Introduction

The filamentous fungus *Aspergillus oryzae* is known to have a high capacity for protein secretion and has therefore been developed for the industrial production of homologous and heterologous proteins (Archer and Peberdy, 1997). In recent years, because of the availability of genetic engineering techniques and the determination of its genome sequence, *A. oryzae* has attracted a great deal of attention as a host for the production of heterologous proteins (Kitamoto, 2002; Machida et al., 2005). However, for most non-fungal proteins the secreted yields are very low, often two to three orders of magnitude lower than homologous proteins (Gouka et al., 1997a). Therefore, the bottlenecks that limit heterologous non-fungal protein production in *A. oryzae* are a subject of great interest (Conesa et al., 2001).

In general, heterologous gene expression can be inhibited by unsuitable codons and differences in base contents between the expression host and the DNA donor. It causes abnormal splicing (Hamann and
Lange, 2006; Haseloff et al., 1997) and premature polyadenylation (Haffani et al., 2000; Jarvis et al., 1997) and affects mRNA stability (Doma and Parker, 2006; Hoekema et al., 1987). These problems have often been attributed to the A/T-richness of the coding region. In several cases, expression level was improved dramatically by codon optimization (Gouka et al., 1997b; Koda et al., 2005; Xiong et al., 2006). Therefore, codon optimization is thought of as a technique to improve the production of heterologous proteins.

Termites are brought to international attention for the purpose of digesting lignocellulose. Reticulitermes speratus is one of the most-extensively investigated termites in terms of its cellulytic systems (Inoue et al., 1997; Ohtoko et al., 2000; Todaka et al., 2007; Tokuda et al., 2004; Watanabe et al., 1998). It maintains anaerobic microbial communities in its guts which are very small but highly structured microenvironments (Ohkuma, 2008). The symbiotic relationship between the termite and protists in its hindgut is a well-known example of symbiosis. A substantial amount of cellulose ingested by termites can be degraded by the endogenous cellulases, and the cellulose not hydrolyzed in the anterior portion of the gut then travels to the hindgut where it can be endocytosed and fermented by the symbiotic protists (Watanabe et al., 1998). The protists are believed to play important roles in the cellulytic system and contribute to decomposition of wood glucan to an extent greater than 90% (Breznak and Brune, 1994). Extensive analysis of the system is needed, however, to explore its potential for industrial application.

In this study, we attempted to express glycoside hydrolase family (GHF) 45 endoglucanase (RsSym45EG1), GHF7 cellbiohydrolase (RsSym7CBH1), and GHF3 β-glucosidase (RsSym3BG1) from symbiotic protists in the hindgut of the termite R. speratus (Todaka et al., 2007) using the A. oryzae expression system. Although the production level was low when the wild-type RsSym45eg1 sequence was used, it was improved after adjustment of the codon usage to that of tef1 encoding the translation elongation factor 1α. Furthermore, we demonstrated that in cells harboring wild-type RsSym45eg1 as well as RsSym7cbh1 and RsSym3bg1 premature polyadenylation occurred and truncated transcripts were generated. These results suggest that the coding regions of RsSym45eg1, RsSym7cbh1, and RsSym3bg1 contain sequences that can be recognized as polyadenylation signals in A. oryzae, consequently limiting the expression level of these cellulases in A. oryzae.

Materials and Methods

Strains, media, and A. oryzae transformation. A. oryzae NS-tApE (niaD ΔsCel-ΔpepE ΔtppA) deficient for the PepE and TppA proteases was used as a recipient strain for transformation (Jin et al., 2007). 5×DPY medium containing 10% dextrin, 5% polypepton, 2.5% yeast extract, 0.5% K2HPO4 and 0.05% MgSO4 · 7H2O (pH 8.0) was used as the medium for heterologous protein expression. Transformation of A. oryzae was carried out using a standard A. oryzae transformation method (Kitamoto, 2002).

Design and synthesis of codon-optimized RsSym45eg1 and RsSym7cbh1. Codon-optimization of RsSym45eg1 (DDBJ accession number AB045171) was conducted so that its nucleotide sequence matched the codon usage of tef1. In the case of RsSym7cbh1, the nucleotide sequence was modified according to the A. oryzae codon usage database provided by Kazusa DNA Research Institute (http://www.kazusa.or.jp/codon/). The synthetic RsSym45eg1 (i.e. RsSym45eg1-co; AB429160) thus designed was constructed by assembly PCR using thirty overlapping 40 base-long oligonucleotides with Pyrobest DNA polymerase (TaKaRa) essentially based on the procedure of PCR-based two-step DNA synthesis (Xiong et al., 2004). First, three reactions, each containing 10 oligonucleotides, were carried out using 1.5 pmol of 8 inner oligonucleotides and 30 pmol of 2 outer oligonucleotides. Conditions of the first PCR were 25 cycles of 98°C for 10 s, 56°C for 30 s, 72°C for 20 s, with 2.5 U Pyrobest DNA polymerase. The three fragments that were amplified in the first PCR were inserted into pCR-Blunt II-TOPO and the sequences were confirmed. After amplification of three fragments by PCR, fusion PCR (Jin et al., 2004) was carried out with these three fragments and full-length RsSym45eg1-co was obtained. Conditions of the second PCR were 25 cycles of 98°C for 10 s, 56°C for 30 s, 72°C for 60 s, with 2.5 U Pyrobest DNA polymerase. In the case of codon optimization of RsSym7cbh1 (BJ979440), four assembly PCRs, each containing 4 or 6 around 100 bases-long oligonucleotides, were carried out to generate RsSym7cbh1-co (AB429161).

Construction of expression vectors. The insert cDNAs (RsSym45eg1, RsSym7cbh1, and RsSym3bg1
and synthetic DNAs (RsSym45eg1-co and RsSym7cbh1-co) encoding the mature region of cellulases without the predicted signal sequence were amplified by PCR using appropriate sets of forward and reverse primers. Corresponding center entry clones were constructed by BP recombination reaction of the MultiSite Gateway system (Invitrogen). To express the cellulases using α-amylase as a carrier, the expression plasmids were constructed using one of these center entry clones: pg5’PFa (carrying the promoter and the coding regions of α-amylase gene (amyB)), either pg3’TaNiaD (amyB terminator and niaD marker) or pg3’scJ (amyB terminator and A. nidulans sc marker), and the destination vector pDEST R4-R3 in the LR recombination reaction of MultiSite Gateway system. The details of the plasmids were described previously (Mabashi et al., 2006). The transformants harboring the expression plasmid integrated at the niaD locus were selected by Southern blot analysis and used for further experiments.

Endoglucanase activity assay. A. oryzae transformants were cultured in 20 ml of 5×DPY medium at 30°C with shaking for 4 days. Endoglucanase activity was determined by measuring the reducing sugars released from carboxymethyl cellulose (CMC; Kokusan Chemical Works) as described previously (Inoue et al., 2005). The background reaction was conducted without the substrate. Reducing sugar produced was measured using tetrazolium blue reagent (Sigma) as previously described (Inoue et al., 1997). One unit (U) of endoglucanase activity was defined as the amount of enzyme releasing 1 µmol of reducing sugar per min.

Purification of RNA. The strains were incubated for 48 h at 30°C in the 5×DPY medium. After harvesting mycelia, they were washed and immediately frozen in liquid nitrogen. Cells were ground with a mortar and pestle and the resulting powder was transferred to a plastic tube. Total RNA was purified using Isogen (Nippon Gene) following the protocol provided by the manufacturer and stored at −80°C until use.

3’ RACE analysis. 3’ RACE analysis was performed using a GeneRacer Kit (Invitrogen). For RT-PCR, genespecific primers were designed to amplify the full-length RsSym45eg1, RsSym45eg1-co, RsSym7cbh1, RsSym7cbh1-co, and RsSym3bg1. As a forward GSP primer, 5’-GCAAGTTTGTACAAAAAGCAGGCTCAA-3’ was used. As forward GSP nested primers, 5’-GGCTCACAAGCGCGCGGTGCGGAC-3’ for RsSym45eg1 and RsSym45eg1-co, 5’-GCGCGGGCTGGCCTT...

GAGATTGG-3’ for RsSym7cbh1 and RsSym7cbh1-co, and 5’-GCGCGGGCGGTGCGAAGTCTGAAGAG-3’ for RsSym3bg1, were used. The PCR products were gel-purified and sequenced.

Quantitative real-time PCR. The first strand cDNA was synthesized from the total RNA of each region with ReverTra Ace reverse transcriptase (Toyobo) using oligo(dT)12-18 (Invitrogen) as a primer. PCR was performed using a LightCycler FastStart DNA Master SYBR Green I (Roche) following the protocol provided by the manufacturer. RsSym45eg1 and RsSym45eg1-co transcripts were monitored using the forward primer (5’-AACACCGGTGCTGTATCTT-3’) and the reverse primer (5’-CCACTTTGTACAAGAAAGCTG-3’). RsSym7cbh1 and RsSym7cbh1-co transcripts were monitored using the forward primer (5’-GATTCGATCT-GACAAGTTC-3’) and the reverse primer (5’-CCACTTTGTACAAGAAAGCTG-3’).

Results

Design and synthesis of codon-optimized RsSym45eg1 sequence for expression in A. oryzae

Using α-amylase as a carrier, GHF45 endoglucanase (RsSym45EG1) cloned from symbiotic protists in the hindgut of the termite R. speratus was expressed in A. oryzae. The KEX2 cleavage site (-Lys-Arg-) was inserted between the C-terminus of α-amylase and the start of the mature region of RsSym45EG1 (lacking the putative signal sequence composed of the first 14 amino acids) to allow release of RsSym45EG1 from the fusion protein. Although the endoglucanase activity was detected in the culture supernatant of transformants, the amount produced was limited and the protein band was hardly detectable upon SDS-PAGE analysis. Since RsSym45eg1 sequence was relatively AT-rich and displayed a different bias from highly expressed genes in A. oryzae, we envisaged that codon-optimization might improve the production of RsSym45EG1. We therefore synthesized a recombinant gene, designated as RsSym45eg1-co, based on the codon bias of tef1 gene encoding translation elongation factor 1α, since it is one of the most abundantly expressed genes in A. oryzae. Thirty synthetic oligonucleotides, each composed of 40 bases, were used to synthesize the recombinant gene by the two-step DNA synthesis method. The entire coding region of RsSym45eg1 was evenly divided into three regions, each of which was individually amplified in the first...
PCR reaction. The amplified fragments were cloned and the nucleotide sequences were determined. After amplification of three fragments by PCR, fusion PCR was carried out with these three fragments to obtain full-length RsSym45eg1-co. As a result, 51% of the codons in the coding sequence was replaced by more suitable codons and the A+T content decreased from 51% to 41%. Using this sequence, the expression plasmid was constructed and transformed into A. oryzae. Transformants harboring a single copy of RsSym45eg1-co, confirmed by Southern blot analysis (Fig. 1), were named AEGO and used for the subse-

![Diagram](image-url)

Fig. 1. Southern blot analysis of RsSym45EG1-expressing transformants.
(A) Schematic structure of RsSym45eg1 and RsSym45eg1-co expression plasmids. 'S' denotes Sal I sites. (B) The genomic structure of the niaD locus of the NS-tApE strain of A. oryzae (top; the defective niaD locus is shown by 'X'), and the transformants harboring one or two copies of the expression plasmids (middle and bottom, respectively). Bars indicate the positions of the hybridization probe (located within the coding sequence of niaD) used in the Southern blot analysis. The sizes of Sal I fragments are shown in kb. (C) Genomic DNAs from AEG2, 3, 4, and 6 strains harboring the wild-type RsSym45eg1, and AEGO3, 4, 6, and 7 strains harboring RsSym45eg1-co, were digested with Sal I (S), and Southern blot analysis was performed. The probe hybridizes to the Sal I fragments of 8.1 kb (host strain), 10.2 and 6.2 kb (transformants containing one copy of the expression plasmid), and 10.2, 8.3, and 6.2 kb (two copies of the expression plasmid). Positive control experiment in the right-most lane was done with the plasmid shown in A.
sequent experiments, together with the AEG strains harboring a single copy of the wild-type RsSym45eg1. No overall growth difference was observed between AEG and AEGO strains.

Expression of RsSym45eg1 and RsSym45eg1-co in A. oryzae

To examine whether codon optimization affected the level of transcript, we performed RT-PCR analysis of RsSym45eg1 or RsSym45eg1-co expressed in AEG and AEGO strains, respectively (Fig. 2A). Interestingly, as shown in Fig. 2B, we found that the transcript level of RsSym45eg1-co was much higher than that of RsSym45eg1. The quantitative RT-PCR experiment confirmed this observation, showing that the transcription level of RsSym45eg1-co was 1.8-fold higher than that of RsSym45eg1 (Fig. 2C). Furthermore, RT-PCR analysis showed that there were three types of transcripts in AEG strain, one with a size similar to that of the transcript in AEGO (ca. 0.9 kb), and two with smaller sizes (ca. 0.7 kb and 0.4 kb). To elucidate the molecular identity of these three species, 3’ RACE analysis was performed (Fig. 3). We found that the upper-most band was the full-length RsSym45eg1 transcript covering the entire coding region of RsSym45eg1, while the two smaller bands were truncated transcripts terminated at around 300 or 600 nucleotides from the start of RsSym45eg1 sequence and followed by a poly A chain. This suggests that RsSym45eg1 contains polyadenylation signal-like sequences recognized by A. oryzae which caused the truncation of the transcripts. Similar 3’ RACE analysis was conducted with AEGO, but no truncated transcript was found, indicating that codon optimization efficiently prevented premature termination of transcription and polyadenylation, thereby elevating the amount of RsSym45eg1-co transcript.

Next, the culture supernatants of AEG and AEGO strains were analyzed for the endoglucanase activity. As shown in Fig. 4, the activities of AEG and AEGO strains were apparently higher than that of the control strain transformed by the empty vector, indicating that they are attributable to the heterologously-expressed R. speratus protist endoglucanase. A slight increase in the endoglucanase activity was observed in the culture supernatant of AEGO strains compared to that of AEG strains. Moreover, when the activity staining was performed using the SDS gel containing CMC which was stained with Congo Red, diffuse bands of 20–40 kDa were detected in the supernatants containing RsSym45EG1-co, whereas less intense or no bands were seen in the RsSym45EG1-containing or control supernatants, respectively (data not shown). Thus, the production of GHF45 endoglucanase was slightly elevated by codon optimization.

Abnormal processing of the RsSym7cbh1 and RsSym3bg1 mRNA in A. oryzae

To explore the possibility that the production of truncated transcripts often occurs when heterologous non-fungal genes are expressed in A. oryzae, 3’ RACE
analysis was conducted in AC and ABG strains harboring cellobiohydrolase (RsSym7CBH1) and β-glucosidase (RsSym3BG1) of \textit{R. speratus} protist, respectively. Interestingly, in both cases the production of truncated transcripts was observed (Fig. 5, A and B). In the AC strain expressing RsSym7CBH1, transcripts truncated at around 370 nucleotides from the start of \textit{RsSym7cbh1} sequence and followed by a poly A chain were found, in addition to the normal transcript containing the entire coding sequence. In the ABG strain, two types of transcripts, truncated at around 40 or 200 nucleotides from the start of \textit{RsSym3bg1} sequence and followed by poly A chains, were found.

The effect of codon optimization on the 3'-end processing and the amount of transcript was next examined. To this end, the synthetic gene of \textit{RsSym7cbh1} (\textit{RsSym7cbh1-co}) in which the codon usage was adjusted to that of \textit{A. oryzae} was prepared and introduced into \textit{A. oryzae}, and the 3'-RACE and quantitative RT-PCR analyses were performed. No truncated products were amplified in the transformant harboring \textit{RsSym7cbh1-co}. Moreover, the amount of transcript of \textit{RsSym7cbh1-co} was 1.5-fold higher than that of \textit{RsSym7cbh1}, suggesting that codon optimization prevented premature polyadenylation and increased the level of \textit{RsSym7cbh1-co} transcript (Fig. 5C). Cellobiohydrolase activity was not, however, detected in the culture supernatant from AC3 or ACO2 (data not shown), presumably because the production level was too low or the cellobiohydrolase produced was degraded by proteases.

**Discussion**

In this study, we attempted to express cellulases from the protists in the hindgut of \textit{R. speratus} using the heterologous protein expression system of \textit{A. oryzae}. Previous reports have shown that codon optimization of foreign genes caused significant improvement in both the mRNA level and the production of proteins in filamentous fungi (Gouka et al., 1997b; Koda et al., 2005; Te'o et al., 2000). In \textit{A. oryzae}, extreme codon bias is seen for the highly expressed genes, such as those encoding translation elongation factor 1α, glyc-
Premature polyadenylation in Aspergillus oryzae

eraldehydes 3-phosphate dehydrogenase, fructose-1, 6-bisphosphate aldolase, and mitochondrial ADP/ATP carrier protein. Over 94% of amino acid residues of these highly-expressed genes are coded by a select 30 of 61 possible coding triplets. In the current study, 51% of the codons in the coding sequence of RsSym45eg1 was replaced by more suitable codons, resulting in the decrease of the A+T content from 51% to 41%. Consequently, the amount of RsSym45eg1 transcript increased by 1.8-fold compared to the strains harboring the wild-type gene, although the impact of codon-optimization on the protein production was only marginal. Similarly, the transcript level of strains harboring codon-optimized RsSym7cbh1-co was 1.5-fold higher than that of wild-type RsSym7cbh1-expressing cells.

Interestingly, strains expressing the wild-type sequence of RsSym45eg1, RsSym7cbh1, and RsSym3bg1 contained truncated transcripts which terminated within the coding region and were followed by poly A chains. This suggests that the wild-type sequences contain polyadenylation signals for A. oryzae. In higher eukaryotes, the polyadenylation signal is a highly-conserved hexanucleotide sequence, AAUAAA, which is present approximately 10–30 nucleotides upstream of the polyadenylation site (Proudfoot, 1991). In addition, a less conserved U- or GU-rich sequence is usually present approximately 1–30 nucleotides downstream.

Fig. 5. Premature polyadenylation in RsSym7cbh1 and RsSym3bg1.

(A and B) The 3’ end regions of the truncated RsSym7cbh1 (A) and RsSym3bg1 (B) transcripts are shown. The stop codons for RsSym7cbh1 and RsSym3bg1 ORFs are located at 1273–1275 and 1531–1533, respectively. Arrowheads indicate polyadenylation sites. The nucleotides modified by codon-optimization are shown in bold. The boxes and underlines indicate AT- and GT-rich sequences, respectively. (C) The transcript levels of RsSym7cbh1 or RsSym7cbh1-co in AC3 or ACO2 strains, respectively, were measured by real-time PCR. The amount of transcripts was normalized based on the internal actin transcript. StaPE, the control strain harboring the empty vector. The data are presented as mean ± standard deviation of three independent experiments. *p < 0.05 (t-test; n=3).
of the polyadenylation site (Proudfoot, 1991). In contrast, in the yeast Saccharomyces cerevisiae a diffuse AU-rich element is found for polyadenylation signal (Zaret and Sherman, 1984). Although the polyadenylation signal in A. oryzae is yet to be determined and no strictly-conserved sequence was found in the upstream region of the polyadenylation sites of wild-type RsSym45eg1, RsSym7cbh1, or RsSym3bg1 transcripts, AT-rich, polyadenylation signal-like and GT-rich sequences were found in the upstream and downstream regions, respectively (Fig. 3, Fig. 5, A and B). In line with the possibility that these sequences contributed to premature polyadenylation, their modification resulted in the disappearance of truncated transcripts, although other regions were also altered in the synthetic gene and similar pairs of AT- and GT-rich sequences are also found in other areas where polyadenylation did not occur. Thus additional factor(s), such as relative distance of AU- and GU-rich sequences, might have influenced the recognition of the site of polyadenylation. Therefore, whether or not the sequences identified in the wild-type RsSym45eg1, RsSym7cbh1, or RsSym3bg1 sequences actually function as cis-elements in polyadenylation has to be examined experimentally, for example, by translocating the candidate sequence elsewhere to test if the polyadenylation occurs there.

Several cases of improper polyadenylation occurring when foreign genes are expressed in heterologous hosts have been reported: cry gene of Bacillus thuringiensis expressed in potato plants (Haffani et al., 2000); Ac transposase of maize expressed in Arabidopsis thaliana (Jarvis et al., 1997); and α-galactosidase of guar expressed in Aspergillus awamori (Gouka et al., 1997b). In eukaryotes, mRNA prematurely polyadenylated within a coding region is likely to be degraded by a mechanism known as 'nonstop mRNA decay,' whereby transcripts without proper termination codon are degraded (Isken and Maquat, 2007). Alternatively, transcripts with stalled ribosome at rare codons are subjected to degradation by the 'no-go decay' mechanism (Doma and Parker, 2006). These two mechanisms are considered to function as the conserved safeguards for the protection of cells from potentially harmful effects of improperly synthesized proteins, and seem to account for the decreased levels of transcripts of RsSym45eg1 and RsSym7cbh1 compared with those of codon-optimized counterparts (Fig. 2C and Fig. 5C). If heterologous expression of foreign genes is frequently accompanied by premature polyadenylation or stalled ribosome at rare codons, then it is expected that removal of potential polyadenylation sequences or codon-optimization would increase the productivity of the proteins through prevention of abortive transcription. In this sense, carefully analyzing the cumulative data as to the sites of polyadenylation would ensure accurate prediction of potential polyadenylation signals and help eliminate one of the limitations in heterologous protein production, although it is likely that other difficulties lie ahead in establishing the production system of heterologous proteins.

Acknowledgments

This work was supported in part by grants from the New Energy and Industrial Technology Development Organization (NEDO) and Bio-oriented Technology Research Advancement Institution (BRAIN), Japan.

References

Hamann, T. and Lange, L. (2006) Discovery, cloning and hetero-


