



- (51) International Patent Classification:
C12P 19/00 (2006.01) *C12N 9/24* (2006.01)
- (21) International Application Number:
PCT/US2015/042126
- (22) International Filing Date:
24 July 2015 (24.07.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/028,737 24 July 2014 (24.07.2014) US
- (71) Applicants: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US). **BP CORPORATION NORTH AMERICA INC.** [US/US]; 501 Westlake Park Boulevard, Houston, TX 77079-2607 (US). **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 N. Torrey Pines Road, La Jolla, CA 92037 (US).
- (72) Inventors: **MARLETTA, Michael, A.**; 1672 Via Corona, La Jolla, CA 92037 (US). **BEESON, William, T.**; 8628 Wildview Court, Indianapolis, IN 46278 (US). **VU, Van, V.**; 8510 Costa Verde Blvd., Apt. 2413, La Jolla, CA 92122 (US). **SPAN, Elise, A.**; 13850 Mango Drive, Apt. 17, La Jolla, CA 92014 (US). **PHILLIPS, Christopher, M.**; 17161 Alva Road, Unit 2824, San Diego, CA 92127 (US).
- (74) Agents: **WARD, Michael, R.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).
- (81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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,

[Continued on next page]

- (54) Title: OXIDATIVE STARCH DEGRADATION BY A NEW FAMILY OF PMOS

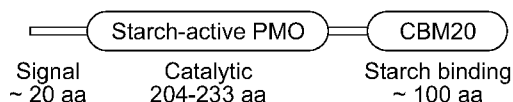


FIG. 2A

(57) Abstract: Described herein are methods for degrading starch-containing material, including the steps of providing a starch-containing material and providing a polypeptide having starch-degrading activity and a copper cofactor. The method further comprises incubating the polypeptide with the starch containing material in the presence of oxygen and a reductant.

*
1 HGYLETIPSRTRLGFEAGIDICPECTILEPVI³⁴AW
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
35 PDLDA³⁵AVGRSGPCGYNARVSYDYNQPC¹⁰²INWGN¹⁰³
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68
69 PV⁶⁹TYT⁷⁰GRYYEVQWCVD¹⁰²ENGDHGGMFTYR¹⁰³ICQ¹⁰⁴
70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102
103 QAL¹⁰³VRKEL¹⁰⁴IPSYLPT¹⁰⁵AEKQAAEDCF¹⁰⁶GL¹⁰⁷LCI¹⁰⁸
104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
137 DV¹³⁷GQ¹³⁸ICGY¹³⁹PD¹⁴⁰CGAC¹⁴¹WRNDWFTCN¹⁴²AF¹⁴³
138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
171 KRR¹⁷¹CGVDNA¹⁷²ELNSCY¹⁷³TS¹⁷⁴AGGYIVIKK¹⁷⁵IK¹⁷⁶PR¹⁷⁷
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204
205 Y²⁰⁵SNHTLL²⁰⁶SF²⁰⁷WNSFQT²⁰⁸SQ²⁰⁹YL²¹⁰CAD²¹¹AI²¹²
206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233

FIG. 2B



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.

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).

(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,

Published:

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

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

OXIDATIVE STARCH DEGRADATION BY A NEW FAMILY OF PMOS**CROSS-REFERENCE TO RELATED APPLICATION(S)**

[0001] This application claims the benefit of U.S. Provisional Application No. 62/028,737, filed on July 24, 2014, which is incorporated herein by reference in its entirety.

SUBMISSION OF SEQUENCE LISTING ON ASCII TEXT FILE

[0002] The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 677792005040SEQLIST.TXT, date recorded: July 23, 2015, size: 242 KB).

FIELD

[0003] The present disclosure relates generally to methods for degrading starch containing material, and more specifically to methods of degrading starch containing material by providing a polypeptide and incubating the polypeptide with the starch containing material in the presence of oxygen and a reductant.

BACKGROUND

[0004] Starch is a carbohydrate polymer that has been shown to be a good source for food products, biofuels and other applications. It is the major energy reserve in plants and the most important energy source in the human diet (Christiansen *et al.*, *FEBS J*, 2009). There are many products that rely on the use of starch, which include food for humans, cattle and other animals. Starch can also be used in the paper industry and to produce dextrin – a substance commonly used in making syrup. Furthermore starch can be fermented to produce biofuel alcohol.

[0005] Developing a sustainable energy industry is of key importance to achieve energy security, large-scale substitution of petroleum-based fuels and reduce carbon footprint (Farrell *et al.*, *Science*, 2006). Biofuels are under intensive investigation due to increasing concerns about energy security, sustainability, and global climate change (Lynd *et al.*, *Science*, 1991).

Bioconversion has been regarded as an attractive alternative to chemical production of fossil fuels (Lynd *et al.*, *Nat Biotech*, 2008; Hahn-Hagerdal *et al.*, *Biotechnol Biofuels*, 2006). Enzymes and organisms that can degrade simple sugars have been known and used for many years.

However, scaling and sustainability problems pose a need for the use of complex biomass polymers to be used for biofuels (Harris *et al.*, *Biochem.* 2010). Thus, enzymes useful for the degradation of complex biomass polymers have been sought after in those organisms that naturally degrade biomass such as *Neurospora crassa* and *Trichoderma reesei*.

[0006] Starch has a complex structure composed of two distinct glucose polymers: amylose, comprising essentially unbranched α -(1 \rightarrow 4)-linked glucose residues, and amylopectin, comprising α -(1 \rightarrow 6) linkages between adjoining straight glucan chains on an α -(1 \rightarrow 4) backbone (Christiansen *et al.*, *FEBS J*, 2009). Because of its complex structure, the enzymatic degradation of starch poses a considerable challenge to the attacking enzymes, as the polysaccharide chains are often poorly accessible to the active sites of the enzymes (Buléon *et al.*, *Int J Biol Macromol*, 1998). Due to its importance and because of its complex structure, new ways for the degradation of starch would be highly beneficial.

[0007] Thus, a need exists for additional enzymes that can degrade starch- containing materials as this would have applications in a wide array of industries, including food & starch-based biofuel industries.

BRIEF SUMMARY

[0008] Fulfilling this need, methods for degrading starch containing material including, but not limited to the steps of providing starch containing material, providing a polypeptide comprising SEQ ID NO: 1, wherein the polypeptide is bound to a copper cofactor; and incubating the starch containing material and polypeptide in the presence of oxygen and a reductant, thereby degrading the starch containing material, are provided.

[0009] In certain embodiments, the polypeptide used in the method includes an amino acid sequence of SEQ ID NO: 6. In other embodiments, the polypeptide includes the amino acid sequence of any of SEQ ID NOs: 8- 37 or SEQ ID No; 87. In further embodiments, the polypeptide described above further include a carbohydrate binding module 20 (CBM20).

[0010] In other embodiments, the polypeptide used in the method includes the amino acid sequence of SEQ ID NO: 7 or the amino acid sequence of any of SEQ ID NOs: 38-86.

[0011] In still other embodiments, the polypeptide used in the method includes an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 6, an amino acid sequence having at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 6, an amino acid sequence encoded by a polynucleotide that hybridizes under medium stringency conditions with the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6 or the complementary strand of the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6, an amino acid sequence encoded by a polynucleotide that hybridizes under high stringency conditions with the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6 or the complementary strand of the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6, an amino acid sequence having a substitution, deletion, and/or insertion of up to 40 amino acids of the amino acid sequence of SEQ ID NO: 6, an amino acid sequence having a

substitution, deletion, and/or insertion of up to 20 amino acids of the amino acid sequence of SEQ ID NO: 6, or an amino acid sequence having a substitution, deletion, and/or insertion of up to 10 amino acids of the amino acid sequence of SEQ ID NO: 6.

[0012] In any of the embodiments described above, the reductant used in the method may be an iron-containing compound. This iron-containing compound may comprise a heme domain and in some embodiments is cellobiose dehydrogenase (CDH). In other embodiments, the iron-containing compound is ferrous sulfate. In other embodiments of the method described above, the reductant is an arene or ascorbic acid.

[0013] In any of the embodiments described above, the method may include additional steps, such as the further step of providing copper, providing amylase, or providing oxygen.

[0014] In any of the embodiments described above, at least 90% of the polypeptide may be bound by copper. In further embodiments, the incubating takes places at a temperature ranging from 25°C to 70°C.

[0015] In another aspect, the methods described herein relate to producing a fermentation product, including the steps of degrading starch-containing material according to the methods described above to form glucose; and culturing the glucose with one or more fermentative microorganisms or a chemical solution under conditions sufficient to produce a fermentation product.

[0016] In another aspect, compositions including a polypeptide including the amino acid sequence of SEQ ID NO: 6 and copper; and a reductant are provided. In some embodiments, the reductant is an iron-containing compound. The iron-containing compound may include a heme domain and may be cellobiose dehydrogenase (CDH). Alternatively, the iron-containing compound is ferrous sulfate. In other embodiments, the reductant is an arene.

In other embodiments, the reductant is ascorbic acid.

BRIEF DESCRIPTION OF THE FIGURES

[0017] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the office upon request and payment of the necessary fee.

[0018] **FIG. 1A** is representative of overall and active site structures of fungal PMOs (PDB ID 2YET), as previously described in Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011. The available structures of PMOs from all three previously known families reveal a conserved fold, including an antiparallel β -sandwich core and a highly conserved monocopper active site on a flat protein surface. Two histidine residues in a motif termed the *histidine brace* coordinate the copper center. The *N*-terminal histidine ligand binds in a bidentate mode, and its imidazole ring is methylated at the *N* ϵ position in fungal PMOs.

[0019] **FIG. 1B** illustrates the structure of cellulose as previously described in Nishiyama *et al.*, *J Am Chem Soc.* 2002 and Nishiyama *et al.*, *J Am Chem Soc.*, 2003. As illustrated, Cellulose and chitin contain long linear chains of $\beta(1\rightarrow4)$ linked glucose units and N-acetylglucosamine units, respectively.

[0020] **FIG. 1C** illustrates the model structure of amylopectin (Perez and Bertoft, *Stärke*, 2010, Popov *et al.*, *Macromolecules*, 2009 and Imberty *et al.*, *J Mol Biol.* 1988.) Hydrogen bonds are shown with green dashed lines. As illustrated, unlike cellulose and chitin, amylose and amylopectin do not form microcrystals; instead, they exist in disordered, single helical, and double helical forms.

[0021] **FIG. 2A and FIG. 2B** illustrate common domain architecture of 43 predicted starch-active PMOs from different fungal species (**SEQ ID NOs: 1-2**). Thirty one have the CBM20 domain.

[0022] **FIG. 3** illustrates data collected from the activity assays of NCU08746. The data show the high performance anion exchange chromatographic (HPAEC) traces of NCU08746 assays under various conditions. Assays contained 5 μ M NCU08746 with 2 mM ascorbic acid and atmospheric oxygen. **Traces A-B**: Maltodextrins (1-7 units) and soluble portion of amylose (average molecular weight \sim 2.8 kDa), respectively, oxidized with Lugol's solution. **Traces C-E**: Assays with 50 mg/mL amylopectin, 5 mg/ml PASC, and 50 mg/mL chitin, respectively. The assays were carried out in 50 mM sodium acetate buffer at pH 5.0 and 42 °C.

[0023] **FIG. 4A** illustrates the effect of NCU08746 (5 μ M) on the rate of oxidation of *MtCDH-2* (1 μ M) incubated with 6 μ M cellobiose at room temperature. The data indicates that this oxidation occurs slowly in the presence of atmospheric oxygen but was significantly enhanced in the presence of NCU08746.

[0024] **FIG. 4B** illustrates NCU08746 (5 μ M) activity assays on amylopectin (50 mg/mL) with 2 mM ascorbic acid (A), with 0.5 μ M *MtCDH-2* and 5 mM cellobiose (B), and with 0.5 μ M *MtCDH-2* only (C). As the data show, oxidation occurs slowly in the presence of atmospheric oxygen but was significantly enhanced in the presence of NCU08746.

[0025] **FIG. 5** illustrates k^3 -weighted EXAFS data of Cu(II)-NCU08746 and its Fourier transform. The boxes highlight the features that arise from the outer-shell atoms of the imidazole ligands. Best fit parameters are provided in Table 5 (Fit 9). The Fourier transform exhibits a strong inner shell feature at $r' \sim 1.5$ Å, a second shell feature at $r' \sim 2.3$ Å, and third shell features between $r' = 2.3$ -3.7 Å. The third shell features in the Fourier transform correspond to

the double-humped feature centered at $\sim 4 \text{ \AA}^{-1}$ in the EXAFS spectrum, which arises from the multiple scattering paths of several imidazole moieties as found in the spectra of many other metalloproteins and model complexes (Vu *et al.*, *J Am Chem Soc*, 2011, Pellei *et al.*, *Dalton Trans*, 2011, Sanyal *et al.*, *J Am Chem Soc*, 1993, Costello *et al.*, *J Biol Inorg Chem*, 2006, D'Angelo *et al.*, *Biochemistry*, 2005). Fitting progress is shown in Table 5. The best fit includes 4 Cu-O/N paths at 1.97 Å, 1 Cu-O/N path at 2.22 Å, 1 Cu-O/N path at 2.42 Å, 2 Cu-C paths at 3.23 Å, and paths from 2.3 imidazole moieties (Table 5).

[0026] **FIG. 6** illustrates a proposed mechanism of NCU08746, analogous to that of cellulose-active PMOs and chitin-active PMOs. NCU08746 is shown here to oxidize the C1 position in both amylose and amylopectin. Illustrated are the mechanisms of NCU08746 involving the cleavages of $\alpha(1\rightarrow4)$ (top) and $\alpha(1\rightarrow6)$ (bottom) linkages via hydroxylation at the C1 position.

[0027] **FIG. 7** illustrates SDS-PAGE analysis of fractions eluted from an amylose resin column (left) and subsequent size exclusion column (S75, GE Healthcare) (right). As illustrated, SDS-PAGE analysis of the fractions from the amylose resin column shows that NCU08746 has relatively strong affinity for amylose.

[0028] **FIG. 8** illustrates fragment mass spectrum of the N-terminal peptide. This data is consistent with a methylated N-terminal histidine. The following were used: **Monoisotopic mass of neutral peptide Mr(calc):** 1203.6400; **Fixed modifications:** Carbamidomethyl (C) (apply to specified residues or termini only; **Variable modifications:** **H1** : Methyl-Histidine; **Ions Score:** 46 **Expect:** 0.00026; **Matches** : 28/80 fragment ions using 74 most intense peaks.

[0029] **FIG. 9** illustrates activity assays of Cu-NCU08746 (5 μM , Enz) with amylopectin (50 mg/mL, AmPe), in the presence (+) or absence (-) of ascorbic acid (2 mM, Asc), air-

saturated oxygen (O_2), and $CuSO_4$ (10 μM , Cu). **Traces A-B:** Maltodextrins (1–7 units) and soluble portion of amylose (average molecular weight ~ 2.8 kDa), respectively, oxidized with Lugol's solution. **Trace C:** AmPe+Enz+Asc+ O_2 . **Trace D:** AmPe-Enz+Cu+Asc+ O_2 . **Trace E:** AmPe+Enz-Asc+ O_2 . **Trace F:** AmPe+Enz+Asc- O_2 . **G:** AmPe+Enz-Asc- O_2 . **Traces H-I:** PASC and chitin, respectively, +Enz+Asc+ O_2 . The assays were carried out in 50 mM sodium acetate buffer at pH 5.0 and 42°C. When both ascorbic acid and oxygen were present, the chromatogram exhibits a set of new peaks (**trace C**) that are not observed in the absence of either reductant or oxygen.

[0030] **FIG. 10** illustrates NCU08746 activity assays with amylopectin from maize. AmPe = 50 mg/mL amylopectin. E = 5 μM Cu-NCU08746. Asc = 2 mM ascorbic acid. Cu = 5 μM $CuSO_4$. As shown, the addition of one equivalent of $CuSO_4$ to the assays of holo NCU08746 had no effect on activity.

[0031] **FIG. 11** illustrates NCU08746 activity assays with corn starch. CoSt = 50 mg/mL corn starch. E = 5 μM Cu-NCU08746. Asc = 2 mM ascorbic acid. Cu = 5 μM $CuSO_4$. As shown, the addition of one equivalent of $CuSO_4$ to the assays of holo NCU08746 had no effect on activity.

[0032] **FIG. 12** illustrates A: The soluble portion of a 50 mg/mL suspension of amylose with average molecular weight of 2.8 kDa (AM2.8). B: AM2.8 oxidized with Lugol's solution. C: A mixture of maltodextrins with 1–7 glucose units. D: C oxidized with Lugol's solution. E: NCU08746 assay with AM2.8 (50 mg/mL suspension). F: NCU08746 assay with amylopectin (injected 20 times more than the assay with AM2.8). The product peaks of NCU08746 do not overlay with the peaks of maltodextrins.

[0033] **FIG. 13** illustrates amylopectin activity assays of partially-apo NCU08746 in the presence of various metal ions. Among the metal ions tested, only Cu(II) increased the oxidative activity of this partially apo NCU08746.

[0034] **FIG. 14** illustrates X-ray absorption near edge spectrum (XANES) of Cu(II)-NCU08746.

[0035] **FIG. 15** illustrates FEFF model for EXAFS fitting. Theoretical scattering paths were calculated with FEFF8.4 (Rehr *et al.*, *J Am Chem Soc*, 1991) using this model. The r and σ^2 values of all significant single and multiple scattering paths of the imidazole moiety were linked together in the fit.

[0036] **FIG. 16** illustrates the activity of truncated NCU08746 lacking the CBM20 domain (Trunc1): **Line A:** 4 μ M full length NCU08746 **Line B:** 4 μ M Trunc1 and 10 μ M CuSO₄. **Line C:** 4 μ M Trunc1 only; **Line D:** 10 μ M CuSO₄ only.

[0037] **FIG. 17** illustrates the activity of the full-length NCU08746 vs. Trunc 1. Numbers correspond to sums of peak areas assigned to malto-aldonic acid products. Assay conditions included 5 μ M fully reconstituted enzyme; 2 mM ascorbic acid, 50 mM sodium acetate pH 5.0 and 50 mg/ml substrate (CoAmPe = amylopectin from corn; CoSt = corn starch; PoAM = potato amylose). The enzyme and substrate were incubated in 96-well plate, 1200 rpm shaking, at 42 C for 4 hours.

[0038] **FIG. 18** illustrates the X-band EPR spectra of wild type NCU08746 and two truncation mutants (Trunc 1 and Trunc 2).

[0039] **FIG. 19** illustrates the activity of full-length NCU08746 enzyme on amylose (source: potato), starch (amylose + amylopectin) or amylopectin from corn. Note the amylose trace is a 1/10 dilution. Assay conditions included 5 μ M fully reconstituted enzyme; 2 mM

ascorbic acid, 50 mM sodium acetate pH 5.0 and 50 mg/ml substrate (CoAmPe = amylopectin from corn; CoSt = corn starch; PoAM = potato amylose). The enzyme and substrate were incubated in 96-well plate, 1200 rpm shaking, at 42 C for 4 hours.

DETAILED DESCRIPTION

[0040] The following description is presented to enable a person of ordinary skill in the art to make and use the various embodiments. Descriptions of specific devices, techniques, and applications are provided only as examples. Various modifications to the examples described herein will be readily apparent to those of ordinary skill in the art, and the general principles defined herein may be applied to other examples and applications without departing from the spirit and scope of the various embodiments. Thus, the various embodiments are not intended to be limited to the examples described herein and shown, but are to be accorded the scope consistent with the claims.

[0041] The present disclosure relates to methods for degrading starch containing material and more specifically, to methods for degrading such material using a new family of polysaccharide monooxygenases (PMOs). In particular, the present disclosure is based, at least in part, on Applicants' surprising discovery of a new function of a novel family of PMOs that are able to degrade starch, *i.e.*, cleave glycosidic bonds of starch.

[0042] PMOs are secreted by a variety of fungal and bacterial species and have been found to degrade chitin and cellulose by oxidizing either the C1 or C4 atom of the $\beta(1\rightarrow4)$ glycosidic bond. Thus far three families of PMOs have been characterized: fungal PMOs that oxidize cellulose; bacterial PMOs that are active either on chitin or cellulose; and fungal PMOs that oxidize chitin. Sequence homology between these three families is very low. Nevertheless,

the available structures of PMOs from all three families reveal a conserved fold, including an antiparallel β -sandwich core and a highly conserved monocopper active site on a flat protein surface. The *N*-terminal histidine ligand binds in a bidentate mode, and its imidazole ring is methylated at the *N* ϵ position in fungal PMOs.

[0043] These currently known PMOs act on substrates with structurally similar structures. Cellulose and chitin contain long linear chains of $\beta(1\rightarrow4)$ linked glucose units and *N*-acetylglucosamine units, respectively. The polymer chains form extensive hydrogen bonding networks, which result in insoluble and very stable crystalline structures. PMOs are thought to bind to the substrate with their flat active site surface, which orients the copper center for selective oxidation at the C1 or C4 position. Some bacterial chitin-binding proteins are cellulose-active PMOs, further suggesting that the set of PMO substrates is restricted to $\beta(1\rightarrow4)$ linked carbohydrate polymers.

[0044] Applicants have discovered that the oxidative mechanism of glycosidic bond cleavage used by PMOs on chitin and cellulose is more widespread than initially expected, also acting on starch despite its structural differences from chitin and cellulose. Starch is made up of amylose and amylopectin. Both amylose and amylopectin contain linear chains of $\alpha(1\rightarrow4)$ linked glucose, while the latter also contains $\alpha(1\rightarrow6)$ glycosidic linkages at branch points in the otherwise $\alpha(1\rightarrow4)$ linked polymer. Unlike cellulose and chitin, amylose and amylopectin do not form microcrystals; instead, they exist in disordered, single helical, and double helical forms. Starch can exist partially in nanocrystalline form, but lacks the flat molecular surfaces as those found in chitin and cellulose.

[0045] Described herein are methods for degrading starch-containing material, including the steps of providing a starch-containing material and providing a polypeptide having starch-

degrading activity and a copper cofactor. The method further comprises incubating the polypeptide with the starch containing material in the presence of oxygen and a reductant.

Polypeptides of the Disclosure

[0046] As used herein a “polypeptide” is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues (*i.e.*, at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues). As used herein, “protein” refers to an amino acid sequence, oligopeptide, peptide, polypeptide, or portions thereof whether naturally occurring or synthetic. “Polypeptide” refers to an amino acid sequence, oligopeptide, peptide, protein, or portions thereof, and the terms “polypeptide” and “protein” are used interchangeably. The polypeptide optionally comprises modified amino acid residues, naturally occurring amino acid residues not encoded by a codon, and non-naturally occurring amino acid residues. The polypeptide can be recombinant or purified. A recombinant polypeptide or recombinant protein refers to polypeptides produced by recombinant DNA techniques. These techniques are known to one skilled in the art and include recombinant polypeptides obtained from cells transformed by an exogenous DNA construct encoding the desired polypeptide.

[0047] The polypeptides of the disclosure may include conserved motifs specific to starch-active PMOs, such as **SEQ ID NOs: 3-5**, either individually or in combination, and a copper cofactor.

[0048] In other embodiments of the present disclosure, the polypeptide includes a conserved sequence of a starch-active PMO catalytic domain, **SEQ ID NO: 1** or **SEQ ID NO: 2**. In another embodiment of the present disclosure, the polypeptide comprises a fragment of the conserved sequence of a starch-active PMO catalytic domain.

[0049] In one embodiment, the polypeptide includes the amino acid sequence of the NCU08746 protein without the signal sequence (**SEQ ID NO: 7**).

[0050] In another embodiment, the polypeptide includes the amino acid sequence of a truncated version of the NCU08746 protein lacking its CBM20 domain (**SEQ ID NO: 6** or **SEQ ID NO: 87**).

[0051] In other embodiments the polypeptide includes the amino acid sequence of a truncated version of the NCU08746 hypothetical protein lacking its CBM20 domain (**SEQ ID NO: 6**) along with a carbohydrate binding domain (CBM) other than the one found in the native sequence. Carbohydrate binding domains present in starch- active enzymes that function as attachment modules between the enzymes and the starch granules or other high molecular weight substrates and have also been suggested to distort the conformation and packing of the polymers, thereby facilitating their degradation. Accordingly, in certain embodiments, the polypeptides described herein include CBM21, CBM48, and/or CBM53 domain(s).

[0052] In embodiment the polypeptide includes the amino acid sequence of a homologous sequence from any species in the Ascomycota phylum. One skilled in the art will recognize that polypeptides can display structural homology and sequence homology. Either or both can be used to predict common mechanisms of action between the polypeptides. Proteins are three dimensional structures for which structure and function are very tightly related. Structure is much more evolutionarily conserved than sequence. Amino acids which are part of the same catalytic domain tend to be conserved, even when not sequential. Likewise, secondary structural elements in a polypeptide are highly conserved as are their arrangement in tertiary structural motifs. Structural alignment programs such as DALI can use the 3D structure of a polypeptide to find proteins with similar folds (Holm *et al.*, *Protein Science*, 1992). Sequence

homology can be used to determine conserved sequential amino acids. Methods for the alignment of sequences and for the analysis of similarity and identity of polypeptide sequences are well-known in the art. In certain embodiments, the polypeptide includes the amino acid sequence of a NCU08746 homolog provided in **SEQ ID NOs: 8 – 86**. **SEQ ID NOs 8 to 37** include a CBM20 domain. **SEQ ID NOs: 38-86** lack a CBM20 domain. In further embodiments, the polypeptides described herein include the amino acid sequence of a homolog lacking a CBM20 domain coupled with a carbohydrate binding motif. In particular embodiments the polypeptide sequence includes the amino acid sequence of the homologue of NCU08746 found in *Myceliophthora thermophila*, GI:347014680 (**SEQ ID NO: 34**).

[0053] Starch has a complex structure composed of two distinct glucose polymers: amylose, comprising essentially unbranched α -(1 \rightarrow 4)-linked glucose residues, and amylopectin, comprising α -(1 \rightarrow 6) linkages between adjoining straight glucan chains on an α -(1 \rightarrow 4) backbone (Christiansen *et al.*, *FEBS J*, 2009). A starch-active PMO catalytic domain has starch glycosidic activity. In some examples, the starch-active PMO catalytic domain cleaves the C1 position in both amylase and amylopectin. In other examples the polypeptide cleaves the C4 and/or C5 positions in amylase and amylopectin. In one embodiment, the polypeptide having a starch-active PMO catalytic domain or a fragment of a starch-active PMO catalytic domain can cleave only the α (1 \rightarrow 4) bond. In another embodiment the polypeptide having a starch-active PMO catalytic domain can attack both α (1 \rightarrow 4) and α (1 \rightarrow 6) linkages. One skilled in the art would recognize that a starch-active PMO catalytic domain might cleave glycosidic bonds in starch in other similar ways not described herein.

Sequence Identity, Alignments, and Variants

[0054] As used herein “sequence identity” refers to the percentage of residues that are identical in the same positions in the sequences being analyzed. As used herein “sequence similarity” refers to the percentage of residues that have similar biophysical / biochemical characteristics in the same positions (*i.e.* charge, size, hydrophobicity) in the sequences being analyzed.

[0055] Methods of alignment of sequences for comparison are well-known in the art, including manual alignment and computer assisted sequence alignment and analysis. This latter approach is a preferred approach in the present disclosure, due to the increased throughput afforded by computer assisted methods. As noted below, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

[0056] In one embodiment the polypeptide comprises an amino acid sequence having at least 99%, at least 98%, at least 97%, at least 96%, at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, or at least 50% sequence identity to the sequences described herein including **SEQ ID NOs: 1-2 and 6-87**.

[0057] The determination of percent sequence identity and/or similarity between any two polypeptide sequences can be accomplished using a mathematical algorithm. Examples of such mathematical algorithms are the algorithm of Myers and Miller, *CABIOS*, 1988; the local homology algorithm of Smith *et al.*, *Adv. Appl. Math.*, 1981; the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.*, 1970; the search-for-similarity-method of Pearson and Lipman, *Proc. Natl. Acad. Sci.*, 1988; the algorithm of Karlin and Altschul, *Proc. Natl. Acad. Sci. USA*, 1990, modified as in Karlin and Altschul, *Proc. Natl. Acad. Sci. USA*, 1993.

[0058] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity and/or similarity. Such implementations include, for example: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the AlignX program, version 10.3.0 (Invitrogen, Carlsbad, CA) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. *Gene*, 1988; Higgins *et al. CABIOS*, 1989; Corpet *et al., Nucleic Acids Res.*, 1988; Huang *et al. CABIOS*, 1992; and Pearson et al., *Meth. Mol. Biol.*, 1994. The BLAST programs of Altschul et al. *J. Mol. Biol.*, 1990, are based on the algorithm of Karlin and Altschul, 1990 *supra*.

[0059] The present disclosure also relates to variations in the polypeptide occurring naturally or by exogenous actions. Variations may be a substitution, deletion or insertion of one or more amino acids resulting in a different sequence when compared with the polypeptide sequence claimed. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains. Amino acid substitutions can be the result of replacing one amino acid with another amino acid. Insertions result when a new amino acid is introduced to the sequence and deletion result when an amino acid is deleted from the native sequence.

[0060] By way of their biochemistry, in the broadest sense, amino acids can be divided into groups based on their chemical properties. For example, one in the art would recognize that hydrophobic amino acids (*i.e.* Ile, Leu, Met, Phe, Trp, Tyr, Val, Ala, Cys, or Pro) are

functionally more similar to each other than they are to hydrophilic amino acids (*i.e.* Gly, Asn, Gln, Ser, Thr, Asp, Glu, Lys, Arg, or His).

[0061] A substitution is considered conservative when it minimally disrupts the biochemical properties of the polypeptide. Conservative substitution tables providing functionally similar amino acids are well known in the art (Henikoff & Henikoff, *Proc. Natl. Acad. Sci. U.S.A.*, 1992). Examples of conservative substitutions include when, positively-charged residues (*i.e.* H, K, and R) are substituted with other positively-charged residues; when negatively-charged residues (*i.e.* D and E) are substituted with other negatively-charged residues; when neutral polar residues (*i.e.* C, G, N, Q, S, and T) are substituted with other neutral polar residues; and when neutral non-polar residues (*i.e.* A, F, I, L, M, P, V, Y, and W) are substituted with other neutral non-polar residues. One skilled in the art would recognize that conserved substitutions are not likely to have major effects on the function.

[0062] In one embodiment the polypeptide comprises an amino acid sequence with substitutions, deletions and/or insertions of up to 40, up to 35, up to 30, up to 25, up to 20, up to 15, up to 10, up to 5 or up to 2 amino acids when compared to the sequences described herein including **SEQ ID NOs: 1-2 and 6-87**.

Metal Cofactor for Polypeptides of the Disclosure

[0063] The polypeptides described herein include a copper cofactor. In some embodiments, the polypeptide is incubated with copper prior to being provided in methods of the disclosure. In other embodiments, methods of the disclosure include a step of providing copper.

[0064] The polypeptide of the present disclosure may be at least 100%, at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at

least 55%, at least 50%, at least 45%, at least 40%, at least 35%, at least 30%, at least 25%, at least 20%, at least 15%, at least 10% or at least 5% bound by copper.

Polynucleotides of the Disclosure

[0065] The present disclosure further relates to polynucleotides that encode the polypeptides described herein.

[0066] Polynucleotides as used herein, can refer to, among other things, genomic DNA, isolated DNA, cDNA, and any and all forms of RNA partaking in the coding and making of the polypeptide. As used herein, the terms “polynucleotide,” “nucleic acid sequence,” “nucleic acid,” and variations thereof shall be generic to polydeoxyribonucleotides (containing 2-deoxy-D-ribose), to polyribonucleotides (containing D-ribose), to any other type of polynucleotide that is an N-glycoside of a purine or pyrimidine base, and to other polymers containing non-nucleotidic backbones, provided that the polymers contain nucleobases in a configuration that allows for base pairing and base stacking, as found in DNA and RNA. Thus, these terms include known types of nucleic acid sequence modifications, for example, substitution of one or more of the naturally occurring nucleotides with analog and inter-nucleotide modifications. As used herein, the symbols for nucleotides and polynucleotides are those recommended by the IUPAC-IUB Commission of Biochemical Nomenclature.

[0067] Polynucleotides encoding the polypeptides of the present disclosure may be prepared by various suitable methods known in the art, including, for example, direct chemical synthesis or cloning. For direct chemical synthesis, formation of a polymer of nucleic acids typically involves sequential addition of 3'-blocked and 5'-blocked nucleotide monomers to the terminal 5'-hydroxyl group of a growing nucleotide chain, wherein each addition is effected by nucleophilic attack of the terminal 5'-hydroxyl group of the growing chain on the 3'-position of

the added monomer, which is typically a phosphorus derivative, such as a phosphotriester, phosphoramidite, or the like. Such methodology is known to those of ordinary skill in the art and is described in the pertinent texts and literature (*i.e.*, in Matteucci *et al.*, *Tetrahedron Lett.*, 1980; U.S. Pat. Nos. 4,500,707; 5,436,327; and 5,700,637). In addition, the desired sequences may be isolated from natural sources by splitting DNA using appropriate restriction enzymes, separating the fragments using gel electrophoresis, and thereafter, recovering the desired polynucleotide sequence from the gel via techniques known to those of ordinary skill in the art, such as utilization of polymerase chain reactions (PCR; *i.e.*, U.S. Pat. No. 4,683,195).

[0068] Polynucleotides homologous to the polynucleotide encoding the polypeptides described herein can be identified by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in references cited below (*i.e.*, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989; Berger and Kimmel, *Guide to Molecular Cloning Techniques*, *Methods in Enzymology*, vol. 152 Academic Press, Inc., San Diego, Calif., 1987; and Anderson and Young, *Quantitative Filter Hybridisation*. In: Hames and Higgins, ed., *Nucleic Acid Hybridisation*, A Practical Approach. Oxford, TRL Press, 73-111, 1985).

[0069] Encompassed by the disclosure are polynucleotide sequences that are capable of hybridizing to the disclosed polynucleotide sequences and fragments thereof under various conditions of stringency (see, for example, Wahl and Berger, *Methods Enzymol.*, 1987; and Kimmel, *Methods Enzymol.*, 1987). Full length cDNA, homologs, orthologs, and paralogs of polynucleotides of the present disclosure may be identified and isolated using well-known polynucleotide hybridization methods.

[0070] In one embodiment the polypeptide includes an amino acid sequence encoded by a polynucleotide that hybridizes under medium stringency conditions with the polynucleotide encoding the amino acid sequence of the polypeptides described herein or the complementary strand of the polynucleotide encoding the amino acid sequence of the polypeptides described herein.

[0071] In another embodiment the polypeptide includes an amino acid sequence encoded by a polynucleotide that hybridizes under high stringency conditions with the polynucleotide encoding the amino acid sequence of the polypeptides described herein or the complementary strand of the polynucleotide encoding the amino acid sequence of the polypeptides described herein.

[0072] With regard to hybridization, conditions that are highly stringent, and means for achieving them, are well known in the art. See, for example, Sambrook *et al.*, 1989 (supra); Berger and Kimmel, 1987 (supra); and Anderson and Young, 1985 (supra).

[0073] Hybridization experiments are generally conducted in a buffer of pH between 6.8 to 7.4, although the rate of hybridization is nearly independent of pH at ionic strengths likely to be used in the hybridization buffer (Anderson and Young, 1985, (supra)). In addition, one or more of the following may be used to reduce non-specific hybridization: sonicated salmon sperm

DNA or another non-complementary DNA, bovine serum albumin, sodium pyrophosphate, sodium dodecylsulfate (SDS), polyvinyl-pyrrolidone, ficoll and Denhardt's solution. Dextran sulfate and polyethylene glycol 6000 act to exclude DNA from solution, thus raising the effective probe DNA concentration and the hybridization signal within a given unit of time. In some instances, conditions of even greater stringency may be desirable or required to reduce non-specific and/or background hybridization. These conditions may be created with the use of higher temperature, lower ionic strength and higher concentration of a denaturing agent such as formamide.

[0074] Stringency conditions can be adjusted to screen for moderately similar fragments such as homologous sequences from distantly related organisms, or to highly similar fragments such as genes that duplicate functional enzymes from closely related organisms. The stringency can be adjusted either during the hybridization step or in the post-hybridization washes. Salt concentration, formamide concentration, hybridization temperature and probe lengths are variables that can be used to alter stringency. As a general guideline, high stringency is typically performed at $T_m-5^{\circ}\text{C}$ to $T_m-20^{\circ}\text{C}$, moderate stringency at $T_m-20^{\circ}\text{C}$ to $T_m-35^{\circ}\text{C}$ and low stringency at $T_m-35^{\circ}\text{C}$ to $T_m-50^{\circ}\text{C}$ for duplex >150 base pairs. Hybridization may be performed at low to moderate stringency ($25-50^{\circ}\text{C}$ below T_m), followed by post-hybridization washes at increasing stringencies. Maximum rates of hybridization in solution are determined empirically to occur at $T_m-25^{\circ}\text{C}$ for DNA-DNA duplex and $T_m-15^{\circ}\text{C}$ for RNA-DNA duplex. Optionally, the degree of dissociation may be assessed after each wash step to determine the need for subsequent, higher stringency wash steps.

[0075] High stringency conditions may be used to select for nucleic acid sequences with high degrees of identity to the disclosed sequences. An example of stringent hybridization

conditions obtained in a filter-based method such as a Southern or northern blot for hybridization of complementary nucleic acids that have more than 100 complementary residues is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

[0076] Hybridization and wash conditions that may be used to bind and remove polynucleotides with less than the desired homology to the nucleic acid sequences or their complements that encode the present transcription factors include, for example: 6X SSC and 1% SDS at 65°C; 50% formamide, 4X SSC at 42°C; 0.5X SSC to 2.0 X SSC, 0.1% SDS at 50°C to 65°C; or 0.1X SSC to 2X SSC, 0.1% SDS at 50°C - 65°C; with a first wash step of, for example, 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with, for example, a subsequent wash step with 0.2 X SSC and 0.1% SDS at 65°C for 10, 20 or 30 minutes.

[0077] For identification of less closely related homologs, wash steps may be performed at a lower temperature, *i.e.*, 50°C. An example of a low stringency wash step employs a solution and conditions of at least 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS over 30 min. Greater stringency may be obtained at 42°C in 15 mM NaCl, with 1.5 mM trisodium citrate, and 0.1% SDS over 30 min. Wash procedures will generally employ at least two final wash steps. Additional variations on these conditions will be readily apparent to those skilled in the art (see, for example, US Patent Application No. 20010010913).

[0078] If desired, one may employ wash steps of even greater stringency, including conditions of 65°C -68°C in a solution of 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS, or about 0.2X SSC, 0.1% SDS at 65° C and washing twice, each wash step of 10, 20 or 30 min in duration, or about 0.1 X SSC, 0.1% SDS at 65° C and washing twice for 10, 20 or 30 min. Hybridization stringency may be increased further by using the same conditions as in the

hybridization steps, with the wash temperature raised about 3°C to about 5°C, and stringency may be increased even further by using the same conditions except the wash temperature is raised about 6°C to about 9°C.

[0079] Polynucleotide probes may be prepared with any suitable label, including a fluorescent label, a colorimetric label, a radioactive label, or the like. Labeled hybridization probes for detecting related polynucleotide sequences may be produced, for example, by oligo-labeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

[0080] Various methods are known to those of skill in the art for identifying similar (*i.e.*, homologs, orthologs, paralogs, etc.) polynucleotide sequences, including phylogenetic methods, sequence similarity analysis, and hybridization methods.

[0081] Phylogenetic trees may be created for a gene family by using a program such as CLUSTAL (Thompson *et al. Nucleic Acids Res.*, 1994; Higgins *et al. Methods Enzymol.*, 1996 or MEGA (Tamura *et al. Mol. Biol. & Evo.*, 2007). Once an initial tree for genes from one species is created, potential orthologous sequences can be placed in the phylogenetic tree and their relationships to genes from the species of interest can be determined. Evolutionary relationships may also be inferred using the Neighbor-Joining method (Saitou and Nei, *Mol. Biol. & Evo.*, 1987). Homologous sequences may also be identified by a reciprocal BLAST strategy. Evolutionary distances may be computed using the Poisson correction method (Zuckerandl and Pauling, pp. 97-166 in *Evolving Genes and Proteins*, edited by V. Bryson and H.J. Vogel. Academic Press, New York, 1965).

[0082] In addition, evolutionary information may be used to predict gene function. Functional predictions of genes can be greatly improved by focusing on how genes became similar in sequence (*i.e.* by evolutionary processes) rather than on the sequence similarity itself

(Eisen, *Genome Res.*, 1998). Many specific examples exist in which gene function has been shown to correlate well with gene phylogeny (Eisen, *Genome Res.*, 1998). By using a phylogenetic analysis, one skilled in the art would recognize that the ability to deduce similar functions conferred by closely-related polypeptides is predictable.

[0083] When a group of related sequences are analyzed using a phylogenetic program such as CLUSTAL, closely related sequences typically cluster together or in the same clade (a group of similar genes). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle, *J. Mol. Evol.*, 1987). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous sequences, or orthologous sequences that share the same function (see also, for example, Mount, *Bioinformatics: Sequence and Genome Analysis* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., page 543, 2001).

[0084] To find sequences that are homologous to a reference sequence, BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the disclosure. BLAST protein searches can be performed with the BLASTX program, score=50, wordlength=3, to obtain amino acid sequences homologous to a protein or polypeptide of the disclosure. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.*, *Nucleic Acids Res.*, 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.*, 1997 *supra*. When utilizing BLAST,

Gapped BLAST, or PSI-BLAST, the default parameters of the respective programs (*i.e.*, BLASTN for nucleotide sequences, BLASTX for proteins) can be used.

[0085] Methods for the alignment of sequences and for the analysis of similarity and identity of polynucleotide sequences are well-known in the art.

[0086] As used herein “sequence identity” refers to the percentage of residues that are identical in the same positions in the sequences being analyzed. As used herein “sequence similarity” refers to the percentage of residues that have similar biophysical / biochemical characteristics in the same positions (*i.e.*, charge, size, hydrophobicity) in the sequences being analyzed.

[0087] Methods of alignment of sequences for comparison are well-known in the art, including manual alignment and computer assisted sequence alignment and analysis. This latter approach is a preferred approach in the present disclosure, due to the increased throughput afforded by computer assisted methods. As noted below, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

[0088] The determination of percent sequence identity and/or similarity between any two polypeptide sequences can be accomplished using a mathematical algorithm. Examples of such mathematical algorithms are the algorithm of Myers and Miller, *CABIOS*, 1988; the local homology algorithm of Smith *et al.*, *Adv. Appl. Math.*, 1981; the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.*, 1970; the search-for-similarity-method of Pearson and Lipman, *Proc. Natl. Acad. Sci.*, 1988; the algorithm of Karlin and Altschul, *Proc. Natl. Acad. Sci. USA*, 1990, modified as in Karlin and Altschul, *Proc. Natl. Acad. Sci. USA*, 1993.

[0089] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity and/or similarity. Such implementations

include, for example: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the AlignX program, version 10.3.0 (Invitrogen, Carlsbad, CA) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. *Gene*, 1988; Higgins *et al. CABIOS*, 1989; Corpet *et al., Nucleic Acids Res.*, 1988; Huang *et al. CABIOS*, 1992; and Pearson et al., *Meth. Mol. Biol.*, 1994. The BLAST programs of Altschul et al. *J. Mol. Biol.*, 1990, are based on the algorithm of Karlin and Altschul, 1990 *supra*.

Methods of Degrading Starch Containing Materials

[0090] The present disclosure relates to methods for degrading starch containing material, including the steps of providing a starch-containing material and providing a polypeptide having starch-degrading activity and a copper cofactor. The method further comprises incubating the polypeptide with the starch containing material in the presence of oxygen and a reductant.

Starch-Containing Materials

[0091] Starch-containing materials include the endosperm of many plants, including the cereal grains. For example, starch containing materials may include, but are not limited to, whole grains, corn, wheat, barley, rye, milo, sago, cassava, tapioca, sorghum, rice, peas, beans, or sweet potatoes. Starch-containing materials should also be understood to include starches with a high amylopectin content sometimes called waxy starches, as well as chemically and physically modified starches, such as for example starches whose acid values have been reduced, starches in which the type and concentration of cations associated with the phosphate groups have been

modified, ethoxylated starches, starch acetates, cationic starches, oxidated starches and cross-linked starches.

Incubation Conditions

[0092] In methods of the present disclosure, the polypeptides described herein are incubated with starch-containing material in the presence of a reductant and oxygen under conditions suitable to degrade the starch-containing material. In some embodiments, the methods include a step of providing reductants and/or oxygen.

[0093] A reductant is an element or compound that donates (or loses) an electron to another chemical species in a reaction. One skilled in the art will recognize that for any redox reaction, many different types of reductants can be successfully used. For example, one skilled in the art will recognize that any iron containing compound, arene, or any other functionally similar reductant can be used. Additionally, one skilled in the art will recognize that many different types of reductants can be used, both organic and inorganic. Examples include metals (*i.e.*, iron, zinc, and magnesium), metal alloys and organic materials (*i.e.*, hydrides).

[0094] In one embodiment of the present disclosure the reductant used is an iron containing compound. More specifically, this iron containing compound can comprise a heme domain. For example, the iron containing compound can be cellobiose dehydrogenase. In another example, the iron containing compound can be ferrous sulfate. One skilled in the art would recognize that any iron containing compound, in which iron is able to transfer an electron can serve as an iron containing reductant. One skilled in the art would recognize that many other metal reductants can also be used. For example other metal reductants include lithium, sodium, magnesium, iron, zinc, and aluminum. Other reductants include formaldehyde, formic acid,

Hantzsch ester, hydrazine, isopropanol, lithium aluminum hydride, lithium tetrahydridoaluminate, lithium triethylborohydride, and silanes.

[0095] In some embodiments, the reductant is either ascorbic acid or gallic acid. The reductant can also be any arene.

[0096] Incubation may occur for any period of time sufficient to degrade the starch-containing material. For example the incubating step may be 10 hours, 9 hours, 8 hours, 7 hours, 6 hours, 5 hours, 4 hours, 3 hours, 2 hours, or 1 hour.

[0097] Incubation may occur for a period of time sufficient to achieve the desired amount of degradation of starch-containing material. For example, incubation may occur for a period of time sufficient to achieve 100%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% degradation of the starch-containing material.

[0098] Incubation may occur with any amount of starch-containing material. For example, the amount of starch-containing material may be 1 mg/mL, 5mg/mL, 10mg/mL, 15 mg/mL, 20 mg/mL, 25 mg/mL, 30 mg/mL, 35 mg/mL, 40 mg/mL, 45 mg/mL, 50 mg/mL, 55 mg/mL, 60 mg/mL, 65 mg/mL, 70 mg/mL, 75 mg/mL, 80 mg/mL, 85 mg/mL, 90 mg/mL, 95 mg/mL, 100 mg/mL.

[0099] Incubation may occur with any amount of polypeptide. For example, the amount of polypeptide used may be 0.001 μ M, 0.005 μ M, 0.01 μ M, 0.05 μ M, 0.1 μ M, 0.5 μ M, 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, or 100 μ M.

[0100] Incubation may occur with any amount of reductant. For example, the amount of ascorbic acid used may be 0.001 mM, 0.005 mM, 0.01 mM, 0.05 mM, 0.1 mM, 0.5 mM, 1 mM,

2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 15 mM, 20 mM, 25 mM, 30 mM, 35 mM, 40 mM, 45 mM, 50 mM, 55 mM, 60 mM, 65 mM, 70 mM, 75 mM, 80 mM, 85 mM, 90 mM, 95 mM, or 100 mM. The amount of CDH used may be 0.001 μ M, 0.005 μ M, 0.01 μ M, 0.05 μ M, 0.1 μ M, 0.5 μ M, 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, or 100 μ M.

[0101] Incubation may occur at a temperature ranging from 25°C to 70°C, for example, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, or 70°C.

[0102] In some embodiments, the methods include a further step of providing amylases. Amylases are enzymes used to break starches into dextrans and to break dextrans into glucose. Suitable amylases (alpha and/or beta) include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Amylases include, for example, alpha-amylases obtained from *Bacillus*, *i.e.*, a special strain of *Bacillus icheniformis*, described in more detail in GB 1,296,839.

Methods of Producing a Fermentation Product

[0103] One aspect of the current disclosure relates to the use of the process in fermentation. The fermentable sugar obtained from the starch-containing material can be fermented by one or more fermenting microorganisms capable of fermenting the sugars directly or indirectly into a desired fermentation product. Fermentation products may include, alcohols (*i.e.*, ethanol, methanol, butanol); organic acids (*i.e.*, citric acid, acetic acid, itaconic acid, lactic acid, succinic acid, gluconic acid); ketones (*i.e.*, acetone); amino acids (*i.e.*, glutamic acid); gases (*i.e.*, H₂ and CO₂); antibiotics (*i.e.*, penicillin and tetracycline); enzymes; vitamins (*i.e.*, riboflavin, B₁₂, beta-carotene); and hormones. For example, the fermentation product can be

ethanol (*i.e.*, fuel ethanol, drinking ethanol, potable neutral spirits, industrial ethanol), products used in the consumable alcohol industry (*i.e.*, beer and wine), leather industry and tobacco industry. The fermentation product can be used in many different ways. For example the fermentation product, such as ethanol, obtained according to the method, may be used as fuel, whether or not blended with gasoline or it may also be used as potable ethanol. "Fermentation" or "fermentation process" refers to any fermentation process or any process comprising a fermentation step. Fermentation processes include those used in the consumable alcohol industry, leather industry, tobacco industry and biofuel industry. The fermentation conditions depend on the desired fermentation product and proper fermenting organism and can be easily determined by one skilled in the art. In the fermentation step, sugars, released from the starch-containing material as a result of enzymatic glycolysis, are fermented to a fermentation product by a fermenting organism, such as yeast. The fermentation product is then further isolated and/or purified. The fermentation can also be carried out simultaneously with the enzymatic saccharification in the same vessel, again under controlled pH, temperature, and mixing conditions. When saccharification and fermentation are performed simultaneously in the same vessel, the method is generally termed simultaneous saccharification and fermentation. Any suitable starch-containing material may be used in the fermentation method (2010/059413 PCT/US2009/062955). "Fermenting microorganism" refers to any microorganism suitable for use in a desired fermentation method. Suitable fermenting microorganisms are able to ferment, which is, convert, sugars, such as glucose, maltose, or maltodextrins directly or indirectly into the desired fermentation product.

[0104] Fermentation systems and culture conditions which can be used are described in WO2009/076676, WO2010/003007, WO 2009/132220, WO 2010/031062, WO2010/031068,

WO 2010/031076, WO2010/031077, WO2010/031079, WO2010/148150, WO2010/005525, WO 2010/078457, WO2010/124146, WO2010/148144, WO2010/148256 and U.S. Patent Application Nos. 12/496,573, 12/560,390, 12/560,317, 12/560,370, 12/560,305, and 12/560,366.

[0105] There are a large number of different types of fermentation methods that are used commercially. The fermenting microorganisms are grown using any known mode of fermentation, such as batch, fed-batch, continuous, or continuous with recycle methods.

[0106] In one embodiment of the present invention, the polypeptide is used to degrade starch to form glucose, which is then cultured with fermentative organisms to produce a fermentation product as described. In another embodiment of the present invention, the polypeptide is used to degrade starch to form glucose, which is then cultured with a chemical solution to produce a fermentation product as described.

[0107] The processes disclosed herein can be performed in the presence of amylase. Amylases are enzymes used to break starches into dextrins and to break dextrins into glucose. Suitable amylases (alpha and/or beta) include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Amylases include, for example, alpha-amylases obtained from *Bacillus*, *i.e.*, a special strain of *Bacillus icheniformis*, described in more detail in GB 1,296,839.

[0108] In one embodiment all the processes described above can be performed in the presence of any amylase. In another embodiment all the processes described above can be performed in the presence of any amylase and a reductant. In another embodiment all the processes described above can be performed in the presence of any amylase and an iron containing compound. In another embodiment all the processes described above can be

performed in the presence of a polypeptide and an amylase and a reductant comprising a heme domain. In yet another embodiment all the processes described herein can be performed in the presence of any amylase and cellobiose dehydrogenase. In another embodiment all the processes described above can be performed in the presence of any amylase and ferrous sulfate. In another embodiment all the processes described above can be performed in the presence of an amylase and an arene. For example, all the processes described above can be performed in the presence of any amylase and ascorbic acid.

Compositions

[0109] The present disclosure also relates to a composition including a polypeptide having conserved motifs specific to starch-active PMOs, including **SEQ ID NOs: 3-5**, either individually or in combination, a copper cofactor, and a reductant. In other embodiments, the polypeptide in the composition includes the amino acid sequence of **SEQ ID NO: 1 or 2**. In still other embodiments, the polypeptide in the composition includes any of the following amino acid sequences: **SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 87 or SEQ ID NOs: 8-86**.

[0110] In one embodiment the composition includes the polypeptide, copper and an iron containing reductant. For example, the iron containing reductant can be a compound with a heme reducing domain, cellobiose dehydrogenase or ferrous sulfate. For example the composition may comprise, any concentration of the polypeptide, whether in its entirety or truncated form including, 0.001 μ M, 0.005 μ M, 0.01 μ M, 0.05 μ M, 0.1 μ M, 0.5 μ M, 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, and 100 μ M; any concentration of CDH, including 0.001 μ M, 0.005 μ M, 0.01 μ M, 0.05 μ M, 0.1 μ M, 0.5 μ M, 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M,

20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, and 100 μ M; and any concentration of a copper, including 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, 100 μ M, 150 μ M, 200 μ M, 250 μ M, 300 μ M, 350 μ M, 400 μ M, 450 μ M, 500 μ M, 550 μ M, 600 μ M, 650 μ M, 700 μ M, 750 μ M, 800 μ M, 850 μ M, 900 μ M, 950 μ M, and 1000 μ M.

[0111] In another embodiment the composition includes the polypeptide, copper and an arene. For example, the composition can comprise a polypeptide, copper and ascorbic acid. For example the composition may comprise any concentration of the polypeptide, whether in its entirety or truncated form including, 0.001 μ M, 0.005 μ M, 0.01 μ M, 0.05 μ M, 0.1 μ M, 0.5 μ M, 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, and 100 μ M; any concentration of ascorbic acid including 0.001 mM, 0.005 mM, 0.01 mM, 0.05 mM, 0.1 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 15 mM, 20 mM, 25 mM, 30 mM, 35 mM, 40 mM, 45 mM, 50 mM, 55 mM, 60 mM, 65 mM, 70 mM, 75 mM, 80 mM, 85 mM, 90 mM, 95 mM, and 100 mM.; and any concentration of a copper, including 1 μ M, 2 μ M, 3 μ M, 4 μ M, 5 μ M, 6 μ M, 7 μ M, 8 μ M, 9 μ M, 10 μ M, 15 μ M, 20 μ M, 25 μ M, 30 μ M, 35 μ M, 40 μ M, 45 μ M, 50 μ M, 55 μ M, 60 μ M, 65 μ M, 70 μ M, 75 μ M, 80 μ M, 85 μ M, 90 μ M, 95 μ M, 100 μ M, 150 μ M, 200 μ M, 250 μ M, 300 μ M, 350 μ M, 400 μ M, 450 μ M, 500 μ M, 550 μ M, 600 μ M, 650 μ M, 700 μ M, 750 μ M, 800 μ M, 850 μ M, 900 μ M, 950 μ M, and 1000 μ M.

[0112] In other embodiments, the compositions described above further include an amylase.

EXAMPLES

[0113] This example describes the identification of a new family of starch-active polysaccharide monooxygenases.

Introduction: Characteristics of polysaccharide monooxygenases

[0114] Polysaccharide monooxygenases (PMOs) are recently discovered fungal and bacterial enzymes that are secreted to the extracellular environment. PMOs are secreted by a variety of fungal and bacterial species (Horn *et al.*, *Biotechnol Biofuels*, 2012, Hemsworth *et al.*, *Curr Opin Chem Biol*, 2013, Tian *et al.*, *Proc Natl Acad Sci U S A*, 2009, Yakovlev *et al.*, *Appl Microbiol Biotechnol*, 2012, Berka *et al.*, *Nat Biotech*, 2011.). They have recently been found to oxidatively degrade chitin (Vaaje-Kolstad *et al.*, 2010 *Science*, 2010, Hemsworth *et al.*, *Nat Chem Biol*, 2014, Forsberg *et al.*, *Biochemistry*, 2014) and cellulose (Forsberg *et al.*, *Biochemistry*, 2014, Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Phillips *et al.*, *ACS Chem Biol*, 2011, Beeson *et al.*, *J Am Chem Soc*, 2012, Forsberg *et al.*, *Protein Sci*, 2011). PMOs have been shown to oxidize either the C1 or C4 atom of the $\beta(1\rightarrow4)$ glycosidic bond on the surface of chitin (Vaaje-Kolstad *et al.*, 2010 *Science*, 2010, Hemsworth *et al.*, *Nat Chem Biol*, 2014) or cellulose (Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Phillips *et al.*, *ACS Chem Biol*, 2011, Beeson *et al.*, *J Am Chem Soc*, 2012), resulting in the cleavage of this bond and the creation of new chain ends that can be subsequently processed by hydrolytic chitinases and cellulases. Several fungal PMOs have been shown to significantly enhance the degradation of cellulose by hydrolytic cellulases (Harris *et al.*, *Biochem.* 2010), indicating that these enzymes can be used in the conversion of plant biomass into biofuels and other renewable chemicals.

[0115] There are three families of PMOs characterized thus far: fungal PMOs that oxidize cellulose (Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Phillips *et al.*, *ACS Chem Biol*, 2011, Beeson *et al.*, *J Am Chem Soc*, 2012) (also known as GH61 and AA9); bacterial PMOs that are active either on chitin (Vaaje-Kolstad *et al.*, *Science*, 2010, Forsberg *et al.*, *Biochemistry*, 2014) or cellulose (Forsberg *et al.*, *Biochemistry*, 2014, Forsberg *et al.*, *Protein Sci*, 2011) (also known as CBM33 and AA10); and fungal PMOs that oxidize chitin (AA11) (Hemsworth *et al.*, *Nat Chem Biol*, 2014). Sequence homology between these three families is very low. Nevertheless, the available structures of PMOs from all three families reveal a conserved fold, including an antiparallel β -sandwich core and a highly conserved monocopper active site on a flat protein surface (**FIG. 1A**) (Hemsworth *et al.*, *Curr Opin Chem Biol*, 2013, Vaaje-Kolstad *et al.*, 2010 *Science*, 2010, Hemsworth *et al.*, *Nat Chem Biol*, 2014, Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Karkehabadi *et al.*, *J Mol Biol*, 2008, Li *et al.*, *Structure*, 2012, Wu *et al.*, *J Biol Chem*, 2013). Two histidine residues in a motif termed the *histidine brace* coordinate the copper center. The *N*-terminal histidine ligand binds in a bidentate mode, and its imidazole ring is methylated at the *N* ϵ position in fungal PMOs (**FIG. 1A**). Considering the conserved structural features, it is not surprising that the currently known PMOs act on substrates with structurally similar structures. Cellulose and chitin contain long linear chains of β (1 \rightarrow 4) linked glucose units and *N*-acetylglucosamine units, respectively (**FIG. 1B**). The polymer chains form extensive hydrogen bonding networks, which result in insoluble and very stable crystalline structures (Nishiyama *et al.*, *J Am Chem Soc*, 2002, Nishiyama *et al.*, *J Am Chem Soc*, 2003, Nishiyama *et al.*, *Macromolecules*, 2011, Sikorski *et al.*, *Biomacromolecules*, 2009). PMOs are thought to bind to the substrate with their flat active site surface, which orients the copper center for selective

oxidation at the C1 or C4 position (Vaaje-Kolstad *et al.*, 2010 *Science*, 2010, Li *et al.*, *Structure*, 2012, Vu *et al.*, *J Am Chem Soc*, 2014). Some bacterial chitin-binding proteins are cellulose-active PMOs, further suggesting that the set of PMO substrates is restricted to $\beta(1\rightarrow4)$ linked carbohydrate polymers.

[0116] Despite these findings suggesting that PMOs act on a limited set of substrates, described herein is the identification of PMOs that contain several key features of previously characterized PMOs, but act on substrates different from cellulose or chitin. A member of one of these novel families of PMOs, NCU08746, was shown to oxidatively cleave amylose, amylopectin, and starch. The NCU08746 family has been designated as starch-active PMOs.

[0117] Both amylose and amylopectin contain linear chains of $\alpha(1\rightarrow4)$ linked glucose, while the latter also contains $\alpha(1\rightarrow6)$ glycosidic linkages at branch points in the otherwise $\alpha(1\rightarrow4)$ linked polymer. Unlike cellulose and chitin, amylose and amylopectin do not form microcrystals; instead, they exist in disordered, single helical, and double helical forms (Perez and Bertoft, *Stärke*, 2010, Popov *et al.*, *Macromolecules*, 2009, Imberty *et al.*, *J Mol Biol.* 1988) (see **FIG. 1C** for example). Starch can exist partially in nanocrystalline form, but lacks the flat molecular surfaces as those found in chitin and cellulose. The discovery of starch-active PMOs shows that this oxidative mechanism of glycosidic bond cleavage is more widespread than initially expected.

Identification of a starch-active fungal PMO family

[0118] To identify potential other PMOs, the signaling peptide presumed to be present in all fungal PMOs was used. All currently known fungal PMOs are predicted to be secreted via the canonical signal peptide-triggered pathway and contain several conserved features, including an absolutely conserved N-terminal histidine residue, a second conserved histidine residue, and the

N/Q/E-X-F/ Y motif. All *Neurospora crassa* proteins were input into SignalP 4.0 program (Petersen *et al.*, *Nat Meth*, 2011). Predicted sequences with an N-terminal histidine residue were compared to orthologues found in protein databases. Potential new PMOs were identified as those sequences having a second conserved histidine residue and an N/Q/E-X-F/Y motif. Structural prediction was also performed using Phyre2 (Kelley *et al.*, *Nat Protoc*, 2009). Using SignalP 4.0 (Petersen *et al.*, *Nat Meth*, 2011), 777 *N. crassa* sequences with predicted signal peptides were identified. Among these predicted secreted proteins, 49 were predicted to contain a N-terminal histidine residue. For 21 of these 49 proteins, the N-terminal histidine was absolutely conserved in all identified fungal homologues. The second histidine residue and the N/Q/E-X-F/Y motif in each of these 21 sequences were identified using multiple sequence alignments with homologous sequences found in public databases. **Table 1** shows *N. crassa* proteins with these PMO conserved features. The majority of these proteins are cellulose active PMOs (in light blue; formerly known as GH61) (Tian *et al.*, *Proc Natl Acad Sci U S A*, 2009, Phillips *et al.*, *ACS Chem Biol*, 2011, Beeson *et al.*, *J Am Chem Soc*, 2012). Three sequences belonging to a family annotated as chitinases are also shown (in aqua); a member of this family was recently characterized as a chitin-active PMO (Hemsworth *et al.*, *Nat Chem Biol*, 2014). The N-terminus of NCU08746 (in yellow) shares weak homology with chitin binding domains and a C-terminal carbohydrate binding module 20 (CBM20) (Christiansen *et al.*, *FEBS J*, 2009). The other sequences in **Table 1** with unknown substrates are not annotated.

Table 1: Predicted *N. crassa* PMOs, based on secreted proteins with a N-terminal histidine residue

Gene ID	SP	MS	Sub	Gene ID	SP	MS	Sub
NCU01867	AIA	HA	Cel	NCU02240	ASA	HT	Cel
NCU08760	AAA	HA	Cel	NCU02916	ASA	HT	Cel
NCU03000	VAA	HG	Cel	NCU07974	ASA	HT	Cel
NCU05969	VAA	HG	Cel	NCU01050	ASA	HT	Cel
NCU03328	TSA	HY	Cel	NCU07760	VAA	HG	Cel
NCU00836	ASA	HY	Cel	NCU07898	VAA	HG	Cel
NCU02344	AQA	HY	Cel	NCU07520	ALA	HS	Cel
NCU05405	ASA	HI	Chi	NCU08746	VDA	HG	St
NCU00822	VQG	HM	Chi	NCU00881	ASA	HT	U
NCU01380	SQA	HM	Chi	NCU02884	ASA	HF	U
NCU02613	SAG	HS	U				

SP = The last three residues of signal peptide.

MS = The first two residues of mature sequence.

Sub = Substrate. Cel = Cellulose. St = Starch. Chi = Chitin.

U = Unknown.

Shading represents different PMO families as determined by sequence homology

[0119] **FIG. 2** shows the domain architecture and the consensus sequence of NCU08746 and homologues. Forty three sequences of this family were found with an iterative protein database search using a Hidden Markov Model algorithm (Christiansen *et al.*, *FEBS J*, 2009). They are all found in the genomes of fungi in the Ascomycota phylum. Unlike chitin- and cellulose-active PMOs, the NCU08746 family has very high sequence identity. The N-terminal domain suggested by this sequence alignment contains 204–233 amino acids; this finding is inconsistent with its original annotation. Given biochemical characterization reported herein, this catalytic domain is designated as a starch active PMO domain.

Expression, purification and verification of NCU08746

[0120] NCU08746 was obtained by expression in *N. crassa* as previously reported (Vu *et al.*, *J Am Chem Soc*, 2014) using the following primer sets:

1a-5' ATATATCCGCGGAGCGCATTTCCGACGTTAAG-3' (**SEQ ID NO: 88**),

1b-5' AACCGAGATGATGGAGAACTTCATTTTGATTTCTGTGATGTG-3' (SEQ ID NO: 89),

2a-5' CCTCCCCACATCACAGAAATCAAAATGAAGTTCTCCATCATC-3' (SEQ ID NO: 90),

2b-5' ATATATCCGCGGGGCGGAAAGAGGAGGAAAG-3' (SEQ ID NO: 91),

3a-5' GAGCTCGACACGGCCTTC-3' (SEQ ID NO: 92),

3b-5' CGTCGGAAATGCGCTGCGGCCGCATATATTGGATTTCTGCGCTGCA-3' (SEQ ID NO: 93), and

4a-5' TTCCTCCTCTTTCCGCCGCGGCCGCATATATAATGGAAGCTCTCAGAGCTCTGC-3' (SEQ ID NO: 94),

4b-5' -AACACCTCCGTCGCCATAAACT-3' (SEQ ID NO: 95).

[0121] This expression method yields recombinant PMOs indistinguishable from the corresponding natively purified enzymes. **NCU08746 (SEQ ID NO: 7)** was purified using an amylose resin column, followed by a size exclusion column. SDS-PAGE analysis of the fractions from the amylose resin column shows that NCU08746 has relatively strong affinity for amylose (**FIG. 8**), supporting a functional role for the predicted C-terminal CBM20 domain.

Furthermore, MS/MS analysis of the trypsin-digested purified NCU08746 yielded nearly 50% sequence coverage, including both the N- and C-termini, confirming the predicted sequence of this protein. Trypsin-digested purified NCU08746 was analyzed with nano LC MS/MS at the Scripps Center for Metabolomics and Mass Spectrometry of The Scripps Research Institute. The mass spectrometry data is consistent with a methylated N-terminal histidine (**FIG. 8 and Table 2**), a feature previously observed for PMOs natively purified from or recombinantly expressed in fungi (Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Phillips *et al.*, *ACS Chem Biol*, 2011, Vu *et*

al., *J Am Chem Soc*, 2014). Five to 10 L filtered culture was concentrated to ~ 100 mL, desalted to 10 mM sodium acetate buffer pH 5.0 (buffer A), and mixed with 10 mM EDTA overnight. The EDTA-treated sample was subsequently loaded onto a pre-equilibrated amylose resin column (10 mL). After sample injection, the column was washed with 15 column volumes of buffer A. Protein was eluted with 10 column volumes of 50 mM maltose in buffer A. Fractions containing NCU08746 were pooled and mixed with 100 μ M CuSO₄ overnight. Subsequently, the CuSO₄-treated sample was then concentrated to approximately 5 mL and purified further with a size exclusion column (Superdex 75, GE Healthcare). To obtain apo NCU08746, the reconstitution with CuSO₄ was replaced with another EDTA treatment.

Table 2

#	b	b ⁺⁺	b ⁰	b ⁺⁺⁺	Seq.	y	y ⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁺⁺	#
1	152.0818	76.5446			H							10
2	289.1033	105.0553			G	1053.5728	527.2900	1036.5462	518.7767	1035.5622	518.2847	9
3	372.1668	186.5870			Y	896.5513	498.7793	979.5247	490.2660	978.5407	489.7740	8
4	485.2507	243.1290			L	833.4880	417.2476	816.4614	408.7343	815.4774	408.2423	7
5	586.2984	293.6528	568.2878	284.6475	T	720.4039	360.7056	703.3774	352.1923	702.3933	351.7003	6
6	699.3824	350.1949	681.3719	341.1896	I	618.3552	310.1817	602.3297	301.6685	601.3457	301.1765	5
7	796.4352	398.7212	778.4246	389.7160	P	508.2722	253.6397	489.2456	245.1264	488.2616	244.6344	4
8	943.5038	472.2554	925.4931	463.2502	F	489.2194	205.1133	392.1928	196.6001	391.2088	196.1081	3
9	1030.5356	515.7715	1012.5251	506.7662	S	282.1510	131.5791	245.1244	123.0659	244.1404	122.5738	2
10					R	175.1190	88.0631	158.0924	79.5498			1

Score	Mr(calc)	Delta	Sequence	SEQ ID NO
45.5	1203.6400	-0.2394	HGYLTIPFSR	6

NCU08746 activity assays and product identification

[0122] **FIG. 3** shows the high performance anion exchange chromatographic (HPAEC) traces of NCU08746 assays under various conditions. The reaction with amylopectin is dependent on reductant (ascorbic acid) and oxygen. When both ascorbic acid and oxygen were present, the chromatogram exhibits a set of new peaks (trace C) that are not observed in the absence of either reductant or oxygen (**FIG. 9**). The same set of peaks is also found in the assays

with corn starch or amylose (**FIGS. 10-12**), but not in the assays with phosphoric acid swollen cellulose (PASC) or chitin (traces D and E in **FIG. 3**, respectively). These results showed clearly that NCU08746 specifically cleaves starch substrates via an oxygen-dependent mechanism. The products of NCU08746 assays were identified by comparison with synthetic standards via HPAEC and mass spectrometry analysis. The product peaks of NCU08746 overlay with the peaks of malto-alonic acids (traces A and B in **FIG. 3**), but not with those of maltodextrins (**FIG. 12**). The mass spectrum of the assay with amylose exhibits a series of peaks 16 amu higher than the corresponding maltodextrin peaks (**Table 3**), which is consistent with the presence of an additional oxygen atom. Together, the HPAEC and mass spectrometry data indicate that NCU08746 oxidizes the C1 position of starch substrates. **FIG. 19** shows that the full-length NCU08746 enzyme exhibits significantly higher activity on amylose (source: potato; Sigma A0512) than starch (amylose + amylopectin) or amylopectin from corn.

[0123] Amylopectin from maize (Sigma, 11020), potato amylose with average molecular weight of 2.8 kDa (AM2.8, TCI Chemicals, A0846) or potato amylose from Sigma (A0512), and corn starch (Sigma, S4126), and chitin (Sigma, C9752) were washed with water by centrifugation several times prior to addition to assays. Phosphoric acid swollen cellulose was prepared from Avicel PH101 (Sigma, 11365) as previously described (Zhang *et al.*, *Biomacromolecules*, 2006). Metal salts were purchased from various sources with analytical grade purity. Cellobiose dehydrogenase from *Myceliophthora thermophila* (MtCDH-2) was purified as previously reported (Phillips *et al.*, *ACS Chem Biol*, 2011). Standard glucose and maltodextrins with degrees of polymerization (DP) of 2–7 were purchased from Sigma. Higher DP soluble maltodextrins were obtained by suspending 100 mg/mL AM2.8 in water and removing the pellet by centrifugation. Malto-alonic acids were prepared by oxidation of

maltodextrins with Lugol's solution using a method previously described (Beeson *et al.*, *J Am Chem Soc*, 2012). Activity assays of the polypeptide were carried out in 100 μ L total volume using 96 well plates, which were shaken at 1000 rpm – 1200 rpm at 42 °C for 4 hours using a microplate shaker. All assays contained 50 mg/mL substrate suspension, except for assays with PASC that contained 5 mg/mL substrate. Typical assays contained 5 μ M holo NCU08746, 2 mM ascorbic acid, and 50 mM sodium acetate pH 5.0. Ascorbic acid was substituted by 0.5 μ M MtCDH-2 in some assays. Cellobiose (5 mM) was added to an assay with MtCDH-2. For assays with partially apo polypeptides (NCU08746), 5 μ M of metal salts were added. Assay samples were mixed with one volume of 0.2 M NaOH and analyzing with high performance anion exchange chromatography (HPAEC) using a DIONEX ICS-5000 system and a gradient previously reported (Philips *et al.*, *ASC Chem Biol*, 2011). The synthetic malto-aldonic acids were run in the same manner. Product analysis with LC-MS was carried out at the Energy Biosciences Institute as previously described ((Philips *et al.*, *ASC Chem Biol*, 2011). The products of these assays were identified by comparison with synthetic standards via high performance anion exchange (HPAEC) and mass spectrometry analysis.

Table 3 Mass spectrometry analysis of NCU08746 assays.

Formula	Theoretical mass (Da) Deprotonated ion	Observed mass (Da) Deprotonated ion	Mass difference (ppm)
C ₁₂ H ₃₂ O ₁₂	357.1038	357.0992	12.98
C ₁₈ H ₃₂ O ₁₇	519.1567	519.1593	-5.05
C ₂₄ H ₄₂ O ₂₂	681.2095	681.2100	-0.74
C ₃₀ H ₅₂ O ₂₇	843.2623	843.2626	-0.33
C ₃₆ H ₆₂ O ₃₂	1005.3151	1005.3160	-0.45
C ₄₂ H ₇₂ O ₃₇	1167.3680	1167.3680	0.40
C ₄₈ H ₈₂ O ₄₂	1329.4208	1329.4190	1.05
C ₅₄ H ₉₂ O ₄₇	1491.4736	1491.4730	0.48

Copper is the native metal cofactor of NCU08746

[0124] NCU08746 reconstituted with excess copper(II) sulfate and purified with size exclusion chromatography contained approximately one copper atom per protein molecule based on ICPAES analysis (**Table 4**). The addition of one equivalent of CuSO₄ to the assays of holo NCU08746 had no effect on activity (**FIGS. 10 and 11**). Treatment of holo NCU08746 with 10mM EDTA resulted in a partially apo form containing approximately 0.12 copper atom per protein molecule. This form was assayed with the addition of one equivalent of various second row metal salts. Among the metal ions tested, only Cu(II) increased the oxidative activity of this partially apo NCU08746 (**FIG. 13**). This data supports that copper is the native metal cofactor of NCU08746.

Table 4. CP-AES results of Cu-NCU08746 (samples E1 and E2) and partially apo-NCU08746 (sample E3).

	Enzyme (μ M)	Expected Cu (ppm)	Co (ppm)	Cu (ppm)	Fe (ppm)	Mg (ppm)	Mn (ppm)	Ni (ppm)	Zn (ppm)
Lab Blank			< 0.010	< 0.010	< 0.010	< 0.031	< 0.010	< 0.002	< 0.002
Blank 1			< 0.010	< 0.010	0.479	< 0.031	< 0.010	0.070	0.008
E1	16.5	1.000	< 0.010	0.915	0.107	< 0.031	< 0.010	0.022	0.028
Blank 2			< 0.010	0.034	0.102	0.151	< 0.010	0.018	0.020
E2	8.25	0.500	< 0.010	0.510	0.102	0.094	< 0.010	0.019	0.046
Blank 3			< 0.010	0.015	< 0.010	< 0.010	< 0.010	< 0.010	0.035
E3	16.5	0.000	< 0.010	0.117	< 0.010	< 0.010	< 0.010	< 0.010	0.061

NCU08746 accepts electrons from cellobiose dehydrogenase

[0125] As NCU08746 shares several conserved features with cellulose-active PMOs, it was expected that the starch-active PMO would have a biological redox partner equivalent to that of cellulose-active PMOs, cellobiose dehydrogenase (CDH). Because such a redox partner had not been identified, the possibility of electron transfer from *Myceliophthora thermophila* CDH-2 (MtCDH-2) to NCU08746 was investigated. CDH utilizes FAD and heme cofactors to rapidly oxidize cellodextrins, including cellobiose, cellotriose, and cellotetraose, while activity on

glucose and maltose is minimal (Henriksson *et al.*, *J Biotechnol*, 2000). Oxidation of reduced MtCDH-2, obtained by incubation with cellobiose, was measured as the decrease in absorbance at 430 nm of the reduced heme cofactor. The rate of oxidation was measured by monitoring the decrease in the absorption at 430 nm. Immediately before data collection, 1 μ M of as-isolated MtCDH-2 in 50 mM sodium acetate pH 5.0 was mixed with 6 μ M cellobiose. To test the effect of NCU08746 on the rate of CDH oxidation, 5 μ M NCU08746 was added to the CDH solution prior to mixing with cellobiose. All the assays were carried out at room temperature and in the presence of atmospheric oxygen. This oxidation occurs slowly in the presence of atmospheric oxygen but was significantly enhanced in the presence of NCU08746 (**FIG. 4A**), indicating that NCU08746 is an efficient electron acceptor for MtCDH-2. Subsequently, NCU08746 activity assays on amylopectin were carried out with MtCDH-2 as the electron donor. In the presence of MtCDH-2 only, NCU08746 exhibited very weak activity on amylopectin (**FIG. 4B, trace C**), which is consistent with the low activity of CDH on glucose and maltodextrins (Henriksson *et al.*, *J Biotechnol*, 2000). In the presence of MtCDH-2 and cellobiose, NCU08746 activity (**FIG. 4B, trace B**) was comparable to that in the presence of excess ascorbic acid (**FIG. 4B, trace A**). Together, these results indicated that CDH can serve as an electron donor for NCU08746 catalysis.

Truncated NCU08746 Activity Assays

[0126] Two truncated NCU08746 lacking the CBM20 domain (**SEQ ID NO: 6**) and (**SEQ ID NO: 87**) at 4 μ M or 5 μ M were assayed under the following conditions: 2 mM ascorbic acid, 50 mg/mL amylose from potato, 50 mM sodium acetate pH 5.0, 42 °C, 4 hours. Trunc1 includes the NCU08746 catalytic domain but lacks CBM20. Trunc2 contains the catalytic domain plus the linker region, but lacks the CBM20. Trunc1 exhibits approximately half the

activity of the full-length enzyme on amylose (**FIG. 17**). The activity of Trunc2 is similar to Trunc1, about half as active as full-length enzyme (data not shown). In initial experiments, the truncated protein (Trunc1) appeared to lose affinity for copper, but still required CuSO_4 for activity on amylose (See **FIG. 16**). However, subsequent experiments (EPR experiments described below & activity assays with new batch of enzyme purified with an improved protocol) suggest that Trunc1 does not in fact lose affinity for copper.

[0127] **FIG. 18** shows X-band EPR spectra of wild type NCU08746 and the two truncation mutants. The three spectra are essentially identical, suggesting that the coordination environment around Cu is the same for the three samples. The EPR experiments were performed with 50 mM MES (pH 5.0), 150 mM NaCl, 20% glycerol. The NCU08746 is ~0.5 mM and is fully reconstituted with copper.

Active site characterization of Cu(II)-NCU08746

[0128] Because NCU08746 contains two domains with a flexible linker region, it is not readily amenable to crystallization. Thus, the copper active site was characterized using X-ray absorption spectroscopy, which provides information on the local structure of the copper center.

[0129] The Cu K-edge X-ray absorption near edge spectrum (XANES) of Cu(II)-NCU08746 is typical of a five- or six-coordinate copper(II) species containing oxygen/nitrogen ligands, which does not exhibit any $1s \rightarrow 3d$ or $1s \rightarrow 4p$ pre-edge transition features (**FIG. 14**) (Sarangi *et al.*, *Coord Chem Rev*, 2013).

[0130] Extended X-ray absorption fine structure (EXAFS) data and the resulting Fourier transform are shown in **FIG. 5**. The Fourier transform exhibits a strong inner shell feature at $r' \sim 1.5 \text{ \AA}$, a second shell feature at $r' \sim 2.3 \text{ \AA}$, and third shell features between $r' = 2.3\text{--}3.7 \text{ \AA}$. The third shell features in the Fourier transform correspond to the double-humped feature centered at

$\sim 4 \text{ \AA}^{-1}$ in the EXAFS spectrum, which arises from the multiple scattering paths of several imidazole moieties as found in the spectra of many other metalloproteins and model complexes (Vu *et al.*, *J Am Chem Soc*, 2011, Pellei *et al.*, *Dalton Trans*, 2011, Sanyal *et al.*, *J Am Chem Soc*, 1993, Costello *et al.*, *J Biol Inorg Chem*, 2006, D'Angelo *et al.*, *Biochemistry*, 2005).

[0131] Fitting progress is described in **Table 5**. Table 5 shows the fitting progress to the EXAFS data of NCU08746. The inner shell feature can be fitted with 5 or 6 Cu-O/N paths, which needs to be split into several subshells. Splitting this shell into two subshells significantly improves the fit quality. Including a third subshell of one Cu-O/N path at $\sim 2.5 \text{ \AA}$ also yields a better fit. The total coordination number deduced by EXAFS analysis is thus consistent with the XANES data. The second shell feature can be simulated with several Cu $\bullet\bullet\bullet$ C paths, which may arise from imidazole moieties or backbone atoms of coordinating residues. The third shell features cannot be fitted with several single scattering Cu $\bullet\bullet\bullet$ C/N paths around 4.3 \AA ; however, it is well simulated when all significant single and multiple scattering paths associated with an imidazole moiety are included in a rigid body model. The coordination number of the imidazole moiety was floated during the fit and was subsequently optimized to a value between 2 and 3. This result indicates that Cu(II)-NCU08746 contains two or three histidine ligands. The best fit includes 4 Cu-O/N paths at 1.97 \AA , 1 Cu-O/N path at 2.22 \AA , 1 Cu-O/N path at 2.42 \AA , 2 Cu-C paths at 3.23 \AA , and paths from 2.3 imidazole moieties (**Table 5**). This result indicates that the copper center of NCU08746 contains two or three histidine ligands.

[0132] Cu(II)-NCU08746 was prepared in 100 mM MES buffer pH 5.0, concentrated and mixed with 20% glycerol to obtain a final concentration of 1.28 mM enzyme. The sample was transferred to XAS solution sample holder. XAS data was collected at Beamline X3B of National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. To minimize

photoreduction of the sample, 8 first scans were collected on 8 different sample spots (0.9 mm × 5.8 mm). Standard procedure was applied to reduce, calibrate, and average data using EXAFSPAK (George and Pickering (Stanford Synchrotron Radiation Laboratory, Stanford Linear Accelerator Center), Stanford University; Stanford, CA, 2000). EXAFS fitting was performed using the OPT function of EXAFSPAK. Theoretical scattering paths were calculated with FEFF8.4 (Rehr *et al.*, *J Am Chem Soc*, 1991) using the model shown in **FIG. 15**. The r and σ^2 values of all significant single and multiple scattering paths of the imidazole moiety were linked together in the fit. ICP-OES analysis of 27 elements was performed at the Research Analytical Laboratory (RAL) at the University of Minnesota–Twin Cities. The samples were prepared according to the instructions from RAL.

[0133] **Table 5.** Fitting results to unfiltered k^3 -weighted EXAFS data of Cu(II)-NCU08746.

Fit #	Cu-O/N			Cu-N/O			Cu-N/O			Cu...C			Cu...Im			F'
	N	R	σ^2	N	R	σ^2	N	R	σ^2	N	R	σ^2	N	R	σ^2	
1	6	1.93	12.0													154
2	4	1.95	7.1	2	2.27	27.5										104
3	3	1.96	4.3	1	2.17	5.6	2	2.44	17.6							93
4	3	1.95	4.9	2	2.23	25.3	1	2.53	5.7							91
5	3	1.95	4.8	1	2.19	15.7	1	2.51	7.14							93
6	4	1.95	6.9	1	2.23	9.2										109
7	4	1.96	7.4	1	2.25	5.1	1	2.49	5.8							96
8	4	1.95	7.6	1	2.28	5.9	1	2.51	3.8	2	3.28	1.5				85
9	4	1.97	6.9	1	2.22	1.6	1	2.42	8.2	2	3.23	6.0	2.2lm^a	n/a	n/a	32
10	4	1.93	7.9	1	2.31	4.9	1	2.52	1.6	2	3.27	0.7	4C ^b	4.35	6.5	75

Discussion

[0134] Bioinformatic, biochemical, and spectroscopic studies reported show that NCU08746 represents a member of a new family of PMOs that cleave starch. Sequence analysis reveals that this family contains the conserved histidine residues that are expected to form the histidine brace motif, as well as the motif N/Q/E-X-F/Y that contains the active site tyrosine

residue in PMOs (**FIG. 1**). MS/MS analysis supports the presence of an N-terminal methylhistidine residue in NCU08746. Activity assays indicate that copper is the native metal cofactor of this enzyme. ICP analysis shows that it contains one Cu atom per protein molecule. Cu K-edge EXAFS analysis indicates that purified NCU08746 contains a copper(II) center with two or three histidine ligands. These features closely resemble those of the active sites in chitin- and cellulose-active PMOs (**FIG. 1A**) (Hemsworth *et al.*, *Curr Opin Chem Biol*, 2013, Vaaje-Kolstad *et al.*, 2010, *Science*, 2010, Hemsworth *et al.*, *Nat Chem Biol*, 2014, Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Karkehabadi *et al.*, *J Mol Biol*, 2008, Li *et al.*, *Structure*, 2012, Wu *et al.*, *J Biol Chem*, 2013).

[0135] The proposed structure for the active site of NCU08746 is similar to that shown in **FIG. 1A** in which the elongated hexacoordinate copper(II) site contains two histidine ligands. The coordination number is consistent with both the XANES features and EXAFS analysis. The coordination sphere deduced from EXAFS fitting contains 4 O/N ligands at 1.97 Å from the copper center and two additional O/N ligands at longer distances, 2.22 and 2.42 Å, which is consistent with the elongated octahedral geometry as predicted by Jahn-Teller effect. The uncertainty in the coordination number deduced by EXAFS analysis does not rule out a third histidine ligand. There are three conserved histidine residues throughout the NCU08746 family, which are also observed in fungal cellulose-active PMOs (Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Karkehabadi *et al.*, *J Mol Biol*, 2008, Li *et al.*, *Structure*, 2012, Wu *et al.*, *J Biol Chem*, 2013, Vu *et al.*, *J Am Chem Soc*, 2014). In fungal cellulose-active PMOs, the N-terminal histidine and the second histidine, around position 70-100, form the histidine brace (Vu *et al.*, *J Am Chem Soc*, 2014); the third histidine, near the end of the catalytic domain, is involved in a conserved hydrogen bonding network at the active site

(Harris *et al.*, *Biochemistry*, 2010, Quinlan *et al.*, *Proc Natl Acad Sci USA*, 2011, Karkehabadi *et al.*, *J Mol Biol*, 2008, Li *et al.*, *Structure*, 2012, Wu *et al.*, *J Biol Chem*, 2013). The length of the catalytic domain and the distribution of the three conserved histidine residues in starch-active PMOs are similar to those of cellulose-active PMOs, suggesting that starch-active PMOs likely adopt a homologous structural arrangement of the conserved histidine residues.

[0136] Results show that NCU08746 oxidatively cleaves starch substrates. There are no other enzymes with such activity reported thus far. Oligosaccharide dehydrogenases (Tessema *et al.*, *Anal Chem*, 1997) and maltose dehydrogenases (Kobayashi *et al.*, *Enzymol*, 1980) are capable of oxidizing the reducing ends of a number of oligosaccharides, including maltodextrins. However, these enzymes have not been reported to cleave glycosidic bonds. Starch-active PMOs introduce a new perspective on starch degradation and metabolism.

[0137] Starch is known to be degraded mainly by alpha-amylase and glucoamylase, two efficient hydrolytic enzymes currently utilized in the starch-based biofuel industry. Starch-active PMOs may help to further the efficiency of these enzymes.

[0138] NCU08746 requires both oxygen and a source of electrons. As shown, ascorbic acid is an efficient electron donor for the NCU08746 reaction. It is possible that oligosaccharide dehydrogenases or maltose dehydrogenases can serve as biological electron donors to starch-active PMOs; however, they have not been well characterized (Tessema *et al.*, *Anal Chem*, 1997, Kobayashi *et al.*, *Enzymol*, 1980). Cellobiose dehydrogenase (CDH), which is widely accepted as the biological partner of cellulose-active PMOs, can donate electrons to NCU08746 as efficiently as ascorbic acid. This result suggests that NCU08746 and other family members may possess some protein structural features similar to those in cellulose-active PMOs that allow the transfer of electrons from CDH to the copper center. It is also possible that CDH is the

biological partner of starch-active PMOs. In the wild, fungi routinely encounter food sources containing mixed polysaccharides, and co-expression of enzymes that cleave diverse substrates would allow for parallel degradation pathways. Transcriptomic and proteomic analysis of the *N. crassa* secretome on different growth substrates would shed light on the potential biological relationship between CDH and starch-active PMOs.

[0139] The discovery and characterization of starch-active PMOs shows that the bioinformatics approach described above can be used to identify new PMO families. There are two more uncharacterized potential PMO families listed in **Table 1**, whose substrates might be different from chitin, cellulose, and starch. This finding suggests the existence of a PMO superfamily with a much broader range of substrates. Unlike the uniform microcrystalline structure of chitin and cellulose (**Fig. 1B**), the higher order structure of starch is complicated and varied. Starch is a mixture of amylose and amylopectin with the relative ratio varying depending on the source.

[0140] Amylose contains $\alpha(1\rightarrow4)$ glycosidic linkages and exists in disordered, single helical, and double helical forms. Amylopectin contains $\alpha(1\rightarrow6)$ bonds in addition to $\alpha(1\rightarrow4)$ bonds, which creates branches (**FIG. 1C**). Double helices can form between amylose and amylopectin in starch, which further complicates the structure (Perez and Bertoft, *Stärke*, 2010, Popov *et al.*, *Macromolecules*, 2009, Imberty *et al.*, *J Mol Biol.* 1988). Recently, a cellulose-active PMO from *N. crassa* (Vu *et al.*, *J Am Chem Soc*, 2014) was found to act on soluble cellodextrins (Isaksen *et al.*, *J Biol Chem*, 2014) and hemicelluloses that contain $\beta(1\rightarrow4)$ linked glucose and glucose derivative units (Agger *et al.*, *Proc Natl Acad Sci U S A*, 2014). The crystalline surface has thus not been required for all previously identified PMOs; however, $\beta(1\rightarrow4)$ linkages of glucose and glucose derivatives have been essential. Our finding emphasizes

that PMOs can act on substrates with higher order structures and glycosidic linkages beyond those found in cellulose and chitin.

[0141] Although Applicants do not wish to be bound by theory, a proposed mechanism of NCU08746 is shown in **FIG. 6**, analogous to that of cellulose-active PMOs and chitin-active PMOs. NCU08746 is shown here to oxidize the C1 position in both amylose and amylopectin. For amylopectin, it is unclear if only the $\alpha(1\rightarrow4)$ bond is cleaved or if NCU08746 can attack both $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages. In addition, although C4 and C6 oxidized products are not detected for NCU08746, oxidation at these positions is relevant to bond cleavage and may be carried out by other starch-active PMOs. C4 oxidation is known to occur with some cellulose-active PMOs (Phillips *et al.*, *ACS Chem Biol*, 2011, Beeson *et al.*, *J Am Chem Soc*, 2012, Vu *et al.*, *J Am Chem Soc*, 2014, Isaksen *et al.*, *J Biol Chem*, 2014) and has recently been implicated in fungal chitin-active PMOs (Hemsworth *et al.*, *Nat Chem Biol*, 2014).

INFORMAL SEQUENCE LISTING

Catalytic domain motif (SEQ ID NO: 1):

HGYXXXPSRX(7,24)GDXCP[ED]CXIXEPVXXWPXXXXAXVGRXGXCGXNARXXXDYNXPXXX[WY]GXXXXXXXXXXXXXXXXXXVXWCXDXNGDHGGXXXXX[IV]CXXQXXXXXXXXXXPXXPXXX[ED]KXAA[EQ]XCFXXGXLXCXXVXGXXCXXXXDCXXXXCXXDXWFX(0,20)[TA]CX(7,13)CXXVDXXXXXSCXTXIXGGXXXXXXXXXXPX(3,4)SXHTLXXXXWNXX[QE]TXQXYXXCXDXXI

X = any amino acid

[] = any one of the characters enclosed in the brackets, e.g., [ED] means one occurrence of E or D

(m,n) = the preceding residue is repeated between m to n times, $m < n$

Catalytic domain motif version 2 (specifying residues for certain X's) (SEQ ID NO: 2)

HGY[LM]XXPSRXXXXX[EQLS]X(1,18)GDXCP[ED]C[TAS]IXEPVXXWPX[VL]XXAXVGR[SNT]GXCGXNARXXXDYNXPXXX[WY]GXXXXXXXXXXXXXXXXXXVXWCXDXNGDHGGXXXXX[IV]CXXQXXXXXXXXXXPXXPXXX[ED]KXAA[EQ]XCFXXGXLXCXXVXGXXCXXXXDCXXXXXCXXDXWFX(0,20)[TA]CX(7,13)CXXVDXXXXXSCXTXIXGG[YF][TPK]XXXXXXXXPX(3,4)SXHTLXXXXWNXX[QE]TXQXYXXCXDXXI

X = any amino acid

[] = any one of the characters enclosed in the brackets, e.g., [ED] means one occurrence of E or D

(m,n) = the preceding residue is repeated between m to n times, $m < n$

Conserved motifs specific to starch-active PMOs:

HGY (SEQ ID NO: 3)

N-G-D-H-G-G (SEQ ID NO: 4)

T-X-Q-X-Y (SEQ ID NO: 5)

Truncated NCU08746 without linker and CBM20 (Trunc1; SEQ ID NO: 6):

HGYLTIPFSRTRLGAEAGLDTCPCECSILEPVTAWPNVTEAKVGRSGPCGYNARVSIDYNQPATNWGNPVTYTAGDTVDVQWCVDHNGDHGGMFSYRICQDQELVNKFLTPGYLPTEAEKQAAEDCFEKGTLPCDVGNGQSCDFSPDCQQGQACWRNDWFTCNAFQADSRRCQGVDNAALGSCFTTIAGGYTVTKKIKIPNYISGHTLLSFRWNSFQTAQVYLSCADIAIVGD

Full-length NCU08746 with CBM20 (without signal sequence) (SEQ ID NO: 7):

HGYLTIPFSRTRLGAEAGLDTCPCECSILEPVTAWPNVTEAKVGRSGPCGYNARVSIDYNQPATNWGNPVTYTAGDTVDVQWCVDHNGDHGGMFSYRICQDQELVNKFLTPGYLPTEAEKQAAEDCFEKGTLPCDVGNGQSCDFSPDCQQGQACWRNDWFTCNAFQADSRRCQGVDNAALGSCFTTIAGGYTVTKKIKIPNYISGHTLLSFRWNSFQTAQVYLSCADIAIVGDSASTTKVSATATTLVTSSKTASASCTPAATVAVTFNHLASTSYGESIKIVGSISQL

GSWSASSGVALSASQYTTSNPLWTATVSLPAGTKFEYKFFVKVSSEGSVAVTWESDPNRSY
TVPQSCAESVAVESSWK

Homologs without CBM20 (SEQ ID NOS: 8-37)

- >gi|347014680|gb|AE062162.1| hypothetical protein MYCTH_2313229
[Myceliophthora thermophila ATCC 42464] (SEQ ID NO: 8)
- >gi|83771852|dbj|BAE61982.1| unnamed protein product [Aspergillus
oryzae RIB40] (SEQ ID NO: 9)
- >gi|211588574|emb|CAP86689.1| Pc20g13600 [Penicillium chrysogenum
Wisconsin 54-1255] (SEQ ID NO: 10)
- >gi|259479708|tpe|CBF70179.1| TPA: conserved hypothetical protein
[Aspergillus nidulans FGSC A4] (SEQ ID NO: 11)
- >gi|517321635|emb|CCT72014.1| related to glucoamylase precursor
[Fusarium fujikuroi IMI 58289] (SEQ ID NO: 12)
- >gi|114190059|gb|EAU31759.1| conserved hypothetical protein
[Aspergillus terreus NIH2624] (SEQ ID NO: 13)
- >gi|119401636|gb|EAW12059.1| chitin binding domain protein [Aspergillus
clavatus NRRL 1] (SEQ ID NO: 14)
- >gi|220695099|gb|EED51442.1| conserved hypothetical protein
[Aspergillus flavus NRRL3357] (SEQ ID NO: 15)
- >gi|342890664|gb|EGU89432.1| hypothetical protein FOXB_00056 [Fusarium
oxysporum Fo5176] (SEQ ID NO: 16)
- >gi|391871417|gb|EIT80577.1| hypothetical protein Ao3042_02852
[Aspergillus oryzae 3.042] (SEQ ID NO: 17)
- >gi|440636040|gb|ELR05959.1| hypothetical protein GMDG_01921
[Pseudogymnoascus destructans 20631-21] (SEQ ID NO: 18)
- >gi|471566263|gb|EMR66611.1| putative chitin binding domain protein
[Eutypa lata UCREL1] (SEQ ID NO: 19)
- >gi|475675494|gb|EMT72560.1| hypothetical protein FOC4_g10001196
[Fusarium oxysporum f. sp. cubense race 4] (SEQ ID NO: 20)
- >gi|477514447|gb|ENH66825.1| hypothetical protein FOC1_g10007481
[Fusarium oxysporum f. sp. cubense race 1] (SEQ ID NO: 21)
- >gi|558860477|gb|ESU10560.1| hypothetical protein FGSG_04704 [Fusarium
graminearum PH-1] (SEQ ID NO: 22)
- >gi|584138283|gb|EWG47625.1| hypothetical protein FVEG_07686 [Fusarium
verticillioides 7600] (SEQ ID NO: 23)
- >gi|587677986|gb|EWZ00304.1| hypothetical protein FOYG_00186 [Fusarium
oxysporum FOSC 3-a] (SEQ ID NO: 24)

>gi|587699742|gb|EWZ46347.1| hypothetical protein FOZG_02496 [Fusarium oxysporum Fo47] (SEQ ID NO: 25)

>gi|587736006|gb|EXA33722.1| hypothetical protein FOVG_15024 [Fusarium oxysporum f. sp. pisi HDV247] (SEQ ID NO: 26)

>gi|115401060|ref|XP_001216118.1| conserved hypothetical protein [Aspergillus terreus NIH2624] (SEQ ID NO: 27)

>gi|116196734|ref|XP_001224179.1| hypothetical protein CHGG_04965 [Chaetomium globosum CBS 148.51] (SEQ ID NO: 28)

>gi|119483138|ref|XP_001261597.1| hypothetical protein NFIA_027740 [Neosartorya fischeri NRRL 181] (SEQ ID NO: 29)

>gi|121711739|ref|XP_001273485.1| chitin binding domain protein [Aspergillus clavatus NRRL 1] (SEQ ID NO: 30)

>gi|317149073|ref|XP_001823115.2| chitin binding domain protein [Aspergillus oryzae RIB40] (SEQ ID NO: 31)

>gi|238494426|ref|XP_002378449.1| conserved hypothetical protein [Aspergillus flavus NRRL3357] (SEQ ID NO: 32)

>gi|255946143|ref|XP_002563839.1| Pc20g13600 [Penicillium chrysogenum Wisconsin 54-1255] (SEQ ID NO: 33)

>gi|367036250|ref|XP_003667407.1| hypothetical protein MYCTH_2313229 [Myceliophthora thermophila ATCC 42464] (SEQ ID NO: 34)

>gi|629662247|ref|XP_007794308.1| putative chitin binding domain protein [Eutypa lata UCREL1] (SEQ ID NO: 35)

>gi|46118386|ref|XP_384880.1| hypothetical protein FG04704.1 [Fusarium graminearum PH-1] (SEQ ID NO: 36)

>gi|67539866|ref|XP_663707.1| hypothetical protein AN6103.2 [Aspergillus nidulans FGSC A4] (SEQ ID NO: 37)

Homologs with CBM20 (SEQ ID NOS: 38-86):

>gi|345561357|gb|EGX44447.1| hypothetical protein AOL_s00188g352 [Arthrotrrys oligospora ATCC 24927] (SEQ ID NO: 38)

>gi|119398566|gb|EAW08996.1| starch binding domain protein [Aspergillus clavatus NRRL 1] (SEQ ID NO: 39)

>gi|259485090|tpe|CBF81866.1| TPA: conserved hypothetical protein [Aspergillus nidulans FGSC A4] (SEQ ID NO: 40)

>gi|114190970|gb|EAU32670.1| hypothetical protein ATEG_07286 [Aspergillus terreus NIH2624] (SEQ ID NO: 41)

>gi|452001327|gb|EMD93787.1| carbohydrate-binding module family 20 protein [Bipolaris maydis C5] (SEQ ID NO: 42)

>gi|627832851|ref|XP_007687387.1| carbohydrate-binding module family 20 protein [Bipolaris oryzae ATCC 44560] (SEQ ID NO: 43)

>gi|451849214|gb|EMD62518.1| carbohydrate-binding module family 20 protein [Bipolaris sorokiniana ND90Pr] (SEQ ID NO: 44)

>gi|578489419|gb|EUN26848.1| carbohydrate-binding module family 20 protein [Bipolaris victoriae FI3] (SEQ ID NO: 45)

>gi|628193342|ref|XP_007709189.1| carbohydrate-binding module family 20 protein [Bipolaris zeicola 26-R-13] (SEQ ID NO: 46)

>gi|615458438|ref|XP_007597106.1| starch binding domain-containing protein [Colletotrichum fioriniae PJ7] (SEQ ID NO: 47)

>gi|429859218|gb|ELA34007.1| starch binding domain protein [Colletotrichum gloeosporioides Nara gc5] (SEQ ID NO: 48)

>gi|310796060|gb|EFQ31521.1| starch binding domain-containing protein [Colletotrichum graminicola M1.001] (SEQ ID NO: 49)

>gi|380491788|emb|CCF35072.1| starch binding domain-containing protein [Colletotrichum higginsianum] (SEQ ID NO: 50)

>gi|477527745|gb|ENH79560.1| starch binding domain protein [Colletotrichum orbiculare MAFF 240422] (SEQ ID NO: 51)

>gi|640921717|gb|KDN66054.1| putative starch binding domain-containing protein [Colletotrichum sublineola] (SEQ ID NO: 52)

>gi|396486898|ref|XP_003842509.1| similar to starch binding domain containing protein [Leptosphaeria maculans JN3] (SEQ ID NO: 53)

>gi|440464203|gb|ELQ33682.1| starch binding domain-containing protein [Magnaporthe oryzae Y34] (SEQ ID NO: 54)

>gi|256727805|gb|EEU41164.1| hypothetical protein NECHADRAFT_83307 [Nectria haematococca mpVI 77-13-4] (SEQ ID NO: 55)

>gi|119412311|gb|EAW22252.1| starch binding domain protein [Neosartorya fischeri NRRL 181] (SEQ ID NO: 56)

>gi|336468006|gb|EGO56169.1| hypothetical protein NEUTE1DRAFT_130221 [Neurospora tetrasperma FGSC 2508] (SEQ ID NO: 57)

>gi|512186566|gb|EPE02356.1| starch binding domain-containing protein [Ophiostoma piceae UAMH 11346] (SEQ ID NO: 58)

>gi|211584231|emb|CAP92263.1| Pc13g11940 [Penicillium chrysogenum Wisconsin 54-1255] (SEQ ID NO: 59)

>gi|525580168|gb|EPS26418.1| starch binding domain- and chitin binding

>gi|630012761|ref|XP_007830366.1| hypothetical protein PFICI_03594
[Pestalotiopsis fici W106-1] (SEQ ID NO: 61)

>gi|630029258|ref|XP_007836777.1| hypothetical protein PFICI_10005
[Pestalotiopsis fici W106-1] (SEQ ID NO: 62)

>gi|311321950|gb|EFQ89318.1| hypothetical protein PTT_14495
[Pyrenophora teres f. teres 0-1] (SEQ ID NO: 63)

>gi|154693787|gb|EDN93525.1| hypothetical protein SS1G_09392
[Sclerotinia sclerotiorum 1980 UF-70] (SEQ ID NO: 64)

>gi|636586673|ref|XP_008024675.1| carbohydrate-binding module family 20
protein [Setosphaeria turcica Et28A] (SEQ ID NO: 65)

>gi|336260435|ref|XP_003345013.1| hypothetical protein SMAC_06790
[Sordaria macrospora k-hell] (SEQ ID NO: 66)

>gi|346972519|gb|EGY15971.1| starch binding domain-containing protein
[Verticillium dahliae VdLs.17] (SEQ ID NO: 67)

>gi|507414638|emb|CCD56877.2| carbohydrate-Binding Module family 20
protein [Botrytis cinerea T4] (SEQ ID NO: 68)

>gi|115386862|ref|XP_001209972.1| hypothetical protein ATEG_07286
[Aspergillus terreus NIH2624] (SEQ ID NO: 69)

>gi|119494507|ref|XP_001264149.1| starch binding domain protein
[Neosartorya fischeri NRRL 181] (SEQ ID NO: 70)

>gi|121704316|ref|XP_001270422.1| starch binding domain protein
[Aspergillus clavatus NRRL 1] (SEQ ID NO: 71)

>gi|154319704|ref|XP_001559169.1| hypothetical protein BC1G_02333
[Botrytis cinerea B05.10] (SEQ ID NO: 72)

>gi|156046276|ref|XP_001589670.1| hypothetical protein SS1G_09392
[Sclerotinia sclerotiorum 1980 UF-70] (SEQ ID NO: 73)

>gi|255937169|ref|XP_002559611.1| Pc13g11940 [Penicillium chrysogenum
Wisconsin 54-1255] (SEQ ID NO: 74)

>gi|302413077|ref|XP_003004371.1| starch binding domain-containing
protein [Verticillium alfalfae VaMs.102] (SEQ ID NO: 75)

>gi|302895994|ref|XP_003046877.1| hypothetical protein NECHADRAFT_83307
[Nectria haematococca mpVI 77-13-4] (SEQ ID NO: 76)

>gi|330929361|ref|XP_003302617.1| hypothetical protein PTT_14495
[Pyrenophora teres f. teres 0-1] (SEQ ID NO: 77)

>gi|38524238|emb|CAE75704.1| related to glucoamylase precursor
[Neurospora crassa] (SEQ ID NO: 78)

>gi|507414638|emb|CCD56877.2| carbohydrate-Binding Module family 20
protein [Botrytis cinerea T4] (SEQ ID NO: 79)

>gi|530478571|gb|EQB57910.1| starch binding domain-containing protein
[Colletotrichum gloeosporioides Cg-14] (SEQ ID NO: 80)

>gi|553138478|gb|EAA34371.2| starch binding domain-containing protein
[Neurospora crassa OR74A] (SEQ ID NO: 81)

>gi|590041582|gb|EXK43440.1| hypothetical protein FOMG_02397 [Fusarium
oxysporum f. sp. melonis 26406] (SEQ ID NO: 82)

>gi|596671031|ref|XP_007276973.1| starch binding domain protein
[Colletotrichum gloeosporioides Nara gc5] (SEQ ID NO: 83)

>gi|628074429|ref|XP_007701868.1| carbohydrate-binding module family 20
protein [Bipolaris sorokiniana ND90Pr] (SEQ ID NO: 84)

>gi|67538586|ref|XP_663067.1| hypothetical protein AN5463.2
[Aspergillus nidulans FGSC A4] (SEQ ID NO: 85)

>gi|86196933|gb|EAQ71571.1| hypothetical protein MGCH7_ch7g978
[Magnaporthe oryzae 70-15] (SEQ ID NO: 86)

Truncated NCU08746 without CBM20 (Trunc2; SEQ ID NO: 87)

HGYLTIPFSRTRLGAEAGLDTCPCECSILEPVTAWPNVTEAKVGRSGPCGYNARV
SIDYNQPATNWNQSPVV
TYTAGDTVDVQWCVDHNGDHGGMFSYRICQDQELVNKFLTPGYLPTEAEKQAAEDCFEKG
TLPCDVGNGQS
CDFSPDCQQGQACWRNDWFTCNQAFQADSRRCQGVDNAALGSCFTTIAGGYTVTKKIKIP
NYISGHTLLSF
RWNSFQTAQVYLSCADIAIVGDSASTTKVSATATTLVTSSKTASAS

CLAIMS

What is claimed is:

1. A method for degrading starch containing material comprising:
 - providing starch containing material;
 - providing a polypeptide comprising SEQ ID NO: 1, wherein the polypeptide is bound to a copper cofactor; and
 - incubating the starch containing material and polypeptide in the presence of oxygen and a reductant, thereby degrading the starch containing material.
2. The method of claim 1, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 6 or SEQ ID NO: 87.
3. The method of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 34.
4. The method of claim 1, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 8- 37.
5. The method of any of claims 1 to 4, wherein the polypeptide further comprises a carbohydrate binding module 20 (CBM20).
6. The method of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 7.
7. The method of claim 1, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 38-86.

8. The method of claim 1, wherein the polypeptide comprises an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 6.

9. The method of claim 1, wherein the polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 6.

10. The method of claim 1, wherein the polypeptide comprises an amino acid sequence encoded by a polynucleotide that hybridizes under medium stringency conditions with the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6 or the complementary strand of the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6.

11. The method of claim 1, wherein the polypeptide comprises an amino acid sequence encoded by a polynucleotide that hybridizes under high stringency conditions with the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6 or the complementary strand of the polynucleotide encoding the amino acid sequence of SEQ ID NO: 6.

12. The method of claim 1, wherein the polypeptide comprises an amino acid sequence having a substitution, deletion, and/or insertion of up to 40 amino acids of the amino acid sequence of SEQ ID NO: 6.

13. The method of claim 1, wherein the polypeptide comprises an amino acid sequence having a substitution, deletion, and/or insertion of up to 20 amino acids of the amino acid sequence of SEQ ID NO: 6.

14. The method of claim 1, wherein the polypeptide comprises an amino acid sequence having a substitution, deletion, and/or insertion of up to 10 amino acids of the amino acid sequence of SEQ ID NO: 6.

15. The method of any of claims 1 to 14, wherein the reductant is an iron-containing compound.
16. The method of claim 15, wherein the iron-containing compound comprises a heme domain.
17. The method of claim 16, wherein the iron-containing compound is cellobiose dehydrogenase (CDH).
18. The method of claim 15, wherein the iron-containing compound is ferrous sulfate.
19. The method of any of claims 1 to 14, wherein the reductant is an arene.
20. The method of any of claims 1 to 14, wherein the reductant is ascorbic acid
21. The method of any of claims 1 to 20, further comprising providing copper.
22. The method of any of claims 1 to 21, wherein at least 90% of the polypeptide is bound by copper.
23. The method of any of claims 1 to 22, further comprising providing amylase.
24. The method of any of claims 1 to 23, further comprising providing oxygen.
25. The method of any of claims 1 to 24, wherein the products of starch degradation comprise glucose and gluconic acid.
26. The method of any of claims 1 to 25, wherein the incubating takes places at a temperature ranging from 25°C to 70°C.

27. A method of producing a fermentation product, comprising
- degrading starch-containing material according to the method of claim 1 to form glucose; and
 - culturing the glucose with one or more fermentative microorganisms or a chemical solution under conditions sufficient to produce a fermentation product.

28. A composition comprising
- a polypeptide comprising the amino acid sequence of SEQ ID NO: 6 and copper; and
 - a reductant.

29. The composition of claim 28, wherein the reductant is an iron-containing compound.

30. The composition of claim 29, wherein the iron-containing compound comprises a heme domain.

31. The composition of claim 30, wherein the iron-containing compound is cellobiose dehydrogenase (CDH).

32. The composition of claim 29, wherein the iron-containing compound is ferrous sulfate.

33. The composition of claim 28, wherein the reductant is an arene.

34. The composition of claim 28, wherein the reductant is ascorbic acid.

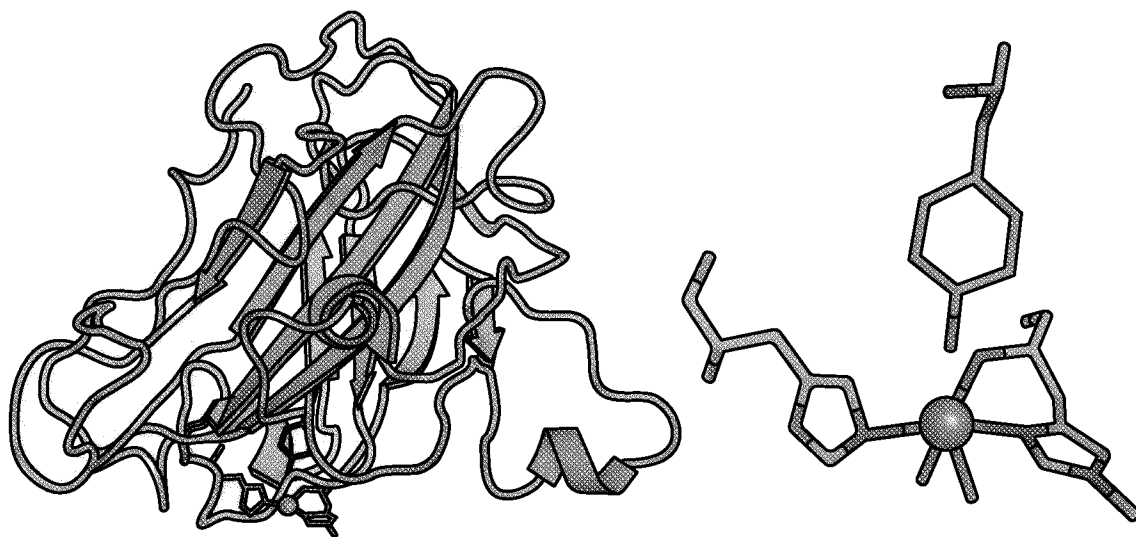


FIG. 1A

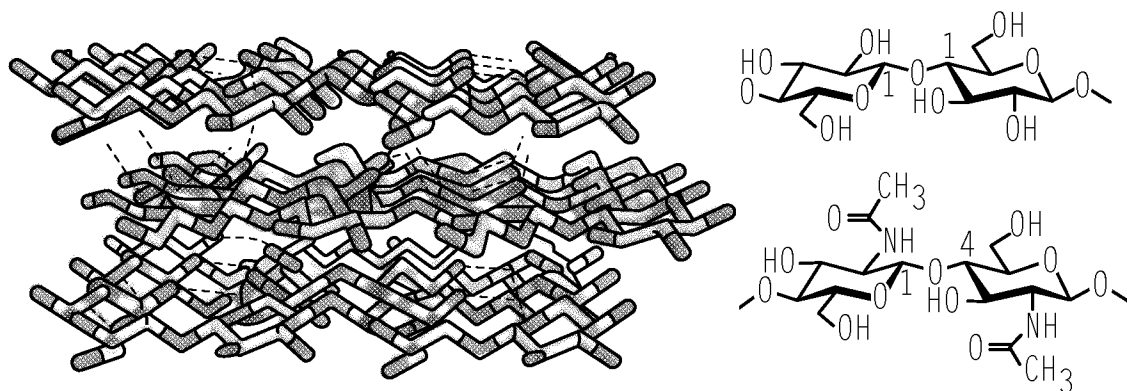


FIG. 1B

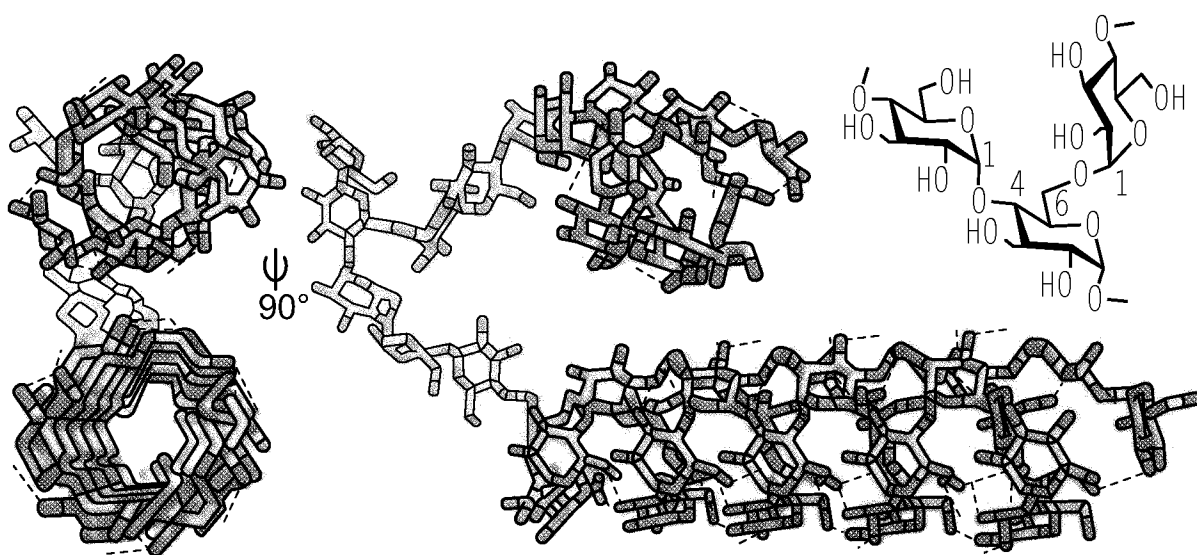


FIG. 1C

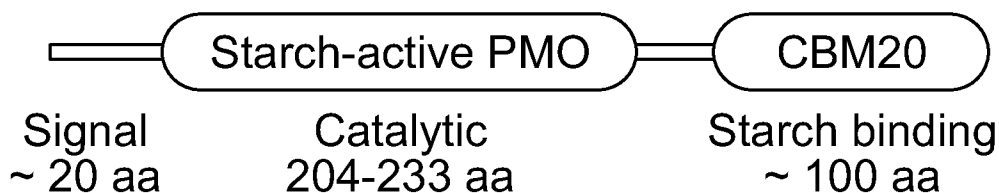


FIG. 2A

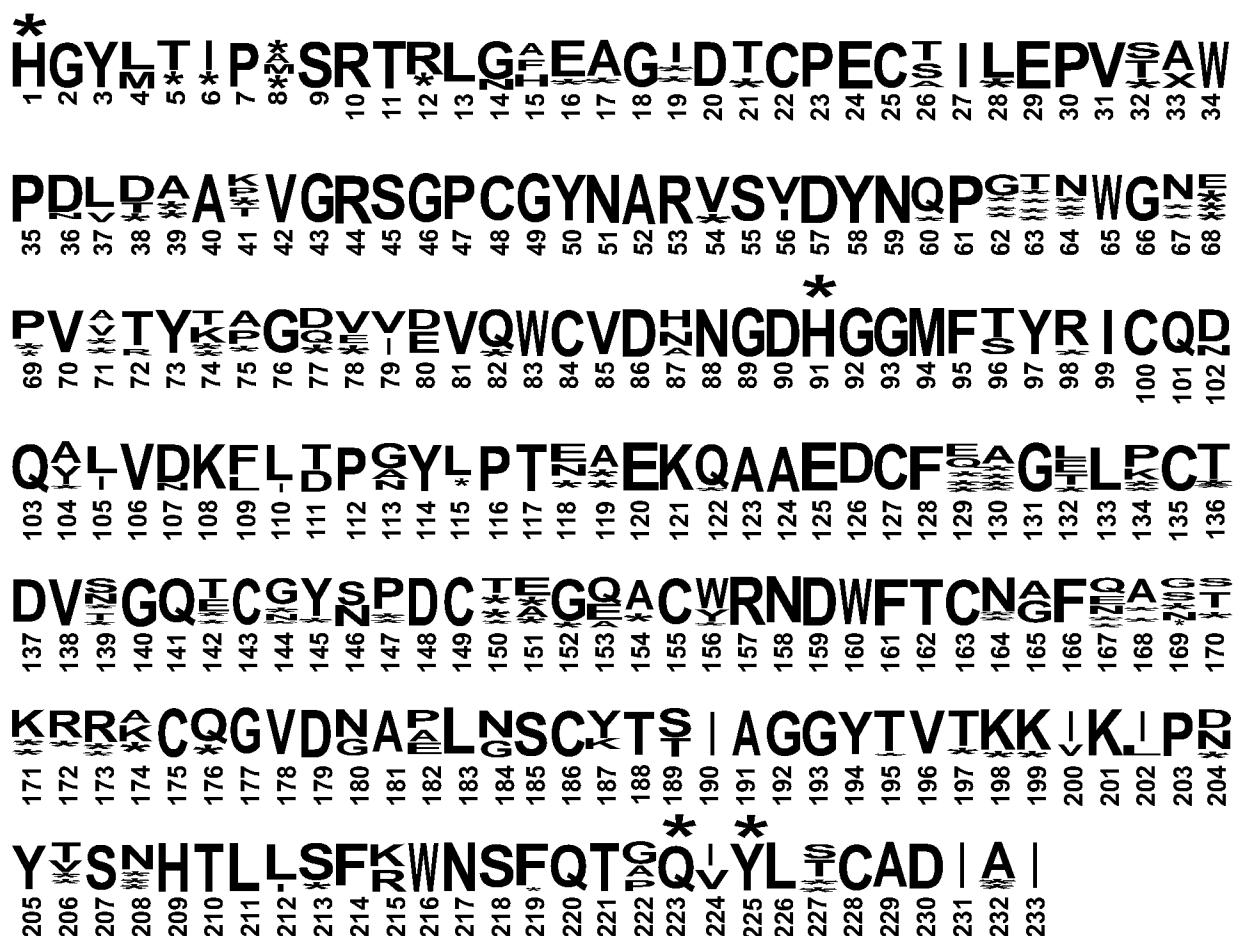


FIG. 2B

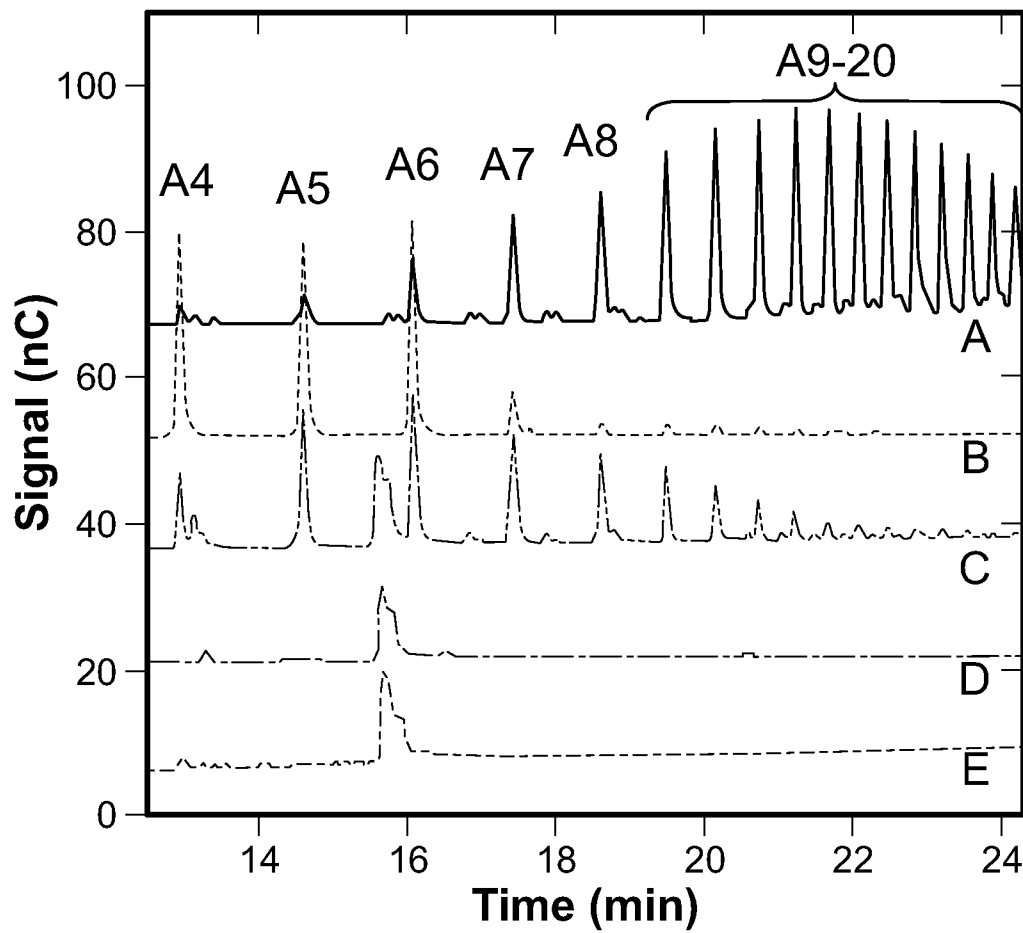


FIG. 3

4/19

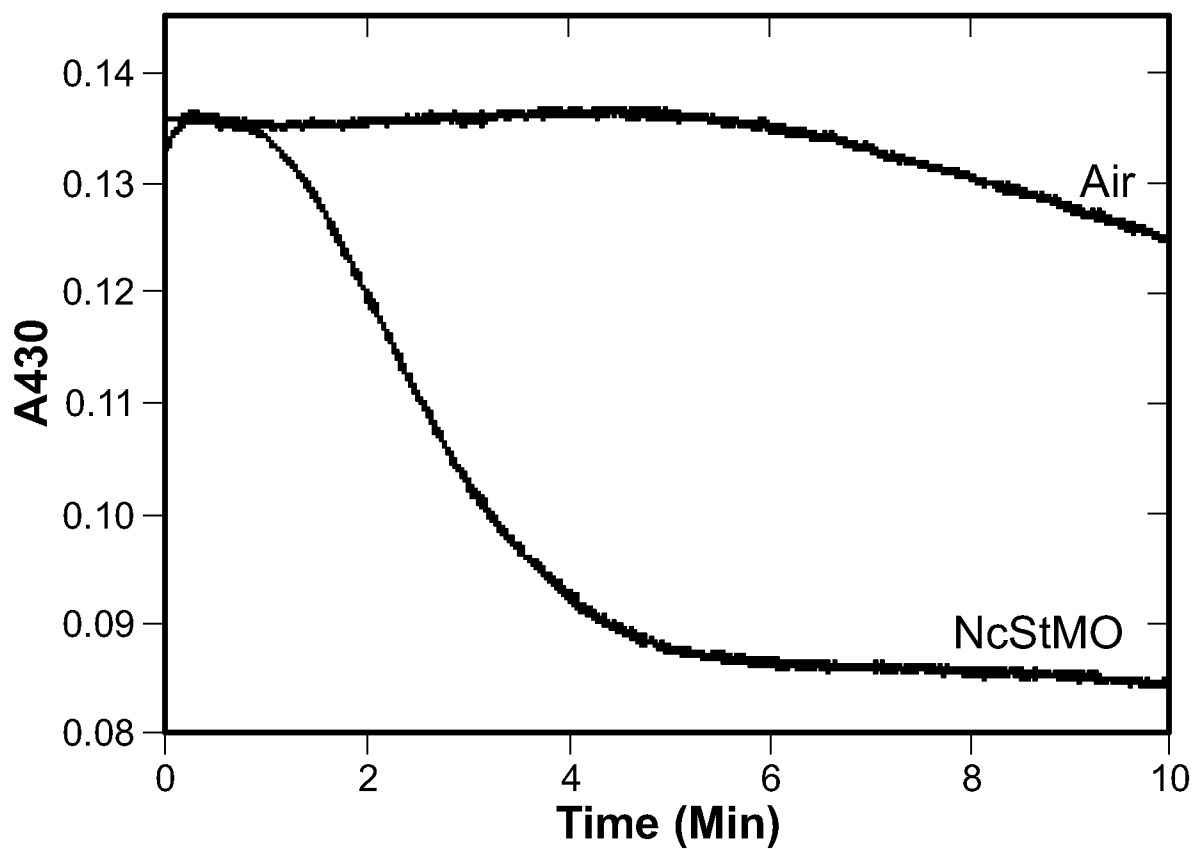


FIG. 4A

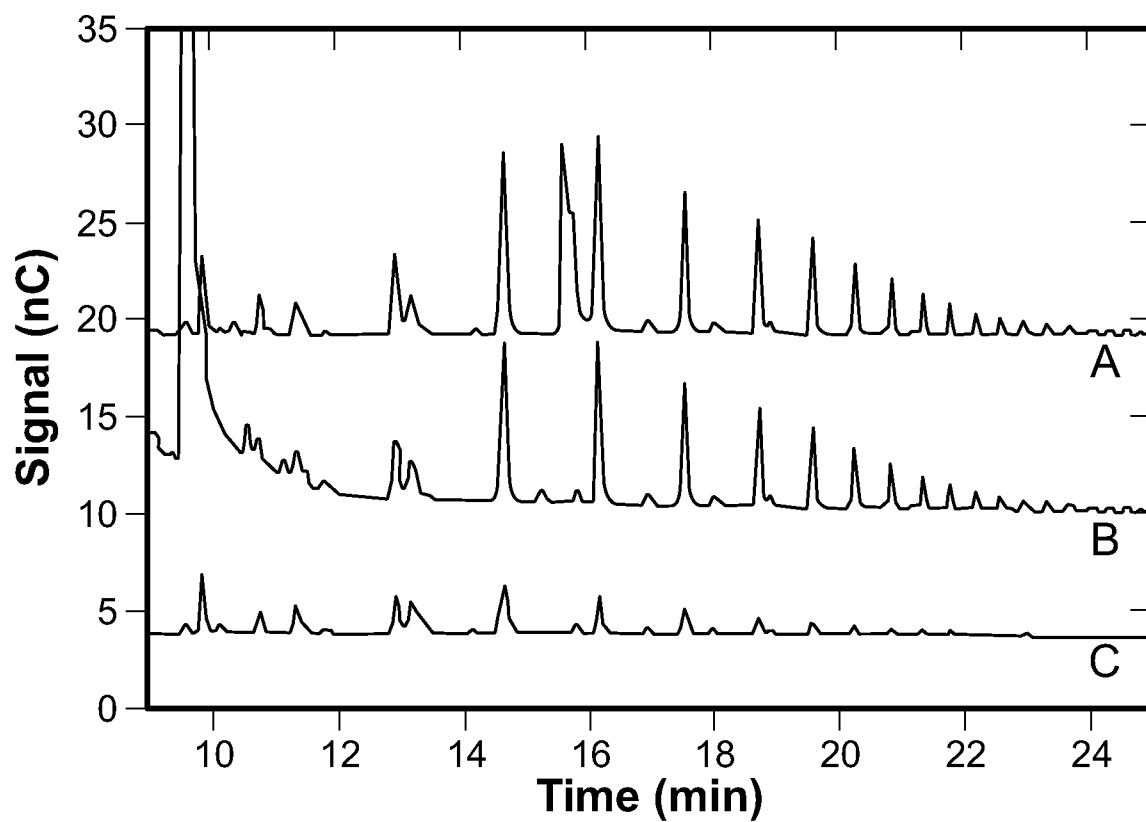


FIG. 4B

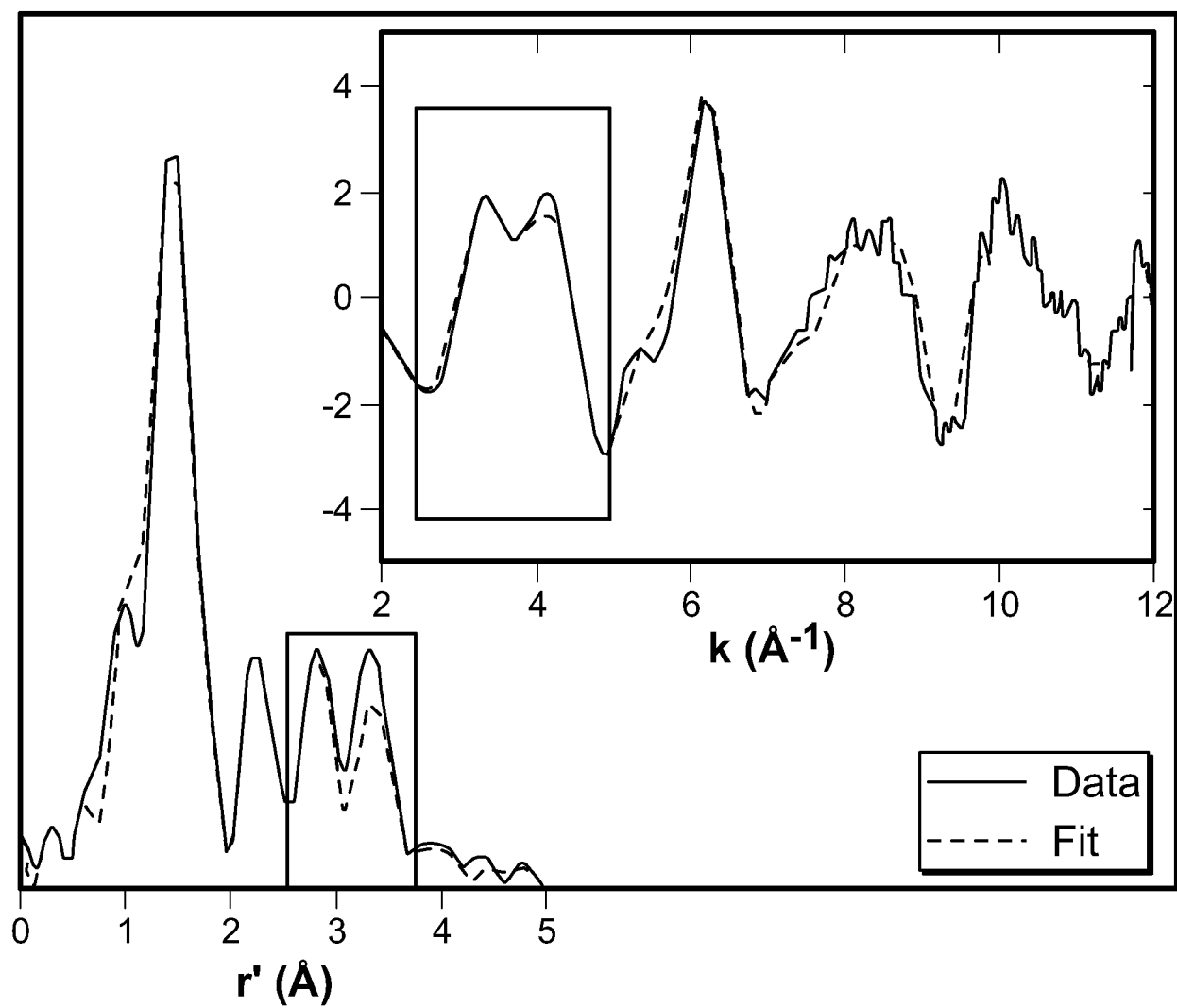


FIG. 5

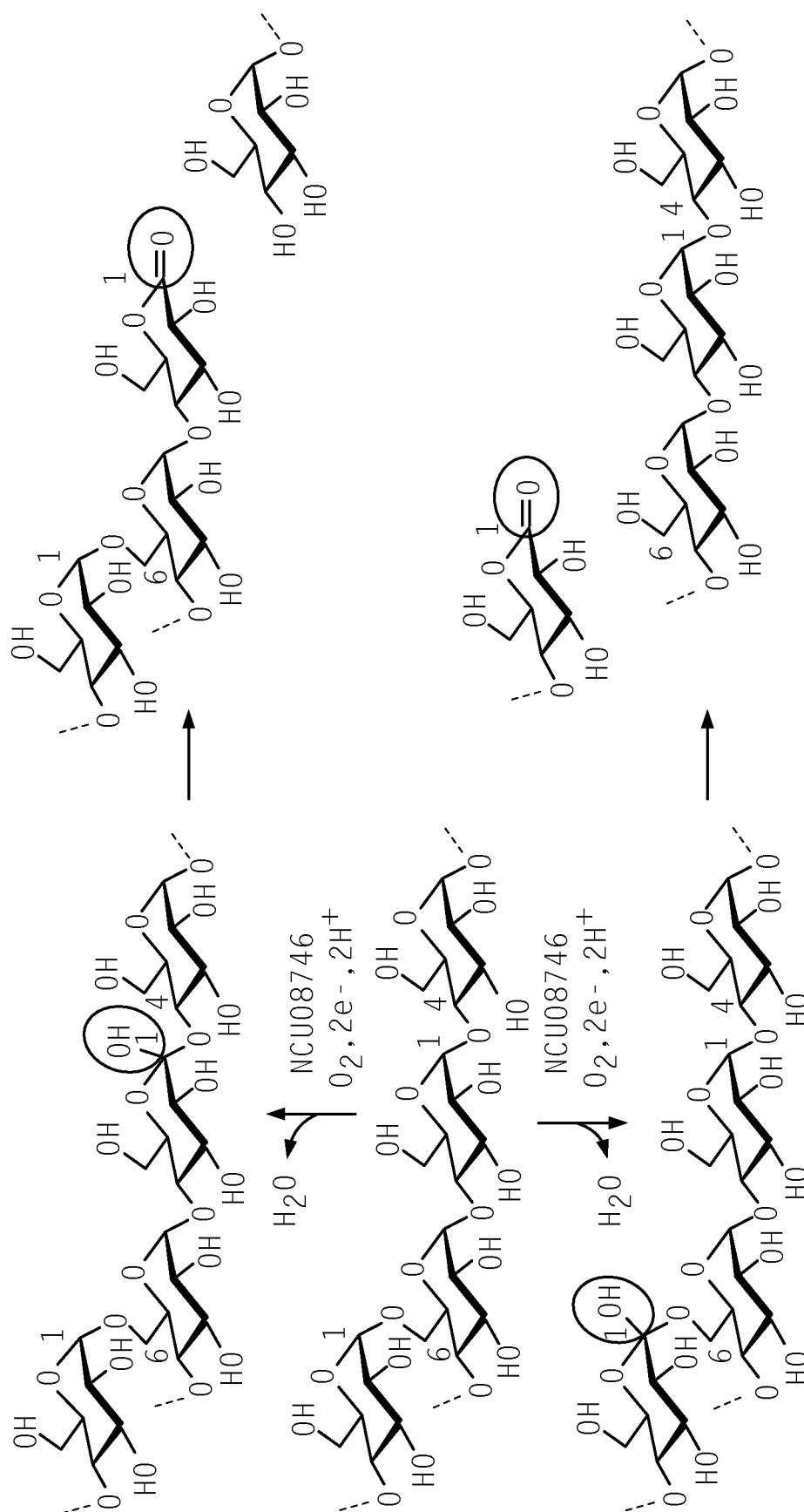
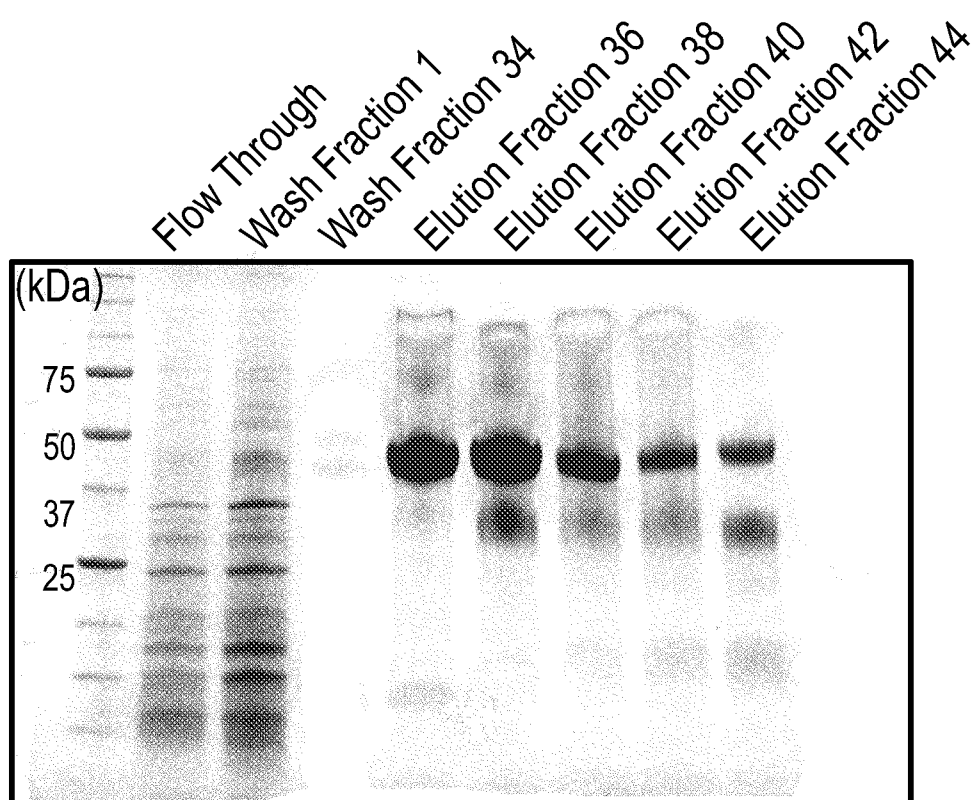


FIG. 6

7/19



NCU08746

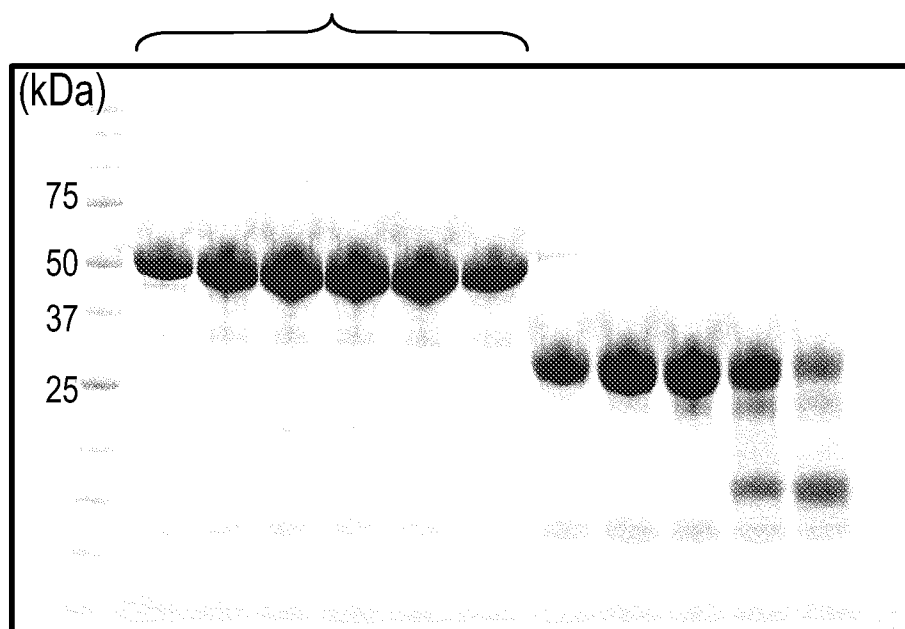


FIG. 7

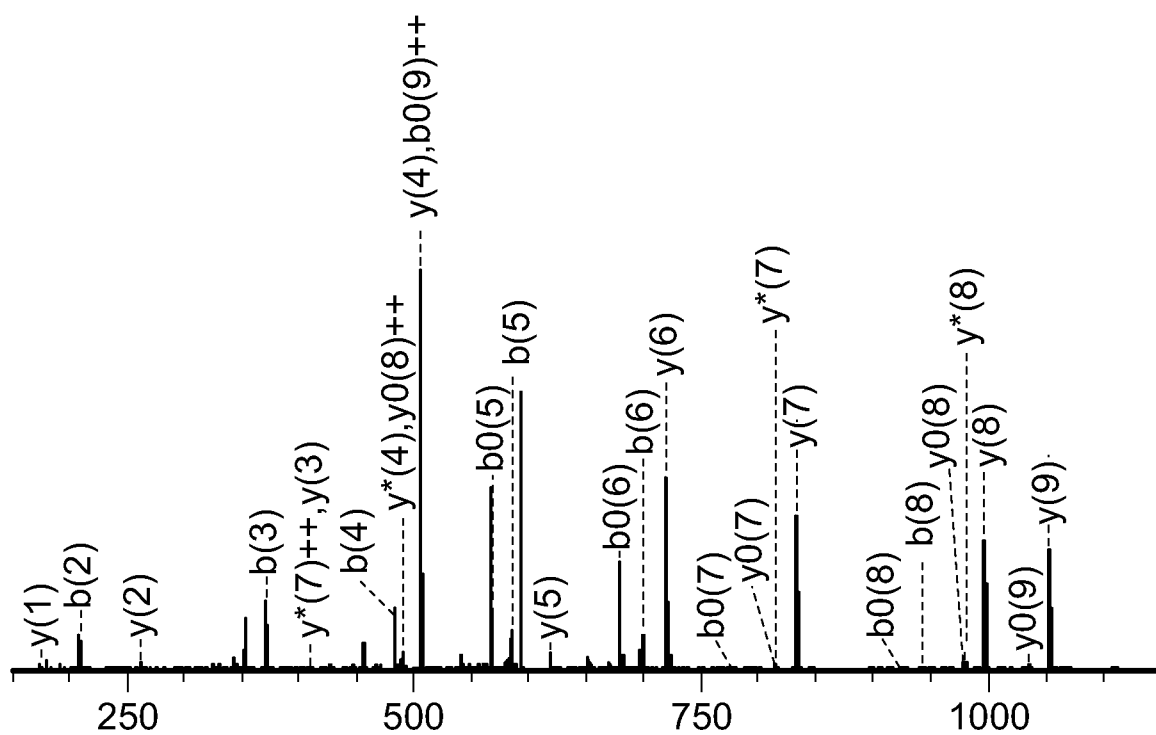


FIG. 8

9/19

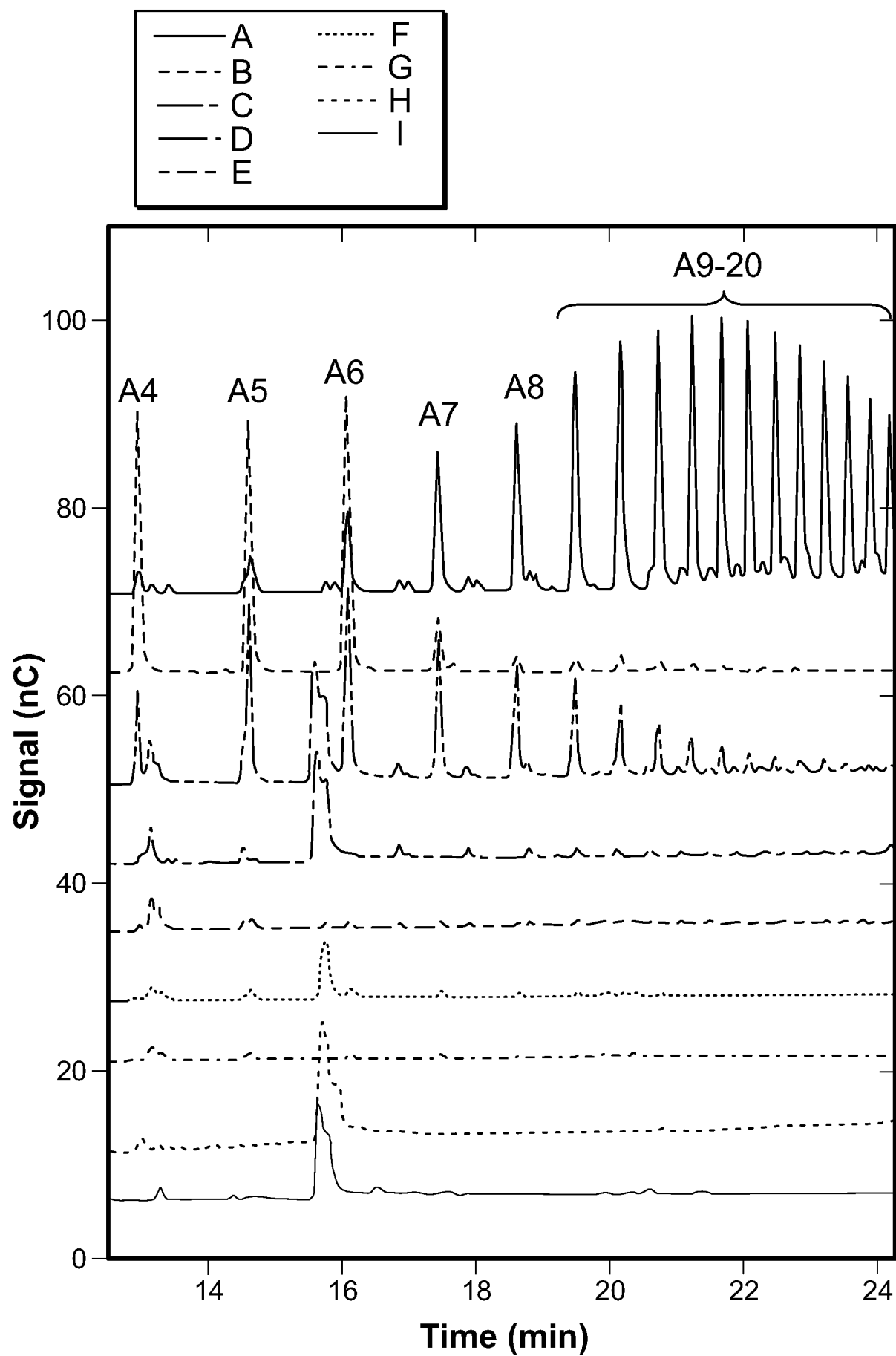


FIG. 9

SUBSTITUTE SHEET (RULE 26)

10/19

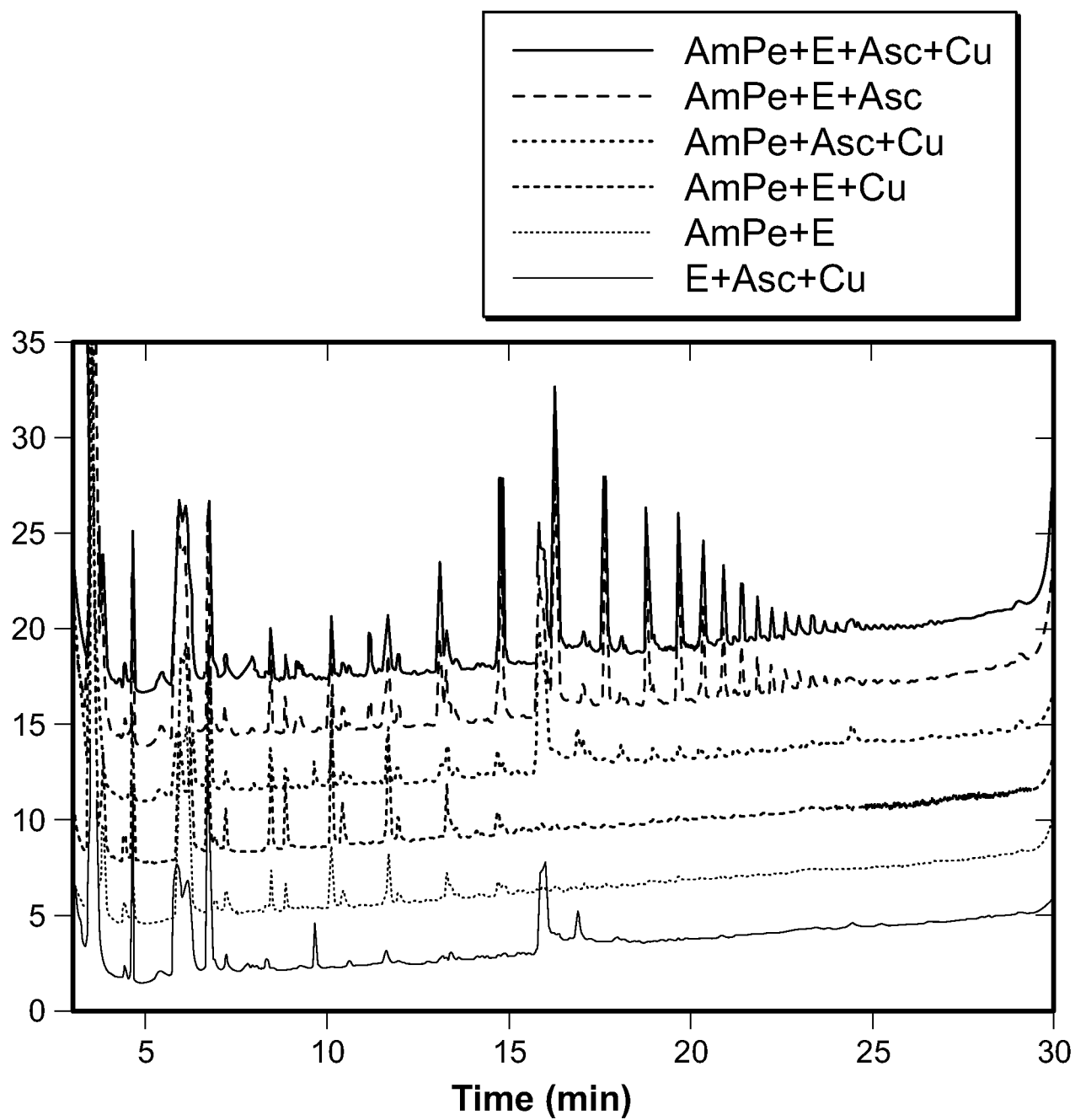


FIG. 10

11/19

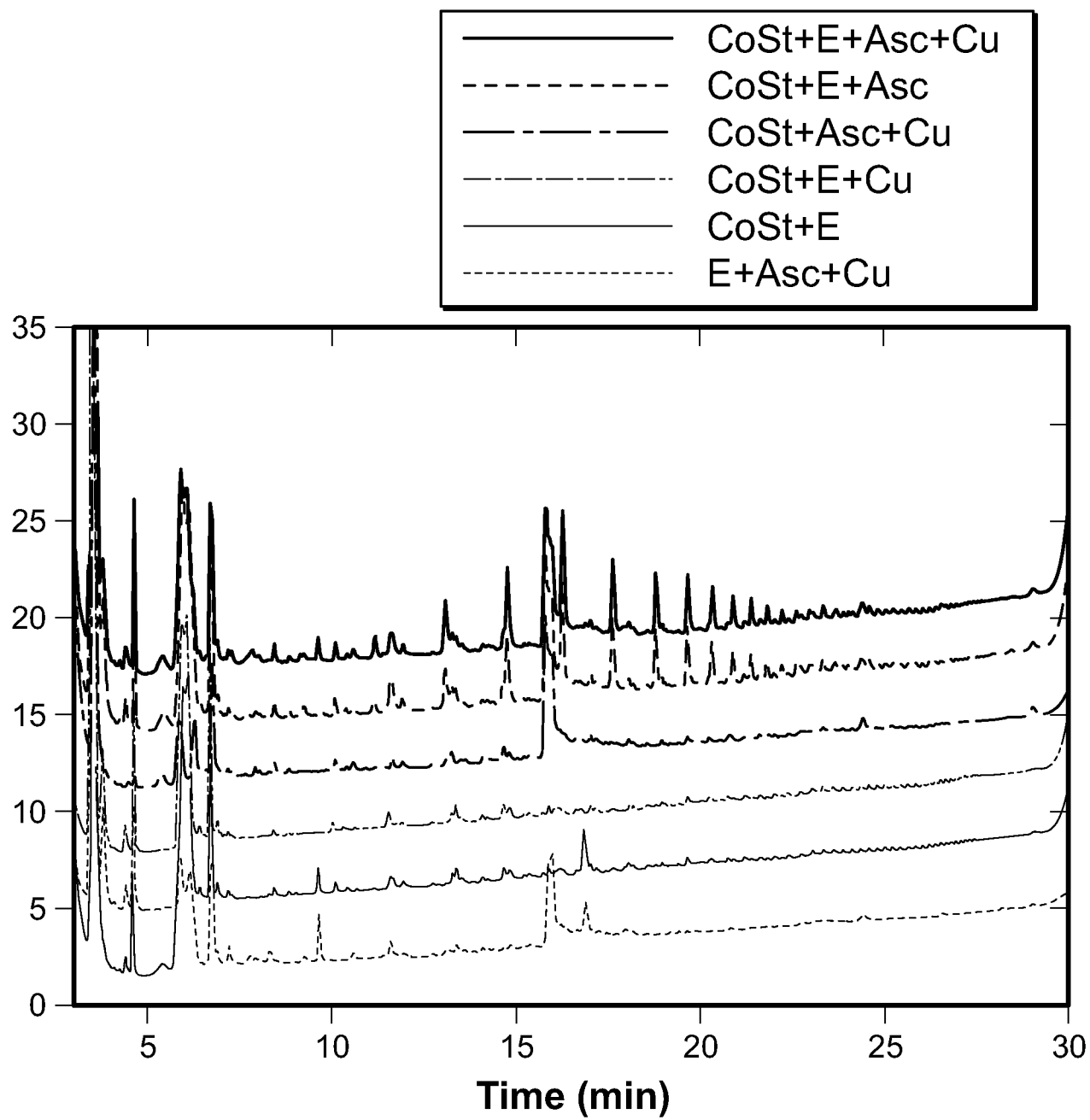


FIG. 11

12/19

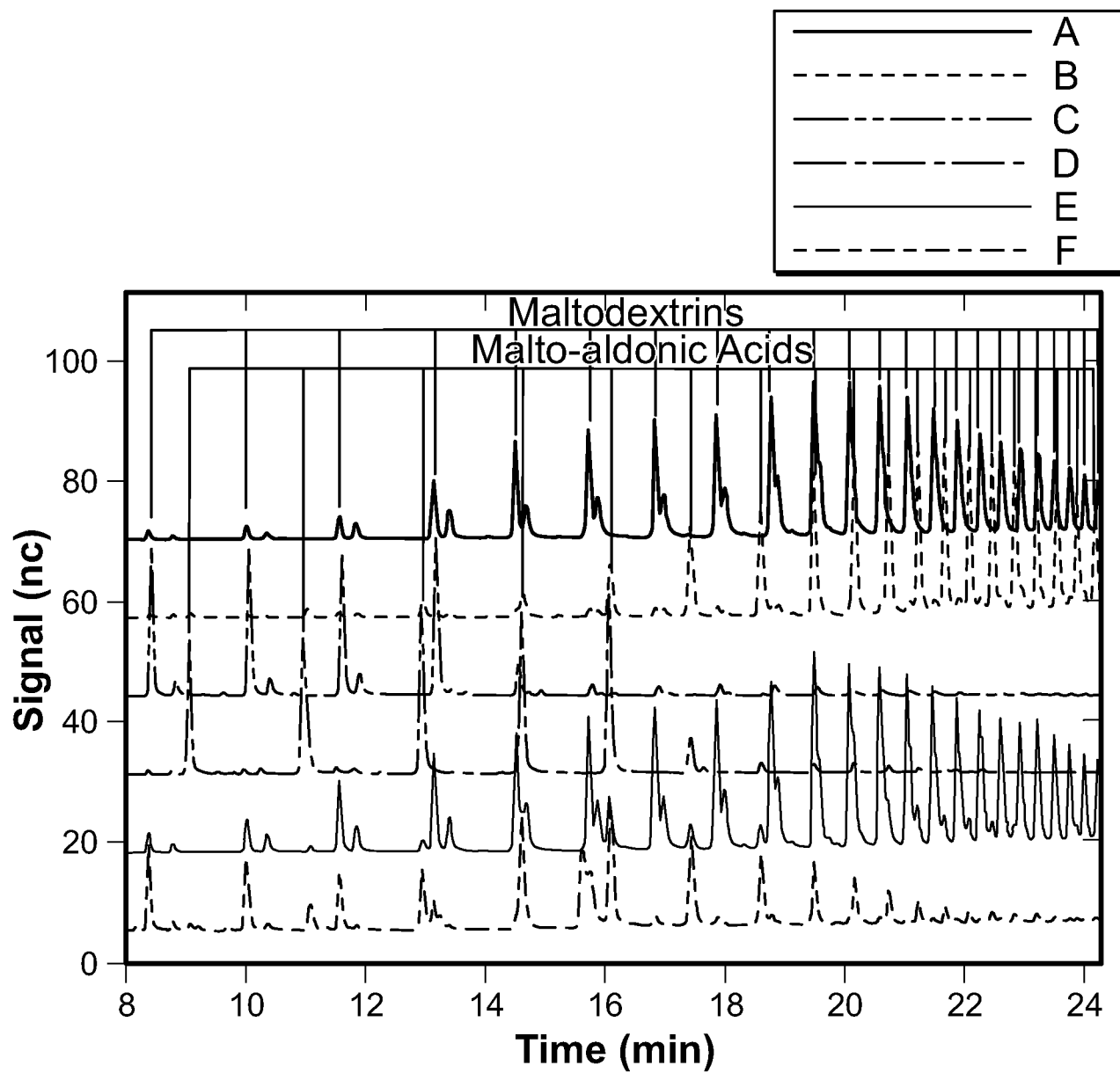
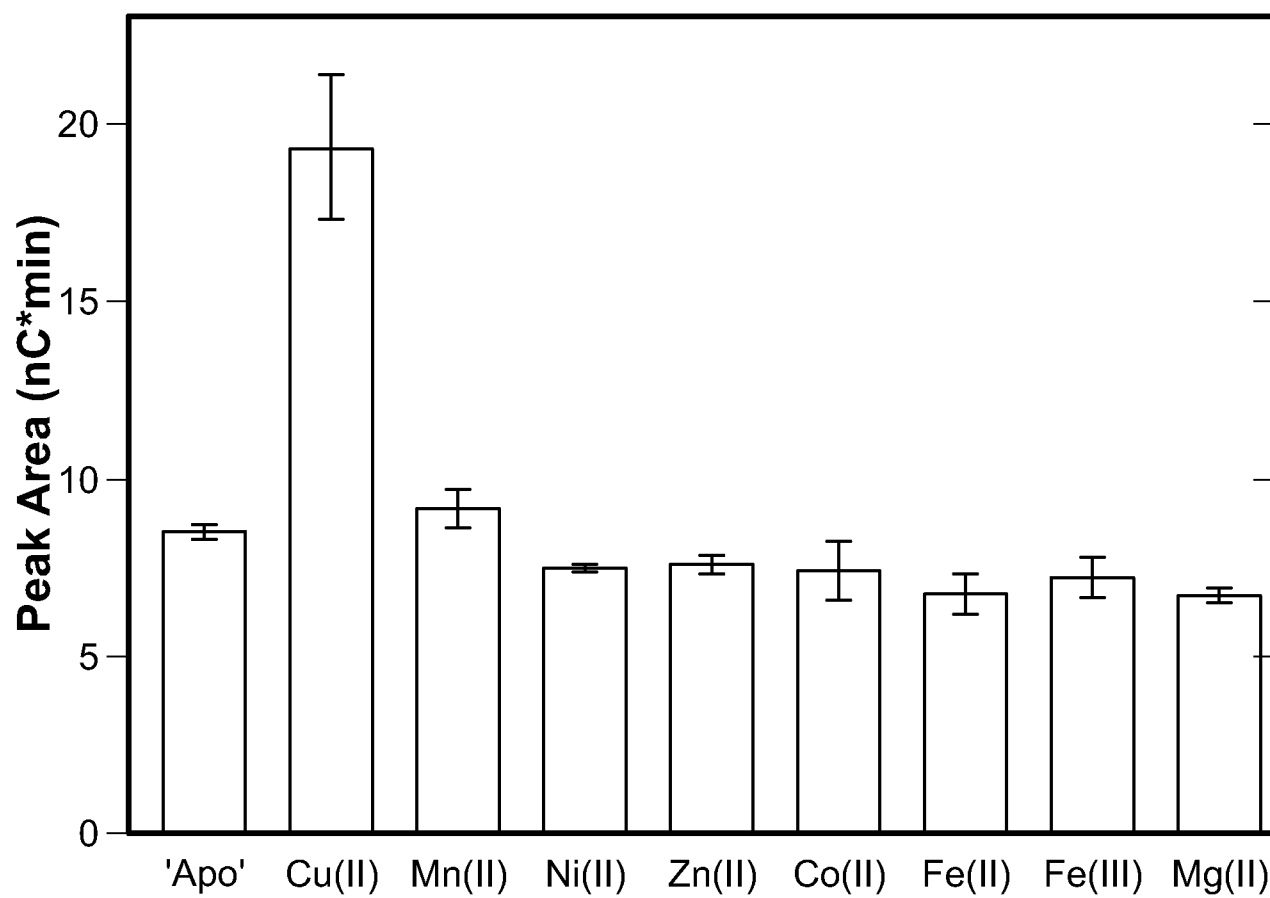


FIG. 12

**FIG. 13**

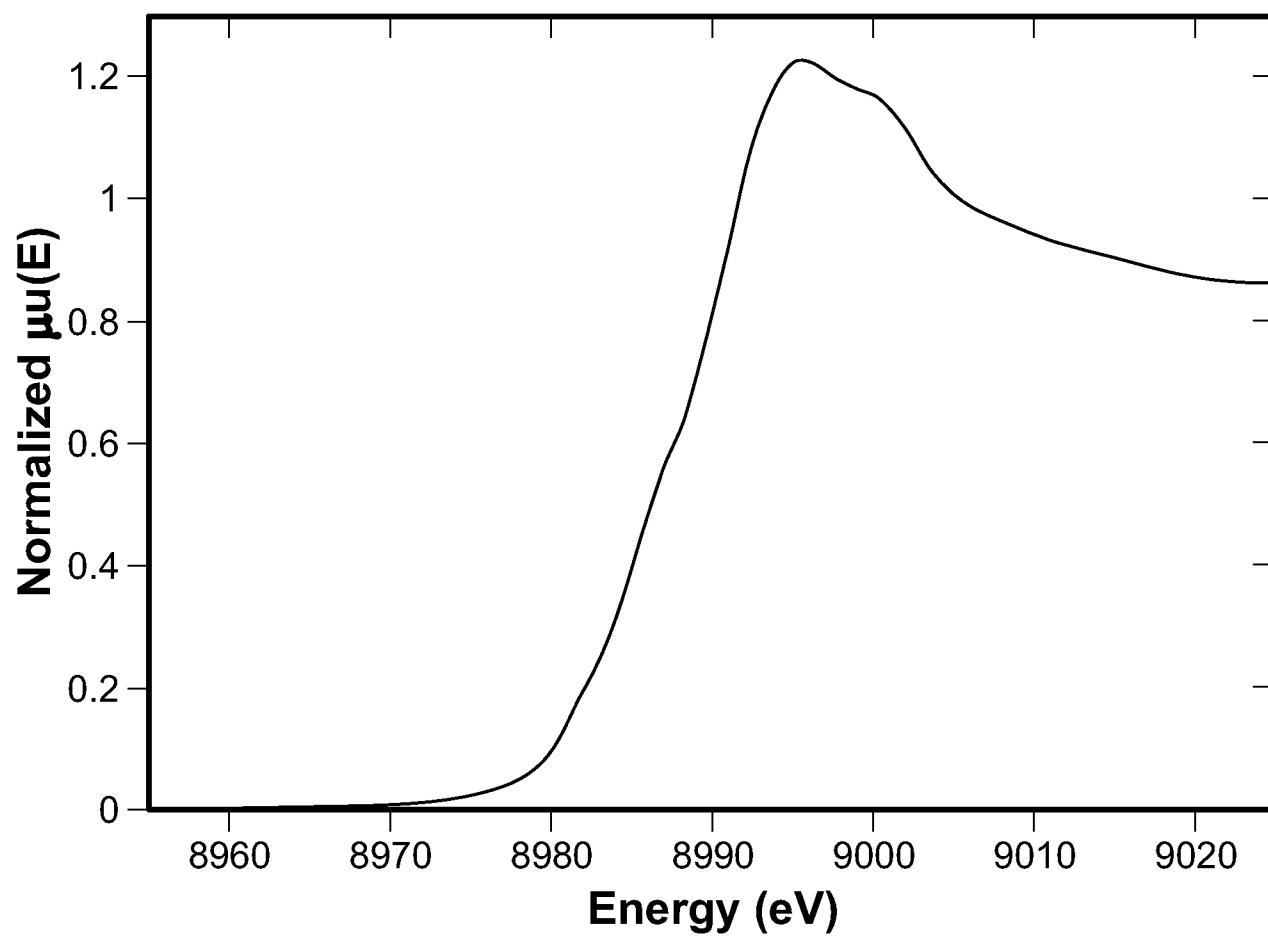


FIG. 14

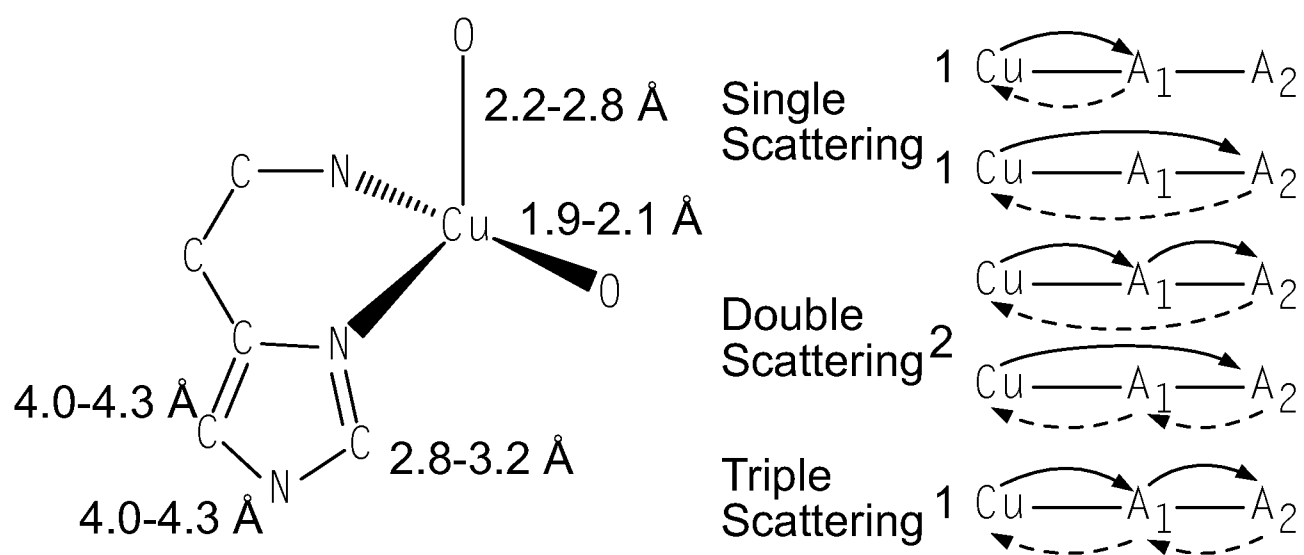


FIG. 15

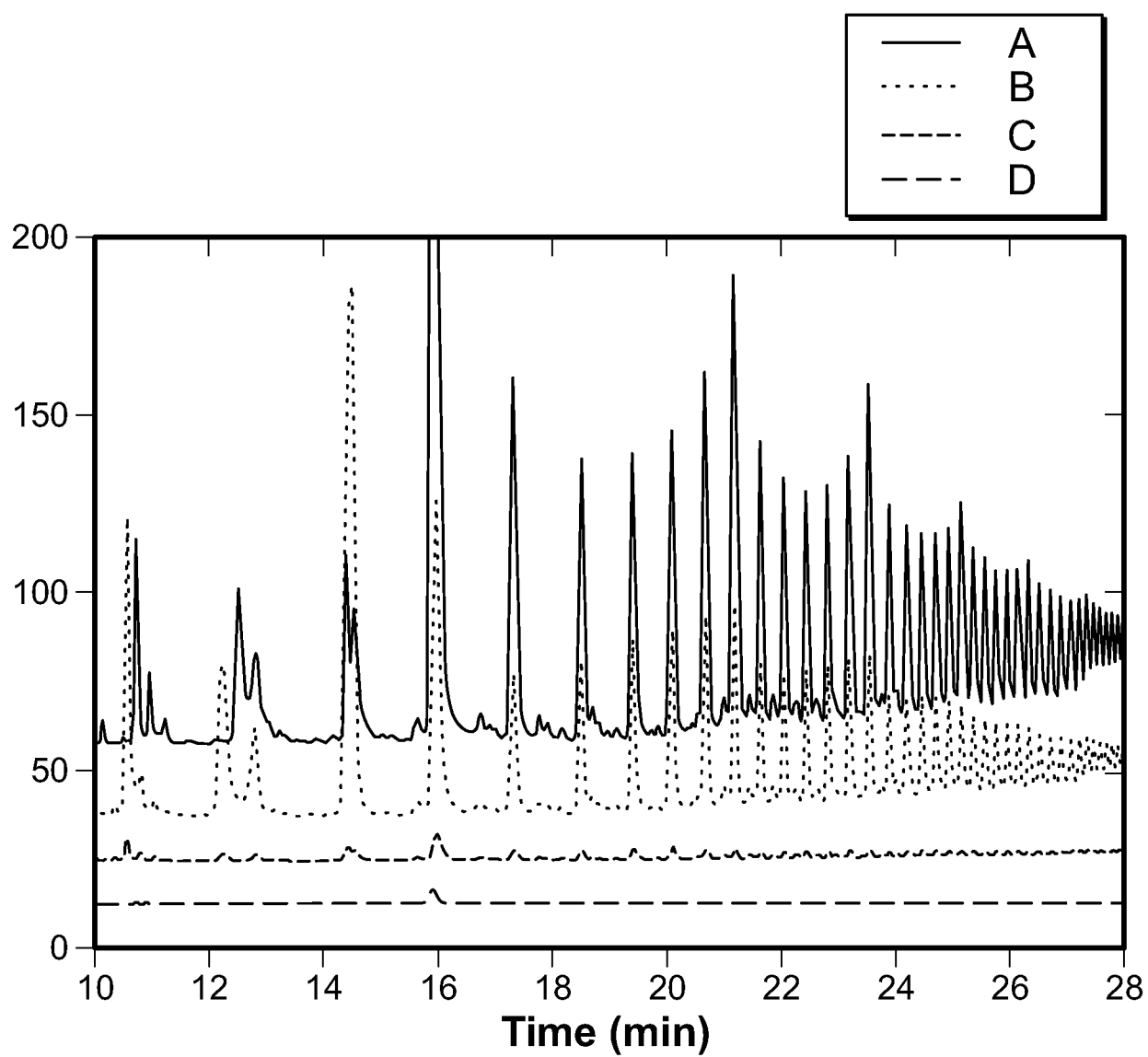
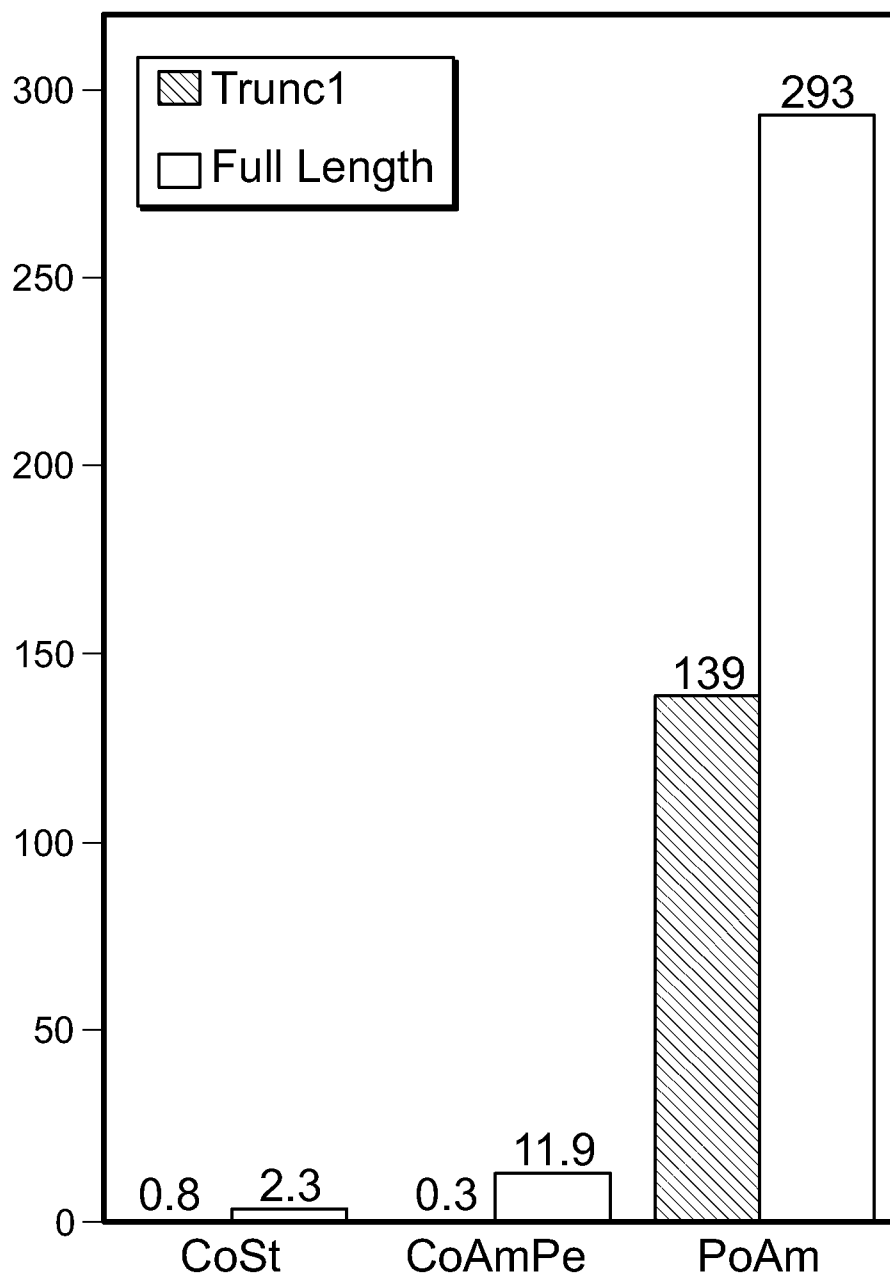


FIG. 16

**FIG. 17**

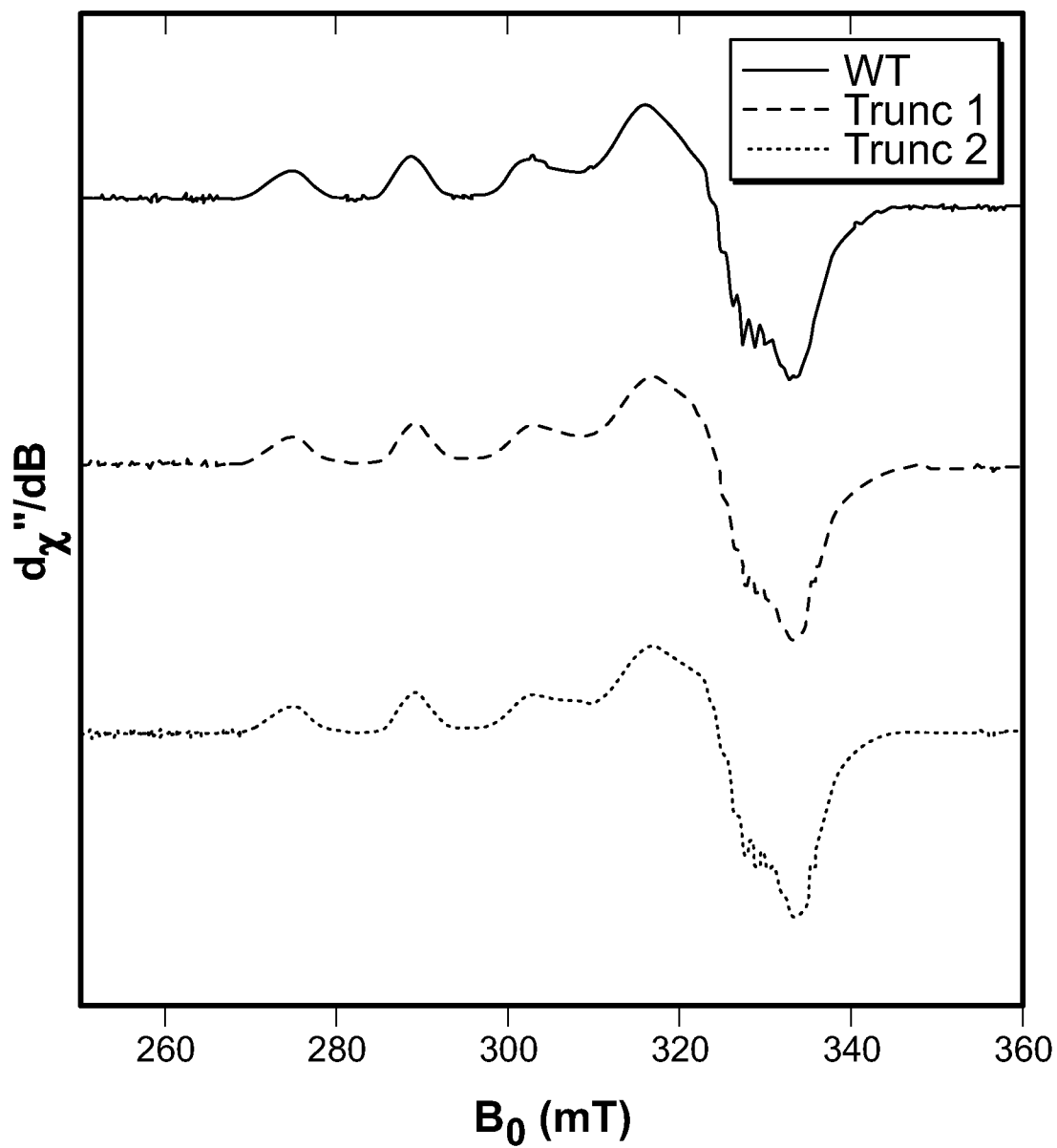


FIG. 18

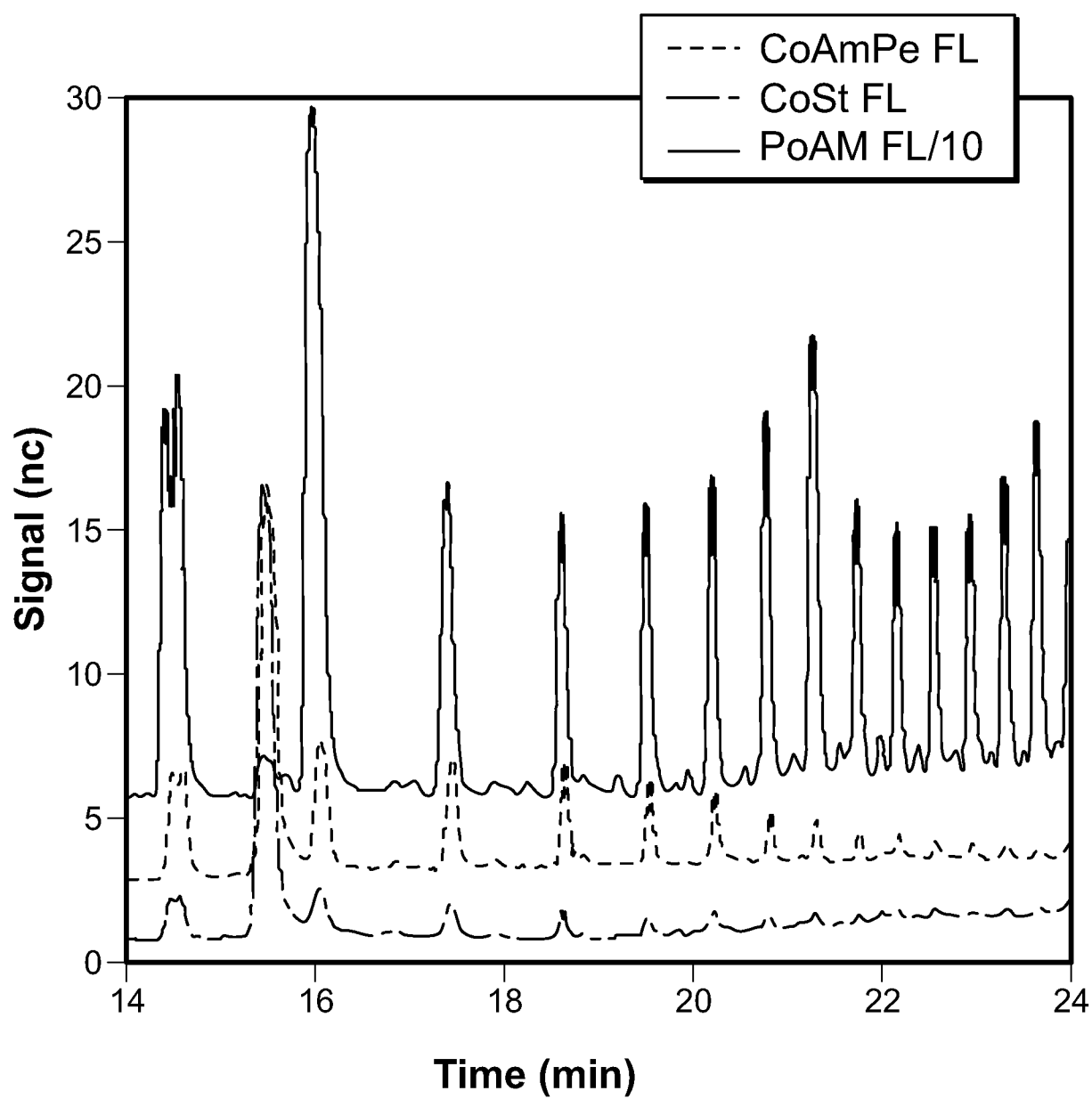


FIG. 19

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/042126

A. CLASSIFICATION OF SUBJECT MATTER
 INV. C12P19/00 C12N9/24
 ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 C12P C12N

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

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

EPO-Internal, FSTA, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/283421 A1 (HARRIS ET AL) 17 November 2011 (2011-11-17) * See SEQ ID NO 2 = SEQ ID NO 81, SEQ ID NO 4 = SEQ ID NO 85, and SEQ ID NO 6 = SEQ ID NO 31 of the Application *	1-34
A	LI ET AL: "Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases", STRUCTURE, vol. 20, 2012, pages 1051-1061, XP028520136, cited in the application * See pages 1051-1052 (Introduction) and page 1053 (Figure 1) * ----- -/--	1-34



Further documents are listed in the continuation of Box C.



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

30 September 2015

Date of mailing of the international search report

09/10/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

Korsner, Sven-Erik

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/042126

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>BEESON ET AL: "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 134, 2012, pages 890-892, XP055027202, cited in the application * See page 892 (last paragraph) *</p> <p>-----</p>	1-34
A	<p>PHILLIPS ET AL: "Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa", ACS CHEMICAL BIOLOGY, vol. 6, 2011, pages 1399-1406, XP002668452, cited in the application * See page 1401 (Figure 4) *</p> <p>-----</p>	1-34
A	<p>VU ET AL: "Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 136, 18 December 2013 (2013-12-18), pages 562-565, XP002745211, cited in the application * See pages 563-564 (Figures 2 and 4); online publication *</p> <p>-----</p>	1-34
A	<p>DATABASE UniProt [Online]</p> <p>16 November 2011 (2011-11-16), BERKA ET AL: "Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris", XP002745212, Database accession no. G2QP40 * See the sequence = SEQ ID NO 34 of the Application *</p> <p>-----</p>	1-34
A	<p>DATABASE UniProt [Online]</p> <p>22 January 2014 (2014-01-22), GALAGAN ET AL: "The genome sequence of the filamentous fungus Neurospora crassa", XP002745236, Database accession no. Q7SCE9 * See the sequence = 19-253 corresponds to SEQ ID NO 6 of the Application *</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-34

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/042126

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MARLETTA ET AL: "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases", 245th National Spring Meeting of the American Chemical Society Abstract 47, 2013, page 1, XP002745213, Retrieved from the Internet: URL: http://acselb-529643017.us-west-2.elb.amazonaws.com/chem/245nm/program/view.php?obj_id=186704&terms=[retrieved on 2015-09-29] * The full oral disclosure might have been more relevant ? *	1-34
A	ELLISON ET AL: "Discovering functions of unnotated genes from a transcriptome survey of wild fungal isolates", MBO, vol. 5, 1 April 2014 (2014-04-01), pages 1-13, XP002745214, * See page 5 (Figure 3/NCU08746); online publication *	1-34
A	SUN ET AL: "Identification of the CRE-1 cellulolytic regulon in Neurospora crassa", PLOS ONE, vol. 6, 2011, pages 1-14, XP055025830, * See page 6 (Figure 6A/NCU08746) *	1-34
X,P	VU ET AL: "A family of starch-active polysaccharide monooxygenases", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, U.S.A., vol. 111, 23 September 2014 (2014-09-23), pages 13822-13827, XP002745216, * Publication of the claimed matter; supporting material available online (see footnote) *	1-34
A,P	BEESON ET AL: "Cellulose degradation by polysaccharide monooxygenases", ANNUAL REVIEW OF BIOCHEMISTRY, vol. 84, 12 March 2015 (2015-03-12), pages 923-946, XP002740953, * See page 939 (Figure 4); online publication *	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/042126

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011283421 A1	17-11-2011	CN 102292444 A	21-12-2011
		EP 2358878 A2	24-08-2011
		EP 2857515 A2	08-04-2015
		ES 2526867 T3	16-01-2015
		US 2011283421 A1	17-11-2011
		WO 2010059413 A2	27-05-2010
