

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies.

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/90413> since 2016-01-28T12:15:51Z

Published version:

DOI:10.1016/j.tibtech.2011.08.003

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



UNIVERSITÀ DEGLI STUDI DI TORINO

This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process - such as editing, corrections, structural formatting, and other quality control mechanisms - may not be reflected in this version of the text. The definitive version of the text was subsequently published in Trends Biotechnol., 30(2), Feb 2012, doi: 10.1016/j.tibtech.2011.08.003.

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

- (1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>), + doi: 10.1016/j.tibtech.2011.08.003

1 **Engineering new metabolic capabilities in bacteria: lessons from**
2 **recombinant cellulolytic strategies.**

3

4 Roberto Mazzoli, Cristina Lamberti and Enrica Pessione

5

6 Dipartimento di Biologia Animale e dell'Uomo. Università di Torino. Via Accademia Albertina 13.

7 10123 Torino. Italy.

8

9 *Corresponding author:* Mazzoli R. (roberto.mazzoli@unito.it)

10

11 **Key words:** cellulases, metabolic engineering, mRNA stability, codon usage, protein secretion,
12 surface anchoring.

13

14

15 **Abstract**

16 Cellulose waste biomass is the most attractive substrate for "biorefinery strategies" producing
17 high-value products (*e.g.* fuels, plastics) by fermentation. However, traditional biomass
18 bioconversions are economically inefficient multistep processes. Thus far, no microorganisms able
19 to perform single-step fermentation into products (consolidated bioprocessings, CBP), have been
20 isolated. Metabolic engineering is currently employed to develop recombinant microorganisms
21 suitable for CBP.

22 The heterologous expression of extracellular proteins (*e.g.* cellulases, hemicellulases) is the
23 key feature of recombinant cellulolytic strategies, conferring cellulolytic ability to microorganisms
24 exhibiting high product yields and titers. Although more and more molecular tools are becoming
25 available, efficient heterologous expression of secreted proteins is still a challenge. The present
26 review summarizes both bottlenecks and solutions of organism engineering for biomass biorefinery
27 strategies.

28

29

30 ***Towards engineered microorganisms for biomass consolidated bioprocessing***

31 Cellulose biomass is the largest waste produced by human activities and the most attractive
32 substrate for “biorefinery strategies” to produce high-value products (*e.g.* fuels, bioplastics, enzymes)
33 through fermentation processes [1-3]. However, so far, no natural microorganisms with the necessary
34 metabolic features for single-step biomass fermentation, *i.e.* consolidated bioprocessings (CBP), have
35 been isolated. Traditional biomass bioconversion processes are economically inefficient multistep
36 processes that require dedicated cellulase production [4]. Research efforts have been aimed at
37 developing recombinant microorganisms that have the characteristics required for CBP [5-7].

38 The heterologous expression of extracellular proteins (*e.g.* cellulases, hemicellulases) is the
39 key feature of recombinant cellulolytic strategies (RCS), as they confer cellulolytic ability to
40 microorganisms with high-value product formation properties [6,8-10]. Although more and more
41 molecular tools and related literature are available, one of the main challenges of metabolic pathway
42 engineering is to find an efficient heterologous protein secretion method. Efficient transformation
43 protocols have been established for few model bacteria. Although the choice of a suitable constitutive
44 or inducible promoter for efficient gene transcription is essential, the latter is only one of several
45 mechanisms, at both mRNA (*i.e.* mRNA stability, translation efficiency) and protein (*i.e.* stability,
46 transport and activity) levels, involved in gene expression in microorganisms [11-14]. Such
47 mechanisms have been optimized in natural organisms through evolution. Those researchers who
48 wish to engineer “new” (*i.e.* recombinant) organisms should modulate heterologous gene expression
49 in order to mimic naturally occurring mechanisms, that evolved through mutation plus selection, or,
50 at least, to obtain functional systems for the envisaged industrial application (Figure 1).

51 This is particularly difficult for RCS since they involve cloning and expression of multiple
52 genes and gene product translocation across the cell envelope and possibly post-translational
53 modifications and anchoring to the cell surface.

54 The present review is aimed at summarizing both the bottlenecks and innovative solutions
55 employed in organism engineering for RCS. Such topics will be detailed in the subsequent sections,
56 after a brief introduction on the native cellulase systems.

57

58 ***Natural cellulolytic systems: structure and regulation***

59 Natural plant degrading microorganisms biosynthesize extracellular multiple enzyme
60 systems. These systems consist of different substrate specificities (*e.g.* cellulases, xylanases,
61 pectinases) and catalytic mechanisms, which can be either free or cell associated [15-18]. Aerobic
62 microorganisms, such as filamentous fungi (*e.g. Trichoderma reesei*) and actinomycete bacteria,
63 generally produce “free” cellulases that do not form stable complexes [15,17-18]. Anaerobic bacteria,
64 such as *Clostridium spp.* and *Ruminococcus spp.*, and fungi (*i.e.* Chytridomycetes) have developed
65 “complexed” cellulase systems called “cellulosomes” [16-17,19] (Box 1).

66 The genes encoding cellulases are either randomly distributed or clustered on the chromosome
67 of cellulolytic microorganisms [15]. The mechanisms regulating cellulase gene expression remained
68 obscure for many years since transcriptional promoters could not be found within large gene clusters.
69 However, the existence of large polycistronic operons has recently been demonstrated in *Clostridium*
70 *cellulolyticum* [20]. The *C. cellulolyticum* 26 kb *cip-cel* cluster of cellulosomal genes consists of at
71 least a 14 kb operon and other smaller transcriptional units that include 1 to 5 genes. It has even been
72 hypothesized that the entire *cip-cel* cluster could be a single operon transcribed as a whole primary
73 mRNA that is then processed into various secondary transcripts, which display different stabilities
74 [20]. Two further operons, *i.e. celC*, consisting of *celC-glyR3-licA*, and *manB-celT*, have been
75 identified in *Clostridium thermocellum* [21]. The promoter of the *celC* operon is repressed by GlyR3,
76 while it is activated when laminaribiose, a β -1,3 glucose dimer, is available. Moreover, a set of six
77 putative alternative σ factors and membrane-associated anti- σ factors, which may play a role in
78 cellulosomal gene regulation, has recently been identified in *C. thermocellum* [22] (Box 2).

79

80 ***Gene expression optimization***

81 *Choice of the promoter*

82 Metabolic engineering by gene manipulation traditionally aims at generating many-fold
83 overexpression of heterologous genes which are considered to be the rate determining step in a
84 pathway [23]. RCS has been performed, in most cases, by cloning heterologous cellulase genes under
85 the control of constitutive promoters in *Bacillus subtilis*, *Clostridium acetobutylicum*, *Lactococcus*
86 *lactis*, *Lactobacillus plantarum*, *Saccharomyces cerevisiae*, and *Zymomonas mobilis* [24-29]. Such a
87 strategy appears more appropriate for microorganisms aimed to biorefineries since it avoids the non
88 negligible supplemental cost of large amounts of specific inducers [5]. Nonetheless, constitutive
89 “uncontrolled” heterologous cellulase biosynthesis may lead to saturation of transmembrane transport
90 mechanisms with inhibitory effects on cell growth and viability [28-30]. Toxicity can therefore be
91 diminished by weakening the promoter strength through rational or random mutagenesis [29-30].
92 Alternatively, inducible promoters could be used to delay protein biosynthesis in a growth phase (*e.g.*
93 mid-log phase) which would be more suitable for both effective protein biosynthesis and reduced
94 toxic effects [28]. Inducible promoters have also been employed to engineer *L. lactis* and *S. cerevisiae*
95 strains with heterologous cellulases, in order to obtaining improved silage fermentation and
96 digestibility of ensiled biomass and amorphous cellulose fermentation to ethanol, respectively [31-
97 33]. As the understanding of cellulase system regulatory networks in natural microorganisms is
98 increasing, it is tempting to mimic such models in recombinant hosts [22]. Furthermore, synthetic
99 biology and metabolix flux analysis will probably play key roles in developing artificial promoters
100 for the fine tuning of heterologous genes and gene networks [23,34].

101

102 *Regulation of mRNA stability*

103 mRNA concentration is a balance between gene transcription and mRNA degradation. The
104 fine tuning of mRNA degradation is actually used by prokaryotes to modulate gene expression, *e.g.*
105 the expression of cellulase genes [11,12,20].

106 The improvement of mRNA stability can be used as a further effective tool to increase the
107 expression of heterologous cellulases, thus eliminating the need for time-consuming promoter
108 screening procedure [35-36]. mRNA 5'-untranslated leader sequences (UTLS) have a 5' stem-loop
109 structure and a ribosome binding site (RBS), and have been reported to contribute to mRNA
110 stabilization in *Bacillus subtilis*, *Escherichia coli* and *Lactobacillus acidophilus* [35,37-38]. Increased
111 amounts of the α -amylase from *Streptococcus bovis* 148 could be biosynthesized in *L. casei* by fusing
112 the UTLS (and the RBS) of the *slpA* gene from *Lactobacillus acidophilus* with the promoter of the
113 gene encoding lactate dehydrogenase of *Lactobacillus casei* [35]. The same strategy has been used to
114 optimize *C. thermocellum* CelA cellulase expression in *Lactobacillus plantarum* [26]. In some cases,
115 the improvement in mRNA stability could be even more effective for the secretion of large amounts
116 of heterologous proteins than using stronger promoters [36].

117

118 *Modulation of translation efficiency*

119 The genome GC content is the primary determinant of the codon and amino acid usage
120 patterns observed in different bacterial groups [39]. The use of amino acids encoded by GC-rich
121 codons increases by approximately 1% for each 10% increase in genomic GC content [39].

122 The GC content compatibility between donor and recipient strains should therefore be taken
123 into account for an efficient heterologous protein translation. In this respect, the heterologous
124 expression of pyruvate decarboxylases (PDC) for the construction of ethanol over-producing strains
125 can be taken as a paradigm. Engineering gram-positive hosts for robust ethanol production has long
126 been limited by the availability of a suitable pool of PDC encoding genes [40]. Since PDC is
127 widespread in plants, yeasts and fungi, but rare in bacteria, the *Zymomonas mobilis pdc* gene has been
128 the workhorse for prokaryote engineering, though with very limited success on gram-positive strains
129 [40]. Talarico and co-workers [40] demonstrated that the levels of heterologous PDC in *B. subtilis*
130 depended on the GC content, *i.e.* the codon usage, of the *pdc* donor strain, although mRNAs were
131 present in similar concentrations. When “donor” strains with a suitable GC content are not available

132 for a given gene, two strategies can be adopted to optimize protein translation: 1) the introduction of
133 accessory tRNA genes to complement the tRNA set of the recipient strain [41,42]; 2) the design of
134 synthetic genes with optimized codon usage, which is obtained by replacing rare codons with optimal
135 codons for the recombinant host without affecting the amino acid sequence of the gene product
136 [25,43-44].

137

138 ***Multiple gene expression: clusters, operons, multiple strains or engineered enzymes ?***

139 The ability of natural microorganisms to degrade plant biomass relies on multiple enzyme
140 systems. Similarly, engineering cellulolytic capabilities in a host implies cloning and expressing
141 multiple genes. In this perspective, two aspects need to be managed: i) the physical arrangement and
142 the coordination of the regulation of such multiple genes (*i.e.* the construction of operons and/or
143 clusters); ii) the carrying capacity of the recipient strain: the higher the number of the required genes,
144 the harder it is to introduce and maintain such large sized heterologous DNA [45].

145 As far as the gene arrangement is concerned, artificial operons are probably the most suitable
146 for industrial process requirements of simple and easily regulated protein systems [29,30]. However,
147 an optimal activity of cellulase systems is obtained for non-equimolar ratios of the different
148 components [10,20,21,27]. The simplest way to obtain non equimolar amounts of heterologous
149 proteins in the same strain is by using different transcriptional promoters [33,46].

150 Furthermore, natural cellulase systems are highly dynamic structures that are able to rapidly
151 adapt to environmental changes, *i.e.* substrate availability, by modifying the subunit composition of
152 the complex. Differential proteomic analysis has proven to be a valuable tool to directly detect
153 cellulase components that are biosynthesized in response to specific cellulosic materials [47-49]. The
154 use of promoters with different regulatory mechanisms and strengths could optimize both the quantity
155 of required subunits and complex composition flexibility.

156 An intriguing strategy to both prevent the cloning of large sized DNA fragments and to obtain flexible
157 enzyme systems has recently been explored in *B. subtilis* and *S. cerevisiae* [24,27]. Designer

158 cellulosomes were assembled by co-culturing recombinant cells expressing different single
159 cellulosomal components (*i.e.* intercellular complementation). Here, the amorphous (*i.e.* phosphoric
160 acid-swollen) cellulose-ethanol bioconversion rate and yield (93% of the maximum theoretical yield)
161 were optimized by adjusting the ratio of each *S. cerevisiae* population [27]. However, it still has to
162 be demonstrated that such a strategy could be manageable once scaled-up to the size of an industrial
163 process.

164 Nature offers a further paradigm to avoid multiple cellulase expression, *i.e.* the multidomain
165 multicatalytic megazymes from the *Caldicellulosiruptor spp.* thermophilic anaerobic gram-positive
166 bacteria [49-50] (Box 3). Such a protein arrangement inspired the design of unconventional and
167 covalent cellulosomes [51]. A panel of enzymes and complex architectures was engineered by
168 combining family 48 and 9 GH domains with efficient CBMs and optional cohesin and/or dockerin
169 modules from *C. cellulolyticum* [51]. A “covalent cellulosome”, consisting of both endoglucanase
170 and exoglucanase modules, two CBMs, a dockerin and a domain of unknown function, was twice
171 more active on crystalline cellulose than the parental free cellulases (Cel48F plus Cel9G). However,
172 this bifunctional protein was 36% less active than “conventional” designer cellulosomes containing
173 Cel48F plus Cel9G plus a miniscaffoldin [51]. Although these results somehow contradict the
174 improved synergy of the megazyme paradigm of *Caldicellulosiruptor spp.*, optimized artificial
175 covalent cellulosomes could probably be designed by increasing catalytic module mobility. The
176 catalytic domains in bi-functional megazymes from hyperthermophilic bacteria are always very
177 distant from each other in the primary sequence, *i.e.* they are separated by at least one carbohydrate
178 binding module (CBM), suggesting that high catalytic domain mobility is essential for efficient
179 substrate degradation [49-50].

180 Detailed understanding of cellulase catalytic mechanisms, with particular regard to
181 interdomain (*i.e.* CBM-catalytic domain interactions) and intermolecular (*i.e.* cellulase mixtures)
182 synergistic interactions in enhancing crystalline cellulose hydrolysis, is essential to engineer enzymes
183 with superior activity on native substrates [52]. Efficient recombinant cellulolytic organisms could

184 be developed by introducing fewer optimized enzymes. Improved enzymatic activity could also
185 compensate for low secretion yields (see next section). Both directed evolution and rational design
186 have been employed to improve cellulase activity on crystalline cellulose, although, so far, these
187 approaches have achieved only moderate success [52,53].

188

189 ***Heterologous protein secretion***

190 The heterologous expression of cellulases is often affected by the bias against their secretion
191 which causes a reduction in or loss of cell viability [28-30,54-55].

192 *E. coli* has been extensively used to express heterologous proteins, although such strategies have
193 mainly been addressed to cytosolic or periplasmic polypeptides [56-58]. Protein secretion in gram-
194 negative bacterial models actually deals with the challenge of translocation across a double membrane
195 system, although a number of secretion pathways (*e.g.*, types I, II, III, IV, V, and VI) have been
196 studied in detail [59]. However, a number of other bacterial models, especially gram positive bacteria
197 (*e.g. B. subtilis* and *L. lactis*), have been optimized for heterologous protein (*e.g.* proteases, α -
198 amylases) secretion [56,60].

199 Most secreted proteins are translocated across the cytosolic membrane by the Sec translocase
200 machinery through a general mechanism that is probably shared by both Gram negative and Gram
201 positive bacteria [for reviews see 57,59,60] (Box 4). The products of genes encoding cellulosomal
202 components of cellulolytic clostridia, including their original signal peptide, could be efficiently
203 secreted by *C. acetobutylicum* and *Lactobacillus plantarum* [26,29]. However, although the *B.*
204 *subtilis* and *E. coli* SecYEG complex subunits exhibit a high sequence similarity, they do not seem
205 to be functionally exchangeable: this indicates that secretory machines have species specificities [57].
206 Furthermore, additional components of the translocation machine (*e.g.* the *E. coli* SecDF/YajC and
207 YidC proteins) are continuously being identified, as well as paralogues of SecA, which are probably
208 involved in the secretion of different protein subsets [57]. These specific factors can be limiting for
209 heterologous protein expression, as was probably the case in the expression of some *C. cellulolyticum*

210 cellulosomal genes in *C. acetobutylicum* [29]. Original cellulase signal peptides have been replaced
211 by signal peptides of efficiently secreted autologous proteins or synthetic sequences to improve
212 secretion efficiency and lower cell toxicity in recombinant hosts. The engineered *sacB* levansucrase
213 signal sequence and the Strep-Tactin octapeptide have been used to express *Clostridium*
214 *cellulovorans* cellulosome components in *B. subtilis* [24,46]. The signal peptide of Usp45, the main
215 secreted protein of *L. lactis*, has been extensively used for heterologous protein secretion in *L. lactis*,
216 e.g. the *C. thermocellum* scaffolding protein CipA [28,56]. Other peptide sequences, located between
217 the signal peptide and the mature protein sequence (propeptides), are essential to either keep the
218 nascent polypeptide in a competent conformation for translocation across the cell membrane or for
219 rapid post-translocation folding which increase secretion efficiency (Box 4) [28,56,60].

220 *C. cellulolyticum* cellulases, with respect to the possibility of being secreted by *C.*
221 *acetobutylicum*, can be divided into two distinct groups: i) enzymes with small catalytic modules (and
222 a dockerin), e.g. Cel5A, Cel8C and Cel9M, can be easily secreted in an active form; ii) more “bulky”
223 cellulases characterized by large catalytic modules (e.g. Cel48F), or possessing additional modules
224 (e.g. Cel9G and Cel9E), are toxic and have resulted in non viable clones [29]. As far as Cel48F is
225 concerned, the unsuitable secretion machinery of *C. acetobutylicum* has been proven to cause cell
226 toxicity, since the same protein could be synthesized in the *C. acetobutylicum* cytoplasm [29]. The
227 secretion of family 48 of cellulosomal glucan hydrolases therefore seems to require specific
228 components that are missing in *C. acetobutylicum* [29]. However, fusion of CBM3a and X2 domains
229 to the Cel48F/Cel9G catalytic module, prevented toxic effects and triggered enzyme secretion [61].

230 Several membrane and periplasmic proteases contribute to the quality control of secreted
231 proteins by removing misfolded or incompletely synthesized polypeptides [60]. Although these
232 systems are essential for high quality protein biosynthesis in natural organisms, they can be among
233 the major bottlenecks of heterolous protein expression. For this reason, *B. subtilis* WB800 and *L.*
234 *lactis* HtrA mutants, which are defective of 8 surface/extracellular proteases of *B. subtilis* and the
235 unique exported housekeeping protease HtrA of *L. lactis*, respectively, have been employed for the

236 efficient secretion of heterologous cellulases [24,28]. Given the high complexity and specificity of
237 the secretion machineries, it is currently difficult to foresee whether a given translocation complex is
238 adapted to secrete a protein of interest. In the case of inefficient protein secretion, the use of weaker
239 or inducible promoters or engineered host secretory system (*e.g.* chaperones, translocation
240 machinery, protein quality check) can diminish the toxic effects on cell growth [29,57].

241

242 *Cell surface anchoring*

243 The assembly and spatial organization of enzymes in naturally occurring cellulosomes
244 constitutes the base of their synergistic activity. Several aspects in cellulosome self-assembly remain
245 to be elucidated with the goal of improving biomass conversion using cellulosomes [62]. Synergistic
246 activity is further enhanced in cellulosomes that are anchored to the cell surface and thus form ternary
247 cellulose-enzyme-microbe (CEM) complexes. CEM complexes benefit from the limited escape of
248 hydrolysis products and enzymes, and minimal distance products must diffuse before the cellular
249 uptake occurs [28]. Furthermore, surface anchoring probably protects enzymes from proteases and
250 thermal degradation [28]. For all these reasons, the assembly of cell surface displayed designer
251 cellulosomes in recombinant microbes is highly desirable.

252 Surface display techniques have been developed for Gram-negative bacteria, with autodisplay
253 probably being the most efficacious technique [58,63]. As far as Gram-positive bacteria are
254 concerned, at least four mechanisms can be exploited for protein surface display either through
255 binding to the cell membrane, *via* transmembrane domains or by covalent linkages to long-chain fatty
256 acids (lipoproteins), or by anchoring to the cell wall through covalent (*via* sortase) or non-covalent
257 (*via* cell wall binding domains) interactions [59,64].

258 Some of these strategies have been recently exploited for the surface display of cellulase
259 components in recombinant microorganisms. Minicellulosomes have covalently been linked to the
260 cell wall of the yeast *S. cerevisiae* using the agglutinin/flocculin display system [27,33,65]. Such cell
261 wall proteins, *e.g.* α -agglutinin and cell wall protein 2, contain a glycosyl phosphatidylinositol (GPI)

262 signal motif and are covalently linked to the cell wall β 1-6 glucan. Miniscaffoldins have been fused
263 with either a GPI signal motif, in order to be covalently linked to the cell wall, or with the C-terminus
264 of the AGA2 protein, which is tethered to the yeast surface *via* non-covalent bonds with the (surface
265 covalently bound) α -agglutinin mating adhesion receptor [27,33,65]. Trifunctional minicellulosome
266 displaying *S. cerevisiae* cells were able to convert amorphous cellulose to ethanol with 62% of the
267 theoretical yield [33].

268 As far as bacteria are concerned, fragments of the scaffolding protein CipA of *C.*
269 *thermocellum* have functionally been displayed on the cell surface of *Lactococcus lactis* by fusing
270 them with the C-terminal anchor motif of the streptococcal M6 protein, a sortase substrate [28] (Box
271 4). Surface-anchored complexes were displayed with efficiencies approaching 10^4 complexes/cell,
272 although significant differences in efficiency were observed among the constructs, depending on their
273 structural characteristics (*i.e.* protein conformation and solubility, scaffold size, and the inclusion and
274 exclusion of non-cohesin modules) [28]. Similarly, engineered scaffoldins and cellulases from *C.*
275 *thermocellum* have covalently been anchored to the *B. subtilis* cell wall by fusing them with the C-
276 terminal sortase sorting signal of *S. aureus* fibronectin binding protein B [66].

277 A non-covalent surface display system for lactic acid bacteria has been developed by fusing a
278 target heterologous protein, *i.e.* the α -amylase, with the C-terminal cA peptidoglycan binding domain,
279 which shows high homology with LysM repeats, of the major autolysin AcmA from *Lactococcus*
280 *lactis* [67].

281

282 ***Post-translational modifications***

283 Signal peptides of cellulases and cellulosome components, as of other secreted proteins, are
284 generally cleaved by signal peptidases during or shortly after translocation across the cytoplasmic
285 membrane [60]. A further post-translational modification, *i.e.* O-glycosylation, of cellulosome
286 components has been reported in *C. thermocellum* and *Bacteroides cellulosolvens* (particularly on the
287 scaffoldin moiety) and hypothesized for the ScaC and the CipA scaffoldins of *Acetivibrio*

288 *cellulolyticus* and *C. acetobutylicum*, respectively [15,54,68-70]. The glycosyl groups may protect
289 the cellulosome against proteases, but may also play a role in cohesin-dockerin recognition and in
290 adhesion to the substrate [71].

291

292 ***Concluding remarks***

293 As far as recombinant cellulolytic strategies are concerned, efficient secretion of designer
294 cellulase systems is still among the most challenging tasks. The high complexity and diversity of
295 protein secretion mechanisms is far to be fully understood. Currently, we cannot predict if a specific
296 cellulase will be secreted in high amounts in a recipient strain or it will result in cell toxicity.

297 However, remarkable progress is continuously being made and recombinant microorganisms
298 that could directly ferment cellulosic substrates to ethanol have recently been reported [31,33]. Even
299 in such cutting edge studies, amorphous, either carboxy-methylated or phosphoric acid-swollen,
300 cellulose was used, while crystalline cellulose could not be metabolized with significant efficiencies
301 by engineered strains. There are still some major gaps in our understanding of the mechanisms by
302 which cellulase systems catalyze crystalline cellulose hydrolysis [52,53]. Synergistic interactions
303 between CBM and catalytic domain and in cellulase mixtures likely play a key role for efficient native
304 plant biomass degradation, but detailed molecular mechanisms need to be clarified. This information
305 is crucial for designing improved enzymes and artificial complexes for biotechnological applications,
306 with particular regard to recombinant strains that are intended for CBP.

307 Researches on natural microorganisms, whose metabolism has been shaped by evolution for
308 cellulolytic lifestyle, indicate that cellulose depolymerization by cellulases is not the only bottleneck
309 of cellulose metabolism [72]. From a metabolic standpoint cellulose cannot be considered as a simple
310 sum of soluble carbohydrate units. Experimental evidences clearly show that the use of cellulose does
311 not result in the same metabolism as soluble sugars, *e.g.* cellobiose. Metabolic flux analysis could be
312 an essential tool to further improve recombinant cellulolytic strains by rational engineering of central
313 metabolic pathways. *In vivo* directed evolution by continuous culture under selective pressure is a

314 very promising alternative approach to optimize cellulose overall metabolism in engineered
315 microorganisms [72,73].

316

317 ***Acknowledgments***

318 We are would like to thank the “Action B - Return of Italian Researches from abroad”
319 convention between the University of Torino and the Piedmont Region for the financial support to
320 our researches.

321

322

323 ***Box 1. Cellulosomes: nanomachines for efficient cellulose degradation***

324 Anaerobic cellulolytic bacteria (*e.g.* belonging to *Clostridium* and *Ruminococcus* genera) and
325 fungi (*i.e.* Chytridomycetes), biosynthesize “complexed” cellulase systems called “cellulosomes”
326 [15-17,19]. In *C. thermocellum*, *C. cellulovorans* and *R. flavifaciens*, it has been demonstrated that
327 these complexes are bound to the cell surface [16]. However, this does not seem to be true for other
328 cellulosome biosynthesizing microorganisms [16].

329 The cellulosome architecture consists of multiple enzyme subunits, with different substrate
330 specificities (*e.g.* cellulases, xylanases, pectinases) and catalytic mechanisms, organized by
331 scaffolding proteins [10,16-17] (Figure 2a). As far as catalytic mechanisms are concerned, glucan
332 hydrolases (GH) can be divided into four classes: 1) endoglucanases, which cut at random internal
333 sites of the polysaccharides and generate oligosaccharides of various lengths; 2) exoglucanases,
334 which act in a processive manner on the reducing or non-reducing ends of polysaccharide chains,
335 liberating either mono- or di-saccharides; 3) processive endoglucanases, that share properties of both
336 endo- and exo-glucanases; 4) β -glucosidases, which hydrolyze soluble di/oligo saccharides to
337 monosaccharides [15, 74]. Cellulosome GHs, apart from the catalytic module, always contain at least
338 one supplementary domain, *i.e.* the dockerin module involved in enzyme interaction with the
339 scaffolding proteins. Furthermore, single or multiple carbohydrate-binding modules (CBM) can be

340 attached to the N or C terminus of catalytic domains through flexible linker-rich regions. CBMs affect
341 polysaccharide binding and hydrolysis, by bringing the catalytic domain into close proximity with
342 the substrate and are particularly important for the initiation and processivity of exoglucanases [15].
343 Other additional modules, such as immunoglobulin-like domains (*e.g.*, for CelE of *C. cellulolyticum*),
344 or fibronectin type III domains (*e.g.* in CbhA of *C. thermocellum*) can be found in GH [15].

345 “Scaffoldins” are large multidomain, multifunctional proteins deputed to: i) recruit catalytic
346 proteins by means of multiple cohesin domains that interact with glucan-hydrolase dockerin domains;
347 ii) improve complex affinity for the substrate and catalytic efficiency *via* carbohydrate binding
348 domains (CBMs). Anchoring scaffoldins provide further function by binding the cellulosome to the
349 cell wall through covalent (sortase mediated) or non-covalent (through surface layer homology
350 domains) interactions [16] (Figure 2a). Generally, scaffoldins do not contain catalytic modules but an
351 exception is ScaA from *Acetivibrio cellulolyticus* that includes a GH9 domain [16].

352 Such a complex architecture enables the enzymatic components to act in a synergistic and coordinated
353 manner *via* intra- and inter-molecular interactions and makes the cellulosomes the most efficient
354 biochemical systems for cellulose degradation [10,61].

355

356 ***Box 2. Cellulosomal genes are activated by alternative σ factors and anti- σ factor borne CBMs in***
357 ***Clostridium thermocellum***

358 The mechanisms by which cellulase gene expression is regulated have long remained an enigma. A
359 set of six putative operons encoding alternative σ factors (homologues to *B. subtilis* σ I) and their
360 cognate membrane-associated anti- σ factors has recently been identified in the *Clostridium*
361 *thermocellum* genome [22]. These proteins likely play essential roles in regulating cellulosomal gene
362 expression in this bacterial strain (Figure 2b, c). Such anti- σ I factors are multimodular proteins that
363 include a strongly predicted transmembrane helix, an intracellular anti- σ domain, and an extracellular
364 module with polysaccharide-related functions, *i.e.* either a CBM, a sugar-binding element, *e.g.* PA14,
365 or a glycoside hydrolase family 10 (GH10) module. Apart from such structural heterogeneity, a

366 unique extracellular carbohydrate sensing mechanism emerges: the presence of extracellular
367 polysaccharides is detected by a corresponding anti- σ factor-borne CBM, GH or PA element (Figure
368 2c). This event triggers conformational changes in the intracellular domain of the anti- σ I factor: this
369 releases the alternative σ factor and enables it to interact with RNA polymerase and promote the
370 transcription of selected cellulosomal genes [22].

371 A similar set of multiple σ I and anti- σ I factors has recently been discovered in another
372 cellulosome-producing bacterium, *Acetivibrio cellulolyticus* CD2 (also belonging to Clostridia) and
373 in the Gram-negative human gut bacterium *Bacteroides thetaiotaomicron* [22]. Apart from these
374 studies, very few information about the molecular mechanisms that modulate cellulosomal gene
375 expression is currently available. Only further researches on other bacterial models will be able to
376 establish if common systems have been evolved by cellulolytic bacteria or if *species* specific solutions
377 are prevalent.

378

379 ***Box 3. Multifunctional megazymes from Caldicellulosiruptor spp.: paradigms to engineer new***
380 ***designer cellulosomes with improved efficiency.***

381 The engineering of cellulolytic capabilities in a heterologous host implies cloning and expressing
382 multiple genes: this constitutes one of major obstacles to the development of efficient recombinant
383 cellulolytic microorganisms. *Caldicellulosiruptor spp.* hyperthermophilic anaerobic gram-positive
384 bacteria have bypassed multiple cellulase expression by synthesizing multidomain multicatalytic
385 megazymes [49,50]. Unlike clostridial cellulosomes, which consist of multiple enzymes containing
386 single glucan hydrolase (GH) domains, *Caldicellulosiruptor spp.* cellulase systems consist of large
387 amounts of a few bifunctional glucan hydrolases with broad substrate specificities [49]. These latter
388 enzymes consist of different permutations of a small set of catalytic modules (*i.e.* GH5, GH9, GH10,
389 GH43, GH44, GH48, and GH74), together with highly conserved family 3 carbohydrate binding
390 modules (CBM3), in a single polypeptide chain [49]. The most abundant enzymes in
391 *Caldicellulosiruptor* supernatants are the bifunctional proteins Athe_1867 (COB47_1673), which

392 consists of a GH9 domain (encoding a endo- 1,4-D-glucanase activity), three CBM3 domains, and a
393 GH48 domain (encoding a processive exoglucanase activity), CelC-ManB (Athe_1865
394 COB47_1669), consisting of a GH9 domain, three CBM3 domains, and a GH5 domain (encoding a
395 mannanase activity), Athe_1857, containing GH10 (likely coding for an endo-1,4-D-xylanase
396 activity) and GH48 domains, and COB47_1671, where the GH10 domain is associated with another
397 GH5 module [49]. Interestingly, when expressed separately, the GH5 and GH10 domains both
398 independently exhibit the same broad substrate specificity, but at decreased hydrolysis rates. Mixing
399 the single enzymes did not completely restore the activity of the full-length version, thus
400 demonstrating the synergistic effects of multidomain proteins [49].

401 Such arrangements suggest an evolution *via* domain shuffling and they could also be interpreted as
402 primitive alternatives to operons [15]. It is possible that the multidomain architecture of
403 *Caldicellulosiruptor* enzymes is an adaptation to high-temperature environments that is characterized
404 by increased enzyme/substrate diffusion rates. This arrangement actually provides an improved
405 synergistic effect due to a closer intramolecular spatial proximity in hyperthermophilic environments
406 that would likely prevent subunit assembly by cohesin-dockerin interactions. Furthermore, multiple
407 CBMs enable stronger binding to the substrate[49]. Such an architecture could inspire protein
408 engineers and lead to advantages associated with designer cellulosomes in recombinant
409 microorganisms through an improved synergism between different catalytic domains.

410

411 ***Box 4. Cellulosomes: from translation into the cytoplasm to surface display***

412 A general mechanism for protein translocation across the cytosolic membrane, which is mainly based
413 on the Sec translocase machinery, is probably shared by both Gram negative and Gram positive
414 bacteria (for reviews see [57,59,60]). The signal peptide of nascent proteins is bound by cytoplasmic
415 ribozymes (homologues to the signal recognition particle, SRP) and transferred to the SecYEG
416 complex *via* membrane-bound SRP receptors. General molecular chaperones (*e.g.* GroEL/GroES and
417 DnaK/DnaJ which are also involved in cytoplasmic protein folding) maintain the nascent polypeptide

418 chain in an “unfolded” translocation-competent conformation and prevent protein aggregation. Other
419 chaperones with more dedicated roles in the secretion of specific proteins (*i.e. B. subtilis* CsaA and
420 ClpX) have also been identified [57]. Polypeptide translocation occurs through the aqueous
421 transmembrane channel that is formed by the integral membrane SecYEG complex and is driven by
422 ATP hydrolysis catalyzed by the peripheral motor domain SecA. The polypeptides that emerge from
423 the Sec translocase are unfolded. It has recently become clear that the rate at which proteins are post-
424 translocationally folded by pro-peptides, peptidyl-prolyl cis/trans isomerases, disulfide isomerases,
425 and metal ions is a key element of their productivity [75]. Class I propeptides are essential for the
426 rapid post-translocational folding of their cognate mature protein, while class II propeptides appear
427 to decrease the rate of intracellular folding, thereby facilitating interactions with chaperones that
428 maintain secretion competence.

429 Cellulosomal complexes can then be anchored to the bacterial cell surface (mainly through non-
430 catalytic scaffolding proteins) by either non-covalent, *via* surface layer homology (SLH) domains, or
431 sortase-catalyzed covalent interactions [16]. Sortases are widely distributed in Gram positive bacteria
432 and recognize proteins that contain a sortase recognition motif (*e.g.* LPXTG, where X is any amino
433 acid). The target protein C-terminal domain usually includes a positively charged tail, a hydrophobic
434 region, which is inserted into the cytosolic membrane, and an LPXTG motif. Cytoplasmic membrane
435 anchored sortases cleave the peptidic bond between T and G of the LPXTG motif and then transfer
436 the N-terminal part of the precursor surface proteins to lipid II, a cell wall precursor that is
437 subsequently incorporated in the peptidoglycan [76].

438

439 **References**

- 440 1. Octave, S. and Thomas, D. (2009) Biorefinery: toward an industrial metabolism. *Biochimie*
441 91, 659-964.
- 442 2. Fitzpatrick, M. *et al.* (2010) A biorefinery processing perspective: treatment of lignocellulosic
443 materials for the production of value-added products. *Bioresour. Technol.* 101, 8915-8922.

- 444 3. Tan, T. *et al.* (2010) Current development of biorefinery in China. *Biotechnol. Adv.* 28, 543-
445 555.
- 446 4. Lynd, L.R. *et al.* (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Curr.*
447 *Opin. Biotechnol.* 16, 577-583.
- 448 5. Adrio, J.L. and Demain, A.L. (2010) Recombinant organisms for production of industrial
449 products. *Bioengineered Bugs* 1, 116-131.
- 450 6. La Grange, D.C. *et al.* (2010) Engineering cellulolytic ability into bioprocessing organisms.
451 *Appl. Microbiol. Biotechnol.* 87, 1195-1208.
- 452 7. Adsul, M.G. *et al.* (2011) Development of biocatalysts for production of commodity
453 chemicals from lignocellulosic biomass. *Bioresour. Technol.* 102, 4304-4312.
- 454 8. Maki, M. *et al.* (2009) The prospects of cellulase-producing bacteria for the bioconversion
455 of lignocellulosic biomass. *Int. J. Biol. Sci.* 5, 500-516.
- 456 9. Elkins, J.G. *et al.* (2010) Engineered microbial systems for enhanced conversion of
457 lignocellulosic biomass. *Curr. Opin. Biotechnol.* 21, 657-662.
- 458 10. Morais, S. *et al.* (2010) Cellulase-xylanase synergy in designer cellulosomes for enhanced
459 degradation of a complex cellulosic substrate. *MBio.* 1, pii: e00285-10. doi:
460 10.1128/mBio.00285-10
- 461 11. Dressaire, C. *et al.* (2009) Transcriptome and proteome exploration to model translation
462 efficiency and protein stability in *Lactococcus lactis*. *PLoS Comput. Biol.* 5, e1000606.
- 463 12. Picard, F. *et al.* (2009) Examination of post-transcriptional regulations in prokaryotes by
464 integrative biology. *C. R. Biol.* 332, 958-973.
- 465 13. Fredrick, K. and Ibba, M. (2010) How the sequence of a gene can tune its translation. *Cell*
466 141, 227-229.
- 467 14. Tyo, K.E., *et al.* (2010) Toward design-based engineering of industrial microbes. *Curr. Opin.*
468 *Microbiol.* 13, 255-262.

- 469 15. Lynd, L.R. (2002) Microbial cellulose utilization: fundamentals and biotechnology.
470 *Microbiol. Mol. Biol. Rev.* 66, 506-577.
- 471 16. Bayer, E.A. *et al.* (2008) From cellulosomes to cellulosomes. *Chem. Rec.* 8, 364-377.
- 472 17. Doi, R.H. (2008) Cellulases of mesophilic microorganisms: cellulosome and noncellulosome
473 producers. *Ann. N. Y. Acad. Sci.* 1125, 267-279.
- 474 18. Dashtban, M. *et al.* (2009) Fungal bioconversion of lignocellulosic residues; opportunities &
475 perspectives. *Int. J. Biol. Sci.* 5, 578-595.
- 476 19. Rincon, M.T. *et al.* (2010) Abundance and diversity of dockerin-containing proteins in the
477 fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS One* 5, e12476.
- 478 20. Maamar, H. *et al.* (2006) Transcriptional analysis of the *cip-cel* gene cluster from *Clostridium*
479 *cellulolyticum*. *J. Bacteriol.* 188, 2614-2624.
- 480 21. Newcomb, M. *et al.* (2011) Co-transcription of the *celC* gene cluster in *Clostridium*
481 *thermocellum*. *Appl. Microbiol. Biotechnol.* 90, 625-634.
- 482 22. Nataf, Y. *et al.* (2010) *Clostridium thermocellum* cellulosomal genes are regulated by
483 extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl. Acad. Sci. U S A.*
484 107: 18646-18651.
- 485 23. De Mey, M. *et al.* (2010) Promoter knock-in: a novel rational method for the fine tuning of
486 genes. *BMC Biotechnol.* 10, 26.
- 487 24. Arai, T. *et al.* (2007) Synthesis of *Clostridium cellulovorans* minicellulosomes by intercellular
488 complementation. *Proc. Natl. Acad. Sci. U S A.* 104, 1456-1460.
- 489 25. Linger, J.G. *et al.*, (2010) Heterologous expression and extracellular secretion of cellulolytic
490 enzymes by *Zymomonas mobilis*. *Appl. Environ. Microbiol.* 76, 6360-6369.
- 491 26. Okano, K. *et al.* (2010) D-lactic acid production from cellooligosaccharides and β -glucan
492 using L-LDH gene-deficient and endoglucanase-secreting *Lactobacillus plantarum*. *Appl.*
493 *Microbiol. Biotechnol.* 85, 643-650.

- 494 27. Tsai, S.L *et al.* (2010) Surface display of a functional minicellulosome by intracellular
495 complementation using a synthetic yeast consortium and its application to cellulose hydrolysis
496 and ethanol production. *Appl. Environ. Microbiol.* 76, 7514-7520.
- 497 28. Wieczorek, A.S. and Martin, V.J. (2010) Engineering the cell surface display of cohesins for
498 assembly of cellulosome-inspired enzyme complexes on *Lactococcus lactis*. *Microb. Cell.*
499 *Fact.* 9, 69.
- 500 29. Mingardon, F. *et al.* (2011). The issue of secretion in heterologous expression of *Clostridium*
501 *cellulolyticum* cellulase-encoding genes in *Clostridium acetobutylicum* ATCC 824. *Appl.*
502 *Environ. Microbiol.* 77, 2831-2838.
- 503 30. Mingardon, F. *et al.* (2005) Heterologous production, assembly, and secretion of a
504 minicellulosome by *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 71,
505 1215-1222.
- 506 31. Jeon, E. *et al.* (2009) Cellulosic alcoholic fermentation using recombinant *Saccharomyces*
507 *cerevisiae* engineered for the production of *Clostridium cellulovorans* endoglucanase and
508 *Saccharomycopsis fibuligera* β -glucosidase. *FEMS Microbiol. Lett.* 301, 130-136.
- 509 32. Ozkose, E. *et al.* (2009) Expression of fungal cellulase gene in *Lactococcus lactis* to construct
510 novel recombinant silage inoculants. *Folia Microbiol. (Praha)* 54, 335-342.
- 511 33. Wen, F. *et al.* (2010) Yeast surface display of trifunctional minicellulosomes for simultaneous
512 saccharification and fermentation of cellulose to ethanol. *Appl. Environ. Microbiol.* 76, 1251-
513 1260.
- 514 34. Lu, T.K. *et al.* (2009) Next-generation synthetic gene networks. *Nat. Biotechnol.* 27, 1139-
515 1150.
- 516 35. Narita, J. *et al.* (2006) Improvement of protein production in lactic acid bacteria using 5'-
517 untranslated leader sequence of *slpA* from *Lactobacillus acidophilus*. Improvement in protein
518 production using UTLS. *Appl. Microbiol. Biotechnol.* 73, 366-373.

- 519 36. Yeh, C.M. *et al.* (2008) Extracellular expression of a functional recombinant *Ganoderma*
520 *lucidum* immunomodulatory protein by *Bacillus subtilis* and *Lactococcus lactis*. *Appl.*
521 *Environ. Microbiol.* 74, 1039-1049.
- 522 37. Daguer, J.P. *et al.* (2005) Increasing the stability of *sacB* transcript improves levansucrase
523 production in *Bacillus subtilis*. *Lett. Appl. Microbiol.* 41, 221–226.
- 524 38. Komarova, A.V. *et al.* (2005) AU-rich sequences within 5' untranslated leaders enhance
525 translation and stabilize mRNA in *Escherichia coli*. *J. Bacteriol.* 187, 1344–1349.
- 526 39. Lightfield, J. *et al.* (2011) Across bacterial phyla, distantly-related genomes with similar
527 genomic GC content have similar patterns of amino acid usage. *PLoS One* 6, e17677.
- 528 40. Talarico, L.A. *et al.* (2005) Construction and expression of an ethanol production operon in
529 Gram-positive bacteria. *Microbiology.* 151, 4023-4031.
- 530 41. Talarico, L.A. *et al.* (2001). Production of the Gram-positive *Sarcina ventriculi* pyruvate
531 decarboxylase in *Escherichia coli*. *Microbiology* 147, 2425–2435.
- 532 42. Raj, K. C. *et al.* (2002). Cloning and characterization of the *Zymobacter palmae* pyruvate
533 decarboxylase (*pdc*): comparison to bacterial homologues. *Appl. Environ. Microbiol.* 68,
534 2869–2876.
- 535 43. Du Plessis, L. *et al.* (2010) Exploring improved endoglucanase expression in *Saccharomyces*
536 *cerevisiae* strains. *Appl. Microbiol. Biotechnol.* 86, 1503-1511.
- 537 44. Angov, E. *et al.* (2011) Adjustment of codon usage frequencies by codon harmonization
538 improves protein expression and folding. *Methods Mol. Biol.* 705, 1-13.
- 539 45. Camps, M. (2010) Modulation of ColE1-like plasmid replication for recombinant gene
540 expression. *Recent Pat. DNA Gene Seq.* 4, 58–73.
- 541 46. Cho, H.Y *et al.* (2004) Production of minicellulosomes from *Clostridium cellulovorans* in
542 *Bacillus subtilis* WB800. *Appl. Environ. Microbiol.* 70, 5704-5707.

- 543 47. Blouzard, J.C. *et al.* (2010) Modulation of cellulosome composition in *Clostridium*
544 *cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and
545 carbohydrate-active enzyme analyses. *Proteomics* 10, 541-554.
- 546 48. Cho, W. *et al.* (2010) Cellulosomic profiling produced by *Clostridium cellulovorans* during
547 growth on different carbon sources explored by the cohesin marker. *J. Biotechnol.* 145, 233-
548 239.
- 549 49. Lochner, A. *et al.* (2011) Label-free quantitative proteomics distinguish the secreted
550 cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*. *Appl.*
551 *Environ. Microbiol.* 77, 4042-4054.
- 552 50. Vanfossen, A.L. *et al.* (2011) Glycoside hydrolase inventory drives plant polysaccharide
553 deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor*
554 *saccharolyticus*. *Biotechnol. Bioeng.* 108, 1559-1569.
- 555 51. Mingardon, F. *et al.* (2007) Exploration of new geometries in cellulosome-like chimeras.
556 *Appl. Environ. Microbiol.* 73, 7138-7149.
- 557 52. Wilson, D.B. (2009) Cellulases and biofuels. *Curr. Opin. Biotechnol.* 20, 295-299.
- 558 53. Bayer, E.A. *et al.* (2007) The potential of cellulases and cellulosomes for cellulosic waste
559 management. *Curr. Opin. Biotechnol.* 18, 237-245.
- 560 54. Sabathé, F. and Soucaille, P. (2003) Characterization of the CipA scaffolding protein and in
561 vivo production of a minicellulosome in *Clostridium acetobutylicum*. *J. Bacteriol.* 185, 1092-
562 1096.
- 563 55. Perret, S. *et al.* (2004) Production of heterologous and chimeric scaffoldins by *Clostridium*
564 *acetobutylicum* ATCC 824. *J. Bacteriol.* 186, 253-257.
- 565 56. Morello, E. *et al.* (2008) *Lactococcus lactis*, an efficient cell factory for recombinant protein
566 production and secretion. *J. Mol. Microbiol. Biotechnol.* 14, 48-58.
- 567 57. Zweers, J.C. *et al.* (2008) Towards the development of *Bacillus subtilis* as a cell factory for
568 membrane proteins and protein complexes. *Microb. Cell Fact.* 7, 10.

- 569 58. Jong, W.S. *et al.* (2010) Extracellular production of recombinant proteins using bacterial
570 autotransporters. *Curr. Opin. Biotechnol.* 21, 646-652.
- 571 59. Desvaux., M. *et al.* (2009) Secretion and subcellular localizations of bacterial proteins: a
572 semantic awareness issue. *Trends Microbiol.* 17, 139-145.
- 573 60. Pohl, S., and Harwood, C.R. (2010) Heterologous protein secretion by *Bacillus* species from
574 the cradle to the grave. *Adv. Appl. Microbiol.* 73, 1-25.
- 575 61. Chanal A., *et al.* (2011) Scaffoldin modules can serve as "cargo" domains to promote the
576 secretion of heterologous cellulosomal cellulases by *Clostridium acetobutylicum*. *Appl.*
577 *Environ. Microbiol.* 2011 Jul 15. [Epub ahead of print] doi:10.1128/AEM.00758-11
- 578 62. Bomble, Y.J. *et al.* (2011) Modeling the self-assembly of the cellulosome enzyme complex.
579 *J. Biol. Chem.* 286, 5614-5623.
- 580 63. Jose, J. and Meyer, F. (2007) The autodisplay story, from discovery to biotechnical and
581 biomedical applications. *Microbiol. Mol. Biol. Rev.* 71, 600–619
- 582 64. Desvaux, M. *et al.* (2006) Protein cell surface display in Gram-positive bacteria: from single
583 protein to macromolecular protein structure. *FEMS Microbiol. Lett.* 256, 1–15.
- 584 65. Lilly, M. *et al.* (2009) Heterologous expression of a *Clostridium* minicellulosome in
585 *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9, 1236-1249.
- 586 66. Anderson, T.D. *et al.* (2011) Assembly of minicellulosomes on the surface of *Bacillus subtilis*.
587 *Appl. Environ. Microbiol.* 77, 4849-4858.
- 588 67. Okano, K. *et al.* (2008) System using tandem repeats of the cA peptidoglycan-binding domain
589 from *Lactococcus lactis* for display of both N- and C-terminal fusions on cell surfaces of lactic
590 acid bacteria. *Appl. Environ. Microbiol.* 74, 1117-1123.
- 591 68. Xu, Q., *et al.* (2003) The cellulosome system of *Acetivibrio cellulolyticus* includes a novel
592 type of adaptor protein and a cell surface anchoring protein. *J. Bacteriol.* 185, 4548-4557.
- 593 69. Xu, Q. *et al.* (2004) Architecture of the *Bacteroides cellulosolvens* cellulosome: description
594 of a cell surface-anchoring scaffoldin and a family 48 cellulase. *J. Bacteriol.* 186, 968-977.

- 595 70. Demain, A.L. *et al.* (2005) Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69,
596 124-154.
- 597 71. Weimer, P.J. *et al.* (2006) Studies of the extracellular glycocalyx of the anaerobic cellulolytic
598 bacterium *Ruminococcus albus*. *Appl. Environ. Microbiol.* 72, 7559-7566.
- 599 72. Desvaux, M. (2006) Unravelling carbon metabolism in anaerobic cellulolytic bacteria.
600 *Biotechnol. Prog.* 22,1229-1238.
- 601 73. Meynial Salles, I. *et al.* (2007) Evolution of a *Saccharomyces cerevisiae* metabolic pathway
602 in *Escherichia coli*. *Metab. Eng.* 9, 152-159.
- 603 74. Wilson, D.B. (2011) Microbial diversity of cellulose hydrolysis. *Curr. Opin. Microbiol.* 14,
604 259-263.
- 605 75. Harwood, C. R., and Cranenburgh, R. (2008). *Bacillus* protein secretion: an unfolding story.
606 *Trends Microbiol.* 16, 73–79.
- 607 76. Hendrickx, A.P. *et al.* (2011) Architects at the bacterial surface - sortases and the assembly of
608 pili with isopeptide bonds. *Nat. Rev. Microbiol.* 9, 166-176.
- 609

610 **Figure legends**

611

612 **Figure 1. The efficient expression of heterologous cellulase genes is a key feature to**
613 **engineer performant recombinant cellulolytic microorganisms.** Recombinant cellulolytic
614 strategies deal with the problem of biosynthesizing and secreting sufficient amounts of
615 heterologous designer cellulase systems for efficient cellulose degradation. The choice of
616 suitable transcriptional promoters and the improvement of mRNA stability and translation
617 efficiency are essential to optimize gene expression. Furthermore, suitable strategies should
618 be adopted to coordinate the expression of the multiple genes required. Nascent proteins need
619 to be maintained in an unfolded conformation so as to be translocated across the cytoplasmic
620 membrane. After translocation, proteins undergo further modifications that include folding,
621 surface anchoring and, possibly, glycosylation. Genetic stability is a further essential
622 requirement for engineered strains that are intended to industrial applications.

623

624 **Figure 2. Simplistic model of a cellulosome that includes only one anchoring scaffoldin**
625 **(a) and proposed mechanism for the cellulosomal gene transcription activation in**
626 ***Clostridium thermocellum* (b, c adapted from [22]).** a) The scaffolding protein (blue) binds
627 the enzymatic components through cohesin-dockerin interactions, enhances the cellulosome
628 affinity for cellulose through the carbohydrate binding modules (CBM), and anchors the
629 cellulosome complex to the cell surface through either non-covalent (by means of multiple S-
630 layer homology domains) or covalent (mediated by sortases) bonds. Apart from the catalytic
631 domains, cellulosomal enzymes include dockerin modules and, possibly, additional domains
632 (e.g. CBM, SLH). b) Extracellular polysaccharides are sensed by a system that consists of
633 alternative σ factors and integral membrane anti- σ factors (pink). The latter proteins include
634 an extracellular carbohydrate binding domain (CBM), a transmembrane helix and an
635 intracellular anti- σ module. c) When the extracellular carbohydrate binding domain interacts

636 with polysaccharides (*e.g.* cellulose), it induces a conformational change in the intracellular
637 anti- σ domain that releases the alternative σ factor. The latter is then able to bind specific
638 promoters (*p*) and trigger cellulosomal gene transcription by RNA polymerase (RNAPol).
639

Figure 1



