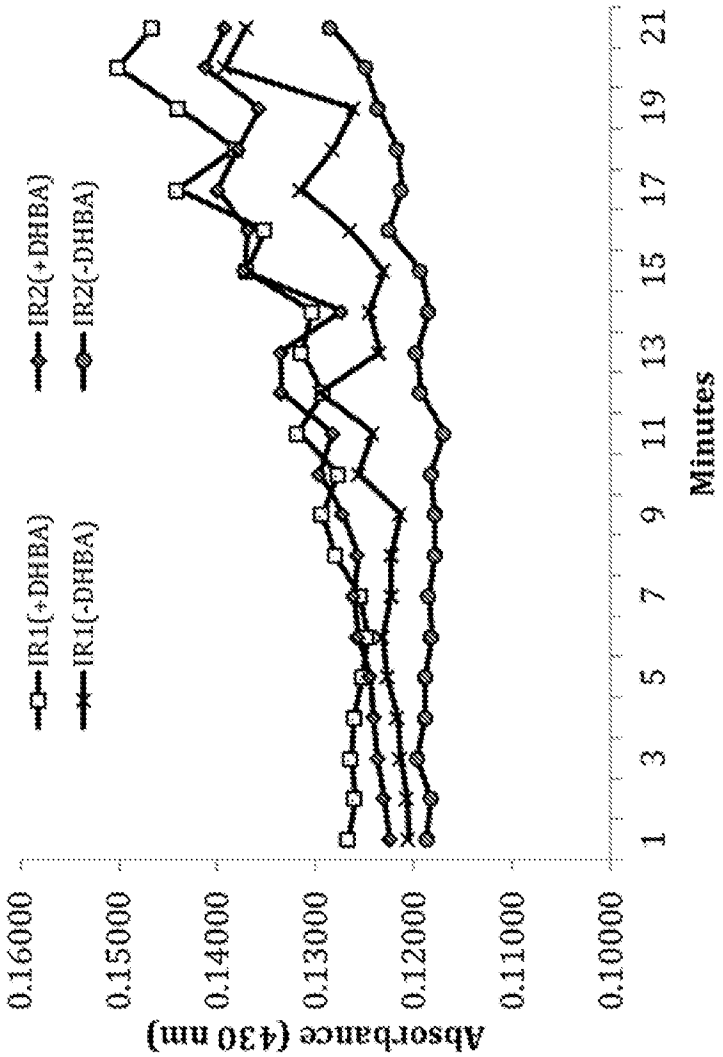

IR1(+DHBA)
IR1(-DHBA)
IR2(+DHBA)
IR2(-DHBA)
Absorbance (430 nm)
0.16000
0.15000
0.14000
0.13000
0.12000
0.11000
0.10000
1
3
5
7
9
11
13
15
17
19
21
Minutes

Figure 7

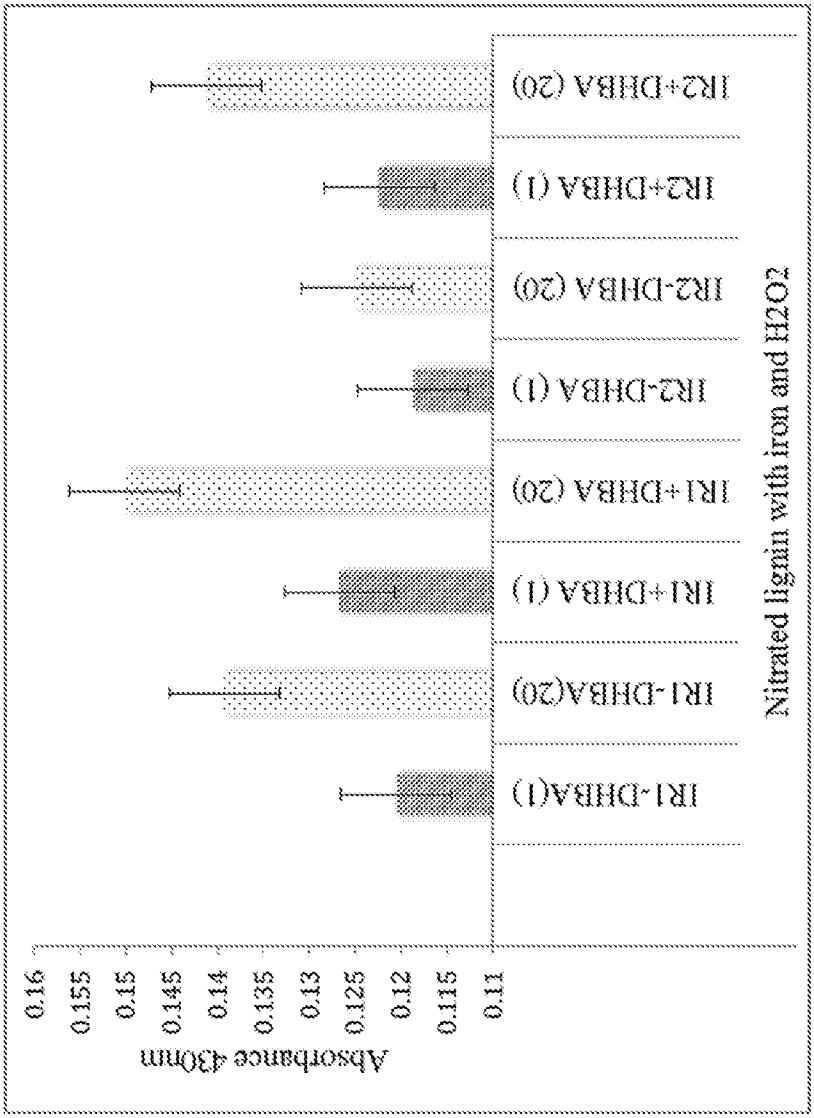
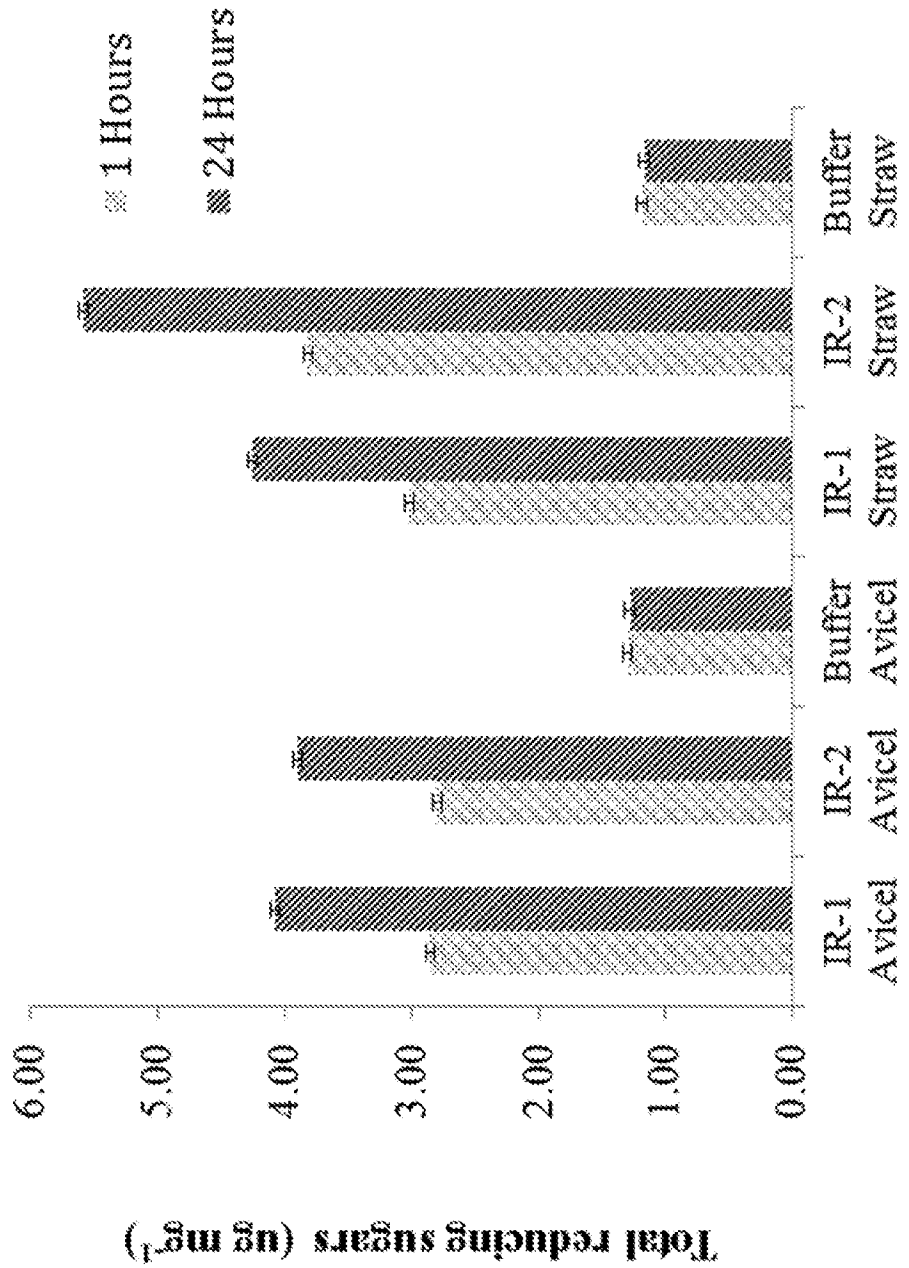

Absorbance 430nm
0.16
0.155
0.15
0.145
0.14
0.135
0.13
0.125
0.12
0.115
0.11
Nitrated lignin with iron and H2O2
IR1-DHBA(1)
IR1-DHBA(20)
IR1+DHBA (1)
IR1+DHBA (20)
IR2-DHBA (1)
IR2-DHBA (20)
IR2+DHBA (1)
IR2+DHBA (20)

Figure 8


1 Hours
24 Hours
IR-1 Avicel
IR-2 Avicel
Buffer Avicel
IR-1 Straw
IR-2 Straw
Buffer Straw
0.00
1.00
2.00
3.00
4.00
5.00
6.00
Total reducing sugars (ug mg-1)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/052219

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P19/02 C07G1/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EASTWOOD DANIEL C ET AL: "The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi", SCIENCE (WASHINGTON D C), vol. 333, no. 6043, August 2011 (2011-08), pages 762-765, XP002761944, cited in the application page 764, right-hand column, paragraph 2 page 765, left-hand column, paragraph 3; figure 3 -/--	1-4,6-8, 10-13,15

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 September 2016	04/10/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Mateo Rosell, A

3

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	& DATABASE UniProt [Online] 21 September 2011 (2011-09-21), "SubName: Full=Carbohydrate-binding module family 1 protein {ECO:0000313¦EMBL:EGO21045.1};", retrieved from EBI accession no. UNIPROT:F8P6H8 Database accession no. F8P6H8 sequence	1-4,6-8, 10-13,15
Y	HYDE SIMON M ET AL: "A mechanism for production of hydroxyl radicals by the brown-rot fungus Coniophora puteana: Fe(III) reduction by cellobiose dehydrogenase and Fe(II) oxidation at a distance from the hyphae", MICROBIOLOGY (READING), vol. 143, no. 1, 1997, pages 259-266, XP002761945, ISSN: 1350-0872 page 261, left-hand column, paragraph 2 - page 263, right-hand column, paragraph 2; figure 7	1-18
Y	WO 2013/004377 A2 (AGRONOMIQUE INST NAT RECH [FR]; INST FRANCAIS DU PETROLE EN NOUVELLE []) 10 January 2013 (2013-01-10) page 1, paragraph [0001] page 3, paragraph [0009] page 38, paragraph [000192] - page 40, paragraph [000199]; example 3 claims 1-6; sequence 1	1-18
Y	WO 2012/138772 A1 (UNIV CALIFORNIA [US]; MARLETTA MICHAEL A [US]; DOUDNA CATE JAMES H [US]) 11 October 2012 (2012-10-11) page 3, paragraphs [0010], [0016] page 4, paragraph [0020] page 6, paragraph [0025] page 7, paragraph [0027] page 8, paragraph [0031]; claims 5, 18-23; sequences 14,16,18	1-18

-/--

3

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	P. KORRIPALLY ET AL: "Evidence from Serpula lacrymans that 2,5-Dimethoxyhydroquinone Is a Lignocellulolytic Agent of Divergent Brown Rot Basidiomycetes", APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 79, no. 7, 1 April 2013 (2013-04-01), pages 2377-2383, XP055302607, US ISSN: 0099-2240, DOI: 10.1128/AEM.03880-12 cited in the application the whole document -----	1-18
A	SHIMOKAWA T ET AL: "Production of 2,5-dimethoxyhydroquinone by the brown-rot fungus Serpula lacrymans to drive extracellular Fenton reaction", HOLZFORSCHUNG 2004 WALTER DE GRUYTER AND CO. DE, vol. 58, no. 3, 2004, pages 305-310, XP002761946, DOI: 10.1515/HF.2004.047 the whole document -----	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2016/052219

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013004377 A2	10-01-2013	EP 2729575 A2	14-05-2014
		US 2014212927 A1	31-07-2014
		WO 2013004377 A2	10-01-2013
WO 2012138772 A1	11-10-2012	AU 2012240229 A1	17-10-2013
		CA 2831954 A1	11-10-2012
		CN 103814128 A	21-05-2014
		EP 2694650 A1	12-02-2014
		JP 2014512181 A	22-05-2014
		US 2014073012 A1	13-03-2014
		US 2016168609 A1	16-06-2016
		WO 2012138772 A1	11-10-2012