



(51) International Patent Classification:

C12N 1/19 (2006.01) C12P 7/06 (2006.01)
C12P 7/10 (2006.01) C12N 15/63 (2006.01)

(21) International Application Number:

PCT/US2009/064491

(22) International Filing Date:

14 November 2009 (14.11.2009)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/115,068 15 November 2008 (15.11.2008) US

(71) Applicant (for all designated States except US): **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 5th Floor, Oakland, CA 94607 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHEN, Wilfred** [US/US]; 2760 Somerset Place, Rowland Heights, CA 91748 (US). **TSAI, Shen-long** [—/US]; 850 Plum Street, Riverside, CA 92507 (US).

(74) Agent: **BAKER, Joseph, R., Jr.**; Gavrillovich, Dodd & Lindsey LLP, 4660 La Jolla Village Drive, Suite 750, San Diego, CA 92122 (US).

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

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

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

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

(54) Title: ENGINEERING OF YEAST FOR CELLULOSIC ETHANOL PRODUCTION

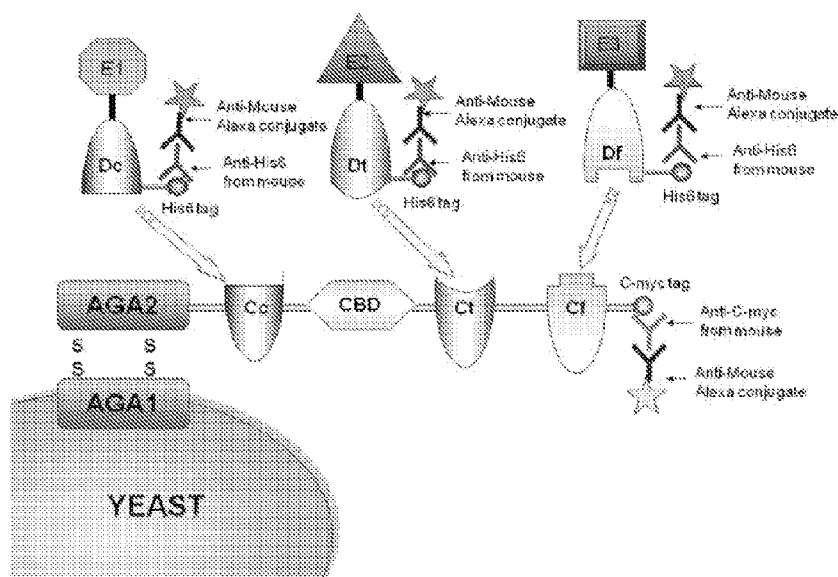


FIGURE 1B

(57) Abstract: The disclosure provides designer cellulosomes for efficient hydrolysis of cellulosic material and more particularly for the generating of ethanol.

ENGINEERED YEAST FOR CELLULOSIC ETHANOL PRODUCTION**CROSS REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims priority under 35 U.S.C. §119 from Provisional Application Serial No. 61/115,068, filed November 15, 2008, the disclosure of which is incorporated herein by reference.

TECHNICAL FIELD

[0002] The disclosure provides designer cellulosomes. The disclosure also provides methods for efficient hydrolysis of cellulosic material and more particularly for the generating of ethanol.

BACKGROUND

[0003] Several billion gallons of renewable fuel must be produced by 2012 with most of that produced as biofuel using renewable biomass. In particular, bioethanol from renewable sources provides an attractive form of alternative energy. It has been estimated that the amount of ethanol needed as transportation fuel will reach 7.5 billion gallons. However, the total capacity of ethanol production in this country is only about 4.2 billion gallons, significantly lower than the required amount.

SUMMARY

[0004] The disclosure provides a synthetic yeast consortium for direct fermentation of cellulose to ethanol with productivity, yield and final concentration close to that from glucose fermentation. The engineering strategy described herein uses the efficiency of hydrolysis and synergy among multi-cellulases. To emulate the success of a natural cellulose hydrolysis mechanism, a complex cellulosome structure is assembled on a yeast cell surface using a constructed yeast consortium, which enables the ethanol-producing strains to utilize cellulose and concomitantly ferment it to ethanol. More importantly, by organizing these cellulases in an ordered structure, enhanced synergy increases the hydrolysis, and thereby the production of ethanol.

[0005] The disclosure provides a culture comprising: a first recombinant yeast strain comprising an anchoring scaffoldin (anScaff); a second recombinant yeast strain comprising an adaptor scaffoldins comprising a plurality of cohesin domains and at least one cellulose binding domain (CBD); and at least one recombinant yeast strain comprising a plurality of secreted dockerin-tagged cellulases. In one embodiment, the yeast strains are cultured under conditions wherein the anchoring scaffoldin, the adaptor scaffoldin comprising the cohesion domains and the plurality of dockerin-tagged cellulases associated to generate an engineered cellulosome. In yet another embodiment, the cellulases are selected from endoglucanases, exoglucanases, β -glucosidase, and xylanase. In a further embodiment, the dockerin-tagged cellulase is engineered to comprise an leader sequence for secretion of the dockerin-tagged cellulase.

[0006] The disclosure provides a recombinant yeast strain comprising a heterologous polynucleotide encoding an anchoring scaffoldin.

[0007] The disclosure also provides a recombinant yeast strain comprising a heterologous polynucleotide encoding an adaptor scaffoldin comprising a plurality of cohesin domains and at least one cellulose binding domain (CBD).

[0008] The disclosure provides a recombinant yeast strain comprising at least one heterologous polynucleotide encoding a secreted dockerin-tagged cellulase.

[0009] The disclosure provides a culture comprising a recombinant yeast strain at least two yeast strains comprising a portion of a functional cellulosome, wherein upon co-culture a functional cellulosome is generated.

[0010] The disclosure also provides a yeast culture comprising at least two recombinant strains of yeast wherein the culture produces a designer cellulosome, and wherein the yeast culture catabolizes cellulosic material to produce a biofuel.

[0011] The disclosure also provides a method of producing a biofuel comprising: culturing the yeast of any as described

above in a fermentation broth comprising a cellulosic material, wherein the microorganism produces the biofuel metabolite.

[0012] The disclosure further provides a method of designing a cellulosome comprising identifying the cellulosic substrate, identifying at least one enzyme useful for degradation of the cellulosic material, recombinantly engineering a dockerin peptide to the enzyme, cloning a polynucleotide encoding the dockerin-linked enzyme into a microorganism, culturing the microorganism in a culture of at least one additional microorganism expressing a scaffoldin having a plurality of cohesion domains and a cellulosic binding domain, wherein the cohesion and dockerin are compatible and culturing the microorganisms to express the scaffoldin and dockerin-linked enzymes.

[0013] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0014] Figure 1A-B shows functional assembly of minicellulosomes on the yeast cell surface. A trifunctional scaffoldin (Scaf-ctf) consisting of an internal CBD flanked by three divergent cohesin (C) domains from *C. thermocellum* (t), *C. cellulolyticum* (c), and *R. flavefaciens* (f) was displayed on the yeast cell surface. Three different cellulases (E1, E2, and E3) fused with the corresponding dockerin domain (either Dt, Dc, or Df) were expressed in *E. coli*. Cell lysates containing these cellulases were mixed with yeast cells displaying Scaf-ctf for the functional assembly of the minicellulosome.

[0015] Figure 2A-D shows phase-contrast and immunofluorescence micrographs of yeast cells displaying minicellulosomes. (A) Cells displaying either scaffoldin Scaf-c, Scaf-ct, or Scaf-ctf. Functional assembly of three

dockerin-tagged cellulases (CelE-Dc [Ec], CelA-Dt [At], or CelG-Df [Gf]) on cells displaying (B) Scaf-ctf, (C) Scaf-ct, or (D) Scaf-c. Cells were probed with either anti-c-Myc or anti-c-His6 serum and fluorescently stained with a goat anti-mouse IgG conjugated with Alexa Fluor 488. Cells displaying only the scaffoldins were used as controls.

[0016] Figure 3 shows fluorescence intensity of cells either displaying scaffoldin Scaf-ctf or with different combinations of dockerin-tagged cellulases (At [A], Ec [E], and Gf [G]) docked on the displayed Scaf-ctf. Cells were probed with either anti-c-Myc or anti-c-His6 serum and fluorescently stained with goat anti-mouse IgG conjugated with Alexa Fluor 488. Whole-cell fluorescence was determined with a fluorescence microplate reader. Cells displaying only Scaf-ctf were used as controls. RFU, relative fluorescence units.

[0017] Figure 4 shows a graph of whole-cell hydrolysis of CMC by different cellulase pairs (CelE-Dc [Ec], CelA-Dt [At], or CelG-Df [Gf]) docked on the displayed Scaf-ctf protein. Cells displaying only Scaf-ctf were used as controls.

[0018] Figure 5A-D show graphs of cellulosome activity. Production of glucose (A) and reducing sugars (B) from the hydrolysis of PASC by free enzymes and by surface-displayed cellulosomes. Reactions were conducted either with different cellulase pairs (CelE-Dc [Ec], CelA-Dt [At], or β -glucosidase-Df [BglA]) docked on the displayed Scaf-ctf protein or with the corresponding purified cellulases. Cells displaying only Scaf-ctf were used as controls. (C) Activity associated with cells and (D) activity in the medium at different initial OD ratios.

[0019] Figure 6A-B shows time profiles of ethanol production. (A) and cellulose hydrolysis (B) from PASC by control strain EBY100 plus free enzymes and yeast cells displaying functional cellulosomes. Fermentations were conducted either with different cellulase pairs (CelE-Dc [Ec], CelA-Dt [At], or β -glucosidase-Df [BglA]) docked on cells displaying Scaf-ctf or with control strain EBY100 plus the corresponding purified cellulases. Cells displaying only Scaf-

ctf were used as controls. The individual enzyme amounts were the same in all cases.

[0020] Figure 7 shows a synthetic consortium for the display of complex cellulosomes.

DETAILED DESCRIPTION

[0021] As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a probe" includes a plurality of such probes and reference to "the primer" includes reference to one or more primers and equivalents thereof known to those skilled in the art, and so forth.

[0022] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0023] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0024] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although any methods and reagents similar or equivalent to those described herein can be used in the practice of the disclosed methods and compositions, the exemplary methods and materials are now described.

[0025] All publications mentioned herein are incorporated herein by reference in full for the purpose of describing and disclosing the methodologies, which are described in the publications, which might be used in connection with the description herein. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors

are not entitled to antedate such disclosure by virtue of prior disclosure.

[0026] Biomass represents an inexpensive feedstock for sustainable bioethanol production. Among the three biological events that occur during the conversion of cellulose to ethanol, *i.e.*, enzyme production, polysaccharide hydrolysis, and sugar fermentation, cellulose hydrolysis is widely recognized as the key step in making bioconversion economically competitive. In addition, it is believed that a significant cost reduction can be achieved when two or more steps are combined, such as in CBP. To achieve this goal, the disclosure provides methods and cellular compositions comprising a functional assembly of a minicellulosome on a yeast cell surface. The minicellulosome was engineered to render the ethanologenic microbe cellulolytic.

[0027] In one embodiment, the disclosure achieves the method and composition by first engineering a chimeric minicellulosome containing three dockerin cohesion pairs from different species on the yeast cell surface. Immunofluorescence microscopy showed the successful translocation of the miniscaffoldin on the yeast cell surface, and the functionality of the cohesin domains was retained by observing the successful assembly of the corresponding dockerin-tagged cellulases. Since the specificity of the dockerin-cohesin pairs is preserved, it is possible to direct any enzymatic subunit to a specified position within a modular scaffoldin by tagging with the designated dockerin.

[0028] The disclosure further demonstrates a synergistic effect on cellulose hydrolysis compared with that of free enzymes.

[0029] In the compositions and methods of the disclosure the displayed minicellulosome retained this key characteristic. Interestingly, the level of synergy increased with an increasing number of cellulases docked on the cell surface. This synergistic effect was preserved even when a new minicellulosome comprising a β -glucosidase (BglA), an

endoglucanase (At), and an exoglucanase (Ec) was assembled on the yeast cell surface.

[0030] The disclosure further demonstrates that the methods and compositions of the disclosure are useful at ethanol production. Cellulose hydrolysis and ethanol production were tested with both free enzymes and a displayed minicellulosome. Independent of the number of cellulases incorporated in the minicellulosome, similar levels of enhancement of cellulose hydrolysis, as well as ethanol production, were detected. The ethanol production achieved, in particular, was more than 2.6-fold higher than that of the culture in which all three cellulases were added as free enzymes. This, when combined with an ethanol yield close to 95% of the theoretical maximum, makes this an efficient process for direct fermentation of cellulose to ethanol.

[0031] Current production processes for using crops such as sugar cane and cornstarch for bioethanol production are well established. However, since the cost of raw materials can be as high as 40% of the overall process, utilization of a cheaper substrate would render bioethanol more competitive with fossil fuel (Zaldivar et al, 2001). Among the different forms of biomass, lignocellulosic biomass is particularly well-suited for energy applications because of its large-scale availability, low cost and environmentally benign production (Lynd et al, 1999). This natural and abundant polymer is found as agricultural waste (wheat straw, corn stalks, soybean residues), industrial waste (pulp and paper industry), forestry residues, and municipal solid waste. Many energy production and utilization cycles based on cellulosic biomass have near-zero greenhouse gas emissions on a life-cycle basis (Lynd et al, 2005).

[0032] The primary obstacle impeding the more widespread production of energy from biomass is the absence of a low-cost technology for overcoming the recalcitrance of these materials (Lynd et al, 2008). For cellulose to be amenable to fermentation, it needs to undergo several treatments to release its monomeric sugars (Zaldivar et al, 2001). Two main

steps are: (1) pretreatments that remove lignin and exposes cellulose for enzymatic degradation, and (2) an enzymatic treatment to generate glucose from cellulose before fermentation. The high cost of cellulases needed for cellulose hydrolysis is one of the major obstacles in the quest for an economically feasible cellulose-based bioethanol process (McBride et al, 2005).

[0033] Although the cost of bioethanol production can become more competitive by combining the hydrolysis and fermentation steps in simultaneous saccharification and cofermentation (SSCF) of both hexoses and pentoses, it has been shown that the overall cost can be even further reduced by 4-fold using a one-step "consolidated" bioprocessing (CBP) of lignocellulose to bioethanol, where cellulase production, cellulose hydrolysis and sugar fermentation can be mediated by a single microorganism or microbial consortium. An ideal microorganism for CBP should possess the capability of simultaneous cellulose saccharification and ethanol fermentation. One attractive candidate is *Saccharomyces cerevisiae*, which is widely used for industrial ethanol production due to its high ethanol productivity and high inherent ethanol tolerance.

[0034] However, due to energetic limitations under anaerobic conditions, only a limited amount of cellulases can often be secreted, resulting in relatively low rates of cellulose hydrolysis. It is believed that, for a process to be viable economically, it must have productivity greater than 1 g/l/hr (Zaldivar et al, 2001). For example, where cellulases are displayed on yeast surface the productivity (0.075 g/l/hr) was more than one order of magnitude lower than strains fermenting glucose.

[0035] Unfortunately, substantial improvement in cellulose hydrolysis may not come from simply increasing the amount of enzymes secreted to the medium or displayed on the surface, which is obviously limited under anaerobic conditions. The key to improving hydrolysis is, perhaps, to increase the catalytic efficiency by maximizing the synergy with limited amount of

enzymes. Recently, it has been demonstrated that the use of ternary cellulose-enzyme-microbe complexes yields much higher rates of cellulose hydrolysis than using binary cellulose-enzyme complexes (Lu et al, 2006). This enzyme-microbe synergy requires the presence of metabolically active *Clostridium thermocellum* displaying cellulosome and appears to be a surface phenomenon involving microbial adhesion onto the cellulose. The 2 to 4-fold synergistic effect observed is significant in decreasing the cost for cellulose hydrolysis.

[0036] An understanding of the role of cellulosomes can be viewed as two distinct mechanisms to tackle the recalcitrant cellulose. Aerobic microbes (such as *Trichoderma reesei*) produce copious amounts of soluble hydrolytic enzymes that synergistically breakdown cellulosic materials (Wilson, 2004; Bayer et al., 2000). In contrast, anaerobic organisms, due to energetic constraints, can only produce a limited amount of enzymes. Therefore, in response, anaerobic organism have become more efficient and have developed an elaborately structured enzyme complex, called a cellulosome, to maximize catalytic efficiency (Bayer et al., 2004; Doi and Kosugi, 2004; Demain et al., 2005). This self-assembled system brings multiple enzymes in close proximity to the substrate, and provides a structure that ensures high local concentration and the correct ratio and orders of the enzymes, thereby maximizing synergy. Consequently, it has a much higher catalytic efficiency than soluble enzymes present in a non-organized fashion. A study showed that the structure endowed an enzyme activity increase of up to 50 times (Johnson et al., 1982).

[0037] Cellulosomes are self-assembled multi-enzyme complexes presented on the anaerobes' cell surface and are dedicated to cellulose depolymerization. The major component of these macromolecule complexes is a structural scaffoldin consists of at least one cellulose binding domain (CBD) and repeating cohesin domains, which are docked individually with a cellulase tagged with a dockerin domain (Fig. 1). The CBD serves as a targeting agent to direct the catalytic domains to

the cellulose substrates. The specific protein-protein, or complementary cohesin-dockerin interaction, provides the mechanism for position-specific self-assembly. Within a given species, the dockerin component appears to bind to all of the cohesins with similar affinity, thus suggesting a random incorporation of the enzymes in the cellulosome. The relative abundance of the catalytic subunits in the cellulosomes is assumed to reflect the level of expression of the corresponding genes as such as in the case of *C. cellulolyticum* using a genetic approach. These cohesin and dockerin modules are species-specific and do not cross interact (Carvalho et al, 2005; Pages et al, 1997). However, recently studies confirmed the presence of other Type II and Type III cohesin/dockerin pairs within a given species in addition to the original Type I mentioned above. These additional pairs are structurally very different and have been shown to possess different specificities to the Type I cohesin/dockerin pairs (Haimovitz et al, 2008).

[0038] The multi-enzyme complex attaches both to the cell envelope and to the substrate, mediating the proximity of the cells to the cellulose (Schwarz, 2001). The ability for substrate-targeting is one of the reasons for increased catalytic efficiency. In addition, the production of cellulosome has a number of advantages over soluble enzymes, for hydrolysis of cellulose:

1. Synergism is optimized due to the proximity of enzyme components,
2. Non-productive adsorption is avoided by the optimal spacing of components,
3. Competitiveness in binding to a limited number of binding sites is avoided by binding the whole cellulosome complex to a single site through a strong binding domain with low specificity,
4. A halt in hydrolysis on depletion of one structural type of cellulose at the site of adsorption is avoided by the presence of other enzymes with different specificity.

[0039] The disclosure provides a recombinant yeast expressing cellulosomal structures. Bioenergetic benefits and synergy are achieved when the cellulosomal structures are displayed onto a microorganism having the ability to ferment biomass to ethanol such as yeast (e.g. *S. cerevisiae*). The recombinant organisms of the disclosure are useful for bioethanol production as fewer enzymes are needed. Additionally, since glucosidase is typically subjected to product inhibition, presence of active glucose-metabolizing cells should further increase the overall hydrolysis rate. The disclosure provides engineered yeast comprising a consortium capable of displaying the highly efficient cellulose-degrading cellulosome structures for one-step CBP of cellulosic materials.

[0040] The functional presentation of various cellulose-binding domains and catalytic subunits in a cellulosome provides improved cellulose hydrolysis over free enzymes as a consequence of the synergistic action among the different components. Because of the modular nature of the cellulosomal subunits the disclosure provides artificial cellulsomes useful in generating biofuels from aerobic organisms at efficiency similar to anaerobic organisms. For example, the disclosure demonstrates that usefulness of a recombinant CBD in yeast using a trifunctional chimeric scaffoldin containing cohesins from a plurality of species. The trifunctional chimeric scaffoldin was constructed and each type of cohesion module was shown to bind specifically to the corresponding dockerin-borne cellulolytic enzymes *in vitro*. The resulting 6-fold improvement in cellulose hydrolysis over similar free enzymes demonstrates that the "designer cellulosome" of the disclosure can be similarly exploited for whole-cell hydrolysis of cellulose and ethanol production when expressed by yeast (e.g., *S. cerevisiae*). The disclosure demonstrates that by displaying a mini-scaffoldin onto the yeast surface a designer cellulosome was obtain. In one embodiment, the resulting yeast cells when tagged with three different dockerin-tagged

cellulases were shown to degrade cellulose up to 3-fold faster.

[0041] In yet another embodiment, the disclosure provides a yeast consortium composed of strains with a surface-display anchoring scaffoldin, strains secreting an adapting scaffoldin, and strains secreting dockerin-tagged cellulases (Fig. 2) for the functional presentation of the complex cellulosome structures.

[0042] The disclosure demonstrates a synthetic yeast consortium for functional presentation of the complex cellulosome structures and to demonstrate the ability for enhanced ethanol production from cellulose. The disclosure provides recombinant yeast for ethanol production comprising a plurality of cohesions and dockerin polypeptides. Currently, the sequences of more than one hundred different cohesions and dockerins from a dozen cellulosome-producing bacteria are known (Hamiovitz *et al.*, 2008). In one embodiment, the disclosure provides the use of 2, 3, 4, 5, 6, 7, or 8 different cohesin/dockerin pairs for the cellulosome assembly.

[0043] The disclosure provides a synthetic yeast consortium for direct fermentation of cellulose to ethanol with productivity, yield and final concentration close to that from glucose fermentation. The engineering strategy described herein uses the efficiency of hydrolysis and synergy among multi-cellulases, rather than focusing on the amount of enzymes produced or used. To emulate the success of a natural cellulose hydrolysis mechanism, a complex cellulosome structure is assembled on a yeast cell surface using a constructed yeast consortium, which enables the ethanol-producing strains to utilize cellulose and concomitantly ferment it to ethanol. More importantly, by organizing these cellulases in an ordered structure, the enhanced synergy will increase the hydrolysis, and thereby the production of ethanol.

[0044] In one embodiment, the disclosure provides a yeast consortium for surface assembly of a mini-cellulosome structure comprising 2, 3 or more cellulases and demonstrates

the feasibility of using the consortium for direct ethanol production from cellulose. In another embodiment, the disclosure provides recombinant yeast for surface-display of one or more of the anchoring scaffoldin, the adaptor scaffoldin, and the dockerin-tagged cellulases.

[0045] In one embodiment, the yeast consortium provides a cellulosome comprising a CelA from *Orpinomyces* sp strain PC-2 (see, e.g., Ljungdahl, 2008; Chen et al, 2006, which are incorporated herein by reference), CelC from *Orpinomyces* sp strain PC-2, CelB, CelD, XynA, and 1 B-glucosidase (Voorhorst, 1995). Sequences for the various dockerins, cohesions and cellulases used in the methods and compositions of the disclosure are readily identifiable by one of skill in the art with reference to readily available databases (e.g., GenBank).

[0046] In one embodiment, a strain of yeast can be genetically engineered by recombinant DNA techniques to express a polynucleotide comprising one or more structural component for producing a cellulosome, a polynucleotide comprising a structural element linked to a cellulose degrading enzyme (e.g., a cellulase), or a combination of any polynucleotide encoding a cellulosome structural polypeptide or enzyme. Where a single yeast or microorganism does not express a full complement of structural and enzyme polypeptide for production of a complete cellulosome, a combination of two or more recombinant yeast (e.g., a consortium) can be used wherein the two or more recombinant yeast express portions of a cellulosome that upon combination generate a full cellulosome capable of degrading a cellulose material.

[0047] For example, in one embodiment, a yeast can be recombinantly engineered to express a trifunctional scaffoldin comprising an internal CBD flanked by three divergent cohesin domains. A first strain recombinant yeast will comprise a plasmid or vector containing one or more (e.g., continuous or separated by linking domains) polynucleotide(s) encoding the CBD flanked by the three divergen cohesin domains. The polynucleotide sequences for a large number of cohesin domains and their corresponding dockerin domains are known in the art.

For example, a search of GenBank will identify numerous sequences for cohesin polynucleotide and polypeptides.

[0048] A second yeast strain can be recombinantly engineered to contain a plasmid or vector comprising sequence encoding one or more dockerin domains linked to a cellulose degrading enzyme. Upon co-culture each strain expresses the corresponding structural or structural-enzymatic polypeptides. The respective dockerin and cohesin domains bind to one another and form a mini-cellulosome. Again, dockerin and enzyme useful in the methods and compositions of the disclosure are recognized in the art and the corresponding sequences are readily identifiable to one of skill in the art performing a simple search on GenBank. For example, the disclosure provides dockerin-enzyme fusion constructs comprising SEQ ID NOs: 5-7. It will be recognized that variants of each of the enzyme domains of the fusion construct may be used, variants of each of the dockerin domains may be used. For example, polypeptides having 80%-99% identity to SEQ ID NO:6 or 8 (or 80-99% [85%, 90%, 95%, 97%, 98% etc.] identity to the respective enzyme or dockerin domains of the construct) can be used in the methods and compositions of the disclosure so long as the dockerin domain is capable of binding to its respective cohesin domain and the enzyme domain is capable of degrading a cellulose material. Methods for determining percent identity are well known in the art.

[0049] A variety of tools are available for the introduction and expression of genes in yeast (such as *S. cerevisiae*), including established vector series (Gietz and Sugino, 1988; Sikorski and Hieter, 1989) and simultaneous or sequential gene integration vectors for the insertion of multiple genes (Wang and Da Silva, 1996; Parekh et al., 1996; Lee and Da Silva, 1997; Lee and Da Silva, 2007).

[0050] Plasmids designed to enable combinatorial plasmid-based testing and seamless transition to genomic integration for fine-tuning of gene number and stable long-term expression can be used. This allows rapid and stable strain construction relative to previous methods. The set of 16 plasmids combines

four different marker genes (flanked by loxP sites), two different promoters, and two different replication origins (2 μ , CEN/ARS). The autonomous vectors allow initial testing of gene combinations on high and/or low copy plasmids. For insertion into the chromosomes, expression polylinker sites adjacent to the selectable markers facilitate polymerase chain reaction amplification of the test gene and selectable marker using primers with outside ends in the desired genomic target sequences. Using this strategy, genes have been successfully inserted into a set of unique locations in the genome with known expression level. The loxP-mediated excision of the selection marker (Sauer, 1987) allows simultaneous marker excision after a group of genes has been integrated, this set of vectors enables both rapid construction and testing of strains, and development of stable engineered strains for use in any complex medium.

[0051] For example, using the methods described herein an engineered consortium for cellulose hydrolysis by intercellular complementation is provided and useable in any given ecosystem. The disclosure demonstrates the feasibility of using a yeast consortium for the surface-display of a mini-cellulosome for efficient cellulose hydrolysis. The disclosure demonstrates the correct assembly of secreted At onto the Scaff#3 in a co-culture system. To enable surface display of scaffoldins without galactose induction, expression of the surface anchor AGA1 (Fig. 1) is placed under a constitutive PGK promoter and the 2 copies of the gene cassette integrated.

[0052] Various enzymes can be used for to degrade the cellulose material. For example, it has been shown that β -glucosidase is useful for complete degradation of cellulose to glucose. Therefore a well known β -glucosidase BglA from *C. thermocelum* was tagged with the dockerin domain (Bf), produced in *E. coli* and incorporated into the cellulosome structures. The result demonstrates the enhancement effect on glucose liberation by assembling BglA into the mini-cellulosome. Therefore, for the initial demonstration, a mini-cellulosome containing an endoglucanase (At), an exoglucanase (Ec) and

BglA are assembled. Secretion of At into the medium using the secretion leader sequence MF α 1 was used; a similar strategy is employed for the secretion of Ec and Bf. Secretion of the structure-enzyme can be confirmed using various methods in the art. For example, secretion of Ec and Bf into the medium can be confirmed by Western blotting using a FLAG tag on Ec and a S-tag on Bf. After confirming secretion, the activity of the secreted fusion constructs can be assayed. For example, cellulase activity can be determined using Avicel or cellobiose as the substrate. Finally, cells secreting either Ec and Bf are co-cultured with cells displaying Scaff#3 and the correct assembly of the cellulases onto the cell surface can be confirmed by immunofluorescence microscopy and the ability to hydrolyze Avicel or cellobiose. After confirming the assembly of individual secreted cellulases, the feasibility of assembling the complete mini-cellulosome is performed. To begin with, a co-culture of different yeast strains are tested in SDC medium. The correct assembly of all three cellulases is confirmed by immunofluorescence using a unique tag presented on each one.

[0053] In yet another embodiment, a single yeast strain capable of secreting all Ec, At and Bf and a strain displaying Scaff#3 is provided by taking advantage of the rapid sequential integration approach. This method also allows precise regulation of expression by controlling the integrated gene copy number (1 to 5). A small-scale shake-flash cultures are used for varying the initially inoculation cell density from a ratio of 10 to 0.1. The specific culturing conditions that result in the highest number of fluorescence cells with all three tags can then be used. The ability of the consortium to hydrolyze Avicel and the corresponding ethanol production can be measured using standard techniques. For example, cells are grown aerobically in SD medium using glucose as the carbon source. The resuspended cells are then used in anaerobic fermentation (SDC medium) using Avicel or phosphoric acid swollen cellulose (PASC) as the carbon source. Samples of the culture can then be obtained for monitoring the expression

level, reducing sugar, intermediates, and cell growth. The ratio of the two different cell populations can be modified to maximize synergy. A coordinated gene expression system leads to the detection of only glucose, whereas accumulation of other products indicate the level of secreted enzymes (or cell population) should be adjusted. For example, if cellobiose is found to accumulate, it indicates that the glucosidase activity is too low relative to the other enzymes and the problem could be easily solved with this strategy by simply increasing the gene dosage for the secreted glucosidase.

[0054] To achieve the cellulosome structure as shown in Fig. 9, five additional cohesin/dockerin pairs can be used. Since most of them are species specific, three additional cohesion/dockerin pairs from *B. cellulosolvens* (bc), *Ace. cellulolyticus* (ac), and *C. acetobutylicum* (cc) can be used. In addition, Type II cohesion/dockerin pairs from *Clostridium thermocellum* (T) and *Clostridium cellulolyticum* (C) are used in conjunction with the Type I pairs (c, t, and f). A feature of the design is the common display of an anchoring scaffoldin, which will enable the surface-display of the complex cellulosome onto all cells except for the strain designed to secrete the adaptor scaffoldins. Since the dockerins used on the cellulolytic enzymes have no cross affinity with the cohesions on the anchoring scaffoldin, no interaction or interference with the translocation machinery is expected. As a result, the resulting consortium will be comprised of cells displaying a functional cellulosome with virtually no carbon source wasted.

[0055] To construct the synthetic consortium, different strains are generated. First, a yeast strain designed to display an anchoring scaffoldin (anScaff) containing the cohesin domains from bc and cc is provided. Synthetic oligos coding for two cohesin domains are synthesized and used for plasmid construction. These two domains are joined by a 10 amino acid linker containing GS repeats flanked by a FLAG tag and displayed on the yeast surface using the same GPI anchor. To add the adapter scaffoldins onto the displayed anScaff, a

yeast strain designed to secrete two separate adaptor scaffoldins is provided. The first adaptor scaffoldin (adScaff#1) comprises an N-terminus bc dockerin flanked by three cohesin domains (t, f, c), one CBD domain, and a C-myc tag. The second adaptor scaffoldin (adScaff#2) comprises an N-terminus cc dockerin flanked by three cohesin domains (ac, type II t and type II c), one CBD domain, and a S-tag. Finally, two different strains are provided to secrete three different dockerin-tagged cellulases each (2 endoglucanases, 2 exoglucanases, one β -glucosidase, and one xylanase) secreted using the MF β 1 leader sequence. A His6 tag is added to each cellulase. In this configuration, the consortium will comprise four strains and three will have the complex cellulosome displayed on the surface.

[0056] In addition to At, Ec, and BglA used above, other enzymes from anaerobic fungi that form cellulosome can be used to demonstrate the complex cellulosome structure. This choice is based on the finding that enzymes from these anaerobic fungi have specific activities much higher than enzymes from other sources (Ljungdahl, 2008). Additionally, several of these enzymes were successfully over-expressed in *S. cerevisiae* (Ljungdahl, 2008; Chen et al, 2006), suggesting over expression of enzymes may not be a significant obstacle. Furthermore, since these enzymes are cellulosomal, the structural feature and folding mechanism are likely compatible to the designed structure as compared to other non-cellulosomal enzymes. For example, a cellulosome of the disclosure can comprise CelA and CelC from *Orpinomyces* sp strain PC-2, are both GH family 6 enzymes, possessing both endo and exoglucanase activities; two family 5 cellulases, CelB and CelD, both endoglucanases, are cloned and fused with an appropriate dockerin; the catalytic domain of XynA, a family GH11 xylanase with extremely high activity, is used as the fifth enzyme and a family 1 β -glucosidase from *Piromyces* sp Strain E2 is used as a sixth enzyme.

[0057] All of these strains are constructed using the rapid and stable multi-gene integration systems described above; 1

to 5 copies of each gene cassette are integrated in order to optimize the required expression. Expression is under the control of the PGK promoter based on its constitutive nature and the high-level of protein expression. The number of anScaff molecules displayed on the cell surface will be determined by measuring the fluorescence intensity of the cell pellet suspended in PBS buffer (pH 7.0) using a fluorometer. A standard curve can be prepared by using known amounts of Alexa Fluor 546-conjugated goat anti-mouse IgG. The number of anscaff displayed will be calculated using $(\text{RFU} \times 0.945 \times 10^7) / 1 \times 10^7$ as reported previously. Similarly, the secretion level for the two anScaffs and the different cellulases can be determined. The secretion of anScaff in the medium can be confirmed first by incubating the medium with Avicel for 1 h. After incubation, Avicel is separated by centrifugation and the bound protein after washing three times with PBS buffer is analyzed by SDS-PAGE and Western blot against either the S-tag or C-myc tag. The amount of cellulases produced is analyzed by SDS-PAGE and enzyme assays using procedures that are already in place. After confirming expression, the four different strains are co-cultured and the correct assembly of the adaptor scaffoldins and cellulases onto the cell surface confirmed by immunofluorescence microscopy and the ability to hydrolyze Avicel. Xylanase activity can be measured as described by Bailey et al., 1992).

[0058] The disclosure demonstrates a synthetic consortium for surface-display of a complex cellulosome by combining cells displaying the anchoring and adaptor scaffoldins with cells secreting the dockerin-borne cellulases. The modular nature of the cellulosome and great diversity of cellulases, and the availability of gene fusion technology provide almost unlimited number of combinations of enzymes and ways to incorporate them into artificial cellulosome as designed to optimize their activities to approach or even surpass the natural cellulosome. It has been shown by Fierobe et al (2002; 2005) that cellulosomes containing different cellulases have significantly different abilities for cellulose hydrolysis;

even the same cellulases fused to different dockerins can result in cellulosomes with substantially different hydrolysis efficiencies (up to 2-3 fold), suggesting the order of enzymes on the cellulosome can directly impact the overall activity. In an effort to create the most efficient combination for cellulose hydrolysis, dockerin replacement among the enzymes can be revised and modulated. All possible dockerin combinations will be created and the resulting activity of the cellulosomes can be compared for hydrolysis. Overlap-extension PCR can be used for replacements of the dockerins. Briefly the reverse primers for the region coding for the C-terminal part of the catalytical domain of each cellulase can be overlapped with the forward primers for the region coding for the N-terminal part of the dockerin domain. After several runs of denature, aligning, and extension in PCR the resultant overlapping fragments will be mixed and combined fragment will be synthesized by using external primers. DNS method as described herein can be used for evaluating the efficiencies of the different cellulosome complexes and their synergy effects as well as glucose liberation.

[0059] After optimizing the activity of the cellulosome, the ability of the consortium to hydrolyze Avicel and the corresponding ethanol production can be analyzed. Initially, growth rates of each individual strain are determined to check for any substantial difference in cell growth. Samples are taken periodically for monitoring the expression level, reducing sugar, intermediates, and cell growth. One can coordinate the four different cell populations so that maximum synergy can be obtained and no enzyme represents a limiting step. This can be ensured by using cellulose hydrolysis and following the hydrolysis products by a carbohydrate analysis method established on a Dionex High-pH Ion-Exchange Chromatograph System with electrochemical detector, which detects oligosaccharides (cellobiose, glucose and the like), allowing one to pinpoint limiting enzyme activities (if any) (endo/exo glucanases or β -glucosidase). A coordinated gene expression system leads to the detection of only glucose,

whereas accumulations of other products indicate the level of secreted enzymes (or cell population) should be adjusted by varying the gene copy number.

[0060] The engineered strains can be evaluated for cellulose hydrolysis and ethanol production under different conditions such as resting and growth conditions in SDC medium. Both small and large-scale (shaker flask/one liter bioreactor) studies can be performed. In resting cell experiments, cells are grown aerobically using glucose as the carbon source. Cells are then washed and used in cellulose anaerobic hydrolysis. Enzyme activity, the integrity of cellulosome, hydrolysis products, glucose, ethanol will be monitored using methods described herein. In studies carried out in a fermentor, a mild agitation can be used to promote mixing of solid cellulose material with cells. Once optimized industrial yeast fermentation process may be used. Different cellulose concentrations can also be used. The rate of glucose generation will be estimated from the experiments and compared to those without cellulosome on the cell surface (but with comparable enzyme expression levels). In studies under growing conditions, the cells will be provided cellulose as the sole carbon source, and other nutrients necessary for growth. Anaerobic conditions are maintained. Cell biomass, enzyme activities, cellulosome integrity, any possible accumulated hydrolysis products including glucose, and ethanol are measured.

[0061] The disclosure provides yeast strains for direct fermentation of cellulose to ethanol, eliminating the need for use of purified cellulases. The methods and compositions of the disclosure provide abundant, low-cost, agriculture residue to be used as raw material for ethanol production. The increased production of ethanol not only reduces pollution to the environment but also the need for imported petroleum as transportation fuel. Collectively, the benefits from the invention include at least efficient, economical, and environmentally friendly conversion of biomass.

[0062] Current biofuel production processes are exclusively based upon the soluble enzyme approach, the less efficient model utilized by aerobic microbes. The construction of cellulosomes on the cell surface as described herein departs from the current model of enzymatic hydrolysis. A difference lies in the extent of synergism. While the secretion or co-display of cellulases (without the cellulosome structure) permits a synergistic use of these enzymes to some extent, both follow the model of aerobic organisms in cellulose hydrolysis. As exemplified by *Trichoderma reesei*, this model requires the production of abundant enzymes without needing to maximize the efficiency and synergism. Since ethanol production is carried out anaerobically, the limited amount of enzyme production makes the high coordination of different enzyme components necessary to maximize the synergy and efficiency.

[0063] As used herein, the term "microorganism" includes prokaryotic and eukaryotic microbial species from the Domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The terms "microbial cells" and "microbes" are used interchangeably with the term microorganism.

[0064] As used herein, the term "polynucleotide" refers to a polymer of nucleic acid bases such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

[0065] The term "carbon source" generally refers to a substrate or compound suitable to be used as a source of carbon for bacterial or simple eukaryotic cell growth. Carbon sources may be in various forms, including, but not limited to polymers, carbohydrates such as cellulosic material including cellulooligosaccharides and lignocellulose, acids, alcohols, aldehydes, ketones, amino acids, peptides, etc. These include, for example, various monosaccharides such as glucose,

oligosaccharides, polysaccharides, saturated or unsaturated fatty acids, succinate, lactate, acetate, ethanol, and the like, or mixtures thereof.

[0066] The term "substrate" or "suitable substrate" refers to any substance or compound that is converted or meant to be converted into another compound by the action of an enzyme. The term includes not only a single compound, but also combinations of compounds, such as solutions, mixtures and other materials which contain at least one substrate, or derivatives thereof. Further, the term "substrate" encompasses not only compounds that provide a carbon source suitable for use as a starting material, but also intermediate and end product metabolites used in a pathway associated with a engineered microorganism as described herein. A "biomass derived sugar" includes, but is not limited to, molecules such as glucose, sucrose, mannose, xylose, and arabinose. Exemplary substrate sources include alfalfa, corn stover, crop residues, debarking waste, forage grasses, forest residues, municipal solid waste, paper mill residue, pomace, sawdust, spent grains, spent hops, switchgrass, and wood chips.

[0067] As used herein, the terms "gene" and "recombinant gene" refer to an exogenous nucleic acid sequence which is transcribed and (optionally) translated. Thus, a recombinant gene can comprise an open reading frame encoding a polypeptide. In such instances, the sequence encoding the polypeptide may also be referred to as an "open reading frame".

[0068] "Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of a gene or genes with which they are operably linked.

[0069] "Operably linked" means that a gene and transcriptional regulatory sequence(s) are connected in such a way as to permit expression of the gene in a manner dependent upon factors interacting with the regulatory sequence(s).

[0070] "Exogenous" means a polynucleotide or a peptide that has been inserted into a host cell. An exogenous molecule can result from the cloning of a native gene from a host cell and the reinsertion of that sequence back into the host cell. In most instances, exogenous sequences are sequences that are derived synthetically, or from cells that are distinct from the host cell.

[0071] The terms "host cells" and "recombinant host cells" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0072] As used herein, a "reporter gene" is a gene whose expression may be assayed; reporter genes may encode any protein that provides a phenotypic marker, for example: a protein that is necessary for cell growth or a toxic protein leading to cell death, e.g., a protein which confers antibiotic resistance or complements an auxotrophic phenotype; a protein detectable by a colorimetric/fluorometric assay leading to the presence or absence of color/fluorescence; or a protein providing a surface antigen for which specific antibodies/ligands are available.

[0073] The term "biosynthetic pathway", also referred to as "metabolic pathway", refers to a set of anabolic or catabolic biochemical reactions for converting (transmuting) one chemical species into another. Gene products belong to the same "metabolic pathway" if they, in parallel or in series, act on the same substrate, produce the same product, or act on or produce a metabolic intermediate (i.e., metabolite) between the same substrate and metabolite end product.

[0074] As used herein, the term "metabolic pathway" includes catabolic pathways and anabolic pathways both natural and engineered i.e. synthetic. Anabolic pathways involve constructing a larger molecule from smaller molecules, a

process requiring energy. Catabolic pathways involve breaking down of larger molecules, often releasing energy. An anabolic pathway is referred to herein as "a biosynthetic pathway."

[0075] Biofuel is any fuel that derives from biomass - organisms, such as plants, fermentation waste, or metabolic byproducts, such as manure from cows. It is a renewable energy source, unlike other natural resources such as petroleum, coal and nuclear fuels. Agricultural products specifically grown for use as biofuels and waste from industry, agriculture, forestry, and households - including straw, lumber, manure, sewage, garbage and food leftovers - can be used for the production of bioenergy.

[0076] Cellulose is a polymer polysaccharide carbohydrate, of beta-glucose. It forms the primary structural component of plants and is not digestible by humans. Cellulose is a common material in plant cell walls and was first noted as such in 1838. Cellulose is the most abundant form of living terrestrial biomass (Crawford, R. L. 1981. Lignin biodegradation and transformation, John Wiley and Sons, New York.). Cellulose is also the major constituent of paper. Cellulose monomers (beta-glucose) are linked together through 1,4 glycosidic bonds.

[0077] A polynucleotide, polypeptide, or peptides may have a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same when comparing the two sequences. Sequence similarity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using the methods and computer programs, including BLAST, available over the world wide web at ncbi.nlm.nih.gov/BLAST. See, e.g., Altschul et al. (1990), Mol. Biol. 215:403-10.

[0078] Exemplary polynucleotide and polypeptides are provided herein. One of skill in the art will recognize that minor modification (e.g., conservative substitutions and the like) can be made to the polypeptide without destroying the biological/enzymatic activity of the polypeptide. Such

modification, variation and the like are within the skill in the art as it relates to molecular biology. Screening for activity of such modified polypeptides is described herein. Accordingly, the disclosure encompass polynucleotide and polypeptides having at least 60%, 70%, 80%, 90%, 95%, 98% or 99% identity to a sequence as set forth herein and having a biological activity similar to the wild-type molecule.

[0079] As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms typically use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons. Codons can be substituted to reflect the preferred codon usage of the host, a process sometimes called "codon optimization" or "controlling for species codon bias."

[0080] Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (see also, Murray et al. (1989) Nucl. Acids Res. 17:477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, typical stop codons for *S. cerevisiae* and mammals are UAA and UGA, respectively. The typical stop codon for monocotyledonous plants is UGA, whereas insects and *E. coli* commonly use UAA as the stop codon (Dalphin et al. (1996) Nucl. Acids Res. 24: 216-218). Methodology for optimizing a nucleotide sequence for expression in a plant is provided, for example, in U.S. Pat. No. 6,015,891, and the references cited therein.

[0081] Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given enzyme of the disclosure. The native DNA

sequence encoding the biosynthetic enzymes described above are referenced herein merely to illustrate an embodiment of the disclosure, and the disclosure includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the enzymes utilized in the methods of the disclosure. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The disclosure includes such polypeptides with different amino acid sequences than the specific proteins described herein so long as they modified or variant polypeptides have the enzymatic anabolic or catabolic activity of the reference polypeptide. Furthermore, the amino acid sequences encoded by the DNA sequences shown herein merely illustrate embodiments of the disclosure.

[0082] The recombinant yeast cellulosome of the disclosure can be engineered based upon the cellulosic material to be metabolized. For example, different cellulases and other enzymes may be engineered into a cellulosome pathway depending upon the source of substrate. Exemplary substrate sources include alfalfa, corn stover, crop residues, debarking waste, forage grasses, forest residues, municipal solid waste, paper mill residue, pomace, sawdust, spent grains, spent hops, switchgrass, and wood chips. Some substrate sources can have a larger percentage of cellulose compared to other source, which may have a larger percentage of hemicellulose. A hemicellulose substrate comprises short, branched chains of sugars and can comprise a polymer of five different sugars. Hemicellulose comprises five-carbon sugars (e.g., D-xylose and L-arabinose) and six-carbon sugars (e.g., D-galactose, D-glucose, and D-mannose) and uronic acid. The sugars are typically substituted with acetic acid. Hemicellulose is relatively easy to hydrolyze to its constituent sugars. When hydrolyzed, the hemicellulose produces xylose (a five- carbon sugar) or six- carbon sugars from hardwoods or softwoods, respectively.

[0083] In some instances the feedstock can be pretreated using heat, acid treatment or base treatment. Possible pre-treatments include the use of dilute acid, steam explosion, ammonia fiber explosion (AMFE), organic solvents (BioCycle, May 2005 News Bulletin, and see: Ethanol from Cellulose: A General Review, P. Badger, p. 17-21 in J. Janick and A. Whipkey (eds.), Trends in New Crops and Uses, ASHS Press, 2002). Typically the pretreatment will be biocompatible or neutralized prior to contact with a recombinant microorganism of the disclosure.

[0084] Proteins or polypeptides having the ability to convert the hemicellulose components into carbon sources that can be used as a substrate for biofuel production includes, for example, cellobiohydrolases (Accessions: AAC06139, AAR87745, EC 3.2.1.91, 3.2.1.150), cellulases (E.C. 3.2.1.58, 3.2.1.4, Accessions: BAA12070, BAB64431), chitinases (E.C. 3.2.1.14, 3.2.1.17, 3.2.1.-, 3.2.1.91, 3.2.1.8, Accessions: CAA93150, CAD12659), various endoglucanases (E.C. 3.2.1.4, Accessions: BAA92430, AAG45162, P04955, AAD39739), exoglucanases (E.C. 3.2.1.91, Accessions: AAA23226), lichenases (E.C. 3.2.1.73, Accessions: P29716), mannanases (E.C. 3.2.1.4, 3.2.1.-, Accessions: CAB52403), pectate lyases (E.C. 4.2.2.2, Accessions: AAG59609), xylanase (E.C. 3.2.1.136, 3.2.1.156, 3.2.1.8, Accessions: BAA33543, CAA31109) and silase (E.C. 3.2.2.-, 2.7.7.7, Accessions: CQ80097S).

[0085] Cellulases are a class of enzymes produced chiefly by fungi, bacteria, and protozoans that catalyze the hydrolysis of cellulose. However, there are also cellulases produced by other types of organisms such as plants and animals. Several different kinds of cellulases are known, which differ structurally and mechanistically. The EC number for cellulase enzymes is E.C.3.2.1.4. Assays for testing cellulase activity are known in the art.

[0086] Polypeptides having xylanase activity are useful for including in synthetic cellulosomes. Xylanases is the name given to a class of enzymes which degrade the linear polysaccharide beta-1,4-xylan into xylose, thus breaking down

hemicellulose. The EC number for xylanase enzymes is E.C. 3.2.1.136, 3.2.1.156, 3.2.1.8. Assays for testing xylanase activity are known in the art.

[0087] Scaffold or structure polypeptides refer to peptides that do not have enzymatic activity, but rather play a role as a building block. For example, scaffold or structure polypeptides as used herein include scaffoldin, cohesion, and cellulose binding polypeptides useful for generating the cellulosome structure for binding of enzyme(s) to the host cell surface, or bind the cellulosic material degrading enzyme(s) to the cellulosic material carbon source. In one example, a recombinant cellulosome of the disclosure can comprise a recombinant scaffold polypeptide and/or a recombinant enzymatic polypeptide. For example, a synthetic cellulosome can have a cellulosic material degrading enzyme domain, and one or more structural domains. Depending on the structural peptide domain the synthetic cellulosome will bind to the carbon source and serve to place the cellulosic material degrading enzyme activity in close physical proximity to the carbon source. In other examples the synthetic cellulosome will have peptide sequences that bind the synthetic cellulosome to the host cell surface function to place the cellulosic material degrading enzyme activity in close proximity to the cell surface.

[0088] Yeast of the disclosure are engineered to convert cellulosic material to a biofuel, such as ethanol, by engineering them to produce both a synthetic cellulosome and cellulose degrading enzymes. Multiple recombinant yeast can be co-cultured to generate the cellulosome, each of a plurality of recombinant yeast producing one or more, but not all, the elements of a function cellulosome.

[0089] One of ordinary skill in the art will appreciate that there are a variety of synthetic cellulosomes that can be made, for example, comprising a variety of degradation enzymes for a specific cellulose or hemicelluloses containing material, a variety of associated scaffoldin and cohesion molecules to generate a recombinant cellulosome having a

desired efficiency or pathway of degrading enzymes provided herein.

[0090] A microorganism (e.g., a yeast) are engineered to express the synthetic cellulosome by constructing a vector containing a scaffoldin domain, a Carbohydrate Binding Domain (CBM), and one or more cellulosic material -degrading enzymes that have been fused with cohesin domains. In one embodiment, a plurality of different recombinant microorganisms are generated each expressing a desired element of a cellulosome. For example, a first recombinant yeast can comprise a polynucleotide encoding a heterologous anchoring protein, a second recombinant yeast comprising a polynucleotide encoding a soluble scaffoldin unit comprising a plurality of cohesion units and a cellulose binding domain, and a third recombinant yeast comprising a polynucleotide encoding an enzyme (e.g., a cellulose or other degradation enzyme) linked to a dockerin subunit. The cells can be recombinant engineered to produce the elements of the cellulosome, wherein a function cellulosome is produced and function in culture.

[0091] The following examples are intended to illustrate but not limit the disclosure. While they are typical of those that might be used, other procedures known to those skilled in the art may alternatively be used.

EXAMPLES

[0092] **Strains, plasmids, and media.** *Escherichia coli* strain JM109 [*endA1 recA1 gyrA96 thi hsdR17 (rK^mK⁺) relA1 supE44 Δ(lac-proAB)*] was used as the host for genetic manipulations. *E. coli* BL21(DE3) [*F⁻ ompT gal hsdSB (rB⁻ mB⁻) dcm lon λDE3*] was used as the production host for cellulase expression. *S. cerevisiae* strain EBY100 [*MATa AGA1::GAL1-AGA1::URA3 ura3-52 trp1 leu2-1 his3-200 pep4::HIS3 prb1-1.6R can1 GAL*] was used for surface display of scaffoldins. All *E. coli* cultures were grown in Luria-Bertani (LB) medium (10.0 g/liter tryptone, 5.0 g/liter yeast extract, 10.0 g/liter NaCl) supplemented with either 100 µg/ml ampicillin or 50 µg/ml kanamycin. All yeast cultures were grown in SDC medium

(20.0 g/liter dextrose, 6.7 g/liter yeast nitrogen base without amino acids, 5.0 g/liter Casamino Acids).

[0093] To display scaffoldins, a gene fragment coding for a scaffoldin containing three cohesins from *C. cellulolyticum*, *C. thermocellum*, and *R. flavefaciens* and one CBD was amplified with plasmid pETscaf6 as the template with forward primer F1NdeI (5'-TATAG**CTAGC**GGCGATTCTCTTAAAGTTACAGT-3' [the boldface portion is a restriction endonuclease site]) and reverse primer R1SalI (5'-ATAT**GTCGAC**GTGGTGGTGGTGGTG-3'). The PCR product was then digested and ligated into the surface display vector pCTCON2 to form pScaf-ctf. Similar procedures, except that the reverse primers were changed to RASalI (5'-ATAT**GTCGAC**ATCTGACGGCGGTATTGTTGTTG-3') and RBSalI (5'-ATAT**GTCGAC**TATATCTCCAACATTTACTCCAC-3'), were used for the construction of pScaf-c and pScaf-ct.

[0094] Plasmids pETec and pETGf, encoding exoglucanase CelE (Ec) and endoglucanase CelG (Gf) of *C. cellulolyticum* fused to the dockerins from *C. cellulolyticum* and *R. flavefaciens*, respectively. Plasmid pETAt, encoding a His6-tagged endoglucanase (CelA) and a dockerin from *C. thermocellum* (At), was obtained by PCR from pCelA with forward primer F2NdeI (5'-ATAT**CATATG**GCAGGTGTGCCTTTTAACACAAA-3') and reverse primer R2XhoI (5'-ATAT**CTCGAG**CTAATAAGGTAGGTGGGG-3'). The amplified fragment was cloned into NdeI-XhoI-linearized plasmid pET24a to form pETAt. Plasmid pBglAf, encoding a His6-tagged dockerin from *R. flavefaciens* fused to a β -glucosidase (BglA) from *C. thermocellum*, was obtained by two-step cloning. First, a gene fragment coding for the His6-tagged dockerin of *R. flavefaciens* was obtained from pETGf by digestion with BamHI and XhoI and ligated into pET24a to form pETDf. The gene fragment of BglA was amplified by PCR from pBglA with forward primer F3NdeI (5'-ATAT**CATATG**TCAAAGATAACTTTCCCAAAA-3') and reverse primer R3BglIII (5'-ATAT**AGATCT**TTAAAACCGTTGTTTTGATTACT-3') and inserted into NdeI-BamHI-linearized pETDf to form pBglAf. A summary of all of the scaffoldins and dockerin-tagged cellulases used in this study is listed in Table 1.

TABLE 1. Scaffoldins and dockerin-tagged cellulases used in this study

Protein name	Description (from N terminus to C terminus)	Host cell	Tag
Scaf-c	Scaffoldin containing a cohesin from <i>C. cellulolytica</i> followed by a CBD	<i>S. cerevisiae</i>	c-Myc
Scaf-ct	Scaffoldin containing a cohesin from <i>C. cellulolytica</i> followed by a CBD followed by a second cohesin from <i>C. thermocellum</i>	<i>S. cerevisiae</i>	c-Myc
Scaf-ctf	Scaffoldin containing a cohesin from <i>C. cellulolytica</i> followed by a CBD followed by a second cohesin from <i>C. thermocellum</i> and a third cohesin from <i>R. flavefaciens</i>	<i>S. cerevisiae</i>	c-Myc
At	Endoglucanase CelA from <i>C. thermocellum</i> fused with its native dockerin	<i>E. coli</i>	c-His ₆
Ec	Exoglucanase CelE from <i>C. cellulolytica</i> fused with its native dockerin	<i>E. coli</i>	c-His ₆
Gf	Endoglucanase CelG from <i>C. cellulolytica</i> fused with a dockerin from <i>R. flavefaciens</i>	<i>E. coli</i>	c-His ₆
BglA	β -Glucosidase BglA from <i>C. thermocellum</i> fused with a dockerin from <i>R. flavefaciens</i>	<i>E. coli</i>	c-His ₆

[0095] A plasmid coding for a trifunctional scaffoldin (Scaf-ctf (SEQ ID NO:1 and 2) consisting of an internal CBD flanked by three divergent cohesin domains from *C. thermocellum* (t), *C. cellulolyticum* (c), and *R. flavefaciens* (f) (Fig. 3) was created for surface display. To further demonstrate the specificity of the different dockerin-cohesin pairs, two smaller scaffoldins, (i) Scaf-c containing a cohesin domain from *C. cellulolyticum* followed by a CBD and (ii) Scaf-ct containing an additional cohesin domain from *C. thermocellum* at the C terminus of the CBD, were generated (Fig. 3). The different scaffoldins were displayed on the yeast cell surface by using the glycosylphosphatidyl-inositol (GPI) anchor linked at the N-terminal side of the scaffoldins. A c-Myc tag was added to the C terminus of each scaffoldin to allow detection with anti-Myc serum.

[0096] Display of scaffoldins on the yeast cell surface.

For the display of scaffoldins on the yeast cell surface, yeast cells harboring pScaf-c, pScaf-ct, or pScaf-ctf were precultured in SDC medium for 18 h at 30°C. These precultures were subinoculated into 200 ml SGC medium (20.0 g/liter galactose, 6.7 g/liter yeast nitrogen base without amino acids, 5.0 g/liter Casamino Acids) at an optical density (OD) at 600 nm of 0.1 and grown for 48 h at 20°C.

[0097] Expression and purification of dockerin-tagged cellulases. *E. coli* strains expressing At, Ec, and Gf were precultured overnight at 37°C in LB medium supplemented with appropriate antibiotics. The precultures were subinoculated into 200 ml LB medium supplemented with 1.5% glycerol and appropriate antibiotics at an initial OD of 0.01 and incubated at 37°C until the OD reached 1.5. The cultures were then cooled to 20°C, and isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 200 μ M. After 16 h, cells were harvested by centrifugation (3,000 X g, 10 min) at 4°C, resuspended in buffer A (50 mM Tris-HCl [pH 8.0], 100 mM NaCl, 10 mM CaCl₂), and lysed with a sonicator. The different

cellulases were purified with a His-binding resin (Novagen) at 4°C.

[0098] Minicellulosome assembly on the yeast cell surface.

To assemble the minicellulosomes, either cell lysates containing dockerin-tagged cellulases or purified cellulases were incubated with yeast cells displaying the scaffoldin for 1 h at 4°C in buffer A. After incubation, cells were washed and harvested by centrifugation (3,000 X g, 10 min) at 4°C and resuspended in the same buffer for further use.

[0099] Immunofluorescence microscopy. Yeast cells

displaying scaffoldins or the minicellulosomes on the surface were harvested by centrifugation, washed with phosphate-buffered saline (PBS; 8 g/liter NaCl, 0.2 g/liter KCl, 1.44 g/liter Na₂HPO₄, 0.24 g/liter KH₂PO₄), and resuspended in 250 µl of PBS containing 1 mg/ml bovine serum albumin and 0.5 µg of anti-c-Myc or anti-c-His immunoglobulin G (IgG; Invitrogen) for 4 h with occasional mixing. Cells were then pelleted and washed with PBS before resuspension in PBS plus 1 mg/ml bovine serum albumin and 0.5 µg anti-mouse IgG conjugated with Alexa 488 (Molecular Probes). After incubation for 2 h, cells were pelleted, washed twice with PBS, and resuspended in PBS to an OD at 600 nm of 1. For fluorescence microscopy (Olympus BX51), 5- to 10-µl volumes of cell suspensions were spotted onto slides and a coverslip was added. Images from Alexa 488 were captured with the QCapture Pro6 software. Whole-cell fluorescence was measured with a fluorescence microplate reader (Synergy4; BioTek, VT) at an excitation wavelength of 485 nm and an emission wavelength of 535 nm.

[00100] Enzyme assays. Carboxymethyl cellulose (CMC) was obtained from Sigma and used as a substrate. Phosphoric acid-swollen cellulose (PASC) was prepared from Avicel PH101 (Sigma) according to the method of Walseth (27). Enzyme activity was assayed in the presence of a 0.3% (wt/vol) concentration of cellulose at 30°C in 20 mM Tris-HCl buffer (pH 6.0). Samples were collected periodically and immediately mixed with 3 ml of DNS reagents (10 g/liter dinitrosalicylic acid, 10 g/liter sodium hydroxide, 2 g/liter phenol, 0.5

g/liter sodium sulfite). After incubation at 95°C for 10 min, 1 ml of 40% Rochelle salts was added to fix the color before measurement of the absorbance of the supernatants at 575 nm. Glucose concentration was determined with a glucose HK assay kit from Sigma.

[00101] **Fermentation.** Fermentation was conducted anaerobically at 30°C. Briefly, yeast cells were washed once with buffer containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 10 mM CaCl₂ and resuspended in SDC medium containing 6.7 g/liter yeast nitrogen base without amino acids, 20 g/liter Casamino Acids, and 10 g/liter PASC as the carbon source. Reducing sugar production and glucose concentration were measured by the methods described above. The amount of residual cellulose was measured by the phenol-sulfuric acid method as described by Dubois *et al.* Ethanol concentration was measured with a gas chromatograph (model 6890; Hewlett Packard) with a flame ionization detector and an HP-FFTP column.

[00102] To probe the surface localization of the scaffoldins, immunofluorescent labeling of cells was carried out using anti-c-Myc sera and Alexa Fluor TM 488 conjugated goat anti-mouse IgG (Molecular Probe) and observed under a fluorescence microscope (Olympus America, Inc., San Diego, CA). Cells displaying the scaffoldin domains (1, 2, or 3) on the surface were brightly fluorescence (Figure 2), while no fluorescence was observed for the control yeast cells. These results demonstrate that a synthetic scaffoldin can be successfully displayed on the surface of a heterologous host (*e.g.*, a yeast).

[00103] To test the functionality of the displayed scaffoldins, three different cellulases fused with a corresponding dockerin domain (either c, t, or f) were expressed in *E. coli* (*i.e.*, an exoglucanase (CelE) from *C. cellulolyticum* fused to a dockerin domain from the same species (Ec), an endoglucanase (CelG) from *C. cellulolyticum* fused to a dockerin domain from *R. flavefaciens* (Gf), and an endoglucanase (CelA) fused to a dockerin domain from *C.*

thermocellum (At) were expressed in *E. coli*). The plasmids used were: (i) pETec containing an exoglucanase CelE from *C. cellulolytica* fused with a dockerin domain from the same species (Ec; see, e.g., SEQ ID NO:3 and 4, Christian Gaudin et al., Journal of Bacteriology, 2000. 182: 1910-1915, incorporated herein by reference); (ii) pETGf containing an endoglucanase CelG from *C. cellulolytica* fused with a dockerin domain from *R. flavefaciens* (Gf; see, e.g., SEQ ID NO:5 and 6; Henri-Pierre Fierobe, et al. The Journal of Biological Chemistry. 2005. 280:16325-16334, incorporated herein by reference); and (iii) pETAt containing an endoglucanase CelA fused with a dockerin domain from *C. thermocellum* (At; see, e.g., SEQ ID NO:7 and 8; Dae-Kyun Chung et al., Biotechnology Letters. 1997, 19:503-506, incorporated herein by reference). A His6 tag was added to the C terminus of each of the dockerin domains for detection of the assembly. Cells displaying scaffoldins on the surface were incubated directly with *E. coli* cell lysates containing At, Ec, or Gf for 1 h to form the cellulosome complex. The presence of each cellulase-dockerin pair on cells displaying Scaf-ctf was confirmed by immunofluorescence microscopy with the anti-His6 antibody (Fig. 2B).

[00104] To demonstrate the specificity of different cohesin-dockerin pairs, similar experiments were performed with cells displaying either Scaf-ct or Scaf-c. In Scaf-ct-displaying cells, fluorescence was detected only in the presence of Ec or At, whereas incubation with Gf did not result in any detectable fluorescence (Fig. 2C). Similarly, in Scaf-c-displaying cells, fluorescence was only observed in the presence of Ec (Fig. 2D). These results confirm that the specificity of the cohesins is preserved even when they are displayed on the cell surface, as only the corresponding dockerin-tagged enzymes are assembled correctly.

[00105] **Functionality of the displayed minicellulosomes.** To demonstrate the functionality of the assembled minicellulosomes, cells expressing Scaf-ctf were first saturated with different combinations of Ec, At, and/or Gf. As

depicted in Fig. 3, a similar level of fluorescence was detected from the c-Myc or c-His6 tag when only one dockerin-tagged enzyme was added, indicating the correct 1:1 binding between the cohesin-dockerin pairs. A corresponding increase in fluorescent intensity was observed when an increasing number of enzymes were docked on Scaf-ctf. This result confirms that the correct 1:1 binding ratio of each dockerin-cohesin pair was preserved even when it was assembled into a three-enzyme minicellulosome on the cell surface (Fig. 3).

[00106] Engineered yeast cells docked with different combinations of cellulases were further examined for functionality in cellulose hydrolysis. Cells were resuspended in Tris buffer containing CMC, and the rate of reducing sugar production was determined. As shown in Fig. 4, cells with any one of the three cellulases docked on the surface showed visible differences in cellulose hydrolysis from the control. The endoglucanase At had the highest rate of hydrolysis, followed by Gf and Ec, a trend consistent with the relatively low activity of the exoglucanase CelE on CMC. The rate of CMC hydrolysis increased in an additive fashion when two of the cellulases were docked on the cell surface, and the highest rate of hydrolysis was observed when all three cellulases were assembled. The additive effect on CMC hydrolysis confirms that the recruitment of cellulases to the displayed scaffoldin has a very minimum effect on their individual functionality.

[00107] Synergistic effect of displayed minicellulosomes.

The synergistic effect on cellulose hydrolysis is an intriguing property of naturally occurring cellulosomes. To test whether the synergistic effect of the minicellulosome structure was preserved when displayed on the yeast cell surface, Avicel hydrolysis was compared with that of purified cellulases. In this case, the amount of each cellulase docked on Scaf-ctf was first determined from the binding experiments. These predetermined amounts of cellulases were then mixed together, and the hydrolysis of Avicel with the cellulase mixture was compared with that of whole cells displaying the functional cellulosome containing the same amount of each

cellulase. As shown in Table 2, the level of reducing sugar production was consistently higher for cells displaying the cellulosome, confirming that synergy was indeed maintained. The level of synergy increased from 1.62 to 2.44 when the number of cellulases recruited in the minicellulosome system increased from one to three. This result suggests the potential to further enhance cellulose hydrolysis by increasing the number of displayed cellulases.

TABLE 2. Amounts of reducing sugars released from Avicel after 24 h of incubation at 30°C either by cells displaying cellulosomes or by the same amount of free enzymes^a

Cellulase pair(s)	Amt of reducing sugars (mg/liter) released from:		Degree of synergy
	Cellulosome	Free enzymes	
At	46.1	28.3	1.62
At + Ec	80.1	37.6	2.13
At + Ec + Gf	132.3	54.2	2.44

^a Reactions were conducted either with different cellulase pairs (CelE-Dc [Ec], CelA-Dt [At], or CelG-Df [Gf]) docked on the displayed Scaf-ctf or with the corresponding purified cellulases. The degree of synergy is defined as the amount of sugar released from the cellulosome over the amount of sugar released from free enzymes.

[00108] **Incorporation of β -glucosidase into the minicellulosome.** Since *S. cerevisiae* is unable to transport and utilize oligosaccharides, directing the complete hydrolysis of cellulose to glucose is essential. To achieve this goal, a β -glucosidase (BglA) from *C. thermocellum* tagged with the dockerin from *R. flavefaciens* was constructed. The resulting dockerin-tagged BglA retained the same specificity and docking efficiency as Gf (Fig. 3). Figure 5 shows the time course of reducing sugar and glucose released from PASC with different enzyme combinations docked on the cell surface. Although 40% of the PASC was hydrolyzed in the presence of the endoglucanase At, 25% of the reducing sugar was further hydrolyzed to glucose.

[00109] In comparison, the presence of the exoglucanase Ec not only enhanced reducing sugar production but also increased glucose production threefold. The addition of BglA further

improved the rate of glucose liberation, although no difference in reducing sugar formation was observed. This result is very significant, as demonstrated, a functional minicellulosome containing all three exoglucanase, endoglucanase, and β -glucosidase activities can be successfully assembled on the surface of a heterologous host cell. The result also confirms a role of β -glucosidase in achieving a higher conversion of cellulose to glucose. The displayed minicellulosome exhibited synergy in both reducing sugar and glucose liberation compared to that of free enzymes.

[00110] Direct fermentation of amorphous cellulose to ethanol. The ability of ethanol fermentation from PASC was examined by using the scaffoldin-displaying strains docked with different cellulases. As shown in Fig. 6, the increase in ethanol production was accompanied by a concomitant decrease in the total sugar concentration. The levels of ethanol production and PASC hydrolysis were directly correlated with the number of cellulases docked on the cell surface. The maximum ethanol production of cells displaying At, Ec, and BglA was 3.5 g/liter after 48 h; this corresponds to 95% of the theoretical ethanol yield, at 0.49 g ethanol/g sugar consumed. Moreover, the glucose concentrations during the fermentation were below the detection limit. This indicates that all of the glucose produced was quickly consumed, resulting in no detectable glucose accumulation in the medium. The level of ethanol production by cells displaying all three cellulases was higher than that of cells displaying only At and Ec, again confirming the importance of β -glucosidase in the overall cellulose-to-ethanol conversion process. More importantly, the synergistic effect of the minicellulosome was also observed, as the ethanol production by a culture with the same amounts of purified At, Ec, and BglA added to the medium was more than threefold lower.

[00111] The feasibility of using secreted cellulases for the direct assembly of functional cellulosome has also been demonstrated. The yeast secretion vector pCEL15 containing the secretion leader sequence MF α 1 was used for inserting the gene

coding for At. Cells carrying the secretion vector were co-cultured with cells displaying scaff#3 for 24 h. The correct assembly of the secreted At onto the cell surface was again verified by immunofluorescence microscopy. The assembled At remained active as demonstrated by the ability to hydrolyze Avicel. By changing the inoculation densities of the two cultures, different levels of At activity associated with cells and remained in the medium were detected. These results confirm the possibility of fine-tuning the assembly of functional cellulosomes on the cell surface using an engineered consortium of cells performing separate functions.

[00112] Overall, the results demonstrated the successful functional assembly of a mini-cellulosome on the yeast surface. The displayed mini-cellulosomes enable the cells to hydrolyze cellulose and grow using cellulose as the sole carbon source. Moreover, the increased cell growth and reducing sugar production with increasing cellulases docked on the surface indicates the potential to further increase the efficiency of cellulose hydrolysis by increasing the number of displayed cellulases via the use of more complex cellulosome structures.

[00113] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

Sequence of SCAF6

```
ATGGGCGATTCTCTTAAAGTTACAGTAGGAACAGCTAATGGTAAGCCTGGCGATACAGTAACAGTTCCTGTTACAT
TTGCTGATGTAGCAAAGATGAAAAACGTAGGAACATGTAATTTCTATCTTGGATATGATGCAAGCCTGTTAGAGGT
AGTATCAGTAGATGCAGGTCCAATAGTTAAGAATGCAGCAGTTAACTTCTCAAGCAGTGCAGCAACGGAACAATC
AGCTTCTGTTCTGGATAACACAATTACAGACGAATTGATAACTGCAGACGGTGTGTTTGCAAATATTAAGTTCA
AATTAAAGAGTGTAAACGGCTAAACTACAACACCAGTAACATTTAAAGATGGTGGAGCTTTTGGTGACGGAATAT
GTCAAAGATAGCTTCAGTTACTAAGACAACGGTAGTGTAAACGATCGATCCGACCAAGGGAGCAACACCAACAAAT
ACAGCTACGCCGACAAATCAGCTACGGCTACGCCACCAGGCCATCGGTACCGACAAACACACCGACAAACACAC
CGGCAATACACCGGTATCAGGCAATTTGAACGGTGAATTCTACAACAGCAATCCTTCAGATACTACTAATCAAT
CAATCCTCAGTTCAGGTTACTAATACCGGAAGCAGTGAATTTGTTCCAACTCACATTGAGATATTATTAT
ACAGTAGACGGACAGAAAGATCAGACCTTCTGGTGTGACCATGTGCAATAATCGGCAGTAACGGCAGCTACAACG
GAATTACTTCAATGTAAAGGAACATTTGTAATAATGAGTTCCTCAACAAATAACGCAGACACCTACCTTGAAAT
AAGCTTTACAGCGGAACTCTTGAACCGGTGCACATGTTTCAGATACAAGGTAGATTTGCAAGAAATGACTGGAGT
AACTATACACAGTCAAATGACTACTCATTCAAGTCTGCTTCACAGTTTGTGTAATGGGATCAGGTAACAGCATACT
TGAACGGTGTCTTGTATGGGGTAAAGAACCCGGTGGCAGTGTAGTACCATCAACACAGCCTGTAACAACACCACC
TGCAACAACAAACACCTGCAACAACAAACACCTGCAACAACAAATACCGCCGTCAGATGATCCGAATGCAATA
AAGATTAAAGTGGACACAGTAAATGCAAAACCGGAGACACAGTAAATATACCTGTAAGATTTCAGTGGTATACCAT
CCAAGGAATAGCAAATGTGACTTTGTATACAGCTATGACCCGAATGTACTTGAGATAATAGAGATAAAACCGGG
AGAATTGATAGTTGACCCGAATCCTGACAAGAGCTTTGATACTGCAGTATATCCTGACAGAAAGATAATAGTATTC
```

CTGTTTGCAGAAGACAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTTGCTACGATAGTAGCGAAAG
 TAAAAATCCGGAGCACCTAACGGACTCAGTGTAATCAAATTTGTAGAAGTAGGCGGATTTCGCAACAATGACCTTGT
 AGAACAGAGGACACAGTTCCTTTGACGGTGGAGTAAATGTTGGAGATATAGGATCCGCCGGTGGTTTATCCGCTGTG
 CAGCCTAATGTAGTTTAGGCGAAGTACTGGATGTTTCTGCTAACAGAACCGCTGCTGACGGAACAGTTGAATGGC
 TTATCCCAACAGTAACAGCTCCAGGCCAGACGGTCACTATGCCCGTAGTAGTCAAGAGTTCAAGTCTTGCAGT
 TGCTGGTGGCAGTTCAAGATCCAGGCGGCGACAGGCGTAAGTTATTGCTCCAGACGGACGGTACGCTTACGGT
 TCAGGCATTGTGTACAATAATAGTAAGTATGCTTTTGGACAGGGTGCAGGTAGAGGAATAGTTGCAGCTGATGATT
 CGGTTGTGCTTACTCTTGCATATACAGTTCCTGCTGATTTGTGCTGAAGGTACATATGATGTCAAGTGGTCTGATGC
 GTTTGTAAGTGATACAGACGGACAGAATATCACAAGTAAGGTTACTCTTACTGATGGCGCTATCATTGTTAAGTAG

Sequence of Ec

ATGCTTGTGGGGCAGGAGATTTGATTGCAAAACCATACCTTTGACAACAGAGTAGGTCTTCCATGGCAGCTGGTTG
 AATCATACCTTGCAAAGGCAAGTTTTGAAATTACATCTGATGGTAAGTACAAGATAACTGCTCAAAGATCGGTTGA
 GGCAGGAAAAGGTGAAAGATGGGATATACAATTCGGTCACAGAGGACTCGCATTGCAACAAGGTCATACATTATACA
 GTAAAGTTTACTGTTTACTGCTAGCAGAGCTTGTAAATTTATCCTAAAATAGGTGACCGGTTGATCCATATGATG
 AATACTGGAATATGAATCAACAATGGAATTTCTGGAATTACAGGCTAATACTCCAAAACTGTAACACAGACATT
 TACACAGACTAAGGGAGATAAGAAGAAGCTTGAATTTGCTTTTACCTTGCTCCCGATAAACTACATCTGAGGCA
 CAGAATCCAGCAAGTTTCCAACCTATAACATATACTTTGATGAAATTTATATTCAGGACCTCAATTTGCAGGAT
 ATACTGAAGATCCACCTGAACCTACTAATGTTGTACGTTTGAATCAGGTAGGTTCTATCCTAATGCTGATAAGAT
 TGCAACAGTAGCAACAAGTTCAACAACCTCAATTAACCTGGCAGTTGGTTAATAGTACTGGAGCAGCTGTTTTAACA
 GGTAATCAACTGTTAAAGGTGCCGACCGTGATCAGGTGATAATGTCCATATCATTGATTTCTCTAGTTACACAA
 CACCTGGTACCGACTATAAGATAGTAACAGATGTATCAGTAACAAAAGCCGGAGACAATGAAAGTATGAAGTTCAA
 TATTGGAGATGACCTTTTACTCAAATGAAATACGATTCATGAAGTATTTCTATCACACAACAGAAGTGCTATTCCA
 ATACAAATGCCATACTGTGATCAATCACAATGGGCACGTCCTGCAGGACACACAACCTGATATACTTGCTCCAGATC
 CAACAAGGATTACAAGGCTAATACACACTTGACGTTACAGGTGGTTGGTATGATGCCGGTGACCATGGTAAAGTA
 TGTTGTTAATGGTGGTATTGCAACCTGGACCGTAATGAATGCATATGAGCGTGCCTACACATGGGTGGAGACACT
 TCAGTTGCTCCATTTAAAGACGTTCTTTAGCAATACCAGAAGCGGAAGTCTATCCTGACATACTGGACGAAGCTC
 GTTACCAGCTCATTAACATGAAACATTATTAATAGTCAGGTTCCAGCAGGAAAGTATGCCGGTATGGCTACCA
 CAAAGCTCATGACGAACGTTGGACAGCTCTTGCTGTACGTCGCCGACAGGATACAATGAAACGTTGGTTGCAGCT
 CCAAGTACAGCAGCTACATTAATCTGGCTGCTATTGCTGCACAAAGTTACAGTCTTTGGAACAGTTTGTATTCTG
 CTTTGCAGACTTGTATTTAAAGACGTTCTTGAAGTACAGCAAGCTGCTAGGGATGCAGCTGTAGCTCATCCAGAAATATATGCAAC
 TATGGAACAGGGTGCCGGTGGTGGAGCATAACGAGACAACCTATGTTCTTGATGATTTCTACTGGGCAGCATGTGAA
 TTGTATGCAACTACAGGCAGTGACAAGTATTTGAACTACATAAAGAGCTCAAAGCATTATCTCGAAATGCCATACAG
 AATTAACAGGCGGTGAGAATACTGGAATTACAGGGGCTTTTACTGGGTTGTACAGCAGGTATGGGAACAATAAC
 ACTTGCAGCTTGTACTTACAAAGCTTCCGGCAGCAGATTTGCTACAGCTAAAGCTAATATTCAAGCTGCAGCTGAT
 AAGTTTCATATCAATTTCAAAGCACAAGGCTATGGTGTACCACTAGAAGAAAAAGTAATTTTCATCTCCTTTTGATG
 CATCTGTTGTTAAAGGTTTCCATGGGGATCAAACCTATTCGTTATTAATGAAGCAATAGTTATGTCATATGCTTTA
 TGAATTCAGCGATGTTAATGGCACAAGAATAATAATATATTAATGGTGTCTTAAACAGCAATGGATTACCTCCTC
 GGACGTAAACCAATATTTCAAAGCTATATACTGGTTATGGTGACAACCCACTTGAAATCCTCATCACCTTTCT
 GGGCATAACAGGCAGACAACACATTCCCAAACCACTCCGGGATGCTGTGTCAGGAGGACCTAATCCGGCTTGCA
 GGATCCTTGGGTTAAGGGTTCAAGGCTGGCAGCCAGGTGAAAGACCTGCTGAAAAATGCTTCATGGACAATATTGAA
 TCTTGGTCAACAAACGAAATAACCATCAACTGGAATGCTCCTCTTGATGGATATCAGCTTACCTTGATGAAAGG
 GGCCAGAGATTGGTGGGTCAGTGACTCCTCAACTAATTTAGGAGATGTTAAGCGCGATGGAAACAGGATGCATT
 GGACTTCGCTGCATTGAAGAAAGCCTTGTAAAGCCAGGATACTTCTACTATAAATGTTGCTAATGCTGATATAAAC
 AAAGATGGTTCTATTGATGCAGTTGACTTTGCATTACTCAAATCATTCTTGTTAGGAAAAATCACACAGTGA

Sequence of Gf

ATGGGAACATATAACTATGGAGAAGCATTACAGAAATCAATAATGTTCTATGAATTCAGCGTTCGGGAGATCTTC
 CGGCTGATAAAGCTGACAACCTGGAGAGACGATTCGGGTATGAAAGACGGTTCTGATGTAGGAGTTGATCTTACAGG
 AGGATGGTACCGATGACGGTGACCATGTGAAATTTAATCTACCTATGTCATATACATCTGCAATGCTTGATGGTCC
 TTATATGAGGATAAGGATGCTTATGATAAGAGCGGTGACAGAAAAATATATAATGGACGGTATAAAATGGGCTAATG
 ATTATTTTATTAATGAATCCGACACCCGGTGTATATTATACCAAGTAGGAGACGGCGGAAAGGACCACTCTTG
 GTGGGGCCCTGCGGAAGTAATGCAGATGGAAGACCGTCTTTTAAAGGTTGACGCTTCTAAAGCCGGTTCTGCAGTA
 TGTGCTTCCACTGCGAGCTTCTCTGGCATCTGCAGCAGTAGTCTTTAAATCCAGTGATCCTACTTATGCAGAAAAGT
 GCATAAGCCATGCAAGAACCTGTTTGATATGGCTGACAAAGCAAAGAGTGATGCTGGTTATACTGCGGCTTCAGG
 CTACTACAGCTCAAGCTCATTTTACGATGATCTCTCATGGGCTGCAGTATGGTTATATCTTGCTACAAATGACAGT
 ACATATTTAGACAAAGCAGAATCCTATGTACCGAATTCGGGTAAAGAACAGCAGACAGATATTATCGCCTACAAGT
 GGGGACAGTGCTGGGATGATGTTCAATTATGGTGTGAGCTTCTTCTTGCAAAGCTTACAAACAAACAATTGTATAA
 GGATAGTATAGAAATGAACCTTGACTTCTGGACAACCTGGTGTAAACGGAACACGTTTCTTACACGCCAAAGGGT
 TTGGCGTGGCTATTCCAATGGGGTTCATTAAAGCATGTACAACCTCAGGCTTTTTTAGCCGGTGTATTATGCAGAGT
 GGAAGGCTGTACGCCATCCAAGTATCTGTATATAAGGATTTCTCAAGAGTCAAATTTGATTATGCACTTGGCAG
 TACCGGAAGAAGTTTTGTTGTCGGATATGGAGTAAATCCTCCTCAACATCCTCATCACAGAACTGCTCACGGTTCA
 TGGACAGATCAATGACTTCACCAACATACCACAGGCATACTATTATGGTGGCTGGTAGGAGGACCGGATAATG
 CAGATGGCTACTGATGAAATAAACAATTATGTCAATAATGAAATAGCCTGCGATTATAATGCCGGATTACAGG
 TGCATTGCAAAAATGTACAAGCATTCTGGCGGAGATCCGATTCCAACTTCAAGGCTATCGAAAAATAACCAAC
 GATGAAGTTATTATAAAGCAGGTTTGAATTAACCTGGCCCTAATCACTGAAATCAAGGCTGTGTTTATAAAC

AGACAGGATGGCCTGCAAGAGTTACGGACAAGATATCATTTAAATATTTTATGGACTTGTCTGAAATTGTAGCAGC
AGGAATTGATCCTTTAAGCCTTGTAACAAGTTCAAATTATTTCTGAAGGTAAGAATACTAAGGTTTCCGGTGTGTTG
CCATGGGATGTTTCAAATAATGTTTACTATGTAAATGTTGATTGACAGGAGAAAAATCTACCCAGGCGGTGAGT
CTGCGTGCAGACGAGAAGTTCAGTTCAGAAATTGCCGCACCACAGGGAAGAAGATATTGGAATCCGAAAAATGATTT
CTCATATGATGGATTACCAACCACCAGTACTGTAAATACGGTTACCAACATACCTGTTTATGATAACGGCGTAAAA
GTATTTGGTAACGAACCCGAGGTGGATCAGAACCAGGACAAAGCTCGTTCCCTACATGGGCGGATACAACTGCG
ACGGCGTTGTAAATGTTGCTGACGTAGTAGTTCTTAACAGATTCCCTCAACGATCCTACATATTCTAACATTACTGA
TCAGGGTAAGGTTAACGCAGACGTTGTTGATCCTCAGGATAAGTCCGGCGCAGCAGTTGATCCTGCAGGCGTAAAG
CTCACAGTAGCTGACTCTGAGGCAATCCTCAAGGCTATCGTTGAACATCACACTTCCTCAATGA

Sequence of At

ATGGCAGGTGTGCCTTTTAAACACAAAATACCCCTATGGTCCTACTTCTATTGCCGATAATCAGTCGGAAGTAACTG
CAATGCTCAAAGCAGAATGGGAAGACTGGAAGAGCAAGAGAATTACCTCGAACGGTGCAGGAGGATACAAGAGAGT
ACAGCGTGATGCTTCCACCAATTATGATACGGTATCCGAAGGTATGGGATACGGACTTCTTTTGGCGGTTTGCTTT
AACGAACAGGCTTTGTTTACGATTATACCGTTACGTAAAATCTCATTTCAATGGAAACGGACTTATGCACTGGC
ACATTGATGCCAACACAATGTTACAAGTCATGACGGCGGCGACGGTGCAGCAACCGATGCTGATGAGGATATTGC
ACTTGCGCTCATATTTGCCGACAAGTTATGGGGTCTTCCGGTGCAATAAACTACGGGCAGGAAGCAAGGACATTG
ATAACAATCTTTACAACCATTGTGTAGAGCATGGATCCTATGTATTAAAGCCCGTGACAGATGGGAGGTTTCAT
CAGTAACAACCCGTCATATTTTGCCTGTCATGGTACAAAAGTGTATGCTCAATATACAGGAGACACAAGATGGAA
TCAAGTGGCGGACAAGTGTACCAAATTGTTGAAGAAGTTAAGAAATACAACAACGGAACCGGCTTGTTCCTGAC
TGGTGACTGCAAGCGGAACTCCGGCAAGCGGTACAGATTACGACTACAAATATGATGCTACACGTTACGGCTGGA
GAACTGCCGTGGACTATTCATGGTTTGGTGACCAGAGAGCAAGGCAAACTGCGATATGCTGACCAAAATCTTTGTC
CAGAGACGGGGCAAAAGGAATCGTTGACGGATACACAATTCAAGGTTCAAAAATTAGCAACAATCACACGCATCA
TTTATAGGACCTGTTGCCGCGCAAGTATGACAGGTTACGATTTGAACTTTGCAAAGGAACCTTATAGGGAGACTG
TTGCTGTAAAGGACAGTGAATATTACGGATATTACGGAACAGCTTGAGACTGCTCACTTTGTTGTACATAACAGG
AACTTCCCGAATCCTTTGAGTGACCTTTCCGGCCAACCGACACCACCGTCGAATCCGACACCTTCATTGCCTCCT
CAGTTGTTTACGGTGATGTAAATGGCGACGGTAATGTTAACTCCACTGATTTGACTATGTTAAAAAGATATCTGC
TGAAGAGTGTTACCAATATAAACAGAGAGGCTGCAGACGTTAATCGTGACGGTGCGATTAACTCCTCTGACATGAC
TATATTAAAGAGATATCTGATAAAGAGCATACCCACCTACCTTATTAG

WHAT IS CLAIMED IS:

1. A culture comprising:
 - a first recombinant yeast strain comprising an anchoring scaffoldin (anScaff);
 - a second recombinant yeast strain comprising an adaptor scaffoldin comprising a plurality of cohesin domains and at least one cellulose binding domain (CBD); and
 - at least one recombinant yeast strain comprising a plurality of secreted dockerin-tagged cellulases.
2. The culture of claim 1, further comprising culturing the yeast strains under conditions wherein the anchoring scaffoldin, the adaptor scaffoldin comprising the cohesion domains and the plurality of dockerin-tagged cellulases associate to generate an engineered cellulosome.
3. The culture of claim 1, wherein the cellulases are selected from endoglucanases, exoglucanases, β -glucosidase, and xylanase.
4. The culture of claim 1, wherein the dockerin-tagged cellulase is engineered to comprise a leader sequence for secretion of the dockerin-tagged cellulase.
5. The culture of claim 4, wherein the leader sequence comprises an MF β 1 leader sequence.
6. The culture of claim 1, wherein the first strain and second strain are the same.
7. A recombinant yeast strain comprising a heterologous polynucleotide encoding an anchoring scaffoldin.
8. A recombinant yeast strain comprising a heterologous polynucleotide encoding an adaptor scaffoldin comprising a

plurality of cohesin domains and at least one cellulose binding domain (CBD).

9. A recombinant yeast strain comprising at least one heterologous polynucleotide encoding a secreted dockerin-tagged cellulase.

10. A culture comprising a recombinant yeast strain of any one of claims 9, 10 or 11.

11. A yeast culture comprising at least two recombinant strains of yeast wherein the culture produces a designer cellulosome, and wherein the yeast culture catabolizes cellulosic material to produce ethanol.

12. The yeast culture of claim 11, wherein the synthetic cellulosome comprises polypeptide selected from amino acid sequences having at least 80% sequence identity to SEQ ID NO:2, 4, 6, or 8.

13. The yeast culture of claim 12, wherein the designer cellulosome comprises a scaffoldin domain, a CBD, a cohesin domain, and a cellulosic material degrading enzyme.

14. A method of producing a biofuel metabolite comprising: culturing the yeast of any one of claims 1, 7, 8, 9, or 11 in a fermentation broth comprising a cellulosic material, wherein the microorganism produces the biofuel metabolite.

15. The method of claim 14, wherein at least 50% of the carbon in the fermentation broth is in the form of cellulosic material.

16. The method of claim 16, wherein at least 80% of the carbon in the fermentation broth is in the form of cellulosic material.

17. A method of designing a cellulosome comprising identifying the cellulosic substrate, identifying at least one enzyme useful for degradation of the cellulosic material, recombinantly engineering a dockerin peptide to the enzyme, cloning a polynucleotide encoding the dockerin-linked enzyme into a microorganism, culturing the microorganism in a culture of at least one additional microorganism expressing a scaffoldin having a plurality of cohesion domains and a cellulosic binding domain, wherein the cohesion and dockerin are compatible and culturing the microorganisms to express the scaffoldin and dockerin-linked enzymes.

18. A method of producing a product from cellulosic material comprising culturing designed cellulosome of claim 17, wherein the cellulosic material is used to produce the product.

1/7

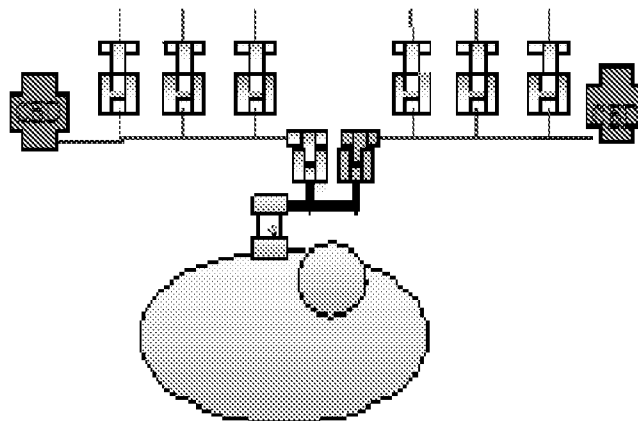


FIGURE 1A

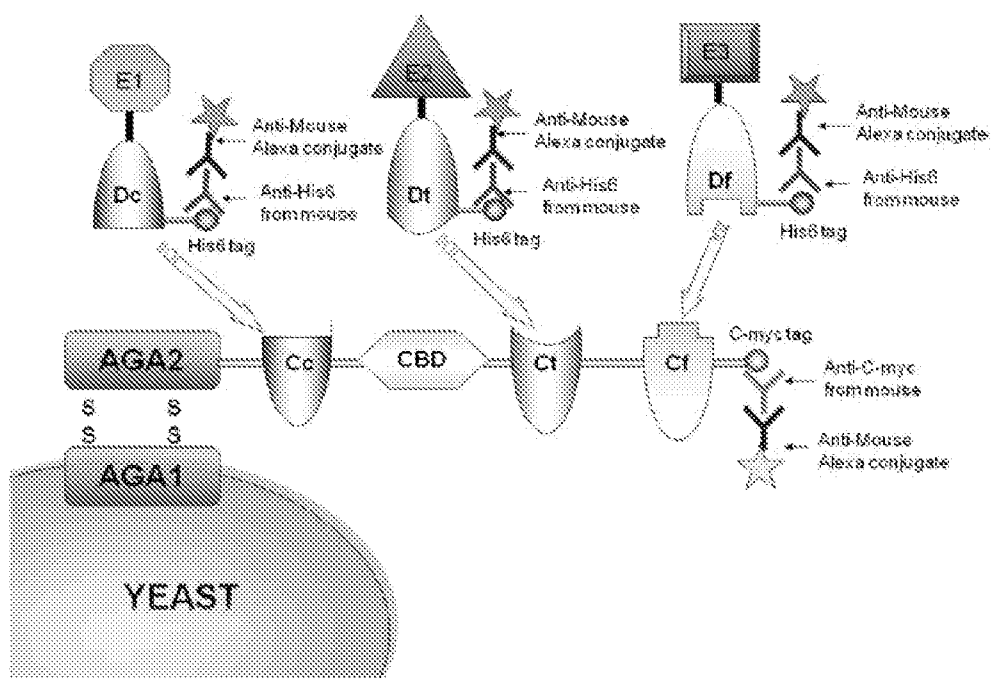


FIGURE 1B

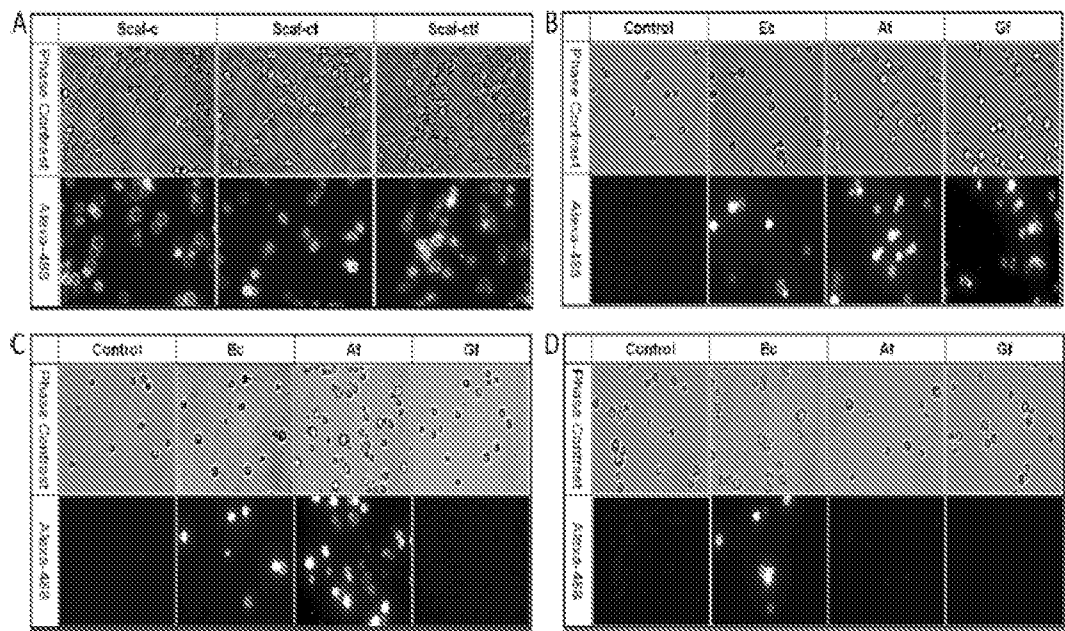


FIGURE 2

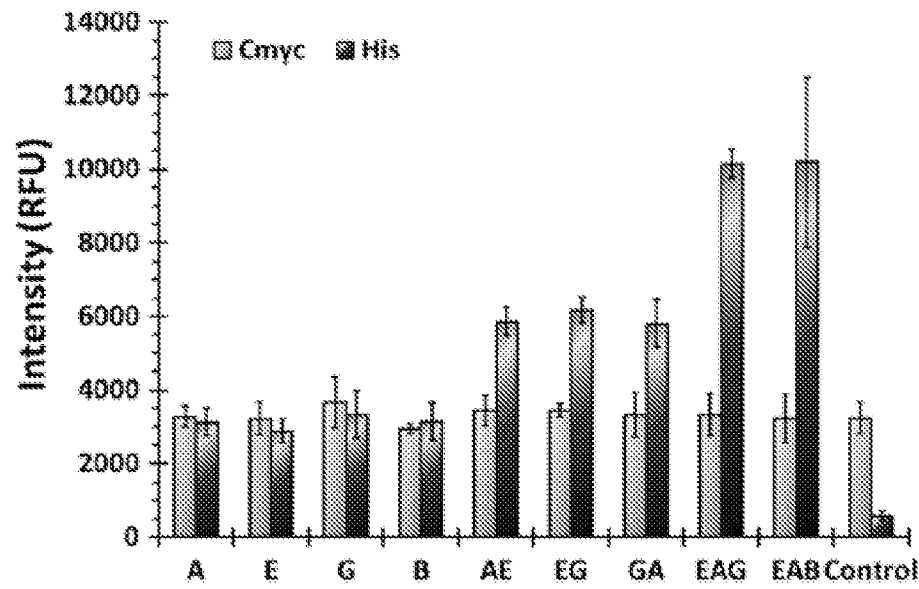


FIGURE 3

3/7

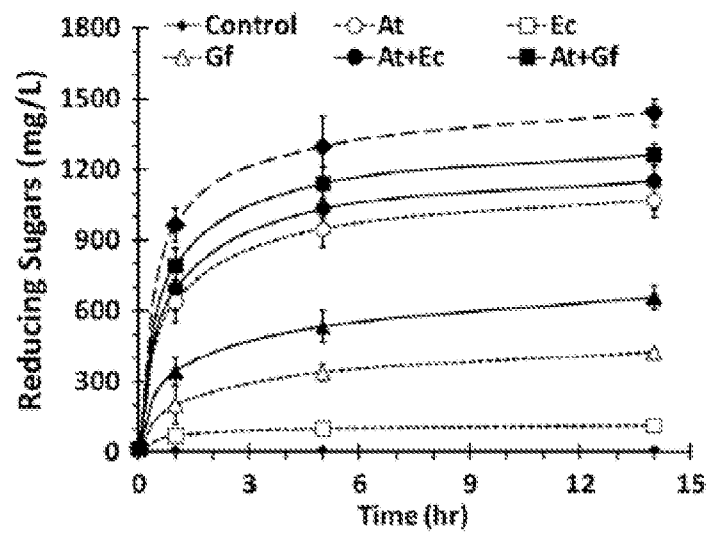


FIGURE 4

4/7

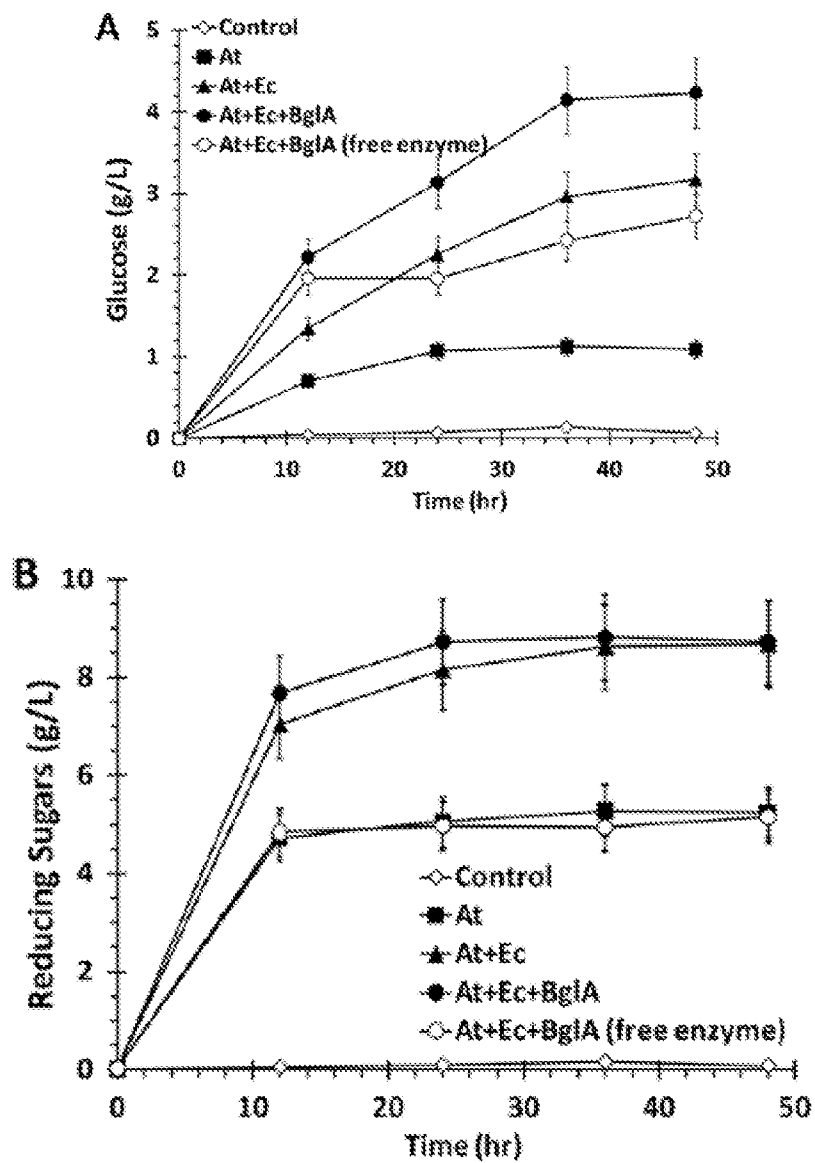


FIGURE 5A-B

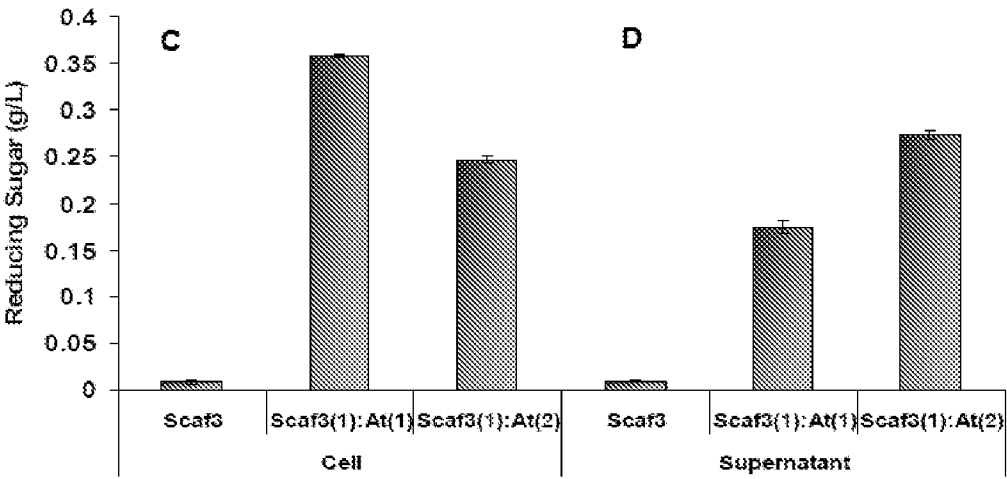


FIGURE 5C-D

6/7

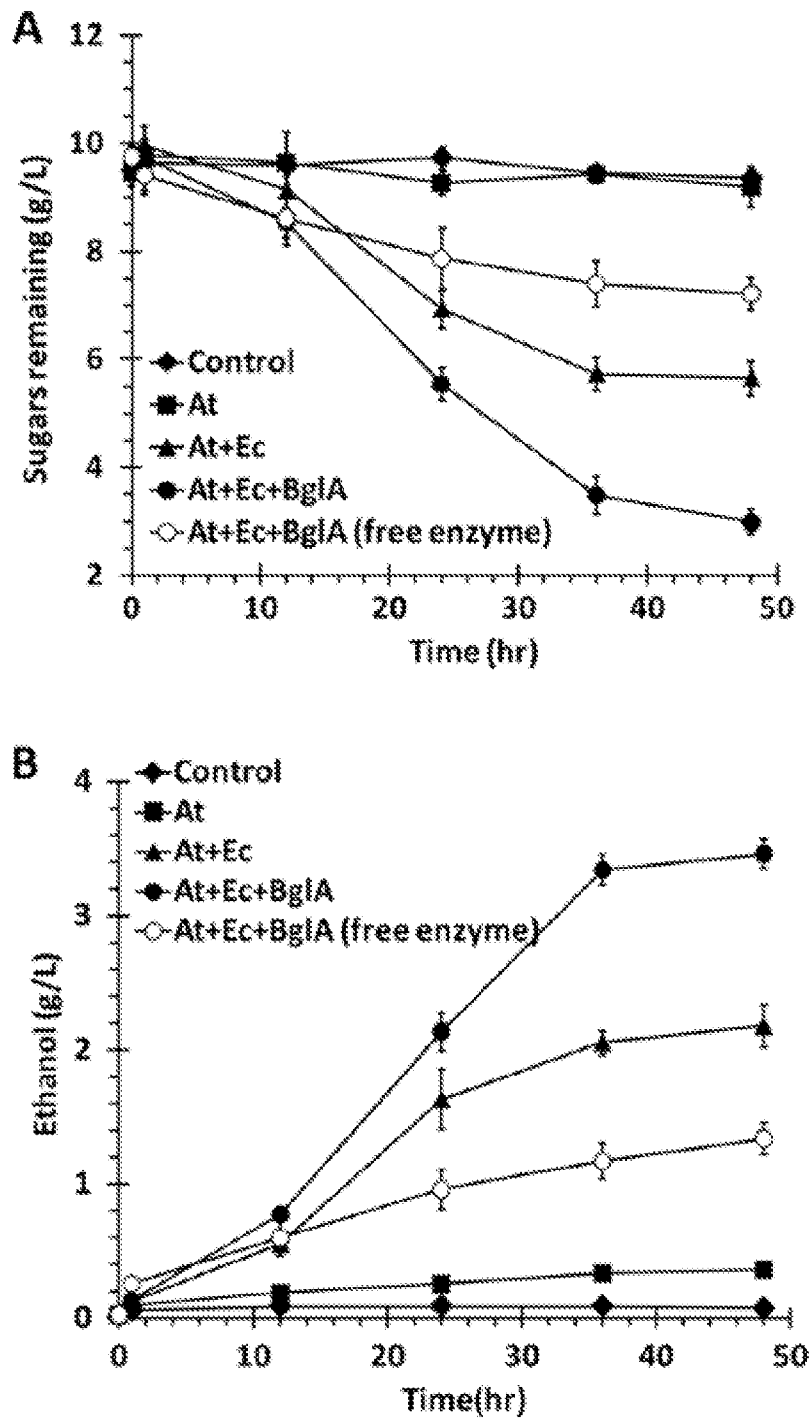


FIGURE 6

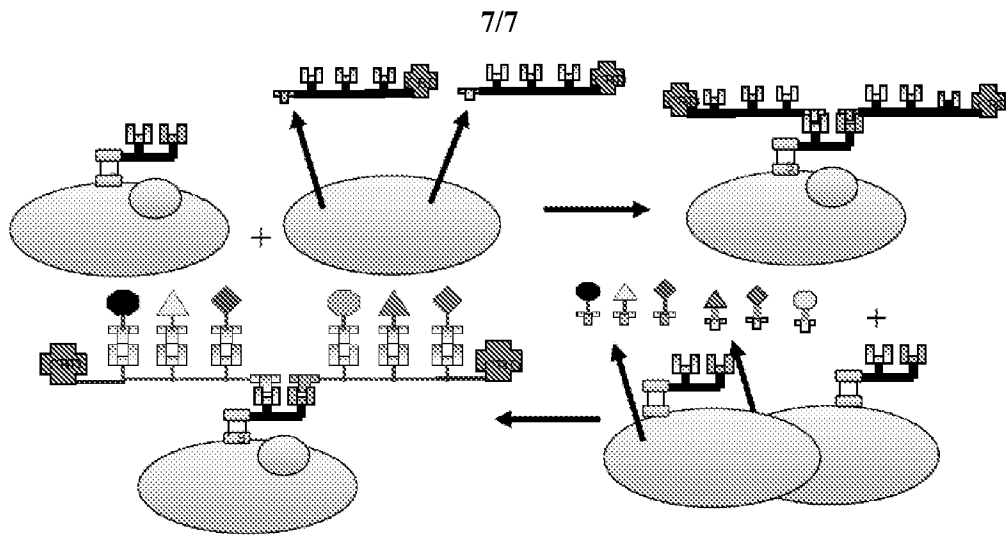


FIGURE 7