Recombinant hyperthermophilic enzyme expression in plants: a novel approach for lignocellulose digestion

Bilal Ahmad Mir[1,2], Ritesh Mewalal[2], Eshchar Mizrachi[2], Alexander A. Myburg[2], and Don A. Cowan[1]

[1] Center for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Private bag X20, Pretoria 0028, South Africa
[2] Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private bag X20, Pretoria 0028, South Africa

Plant biomass, as an abundant renewable carbon source, is a promising alternative to fossil fuels. However, the enzymes most commonly used for depolymerization of lignocellulosic biomass are expensive, and the development of cost-effective alternative conversion technolo-gies would be desirable. One possible option is the heterologous expression of genes encoding lignocellulose-digesting enzymes in plant tissues. To overcome simultaneously issues of toxicity and incompatibility with high-temperature steam explosion processes, the use of heterologous genes encoding hyperthermophilic enzymes may be an attractive alternative. This approach could reduce the need for exogenous enzyme additions prior to fermentation, reducing the cost of the complete processing operation. This review highlights recent advances and future prospects for using hyperthermophilic enzymes in the biofuels industry.

Hurdles to making use of lignocellusic biomass

Growing concerns over increasing energy consumption, global warming and depleting fossil fuel reserves have led to resurgence in the development of alternative and renewable energy sources that can displace fossil transportation fuel and other petroleum-based products. In response, many countries including the USA and Brazil, the two leaders of the global biofuel industries, have initiated extensive research programs on the development of economically and environmentally sustainable biofuels industries. First-generation biofuels (see Glossary) derived from starch and sugar-based raw materials currently provide most of the bioethanol (3.0% of global transportation fuel) in the world [1], but meet only a limited fraction of global fuel requirement and compete with animal and human food markets. Consequently, there is an increased interest in alternative (second-generation) energy sources that can contribute significantly to the future global energy supply without competing with increasing food demand. Lignocellulosic plant biomass is an attractive alternative

Corresponding author: Cowan, D.A. (don.cowan@up.ac.za).
Keywords: lignocellulosic biomass; hyperthermophilic enzymes; heterologous expres-sion; second-generation biofuels.

Glossary

Biomass: biomass is biological material derived from living or recently living organisms. In the context of biomass for energy this term more often refers to plant-based material, but biomass can equally apply to both animal- and vegetable-derived material.
Carbon neutral: removal of as much CO_2 from the atmosphere by a particular activity as it emits into it. Lignocellulosic biomass is the most abundant carbon neutral compound.
Codon harmonization: an algorithm that best approximates codon usage frequencies from the native host and adjusts these for use in the heterologous system in order to improve the heterologous protein expression.
Codon optimization: systematic alteration of codons in recombinant DNA to be expressed in a heterologous system to match the pattern of codon usage in the organism used for expression without changing the amino acids of the synthesized protein. The intention is to enhance yields of the expressed protein by increasing the translational efficiency of gene of interest.
Consolidated bioprocessing (CBP): CBP of lignocellulose to bioethanol refers to the combination of two or more of the four processes required for the conversion (production of saccharolytic enzymes, hydrolysis of the polysaccharides present in pretreated biomass, fermentation of hexose sugars, and fermentation of pentose sugars) in one reactor. CBP is gaining increasing recognition as a potential alternative to the use of sequential, independent unit operations. **First-generation biofuels:** first-generation biofuels are those directly derived from biomass that is generally edible (produced primarily from food crops such as grains, sugar beet, and oil seeds), and which contains readily hydrolysable sugar polymers (starch from maize) or directly fermentable sugars (sucrose from sugar beet).
Lignocellulose: lignocellulose is a generic term for describing the main constituents in most plants, namely cellulose, hemicelluloses, and lignin. Lignocellulose is a complex matrix, comprising many different polysaccharides, phenolic polymers, and proteins. Cellulose, the major component of cell walls of land plants, is a glucan polysaccharide.
Mesophilic enzymes: mesophilic enzymes are defined as those isolated and/or cloned from mesophilic microorganisms (i.e., organisms growing optimally between 20 and 37°C, and maximally at below 50°C).
Second-generation biofuels: second-generation biofuels are defined as fuels produced from a wide array of different feedstocks that are primarily composed of lignocellulosic materials. The lignocellulosic biomass may include materials such as agricultural residues (corn stover, crop straws, and bagasse), herbaceous crops (alfalfa and switch grass), short rotation woody crops, forestry residues, and waste paper and other wastes (municipal and industrial).
Thermotolerant enzymes: enzymes, from whatever source, which show a higher degree of thermostability than homologous mesophilic enzymes.
Transit peptides: transit peptides are responsible for the transport of a protein encoded by a nuclear gene to a particular organelle. The commonly used transit (signal) peptides include tobacco pathogenesis related protein 1a (Pr1a), Soybean variant-specific protein (VSPβ), Rubisco activase (RA), light harvesting chlorophyll *alb*-binding protein (CAB), Rubisco small subunit (RS), KDEL, LPH, tobacco calreticulin signal peptide (CAL), as shown in Table S1 in the supplementary material online.

for the large-scale production of economically and environmentally sustainable biofuels [2–4]. The main component of plant biomass used as second-generation feedstock is the plant cell wall, which constitutes 70–80% of lignocellulosic polymers. Lignocellulose feedstock can be obtained from wood, perennial grasses, agricultural residue, forestry wastes, and residues from biorefineries and pulp mills. It is the most abundant renewable and carbon-neutral source on earth with >100 billion tons produced annually as a waste product [5,6].

The plant secondary cell wall, the main source of lignocellulosic biomass, is a heterogeneous complex of different polymers (predominantly cellulose, hemicellulose, and lignin). The crystallinity and surface area accessibility of cellulose, protection of cellulose by lignin, the heterogeneous character of biomass particles, and cellulose sheathing by hemicellulose all contribute to the recalcitrance of lignocellulosic biomass to hydrolysis, and collectively represent a technical barrier to commercial deployment of cellulosic biofuel [7–13]. Typically, lignocellulosic biomass is preprocessed (Figure 1A) by steam explosion, solvent treatment, or acid or alkaline treatment to reduce its recalcitrance. The resultant material can then undergo enzymatic saccharification to generate fermentable mono- and oligosaccharides and eventually generate ethanol (Figure 1B).

The major economic burden of biomass refineries include the pretreatment processing of the lignocellulosic biomass and the cost of production of hydrolytic enzymes [14,15]. In order to reduce these cost elements, researchers have focused on genetic engineering of plants for desirable lignin content or composition by partial removal or redistribution of lignin. The objective of such structural re-engineering is to loosen the rigid cell wall structure and expose cellulose and hemicelluloses for enhanced saccharification [11,12,16–20]. However, lignin genetic engineering generally results in detrimental effects on plant growth and development, and such approaches do not exclude the need for exogenous enzyme addition prior to fermentation [13,20]. There is also extensive research on the genetic engineering of microbial strains to enhance saccharification and fermentation capabilities including expression of lignocellulose digesting enzymes in existing fermentation strains (the consolidated bioprocess) [21–25], expanding the catabolic capacity of strains (typically by engineering C5 sugar metabolic pathways: i.e., inserting the genes encoding the enzymes of the pentose sugar degradation pathway) [26–29] and by engineering carbon flow to produce specific end-products in organisms with intrinsically broad metabolic capabilities [30].

A promising alternative to the exogenous addition of lignocellulose-degrading enzymes in the prefermentation pipeline is the heterologous expression of genes encoding lignocellulose-digesting enzymes in growing plant tissues (Figure 2). The conventional pipeline requires expensive up-front capital equipment and inevitably causes sugar degradation, resulting in reduced sugar yield and the formation of compounds that are toxic in fermentation, necessitating additional steps for detoxification, separation, and neutralization. The *in planta* approach eliminates the need for exogenous enzyme addition prior to fermentation. This results in a consolidated process that consists of slightly milder pretreatment (90–120°C) followed by enzymatic hydrolysis without interstage washing of the pretreated biomass. Development of hyperthermophilic fermentation microbes would confer an additional advantage when using the *in planta* approach. Plants are increasingly being investigated as potential 'green bioreactors' for the cost-effective and large-scale production of biologically active recombinant enzymes [31–36], but the production of cellulolytic enzymes using transgenic plants has mostly been restricted to mesophilic and thermotolerant enzymes [37–45]. To the best of our knowledge, the

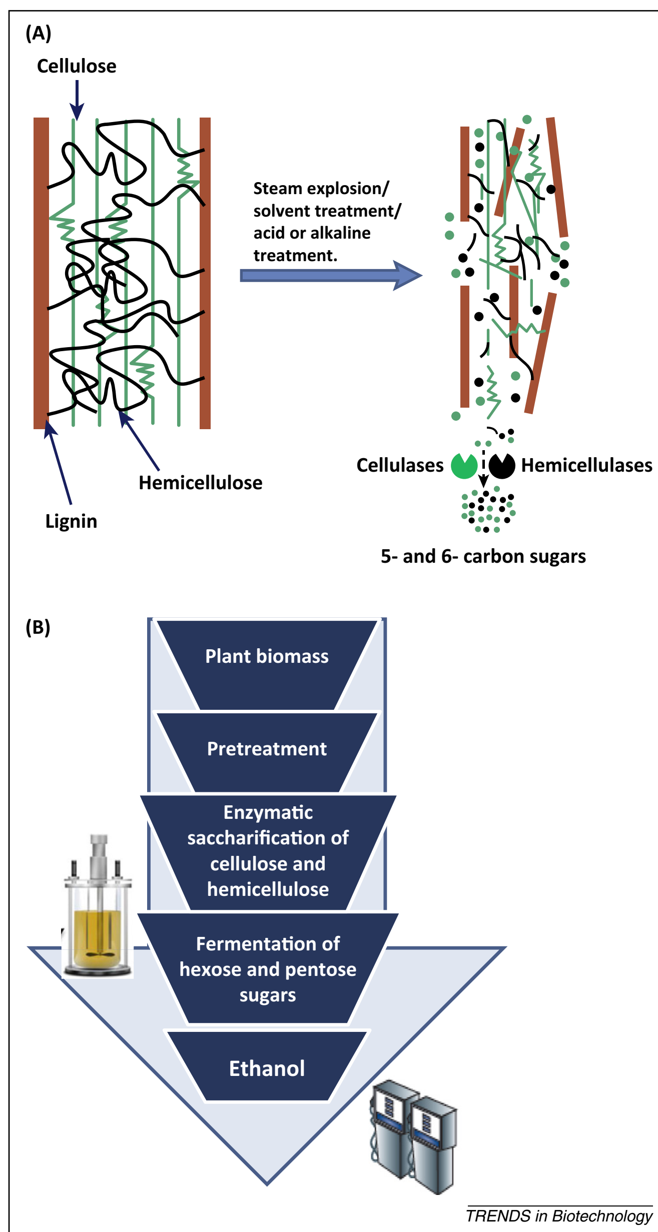

Figure 1. Representation of the generalized components of plant secondary cell walls (primarily cellulose, hemicellulose, and lignin). **(A)** Cellulose microfibrils (green) providing a structural framework of the wall are associated with a coating of hemicellulosic polysaccharides (black) that hydrogen-bond to the microfibril surface and span the distance between fibrils, making a polysaccharide complex that is encased by lignin (brown), a polyphenolic polymer. Polymers disintegrate after pretreatment and subsequent conversion into 5- and 6-carbon sugars by cellulases and hemicellulases. **(B)** Progression from plant biomass to ethanol production.

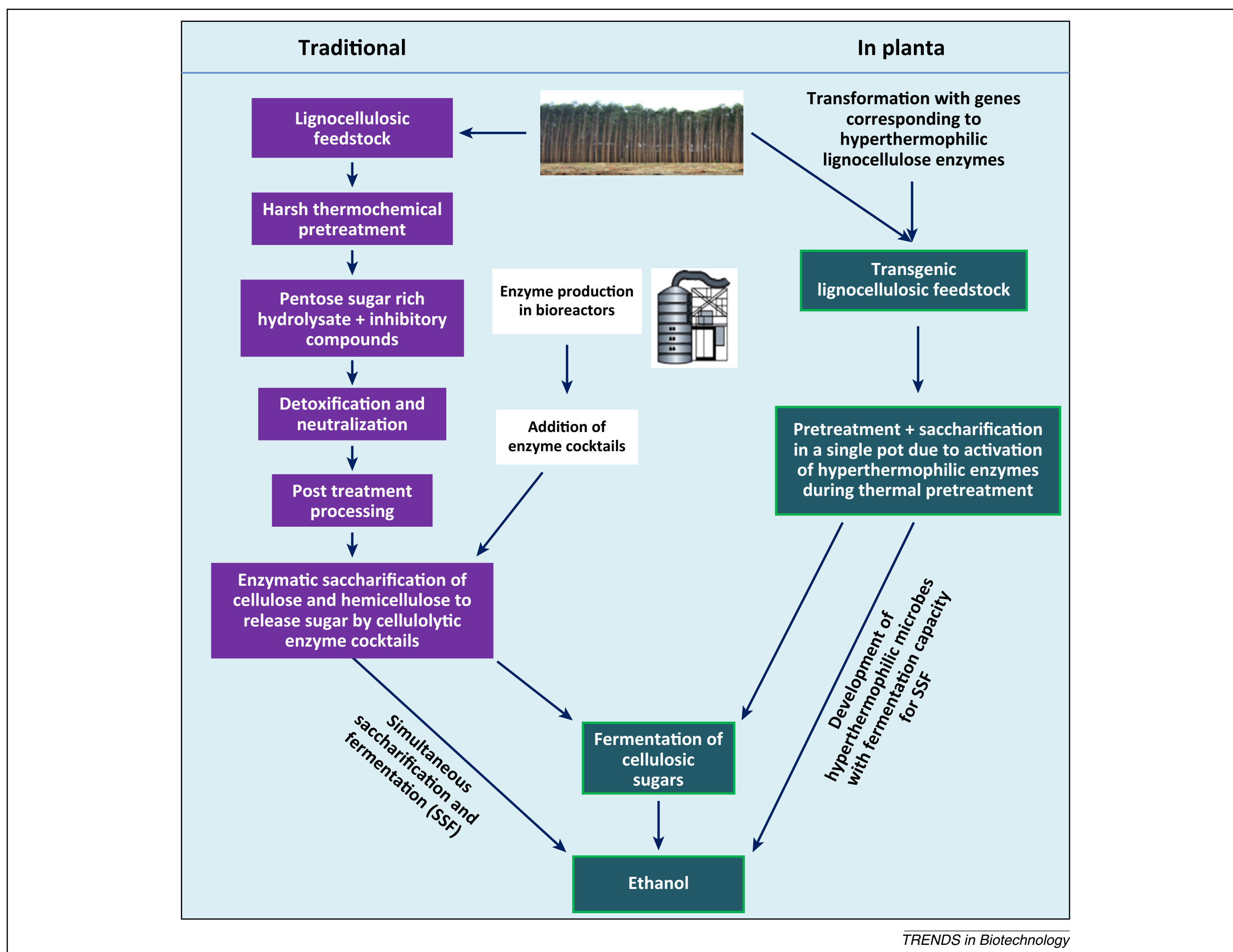

Figure 2. Comparison of a traditional pre-fermentation pipeline (purple boxes) and an approach based on *in planta* expression of hyperthermophilic lignocellulose digesting enzymes combining pretreatment and saccharification in a single reactor (green boxes).

effects of such expression on the efficiency of biomass conversion have been reported in few instances [15].

An immediate and obvious hurdle to the successful *in planta* expression of functional lignocellulose-degrading enzymes would be *in vivo* toxicity, where the potential for the recombinant enzymes to disrupt the physical structure of developing plant tissues would be, at least in theory, high. Two alternative technologies to reduce *in vivo* toxicity may be considered: (i) compartmentalization of the recombinantly expressed enzymes in order to prevent access to substrates inside the cell; and (ii) the expression of effectively inactive enzymes that are subsequently activated as temperatures increase during pretreatment. Another major hurdle to effective implementation of an *in planta* enzyme expression strategy involves the severe physicochemical conditions (steam explosion, acid, alkaline, or solvent treatment) that are required in preprocessing technologies but are incompatible with the functionality of most enzymes. These apparent hurdles can potentially be overcome by using recombinant extremophilic enzymes, cloned from hyperthermophiles and/or acidophiles and/or alkaliphiles, which are, to a greater or lesser extent, compatible with the harsh preprocessing conditions.

Expression of heterologous enzymes in plants

Over the past decade a variety of genes encoding lignocellulose digesting enzymes from cellulose-utilizing microorganisms have been successfully expressed and targeted to subcellular compartments in different host plant species [31,36]. Heterologous expression of these enzymes in plants has several potential advantages over other expression systems (e.g., *Escherichia coli* and yeast). These include better post-translational processing, well-established production and harvesting methods, and the ability to scale up to meet market demands in an economically and environmentally sustainable manner [15,40–44]. The greatest potential benefit offered by plants as expression systems for lignocellulose-digesting enzymes is that they can serve as the source of the cellulosic biomass required for sugar fermentation. *In planta* expression also reduces, or even obviates, the independent cost of production and addition of the recombinant hydrolytic enzymes, which requires sophisticated and capital equipment

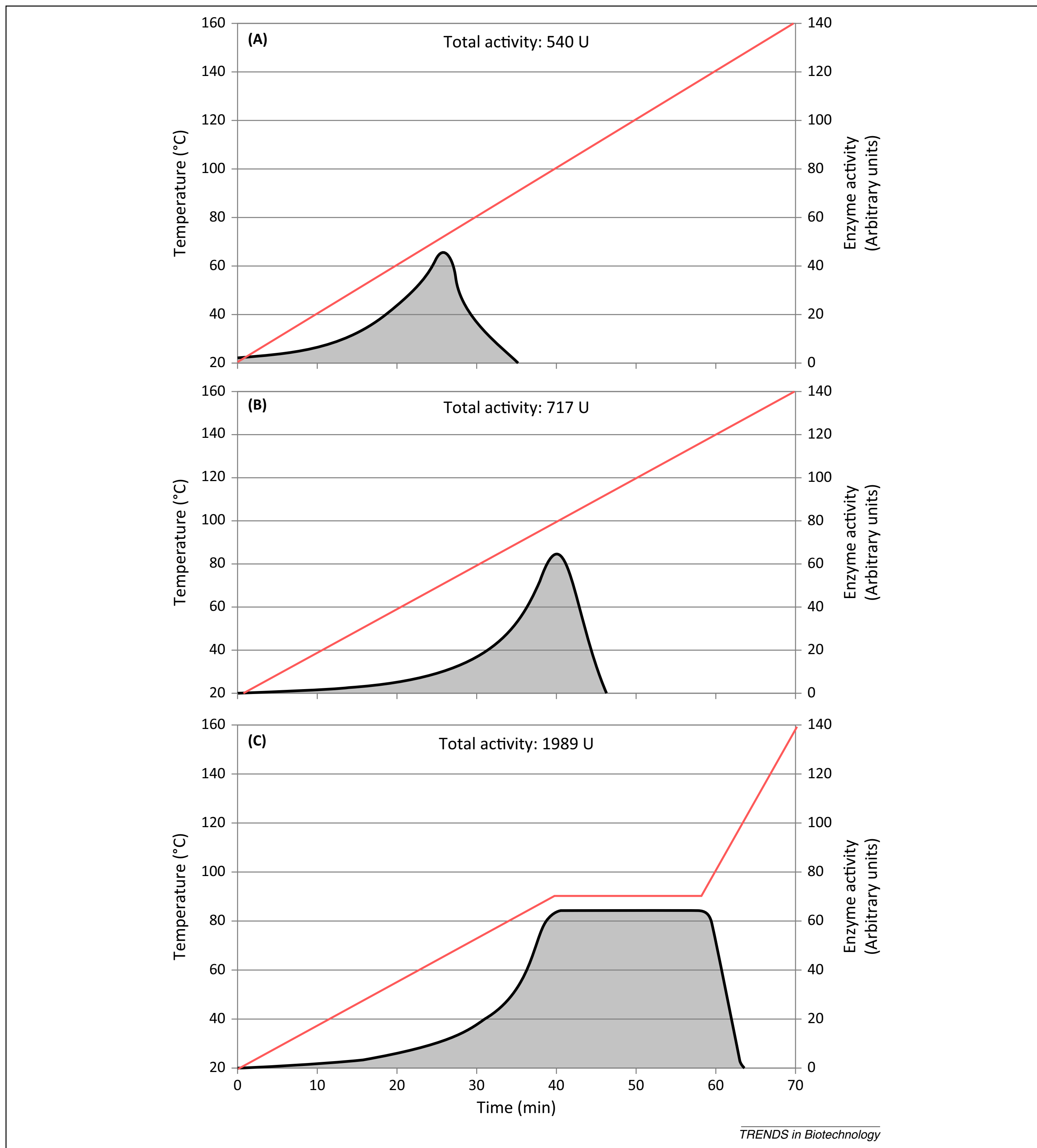

Figure 3. Hypothetical activity parameters of *in planta* expressed mesophilic and hyperthermophilic enzymes during steam explosion of macerated biomass. **(A)** Mesophilic enzyme activity profile. **(B)** Hyperthermophilic enzyme activity profile. **(C)** Hyperthermophilic enzyme activity profile for stepped steam explosion heating process. Activity profiles are hypothetical temperature-optima curves for enzymes obeying Arrhenius' law, for T_{opt} values of 65°C (mesophilic enzyme) and 100°C (hyperthermophilic enzyme). T_{inact} values 10–15°C > T_{opt}. Details can be found in [63,64,67,68]. *Abbreviations*: T_{inact}, temperature at which enzyme are 50% inactive; T_{opt}, optimum temperature for activity.

[15,38]. *In planta* expression of recombinant lignocellulose-digesting genes has involved a range of different promoters and other regulatory elements and targeting of the enzymes to the specific organelles via different transit peptides for accumulation and storage or targeting to the cell wall for immediate modification of the polysaccharides (Table S1 in the supplementary material online).

The relatively modest level of accumulation of these enzymes in plant tissues is a major obstacle for the commercial deployment of this technology. Recombinant

enzymes expressed in host plants often represent $<10\%$ of the total soluble protein (TSP), but would have to be increased to $\sim30\%$ for effective hydrolysis of lignocellulosic tissues without the need for addition of exogenous enzymes [8]. Although *in planta* expression of thermophilic endoglucanases has been achieved in the range of 10–26% TSP without any detrimental effects on transgenic plants [37–48], yields of 0.1–5% TSP are more typical, and several studies have described negative phenotypes associated with *in planta* expression of recombinant mesophilic cellulolytic enzymes [32,49].

Molecular factors affecting heterologous expression

The various factors that affect the expression of a particular heterologous gene and subsequent production of its protein in plant cells include: (i) transcriptional determinants such as transgene copy number, site of integration of the T-DNA in the chromosome, and promoter activity [50]; (ii) post-transcriptional factors including, 5′-untranslated leader sequence, 3′-polyadenylation, and mRNA stability [51,52]; and (iii) translational/post-translational factors such as protein stability, modification, and trafficking [52,53].

Promoters and enhancers (*cis*-acting regulatory elements) play a crucial role in regulating spatial and temporal patterns of transgene expression at particular stages or in response to particular conditions [50–55]. Several promoters and other *cis*-acting elements including signal peptides of plant, plant pathogen, or bacterial origin used in heterologous expression vectors have been described (Table S1 in the supplementary material online). The cauliflower mosaic virus (CaMV) 35S promoter is extensively used in plant expression vectors (reviewed in [31,44]). Multiple copies of enhancer sequences from highly active promoters can be stacked to boost transgene expression [32–34,50,51].

RNA processing such as capping, splicing, and polyadenylation have also been shown to affect the expression levels of heterologous genes [50]. Additionally, 3′-untranslated sequences downstream of the stop codon playing a critical role in processing have been included along with signal sequences targeting and stabilizing the message for polyadenylation (Table S1 in the supplementary material online). Polyadenylation strongly enhances the heterologous expression of microbial genes in plants [33,34,40–42]. Sequences that are located immediately around the translational start site, termed Kozak sequences [56], can also be modified to fit the consensus initiation sequence of plants in order to achieve higher rates of translation.

Subcellular localization and expression

Targeting proteins to specific subcellular sites/compartments of plant cells (e.g., apoplast, chloroplast, mitochondria, vacuoles, and endoplasmic reticulum) has also been used to increase the expression level and accumulation of recombinant enzymes in plants. N-terminal or C-terminal transit peptides generally possess the necessary information for the correct targeting of proteins to subcellular organelles. Subcellular targeting of heterologous proteins generally yields higher accumulation levels in transgenic plants than expression in the cytosol [32–42,47,51]. When expressed in subcellular compartments, heterologous proteins typically display correct folding and activity, appropriate glycosylation, reduced degradation, and increased stability, as compared to expression and accumulation in the cytosol [8,44].

Although there are numerous reports of cellulolytic enzyme production *in planta*, relatively few describe transgenic plants expressing cellulolytic enzymes in specific subcellular compartments. *Acidothermus cellulolyticus* E1 protein/enzyme accumulated to a higher level (2.6% of TSP) in potato leaves when targeted to the chloroplast, compared to the vacuole (0.8% of TSP) [57]. This enzyme was also expressed at a higher level (2.6% TSP) when targeted to the apoplast using plant-based signal peptides [57]. Expression of *A. cellulolyticus* E1 and E1cd proteins in the tobacco leaf apoplastic space was superior to either the chloroplast or the cytosol [41,51]. Several other thermophilic enzyme genes have been successfully and functionally expressed in plant apoplasts (Table S1 in the supplementary material online).

Another factor that dictates the choice of subcellular targeting is pH compatibility. For example, the pH of the mitochondrial matrix and cytosol is approximately 7.8 and 7.5, respectively, whereas chloroplasts exhibit diurnal pH fluctuations from 7.5 to 8.0. Although enzymes with pH-stability and -activity characteristics compatible with the subcellular organelle may be readily selected, in reality the optimal pH-activity characteristics for the digestion phase of the process is a more critical determinant of enzyme selection.

Limitations to expression

The heterologous expression of mesophilic lignocellulose-degrading enzymes *in planta* can potentially lead to physiological problems related to intracellular toxicity, particularly as a result of the degradation of structural elements such as cell wall biopolymers. Such damage may have a strong negative influence on plant growth and development at normal growth temperatures, and such transgenic plants may be particularly vulnerable to environmental stresses [36,58,59]. A further major limitation of *in planta* expression of mesophilic enzymes as an alternative to exogenous addition of enzyme cocktails is that mesophilic enzyme performance parameters (e.g., activity and stability) are not compatible with virtually any of the standard pretreatment conditions (which include high temperatures, and/or extreme pH values and/or organic solutes) (Figure 2).

Hyperthermophilic enzymes

Hyperthermophilic microorganisms, present in the bacterial and archaeal domains, have attracted great interest in the scientific community for their ability to express enzymes that can tolerate high temperatures, extreme pH values, and strong salt solutions [60]. These organisms are typically isolated from environments with temperatures in the range of 80–115°C, including continental solfataras, deep geothermally heated oil-containing reservoirs, shallow marine and deep-sea hot sediments, and deep sea hydrothermal vents [61]. Enzymes from hyperthermophilic microorganisms generally show activity and stability characteristics that are compatible with the ther-

mal origins of the organisms: most are active in the 80–120°C temperature range [61] and many tolerate a wide range of pH values (0.5–9.0) [60]. In line with Arrhenius' Law of temperature–activity relations, most hyperthermophilic enzymes (temperature optimum >90°C) show little or no activity at mesophilic temperatures. Inactivity at mesophilic temperatures, coupled with thermal activation may be a usable property of hyperthermophilic enzymes in lignocellulose digestion.

A large number of hyperthermophilic enzymes capable of hydrolyzing lignocellulosic biomass have been cloned, characterized, and expressed in *E. coli* and other hosts [60,62–64], but only a few have been successfully expressed in various host plants (Table S1 in the supplementary material online). Comparative lignocellulosic-digestion studies with mesophilic and hyperthermophilic cellulases and hemicellulases have demonstrated that the functionally homologous hyperthermophilic enzymes can increase saccharification yields and shorten incubation times, and may eliminate the risk of downstream contamination by mesophilic microorganisms competing for fermentable sugars [65,66].

In planta expression of hyperthermophilic enzymes

The underlying principle of *in planta* expression of hyperthermophilic enzymes in woody tissues, irrespective of how such enzymes are targeted, is that (i) the low activities of hyperthermophilic enzymes at mesophilic temperatures reduce functional toxicity (the enzymes are effectively 'switched off' at ambient temperatures); and (ii) the high temperature stability and activity of such enzymes is, at least partly, compatible with severe preprocessing conditions, such as steam explosion (Figure 3). Enzyme activities are effectively switched on as processing temperatures rise, and the integrated activity (a function of both the longevity of enzyme survival and the specific activities achieved) should, theoretically, be much greater for hyperthermophilic enzymes than for mesophilic enzymes [35,47,67]. This effect could be further enhanced by modifying the heating profile so as to optimize the activity of the particular recombinant enzyme. The additional advantage of using a hyperthermophilic enzyme is that optimum activity occurs at a point at which some thermal disordering of the cellulose fibrils has already occurred, increasing the susceptibility of the cellulose and hemicellulose substrates to hydrolytic attack.

Recent attempts at heterologous expression

In general, the most successful attempts to express hyperthermophilic enzymes in plant tissues (Table S1 in the supplementary material online) have come from targeted expression in the chloroplast [47,48]. For example, a hyperthermophilic endoglucanase (Cel5A) gene was targeted to *Arabidopsis* chloroplasts using the N-terminal transit peptides of nuclear-encoded plastid proteins, the light-harvesting chlorophyll a/b-binding protein (CAB), Rubisco small subunit (RS), and Rubisco activase (RA). An expression cassette containing the RS signal peptide driven by the RbcS promoter yielded the highest level of accumulation (5.2% of TSP) of Cel5A in tobacco chloroplasts, and the recombinant enzyme remained stable throughout the life cycle of the transgenic plants [48]. In a direct comparison of cytosolic and chloroplast expression of a hyperthermophilic cellulase, Bg1B, the highest enzyme activity and accumulation (5.8% of TSP) was achieved when Bg1B was directed to the chloroplast using the RA signal peptide and the RbcS promoter [35].

The expression of a hyperthermophilic endoglucanase (SSO1354) has been successfully targeted to the apoplast using the plant codon optimized murine antibody mAb24 (LPH) signal peptide driven by the widely used CaMV double 35S promoter with a polyadenylation signal (pA35S) at the C terminus and a 5′-untranslated region (UTR) translational enhancer from the *Petroselinum hortense* chalcone synthase gene combined with an endoplasmic reticulum retention sequence (KDEL) [67]. However, yields of 1.1% TSP are lower than have been reported for other cellulolytic enzymes in plant expression systems (Table S1 in the supplementary material online), which can be improved by codon optimization of the gene corresponding to this enzyme.

A recent systematic study on transgenic maize (*Zea mays*) expressing various mesophilic and thermophilic lignocellulose-digesting enzymes, where total plant biomass was harvested and used as a feedstock for sugar production, has demonstrated the potential of this approach [15]. Transgenic maize biomass produced up to 1.4-fold higher glucose and 1.7-fold higher xylose yields than control plants treated with different exogenous enzyme loadings during enzymatic hydrolysis at temperatures below 75°C. Transgenic maize also showed 55% improvement in ethanol production compared to control plants. However, even at the highest temperatures applied, the addition of exogenous enzyme prior to fermentation was still required.

In a similar study, on the autohydrolysis of plant polysaccharides using recombinant hyperthermophilic enzymes expressed in transgenic tobacco, a significant increase in the release of glucose from transgenic plant tissues, as well as from purified polysaccharide substrates was observed after incubation at high temperatures [68]. Recently, a tenfold increase in the activity of thermophilic bacterial endoxylanases (XynA and XynB) was achieved by apoplastic expression compared to chloroplast expression in tobacco and *Arabidopsis* [46]. The highest activity in *Arabidopsis* was reported in dry stems, and long-term storage (>6 months) of dried stems at room temperature did not reduce enzyme activity [46].

Directions for improvement

Despite successful demonstration of the *in planta* expression of heterologous enzymes, several issues require further development. The most obvious is the relatively low level of *in situ* protein expression/accumulation observed in most studies (Table S1 in the supplementary material online), in which an order-of-magnitude increase in expression level is required for complete conversion of biomass [8,33]. In addition to the obvious requirement for careful codon optimization [8,46,59] and/or codon harmonization [69], the further development of combinations of transit peptides and promoters (strong promoters and chimeric promoters), with or without promoter engineering, may be

effective strategies [52,70]. Carbohydrate-binding modules can also be used to target better the recombinant enzymes and ultimately improve biomass digestion efficiency [45]. Other factors that might potentially contribute to the optimization of *in planta* production and function of hyperthermophilic enzymes include UTR optimization and optimization of Kozak sequences. The optimized UTR and Kozak sequences promote RNA stability and translation efficiency to facilitate high-level protein expression. Choosing the appropriate subcellular location for heterologous gene expression could be a deciding factor in improving performance, particularly in woody tissues where the bulk of cells undergo programmed cell death at the completion of xylogenesis. The frequent selection of the apoplast as a preferred compartment for expression is based on the assumption that this compartment is the most 'spacious' and is therefore capable of accumulating large quantities of heterologous proteins. Other options, such as targeting of enzymes to multiple cell compartments, could potentially result in higher accumulation levels, but may not be a viable option for woody tissues. This has yet to be tested.

Crossing high-expressing transgenic lines arising from independent transformation events can boost transgene copy number and expression levels [12]. Transgenic lines containing multiple transgene copies from different events appear to have a reduced risk of gene silencing, compared to those engineered by multiple gene insertions during the same transformation events [12,52].

Concluding remarks and future perspectives

Lignocellulosic biomass, comprising the most abundant renewable and carbon-neutral organic polymer is, at least in theory, an attractive feedstock for the production of economically and environmentally sustainable bioproducts. Despite considerable progress in the development of lignocellulosic biomass conversion techniques, the quest to develop cheaper conversion processes remains a high research priority worldwide. Of the many processes under development, the targeted heterologous expression of hyperthermophilic enzymes in lignocellulosic feedstock crops potentially offers an elegant technology compatible with other unit operations in biomass processing pathways.

To date, most studies on the heterologous expression of cellulases have been performed on non-feedstock model plants. However, *in planta* expression of lignocellulose-degrading enzymes in fast-growing lignocellulosic feedstock crops, such as perennial grasses (*Miscanthus* and switchgrass) and trees (*Eucalyptus*, *Populus*, and *Salix*), has some capacity to develop these species as major energy and biomaterials feedstocks for the future.

The commercial potential of *in planta* expression in the biofuels sector has not gone unnoticed. Commercial research in and patent protection of transgenic technologies around plant improvement for degradation of lignocellulose is an active field, with Syngenta Biotechnology Inc. most actively patenting [71–73].

It is worth considering the conceptual, technological, or economic barriers that might prevent the effective implementation of heterologous thermophilic cellulase expression in plants. The effective conversion of cellulosic biomass into fermentable sugars is a complex process and requires the synergistic action of multiple primary and accessory enzymes. A successful *in planta* expression system for complete deconstruction of the lignocellulosic substrate is therefore likely to require high-level expression of multiple heterologous genes in a single feedstock crop, with individual expression levels tailored to achieve the highest levels of functional synergy. It is clear from a review of the relevant literature that considerable further improvement of vector development, cellular and subcellular targeting and expression control is required.

Enzymes from hyperthermophiles operate optimally at temperatures a few tens of degrees from their denaturation temperatures. Although protein-engineering strategies can offer some increases in thermostability even in these enzymes, there are clearly limits to molecular stability (e.g., at 120–130°C [60,61] at temperatures well below the normal maximum operating temperatures for steam explosion; 160–220°C). However, the inclusion of an intermediate 'temperature plateau' in the heating phase of the steam explosion profile (Figure 2), at or near the optimum temperature for activity of the *in planta*-expressed hyperthermophilic enzyme should offer substantial process benefits.

There are also known energy costs associated with the sequential processing of lignocellulosic biomass macerates by steam explosion, cooling, enzyme digestion, and fermentation. It is likely that there is a direct economic benefit in operating the latter two unit operations at elevated temperatures. The use of exogenous hyperthermophilic enzymes for lignocellulosic digestion (at 78°C) [74] and the engineering of thermophilic bacteria for bioethanol production from lignocellulosic hydrolysates (at 65°C) [75] have both been recently reported.

An alternative (or complementary) approach would be the engineering of the lignocellulosic substrate in order to enhance its susceptibility to digestion. Increasing cellulose synthesis and the ratio of cellulose to hemicellulose, and engineering of plants for desirable lignin content or composition by partial removal or redistribution of lignin, are obvious strategies to improve the economical feasibility of bioenergy crops [9,11,76,77]. However, there are clearly limits to such engineering, dictated by the need for optimal plant development, structural integrity, and defense [13,20]. The other limitation is resistance, in some countries, to the commercial use of genetically modified crop species.

Acknowledgments

The authors gratefully acknowledge the University of Pretoria and National Research Foundation (NRF), South Africa for providing postdoctoral fellowship support to BAM.

References

1 Sawin, J.L. *et al.* (2013) Renewables 2013 global status report. In *Renewable Energy Policy Network for the 21st Century*. (REN21)

2 Lynd, L.R. *et al.* (2008) How biotech can transform biofuels. *Nat. Biotechnol.* 26, 169–172

3 Carroll, A. and Somerville, C. (2009) Cellulosic biofuels. *Annu. Rev. Plant Biol.* 60, 165–182
4 Somerville, C. *et al.* (2010) Feedstocks for lignocellulosic biofuels. *Science* 13, 790–792
5 Wilson, D.B. (2008) Three microbial strategies for plant cell wall degradation. *Ann. N. Y. Acad. Sci.* 1125, 289–297
6 Sainz, M.B. (2011) Commercial cellulosic ethanol: the role of plant-expressed enzymes. In *Biofuels Global Impact on Renewable Energy, Production Agriculture and Technological Advancement* (Tomes, D. *et al.*, eds), pp. 237–264, Springer
7 Himmel, M.E. *et al.* (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807
8 Sticklen, M. (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat. Rev. Genet.* 9, 433–443
9 Vanholme, R. (2008) Lignin engineering. *Curr. Opin. Plant Biol.* 11, 278–285
10 Vanholme, R. *et al.* (2010) Potential of *Arabidopsis* systems biology to advance the biofuel field. *Trends Biotechnol.* 28, 543–547
11 Vanholme, R. *et al.* (2012) Metabolic engineering of novel lignin in biomass crops. *New Phytol.* 196, 978–1000
12 Vanholme, B. *et al.* (2013) Towards a carbon negative sustainable bio-based economy. *Front. Plant Sci.* 4, 174 1–17
13 Yang, F. *et al.* (2013) Engineering secondary cell wall deposition in plants. *Plant Biotechnol. J.* 11, 325–335
14 Ragauskas, A.J. *et al.* (2006) The path forward for biofuels and biomaterials. *Science* 311, 484–489
15 Zhang, D. *et al.* (2011) Consolidated pretreatment and hydrolysis of plant biomass expressing cell wall degrading enzymes. *Bioenergy Res.* 4, 276–286
16 Chen, F. and Dixon, R.A. (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–761
17 Chen, Y. *et al.* (2012) Reducing acid in dilute acid pretreatment and the impact on enzymatic saccharification. *J. Ind. Microbiol. Biotechnol.* 39, 691–700
18 Fu, C. *et al.* (2011) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3803–3808
19 Mansfield, S.D. *et al.* (2012) Designed for deconstruction poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. *New Phytol.* 194, 91–101
20 Ding, S.Y. *et al.* (2012) How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? *Science* 338, 1055–1060
21 Lynd, L.R. *et al.* (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 16, 577–583
22 Ilmen, M. *et al.* (2011) High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 4, 1–15
23 Chen, X. *et al.* (2011) Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnol. Biofuels* 4, 1–12
24 Mazzoli, R. *et al.* (2012) Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. *Trends Biotechnol.* 30, 111–119
25 Hong, K.K. and Nielsen, J. (2012) Metabolic engineering of *Saccharomyces cerevisiae* a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* 69, 2671–2690
26 Nair, N.U. and Zhao, H.M. (2010) Selective reduction of xylose to xylitol from a mixture of hemicellulosic sugars. *Metab. Eng.* 12, 462–468
27 Young, E.M. (2012) A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. *Metab. Eng.* 14, 401–411
28 Young, E.M. *et al.* (2010) Optimizing pentose utilization in yeast: the need for novel tools and approaches. *Biotechnol. Biofuels* 3, 1–12
29 Hasunuma, T. (2011) Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 10, 1–13
30 Chang, T. and Yao, S. (2011) Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives. *Appl. Microbiol. Biotechnol.* 92, 13–27
31 Taylor, L.E. *et al.* (2008) Heterologous expression of glycosyl hydrolases *in planta*: a new departure for biofuels. *Trends Biotechnol.* 26, 413–424
32 Ziegelhoffer, T. (2009) Expression of *Acidothermus cellulolyticus* E1 endo-β-1,4-glucanase catalytic domain in transplastomic tobacco. *Plant Biotechnol. J.* 7, 527–536
33 Gray, B.N. *et al.* (2009) High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. *Biotechnol. Bioeng.* 102, 1045–1054
34 Gray, B.N. *et al.* (2011) An efficient downstream box fusion allows high-level accumulation of active bacterial β-glucosidase in tobacco chloroplasts. *Plant Mol. Biol.* 76, 345–355
35 Jung, S.R. *et al.* (2010) Expression of thermostable bacterial β-glucosidase (BglB) in transgenic tobacco plants. *Bioresour. Technol.* 101, 7144–7150
36 Obro, J. *et al.* (2011) Enzymatic modification of plant cell wall polysaccharides. *Annu. Plant Rev.* 41, 367–388
37 Ziegler, M.T. *et al.* (2000) Accumulation of a thermostable endo-1,4-b-D-glucanase in the apoplast of *Arabidopsis thaliana* leaves. *Mol. Breed.* 6, 37–46
38 Hood, E.E. *et al.* (2007) Sub-cellular targeting is a key condition for high-level accumulation of cellulase protein in transgenic maize seed. *Plant Biotechnol. J.* 5, 709–719
39 Sun, Y. *et al.* (2007) Expression and characterization of *Acidothermus cellulolyticus* E1 endoglucanase in transgenic duckweed *Lemna minor* 8627. *Bioresour. Technol.* 98, 2866–2872
40 Yu, L.X. *et al.* (2007) Expression of thermostable microbial cellulases in the chloroplasts of nicotine-free tobacco. *J. Biotechnol.* 131, 362–369
41 Ziegelhoffer, T. *et al.* (2001) Dramatic effects of truncation and sub-cellular targeting on the accumulation of recombinant microbial cellulase in tobacco. *Mol. Breed.* 8, 147–158
42 Chou, H.L. *et al.* (2011) High level expression of *Acidothermus cellulolyticus* β-1, 4-endoglucanase in transgenic rice enhances the hydrolysis of its straw by cultured cow gastric fluid. *Biotechnol. Biofuels* 4, 1–13
43 Jiang, X. *et al.* (2011) Expressions of thermostable bacterial cellulases in tobacco plant. *Biotechnol. Lett.* 33, 1797–1803
44 Jung, S.K. *et al.* (2012) Heterologous expression of plant cell wall degrading enzymes for effective production of cellulosic biofuels. *J. Biomed. Biotechnol.* 405842, 1–10
45 Byrt, C.S. *et al.* (2012) Plant carbohydrate binding module enhances activity of hybrid microbial cellulase enzyme. *Front. Plant Sci.* 3, 1–8
46 Borkhardt, B. *et al.* (2010) Autohydrolysis of plant xylans by apoplastic expression of thermophilic bacterial endo-xylanases. *Plant Biotechnol. J.* 8, 363–374
47 Kim, J.Y. *et al.* (2011) Production of hyperthermostable GH10 xylanase Xyl10B from *Thermotoga maritime* in transplastomic plants enables complete hydrolysis of methyl glucuronoxylan to fermentable sugars for biofuel production. *Plant Mol. Biol.* 76, 357–369
48 Kim, S. *et al.* (2010) *Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase *Cel5A* in chloroplasts of transgenic tobacco plants. *Transgenic Res.* 19, 489–497
49 Abdeev, R.M. *et al.* (2003) Expression of a thermostable bacterial cellulase in transgenic tobacco plants. *Genetika* 39, 376–382
50 Desai, P.N. *et al.* (2010) Production of heterologous proteins in plants: strategies for optimal expression. *Biotechnol. Adv.* 28, 427–435
51 Dai, Z. *et al.* (2005) Optimization of *Acidothermus cellulolyticus* endoglucanase (E1) production in transgenic tobacco plants by transcriptional, post-transcription and post-translational modification. *Transgenic Res.* 14, 627–643
52 Streatfield, S.J. (2007) Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol. J.* 5, 2–15
53 Yang, L. *et al.* (2009) The 3′-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice. *Biotechnol. Lett.* 31, 1625–1631
54 Harrison, M.D. *et al.* (2011) Accumulation of recombinant cellobiohydrolase and endoglucanase in the leaves of mature transgenic sugar cane. *Plant Biotechnol. J.* 9, 884–896
55 Hwang, M.S. *et al.* (2012) Bipartite and tripartite *Cucumber mosaic* virus-based vectors for producing the *Acidothermus cellulolyticus* endo- 1, 4-β-glucanase and other proteins in non-transgenic plants. *BMC Biotechnol.* 12, 1–13
56 Kozak, M. (1984) Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* 308, 241–246

57 Dai, Z. *et al.* (2000) Expression of *Acidothermus cellulolyticus* endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects. *Transgenic Res.* 9, 43–54

58 Wilson, D.B. (2009) Cellulases and biofuels. *Curr. Opin. Biotechnol.* 20, 295–299

59 Brunecky, R. *et al.* (2011) *In planta* expression of *A. cellulolyticus Cel5A* endocellulase reduces cell wall recalcitrance in tobacco and maize. *Biotechnol. Biofuels* 4, 1–10

60 Unsworth, L.D. *et al.* (2007) Hyperthermophilic enzymes- stability, activity and implementation strategies for high temperature applications. *FEBS J.* 274, 4044–4056

61 Gerday, C. and Glansdorff, N., eds (2007) *Physiology and Biochemistry of Extremophiles*, ASM Press

62 Girfoglio, M. *et al.* (2012) Cellulose degradation by *Sulfolobus solfataricus* requires a cell anchored endo-β-1, 4-glucanase. *J. Bacteriol.* 194, 5091–5100

63Graham,J.E.*etal.*(2011)Identificationandcharacterizationofa multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2, 1–9

64 Verma, D. *et al.* (2013) Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (*Mxyl*) retrieved from compost-soil metagenome. *PLoS ONE* 8, 1–8

65 Kumar, P. *et al.* (2009) Methods for pretreatment of lignocellulosic biomassforefficienthydrolysisandbiofuelproduction.*Ind.Eng. Chem. Res.*48,3713–3729

66 McClendon, S.D. *et al.* (2012) *Thermoascus aurantiacus* is a promising source of enzymes for biomass deconstruction under thermophilic conditions. *Biotechnol. Biofuels* 5, 1–9

67 Klose, H. *et al.* (2012) Hyperthermophilic endoglucanase for *in planta* lignocellulose conversion. *Biotechnol. Biofuels* 5, 1–9

68 Montalvo-Rodriguez, R. *et al.* (2000) Autohydrolysis of plant polysaccharides using transgenic hyperthermophilic enzymes. *Biotechnol. Bioeng.* 70, 151–159

69 Angov, E. *et al.* (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* 35, e2189

70 Zou, G. *et al.* (2012) Construction of a cellulase hyper-expression system in *Trichoderma reesei* by promoter and enzyme engineering. *Microb. Cell Fact.* 11, 1–12

71 Basu, S.S. *et al*. (2009) Syngenta Biotechnology Inc. Self-processing plants and plant parts. AU2004318207B2

72 Dandekar, A.M. (2011) University of California. Production of cellulase enzymesinplanthostsusingtransientagroinfiltration.AU2009282874A1

73Lubieniechi,S.*etal.*(2013)Recentpatentsongeneticmodificationof plants and microbes for biomass conversion to biofuels. *Recent Patents DNA Gene Seq.* 7, 25–35

74 Ketaeva, I. *et al.* (2013) Carbohydrate and lignin are simultaneously solubilized from un-pretreated switchgrass by microbial action at high temperature. *Energy Environ. Sci.* 6, 2186–2195

75 Cripps, R.E. *et al.* (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab. Eng.* 11, 398–408

76 Sticklen, M. (2006) Plant genetic engineering to improve biomass characteristics for biofuels. *Curr. Opin. Biotechnol.* 17, 315–319

77 Guerriero, G. *et al.* (2010) What do we really know about cellulose biosynthesis in higher plants? *J. Integr. Plant Biol.* 52, 161–175

78 Kimura, T. *et al.* (2003) Stable expression of a thermostable xylanase of *Clostridium thermocellum* in cultured tobacco cells. *J. Biosci. Bioeng.* 95, 397–400

79 Biswas, G.C.G. *et al.* (2006) Expression of biologically active *Acidothermus cellulolyticus* endoglucanase in transgenic maize plants. *Plant Sci.* 171, 617–623

80 Herbers, K. *et al.* (1995) A thermostable xylanase from *Clostridium thermocellum*expressedathighlevelsintheapoplastoftransgenic tobacco hasnodetrimentaleffectsandiseasilypurified.*Nat. Biotechnol.*13,63–66

81Nigorikawa,M.*etal.*(2012)Enhancedsaccharificationofricestrawby overexpression of rice exo-glucanase. *Rice* 5, 1–7

82 Oraby, H. *et al.* (2007) Enhanced conversion of plant biomass into glucose using transgenic rice-produced endoglucanase for cellulosic ethanol. *Transgenic Res.* 16, 739–749

83 Patel, M. *et al.* (2000) Transgenic barley expressing a fungal xylanase gene in the endosperm of the developing grains. *Mol. Breed.* 6, 113–124

84 Ransom, C. *et al.* (2007) Heterologous *Acidothermus cellulolyticus* 1,4-β-endoglucanase E1 produced within the corn biomass converts corn stover into glucose. *Appl. Biochem. Biotechnol.* 137, 207–219

85 Sun, J.L. *et al.* (1997) High expression of the xylanase B gene from *Clostridium stercorarium* in tobacco cells. *J. Ferment. Bioeng.* 84, 219–223

86 Ziegelhoffer, T. *et al.* (1999) Expression of bacterial cellulase genes in transgenic alfalfa (*Medicago sativa* L.), potato (*Solanum tuberosum* L.) and tobacco (*Nicotiana tabacum* L.). *Mol. Breed.* 5, 309–318

Appendix A. Supplementary data

Table S1 Summary of the *in planta* expression of recombinant thermophilic and hyperthermophilic enzymes.

Enzyme/ gene	Source	Growth temperature of source organism (°C)	Host plant	Promoter	Transit peptide	Storage compartment	Other regulatory element	Yield (TSP)	Refs.
	Thermophilic								
E1	*Acidothermus cellulolyticus*	55	Tobacco, Alfalfa	CaMV35S	Pr1a, CAL	Apoplast Chloroplast, Cytosol, ER	TMV Ω	ND	[41,86]
E1 (cd)	*A. cellulolyticus*	55	Tobacco	CaMV35S	Pr1a	Apoplast	D1 of psbA	12%	[32]
E1	*A. cellulolyticus*	55	Potato	RbcS-3C, Mac	RbcS-2A, KDEL	Chloroplast, ER, Apoplast	5'UTR, mas, RbcS-3C	2.6%	[57]
E1	*A. cellulolyticus*	55	Maize	CaMV35S	Pr1a	Apoplast	TMV Ω	1.13%	[84]
E1	*A. cellulolyticus*	55	Maize	Glob-1	KDEL, Pr1a, VTS	ER, Vacuole Apoplast	---	15-18%	[38]
E1	*A. cellulolyticus*	55	*Arabidopsis*	CaMV35S	Pr1a	Apoplast	TMV Ω, NOS	26%	[37]
E1 (Cel5A)	*A. cellulolyticus*	55	Maize, Tobacco	CaMV35S	VSPβ	Apoplast	---	ND	[59]
E1	*A. cellulolyticus*	55	Maize	CaMV35S	Pr1a	Apoplast	TMV Ω, 3'NOS	2.1%	[79]
EGA	*A. cellulolyticus*	55	Maize	---	---	---	---	ND	[15]
XynA	*Dictyoglomus thermophilum*	78	Maize	---	---	---	---	ND	[15]
XynA & XynB	*D. thermophilum*	78	*Arabidopsis*	CaMV35S	Barley α-amylase	Apoplast	Tnos	ND	[46]

					SP				
E1	*A. cellulolyticus*	55	Rice	CaMV35S	Pr1a	Apoplast	TMV Ω	4.9%	[82]
E1	*A. cellulolyticus*	55	Rice	Mac	Pr1a	Apoplast	---	6.1%	[42]
E1	*A. cellulolyticus*	55	Duckweed	CaMV35S	---	Cytosol	---	0.24%	[39]
Cel6A	*Thermpbifida fusca*	55-60	Tobacco	Prrn, *psbA*	---	Chloroplast	3' UTR (*psbA*), 5' UTR (*rbcL)*	2-3%	[40]
Cel6B	*T. fusca*	55-60	Tobacco	Prrn, *psbA*	---	Chloroplast	3' UTR (*psbA*), 5' UTR (*rbcL)*	4%	[40]
Cel6A	*T. fusca*	55-60	Tobacco	CaMV35S S	CAL, HDEL	Apoplast, Chloroplast	5' TEV-UTL	ND	[43]
Cel6B	*T. fusca*	55-60	Tobacco	CaMV35S S	CAL, HDEL	Apoplast, Chloroplast	5' TEV-UTL	ND	[43]
Cel6A-Cel6B fusion	*T. fusca*	55-60	Tobacco	CaMV35S S	CAL, HDEL	Apoplast, Chloroplast	5' TEV-UTL	ND	[43]
Cel6A	*T. fusca*	55-60	Alfalfa, Potato	Prrn*, psbA*	---	Cytosol	---	0.01%	[86]
Cel6A	*T. fusca*	55-60	Tobacco	Prrn, *psbA*	---	Chloroplast	TetC	10.7%	[33,34]
CelE1	*Clostridium thermocellum*	60	Tobacco	CaMV35S S	Carrot extension SP	Apoplast	pA	ND	[49]
Xyn A	*C. thermocellum*	60	Rice	CaMV35S	---	Cytosol	5' UTR of soyabean Lox gene	ND	[78]
Xyn A	*C. thermocellum*	60	Barley	GluB-1, Hor2-4	---	Endosperm	5' UTR of GluB-1 of rice, 3' UTR of rbcS	ND	[83]

Xyn B	*C. thermocellum*	60	Tobacco	CaMV35S	---	---	---	5%	[85]
Xyn Z	*C. thermocellum*	60	Tobacco	CaMV35S	Pr1a	Apoplast	---	4.1%	[80]
EXG1, ENG1, BEG1	*Oryza sativa*	--	Rice	CaMV35S	---	---	NOS terminator	ND	[81]
	Hyperthermophilic								
Bg1B	*Thermotoga maritima*	80	Tobacco	T7, RbcS	RA	Cytosol, Chloroplast	NOS terminator	4.5% 5.8%	[35]
Cel 5A	*T. maritima*	80	*Arabidopsis*	RbcS	RA, CAB, RS	Chloroplast	---	5.2%	[48]
Xyl10B	*T. maritima*	80	Tobacco	Tobacco Prrn		Chloroplast	3'UTR of tobacco psbA gene	12-15%	[47]
EGSA (SSO1358)	*Sulfolobus solfataricus*	88	Tobacco	CaMV35S S	KDEL, LPH	Apoplast, ER	5'UTR, pA35S	1.1%	[67]
malA	*S. solfataricus*	88	Tobacco	---	---	---		0.04%	[68]
lacS	*S. solfataricus*	88	Tobacco	---	---	---	---	0.15%	[68]

TSP, total soluble protein; ER, endoplasmic reticulum; Mac promoter, hybrid promoter of manopine synthase promoter and *Cauliflower mosaic virus* 35S promoter enhancer region; CaMV35S, *Cauliflower Mosaic Virus* 35S promoter; Pr1a, pathogenesis related protein 1a; VSPβ, Soybean variant-specific protein; RA, Rubisco activase; CAB, light harvesting chlorophyll *alb*-binding protein; RS, Rubisco small subunit; TMV, translational enhancer (Ω); *Alfalfa Mosaic Virus* 5' untranslated leader; polyadenylation signal of nopaline synthase; 3' UTR of *psbA g*ene; Rubisco large subunit gene *(rbcL)* 5' UTR; CAL, calreticulin signal peptide from carrot; BAASP, barley α-amylase signal sequence; pGlob-1= Globulin-1; VTS, vacuole targeting sequence from maize; E1, endoglucanase from *Acidothermus cellulolyticus*; ND, not determined;