Title: Novel expression system for lytic polysaccharide monooxygenases

Gaston Courtade[a], Simone Balzer Le[b], Gerd Inger Sætrom[a], Trygve Brautaset[a], Finn L. Aachmann[a, *]

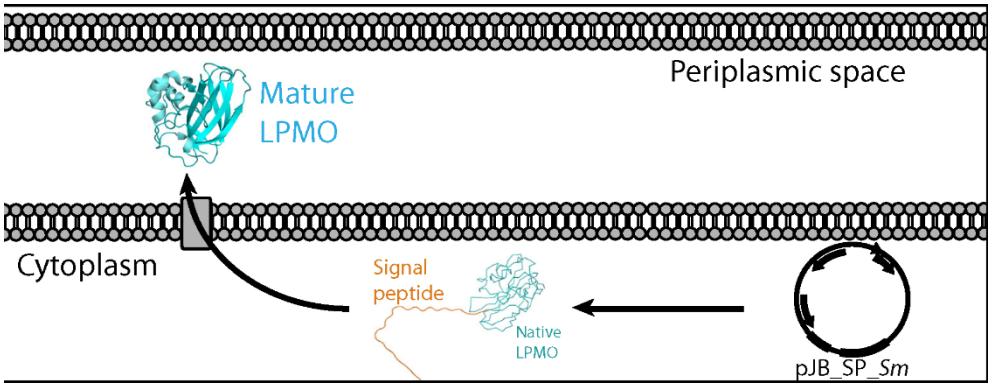
[a] NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491 Trondheim, Norway

[b] Department of Biotechnology and Nanomedicine, SINTEF Materials and Chemistry, N-7034 Trondheim, Norway

* Corresponding author: Sem Sælands vei 6, Kjemi IV 4-140, NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491 Trondheim, Norway. Email: finn.l.aachmann@ntnu.no

Highlights

- An LPMO expression system based on the XylS/*Pm* expression cassette was constructed.
- The signal sequence of *Sm*LPMO10A resulted in efficient translocation of LPMOs.
- The expression system is well suited for producing isotopically enriched LPMOs.
- Successful LPMO production in high cell-density cultivations.



Abstract

Lytic polysaccharide monooxygenases (LPMOs) are key enzymatic players of lignocellulosic biomass degradation processes. As such, they have been introduced in cellulolytic cocktails for more efficient and less expensive lignocellulose saccharification. The recombinant production of LPMOs in bacteria for scientific investigations using vectors typically based on the T7 and *lac*UV5 promoters has been hampered by low yields. Reasons for this have been catabolite repression when producing the proteins in defined media with glucose as the sole carbon source, as well as the lack of an inducible expression system that allows controlled production of LPMOs that are correctly processed during translocation to the periplasmic space. A cassette vector design containing the XylS/*Pm* system was constructed and evaluated, showing that the expression cassette could easily be used for exchanging LPMO coding genes with or without signal sequences. The cassette was shown to reliably produce mature (translocated) LPMOs under controlled conditions that were achieved by using a low dosage (0.1 mM) of the *Pm* inducer *m*-toluic acid and a low (16 °C) cultivation temperature after induction. Furthermore, the signal sequences of five bacterial LPMOs were tested, and the signal sequence of LPMO10A from *Serratia marcescens* was found to give highest levels of recombinant protein production and translocation. The LPMO expression cassette was also evaluated in cultivations using defined media with glucose as the sole carbon source with a product yield of 7-22 mg per L of culture in shaking flasks. The integrity of the recombinant

proteins were analyzed using NMR spectroscopy, showing that the system produced correctly processed and folded LPMOs. Finally, high cell-density cultivations of the recombinant strains were carried out, demonstrating stable protein production levels at similar relative yields (42-1298 mg per L of culture; 3.8-11.6 mg per OD_{600nm} unit) as in shaking flasks, and showing the scale-up potential of the system.

Keywords Heterologous expression system, XylS/*Pm*, protein production, lytic polysaccharide monooxygenase (LPMO), high cell-density cultivation (HCDC)

1. Introduction

Cellulose in lignocellulosic biomass (or chitin in fungi and the exocuticle of arthropods) forms crystalline structures that make the polysaccharide resistant to enzymatic hydrolysis. Additional obstacles to enzymatic degradation are introduced by co-polymeric structures with lignin and hemicelluloses such as xylan, glucomannan or xyloglucan. In order to increase the efficiency of biomass utilization, the degradation process normally requires several different enzymes, including hydrolases and the recently discovered lytic polysaccharide monooxygenases (LPMOs)[1]. LPMOs synergize with hydrolytic polysaccharide degrading enzymes to enhance biomass deconstruction, making them a cornerstone of cost-efficient biorefining processes[1,2,10]. They are bacterial and fungal copper-dependent redox enzymes that catalyze oxidative cleavage of β(1→4) glycosidic linkages and are classified in the auxiliary-activity families AA9, AA10, AA11 and AA13[2–9]. The catalytic site of the LPMOs is centered around a copper ion, which is coordinated by the N-terminal His1 (N^H), its side-chain ($N^{\delta 1}$) and the side-chain ($N^{\varepsilon 2}$) of another His. The premature LPMO is synthesized (translated) with an N-terminal signal peptide, which is cleaved off upon translocation to the periplasmic space to produce the native LPMO domain with an N-terminal His that is able to bind copper to form its active site[2,11].

Bacterial LPMOs have been mainly produced recombinantly using the pRSET and pET vector systems[2,12,13]. These vectors harbor a T7 or a T7lac promoter recognized by T7 RNA polymerase, which is in turn under transcriptional control of the inducible *lac*UV5 promoter in *Escherichia coli* BL21(DE3) or its derivatives[14,15]. It has been observed that induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) results in a mixture of native and premature (non-cleaved signal peptide) LPMO and cell lysis. One solution to this issue has been to rely on the leaky expression of the *lac*UV5 promoter[13] in the absence of IPTG, resulting in LPMO yields in the order of 3-9 mg per L of culture[16,17]. While the leak expression approach has worked well for small-scale batch production of protein in complex media like lysogeny broth (LB), it poses a problem when growing cells in defined media with glucose as the sole carbon source (a requirement to produce isotopically labeled proteins for use in nuclear magnetic resonance (NMR) investigations). A high glucose concentration in the growth medium leads to catabolite repression of the *lac*UV5 promoter, resulting in downregulation of expression of the T7 RNA polymerase encoding gene, and thus low expression of the target LPMO genes[18]. Furthermore, systems based on "leak expression" lack proper control of protein production, leading to large batch-to-batch variations during in our lab.

A solution to this problem is to use an expression system that is not influenced by the carbohydrate metabolism of the cell. One such expression system is based on the XylS/*Pm* promoter/regulator system[19], which uses benzoic acid derivatives (e.g. *m*-toluic acid) as inducers for recombinant gene expression. XylS is a positive transcriptional regulator that upon binding *m*-toluic acid becomes activated and binds to *Pm*, stimulating transcription. Expression vectors relying on the XylS/*Pm* system have several advantages that include: (1) regulation that is independent of the carbohydrate metabolism of the host[20], (2) tight regulation with minimal leak expression in absence of inducer[21], (3) dose-dependent induction[22], (4) broad host-range compatibility by using the RK2 replicon[19], and (5) uniform

expression in the bacterial population as the inducer molecule enters the cell by passive diffusion (no uptake system required)[21]. Recently, the use of the XylS/*Pm* promoter/regulator system for the production of an isotopically labeled LPMO, *Bl*LPMO10A, was demonstrated[24].

In addition to a regulatable promoter, the choice of an appropriate translocation signal sequence is of paramount importance due to the need for correct processing of the premature protein. This ensures the presence of His as the first amino acid in the N-terminus, as the signal peptide is cleaved during translocation to the periplasmic space[25]. Moreover, signal sequences can influence the total expression levels of recombinant proteins[26].

The aim of the present study was thus to design, develop and evaluate a standardized expression system for high-level and functional production and translocation of bacterial LPMOs in isotope labeled media based on the XylS/*Pm* system.

2. Materials and Methods

2. 1 Biological materials and DNA manipulations

All bacterial strains, vectors, primers and LPMO encoding genes used in this study are described in Table 1. *E. coli* strain DH5α was used as a general cloning host, *E. coli* T7 Express and *E. coli* RV308 were used for recombinant protein production in small-scale shake flask cultivations, while *E. coli* RV308 was used as the host for LPMO production under high cell-density cultivations (HCDC) in fermentors. Chemically competent *E. coli* cells were transformed using a heat-shock protocol and grown at 37 °C in liquid or solid LB. All recombinant cell cultures were supplemented with 100 µg/mL ampicillin. Standard recombinant DNA techniques were used, unless stated otherwise, according to Sambrook and Russel[27]. Plasmid DNA was isolated using the Wizard® Plus SV Minipreps DNA purification

system (Promega) or NucleoBond® Xtra Midi (Macherey-Nagel). Linear DNA fragments were purified from agarose gels using the Zymoclean™ Gel DNA Recovery kit (Zymo Research). PCR were run using standard conditions, with the primers listed in Table 1, and the products were purified using the QIAquick® PCR purification kit (Qiagen).

2.2 Bioinformatics

Secondary structure analysis of the signal peptide was performed using the SignalP 4.1 Server[28]. Protein concentrations in the final NMR samples were determined by measuring their A_{280nm} and deducing the protein concentration based on the theoretical extinction coefficient and molecular weight calculated using the ProtParam tool (web.expasy.org/tools/protparam/)[29].

2.3 Construction of the LPMO expression vectors

Totally six different LPMO encoding genes were investigated and they were expressed with their native signal sequence (nSP) or fused to the signal sequence of the *Sm*LPMO10A gene from *Serratia marcescens* (SP), except *Bl*LPMO10A, which was only expressed with SP. Plasmid pGM29[26] without the *gm-csf* insert was used as a vector backbone for the construction of LPMO expression plasmids, and in total ten different LPMO expression vectors were constructed (Table 1). Initially, plasmid pUC57_SP_*Sm* was completely digested with NdeI and HindIII (New England Biolabs) and plasmid pGM29 was partially digested with NdeI and HindIII. The resulting 8.2 kb fragment derived from pGM29 and the 0.6 kb fragment from pUC57_SP_*Sm* were purified, ligated and transformed to *E. coli*. The resulting vector, pJB_SP_*Sm*, could then be used for one-step cloning of the LPMO coding regions using the NcoI/NotI restriction sites, and alternatively the LPMO coding region together with a signal sequence by using the NdeI/NotI sites (Figure 1). As controls, we constructed one vector with the alternative OmpA signal sequence, and the vector pAT64 was

included as a negative control. Target LPMOs and signal sequences were cloned using PCR, except for pJB_SP_*Jd*, which was cloned by sequence- and ligation- independent cloning (SLIC) as described by Jeong et al.[30].

2.4 Production and purification of the recombinant LPMOs from shake flask cultivations

LPMO production and purification, recombinant cells were grown in ^{15}N-labeled M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl) supplemented with 98 % $(^{15}NH_4)_2SO_4$, 4 g/L glucose, 10 mL Bioexpress Cell Growth Media (Cambridge Isotope Laboratories, Tewksbury, MA, USA), 5 mL Gibco™ MEM Vitamin Solution (100x), 2 mM $MgSO_4$ and 10 mL Trace Metal solution (0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, 1 g/L $CaCl_2$) or in 2xLB medium (20 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl).

Pre-cultures were made by inoculating 10 mL LB with recombinant cells followed by incubation at 30 °C and 225 rpm overnight. Main cultures were made in shaking flasks by inoculating 500 mL ^{15}N-labeled M9 medium or 2xLB (100 mL or 500 mL) with 1 % pre-culture and incubating at 30 °C and 225 rpm to $OD_{600\ nm}$ ~0.8. The culture was cooled on ice for 5 min and then induced with 0.1 mM *m*-toluic acid, followed by incubation at 16 °C and 225 rpm for 20 h. Cells were harvested by centrifugation for 10 min at 4000 x g and 4 °C and periplasmic fractions were prepared by the osmotic shock method as follows: The pellet was resuspended in 30 mL spheroplast buffer (1 M Tris-HCl, 0.5 M sucrose, 0.5 mM EDTA, pH 7.5 or 8.5) with half-a-tablet cOmplete™ ULTRA protease inhibitor (Roche). After 5 min incubation on ice followed by a 10 min centrifugation at 5000 x g and 4 °C, the pellet was incubated for 10 min at room temperature and then resuspended on ice in 25 mL ice-cold dd-H_2O supplemented with half-a-tablet cOmplete™ ULTRA protease inhibitor (Roche). 45 seconds after the pellet was fully suspended, 1.25 mL ice-cold $MgCl_2$ (20 mM) were added to

the suspension, which was centrifuged for 10 min at 15000 x g and 4 °C. The supernatant was filtered using a filter (0.22 µm pore size) prior to further protein purification. For total protein extraction experiments, the whole harvested pellet was resuspended in 25 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 1 M $(NH_4)_2SO_4$) and the suspension was sonicated using a Branson Sonifier equipped with a microtip.

*Cj*LPMO10A and *Bl*LPMO10A were purified by loading periplasmic extracts in Buffer A (50 mM Tris-HCl pH 8.5) onto a 5 mL HiTrap® DEAE FF anion exchanger (GE Life Sciences) connected to an ÄKTA Prime Plus FPLC system (GE Life Sciences). Recombinant LPMOs were eluted by using a linear salt gradient (0-500 mM NaCl) over 200 min at a flow-rate of 4.5 mL/min. The fractions containing LPMOs were pooled and concentrated using protein concentrators (10 kDa cut-off, Sartorius). For further purification, the *Cj*LPMO10A sample was loaded onto a HiLoad® 16/600 Superdex® 75 pg size-exclusion column (GE Life Sciences), with a running buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl) and a flow-rate of 1 mL/min.

*Sm*LPMO10A was purified by chitin affinity chromatography as described previously[13], using chitin beads (New England BioLabs) and a buffer consisting of 50 mM Tris-HCl pH 8.0 and 1 M $(NH_4)_2SO_4$. The protein was eluted with 20 mM sodium acetate buffer pH 3.5 using a flow-rate of 2 mL/min. SDS-PAGE gels were run under denaturing conditions using ClearPAGE™ 12 % gels and ClearPAGE™ SDS-R Running buffer (C.B.S. Scientific) followed by staining with InstantBlue™ (C.B.S. Scientific). Precision Plus Protein™ standards (Bio-Rad Laboratories) were used for the identification of target proteins.

Table 1. LPMO genes, vectors, primers and cells used in this study. SP is the signal sequence from *Sm*LPMO10 while nSP is the native and unmodified signal sequence for each protein. All six LPMOs are of bacterial origin and were chosen because of their relevance in recent

publications[24,31,32], and to include both chitin-active and cellulose-active enzymes. Restriction sites in the primer sequences are underlined (NdeI: CATATG, NcoI: CCATGGT, NotI: GCGGCCGC, HindIII: AAGCTT).

Source organism and LPMO	Plasmid	Description	Source
Serratia marcescens *Sm*LPMO10A	pUC57_SP_*Sm*	pUC57 plasmid containing *Sm*LPMO10A from *S. marcescens* with SP	GenScript
	pJB_SP_*Sm*	pJB plasmid containing *Sm*LPMO10A from *S. marcescens* with SP	This study
Bacillus licheniformis *Bl*LPMO10A	pJB_SP_*Bl*	pJB containing *Bl*LPMO10A from *B. licheniformis* with SP	Ref. 24
Jonesia denitrificans *Jd*LPMO10A	pRSET_nSP_*Jd*	pRSET containing *Jd*LPMO10A from *J. denitrificans* with nSP	Ref. 33
	pJB_nSP_*Jd*	pJB containing *Jd*LPMO10A from *J. denitrificans* with nSP	This study
	pJB_SP_*Jd*	pJB containing *Jd*LPMO10A from *J. denitrificans* with SP	This study
Streptomyces coelicolor *Sc*LPMO10C	pJB_nSP_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with nSP	This study
	pJB_SP_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with SP	This study
	pJB_OmpA_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with OmpA	This study
Micromonospora aurantiaca *Ma*LPMO10B	pUC57_nSP_*Ma*	pUC57 containing *Ma*LPMO10B from *M. aurantiaca* with nSP	GenScript
	pJB_nSP_*Ma*	pJB containing *Ma*LPMO10B from *M. aurantiaca* with nSP	This study
	pJB_SP_*Ma*	pJB containing *Ma*LPMO10B from *M. aurantiaca* with SP	This study
Cellvibrio japonicus *Cj*LPMO10A	pUC57_nSP_*Cj*	pUC57 containing *Cj*LPMO10A from *C. japonicus* with nSP	GenScript
	pJB_nSP_*Cj*	pJB containing *Cj*LPMO10A from *C. japonicus* with nSP	This study
	pJB_SP_*Cj*	pJB containing *Cj*LPMO10A from *C. japonicus* with SP	This study
	pGM29	RK2 based plasmid harboring XylS/*Pm* promoter/regulator system, *bla* as reporter gene	Ref. 26
	pAT64	A vector the XylS/*Pm* system controlling the expression of a interferon-α-coding gene with the PelB signal sequence	Ref. 26

Primer	Primer sequences (5′- 3′)
Jd_fwd_nSP	TACTTCATATGAAGAAGAGAAAGTTGAGAGC
Jd_fwd_SP	CAACAAGCCAACGCCCATGGTTGGGTGACAGATCC
Jd_rev	CTAGCAAGCTTGCGGCCGCTTATGAGACCACAACATCCATAC
Sc_fwd_nSP	ATGAACATATGAACAAAACTTCCC
Sc_fwd_SP	CAATGCCCATGGTAGCGTGGTTGATCCG

Sc_rev	AAGCTTGCGGCCGCTTAGCCGCCGTCGAAG
Ma_fwd_nSP	TACTTCATATGTCAACGCCGTATCG
Ma_fwd_SP	CAACGCCCATGGTAGCGTGGTTGATCCGG
Ma_rev	AAGCTTGCGGCCGCTTAGCCAAAGTCAACATCGCTG
Cj_fwd_nSP	TACTTCATATGTTCAATACCCGTCACC
Cj_fwd_SP	CAACGCCCATGGTTATGTGAGCTCTCCGAAAAG
Cj_rev	AAGCTTGCGGCCGCTTAGCCGAAATCAACGTCAATG

Strains	**Description**	**Source**
E. coli T7 Express, a BL21(DE3) derivative	F- λ- *fhuA2* [*lon*] *ompT lacZ::*T7 gene 1 *gal sulA*11 Δ(*mcrC-mrr*)*114::IS10 R*(*mcr*-73::*miniTn10-TetS*)*2 R*(*zgb-210::Tn10)(*TetS) endA1 [*dcm*]	New England Biolabs
E. coli RV308	*lacI*^{q}-, *su*-, Δ*lacX*74, *gal IS II::OP308*, *strA* K12 derivative used for industrial protein production. ATCC strain 31608, deposited by Genentech	ATCC31608
E.coli DH5α	λ^{-} φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17*(r_K^- m_K^-) *supE44 thi-1 gyrA relA1*	Bethesda Research Laboratories

2.5 NMR Spectroscopy

NMR spectra of 0.5-1.2 mM target LPMO samples in 25 mM sodium phosphate buffer pH 5.5 with 10 mM NaCl were recorded at 25°C on a Bruker Ascend 800 MHz spectrometer Avance III HD equipped with a 5-mm Z-gradient CP-TCI (H/C/N) cryogenic probe at the NT-NMR-Center/Norwegian NMR Platform (NNP). NMR data were processed using Bruker TopSpin version 3.5.

2.6 High cell-density cultivations (HCDC)

E. coli RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*) were cultivated in bioreactors to reach high cell-density. All fermentations were performed in 3 L Applikon bioreactors with a 0.75 L working volume. The pH was maintained at pH 6.8 by addition of 30 % (w/v) NH_3 (aq). The dissolved oxygen was maintained at 20 % (v/v) by adjusting the airflow from 0.35 to 1.5 L/L of medium per min.

Fermentations of both strains were carried out using a 3xLB + glycerol medium (30 g/L tryptone, 15 g/L yeast extract, 10 g/L NaCl and 5 g/L glycerol) and Hf medium (main culture medium described previously[34]) supplemented with 100 µg/mL ampicillin. Fermentations on

3xLB were performed in batch, where the main culture medium was inoculated with pre-culture (in LB medium) to give a starting OD_{600nm} ~0.01 and incubated at 30 °C to OD_{600nm} ~3. After induction with 0.5 mM *m*-toluic acid, the incubation temperature was reduced to 16 °C.

Fermentations in Hf medium were carried out in fed-batch mode. Pre-culture medium[34] was inoculated with a starter culture (prepared in LB medium To start the HCDC, Hf was inoculated with pre-culture to a give a starting OD_{600nm} ~0.05 and the fermentation was divided into three phases as follows: (i) Batch phase. Glucose (25 g/L) was added to the freshly inoculated main culture medium, which was incubated at 30 °C until OD_{600nm} ~35. (ii) Exponential feeding phase. Growth was controlled to a specific growth rate (μ) of 0.25 h^{-1} by feeding with 0.6 g/L $MgSO_4$ and 10 g/L glucose. The starting feeding rate was 14 g/h, which was increased exponentially to 35 g/h, and thereafter remained constant. (iii) Induction phase. When OD_{600nm} ~90 had been reached, the culture was induced with 0.5 mM *m*-toluic acid, the incubation temperature was reduced to 16 °C, and the feeding rate was reduced to 5 g/h. Cell concentration was measured as OD_{600nm}. The protein concentration in the HCDC was estimated by the following band densitometry analysis. Samples were taken from the cultures at different time points and analyzed on SDS-PAGE. Digital images of the SDS-PAGE gels were captured with the ChemiDoc™ XRS+ system (Bio-Rad Laboratories) in transillumination mode and the images were analyzed using the Image Lab™ software (Bio-Rad Laboratories). The protein concentration in the bands at ~20 kDa (i.e. the molecular weight of LPMOs) was estimated by the ratio between their intensity and the intensity of the Bio-Rad Precision Plus Protein™ Dual Color Marker standard band at 20 kDa, which has a known concentration of 15 ng/μl. Cultures were harvested by centrifugation at 4500 x g for 10 min. Periplasmic extraction and purification were carried out using the same methodology as described in the previous section.

3. Results and discussion

The need for a reliable expression system for the production of LPMOs led to the design and construction of a plasmid vector based on the XylS/*Pm* expression cassette. In order to evaluate the performance of the system, five bacterial (AA10) LPMOs were cloned in the newly designed vector together with their native signal sequence. Moreover, six LPMO coding regions were combined with the signal sequence from *Sm*LPMO10A (Table 2) to test the effect of the signal peptide for translocation and overall production of mature LPMOs. In total, eleven vectors were constructed, and LPMO production levels were evaluated for all the respective recombinant *E. coli* strains.

3.1 Construction of a new LPMO expression cassette based on the XylS/*Pm* expression system

A novel, carbon-source independent expression cassette pJB_SP_*Sm* was developed with the aim of producing correctly processed LPMOs from the *Pm* promoter in *E. coli* using isotopically enriched media (Figure 1). The result is a cassette design, allowing the LPMO coding region to be easily exchanged by using a combination of restriction enzymes for linearization of the backbone. Insertion of an LPMO gene while keeping the signal sequence of *Sm*LPMO10A is possible with NcoI and NotI. Inserting an LPMO gene with its own signal sequence requires NdeI and NotI, while a combination of NdeI and NcoI exchanges the signal sequence region only. Many bacterial signal sequences like OmpA, PelB, MalE end with the amino acids Ala-X-Ala, and so do the LPMO signal sequences included herein (Table 2). Using the NcoI site, the transition from Ala in the signal sequence and the His in the start of the LPMO gene (Figure 1) should ensure that the N-terminus of the LPMO is processed correctly during translocation.

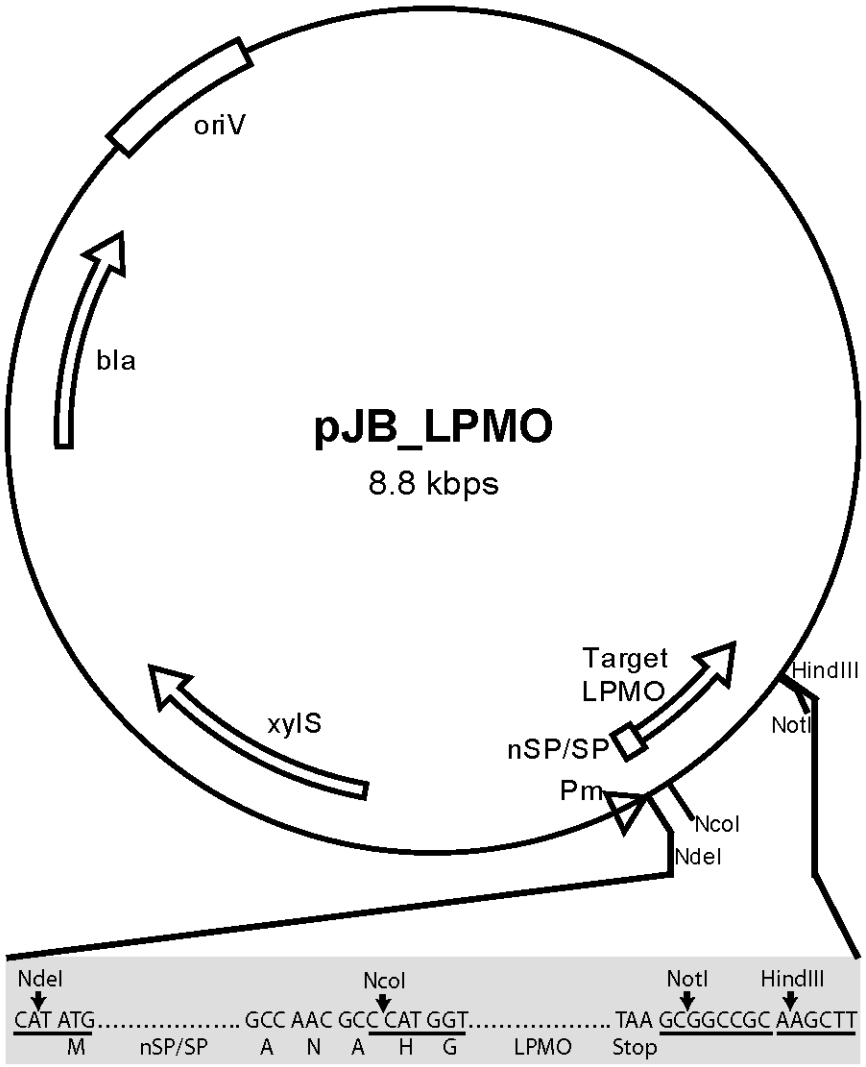


Figure 1. Vector map of the LPMO expression plasmid pJB_SP_*Sm*. The RK2 replicon (*oriV* and *trfA271C*), the β-lactamase *bla* gene, the *xylS* gene, and the *Pm* promoter region are depicted. The cloning sites for the native signal sequence (nSP) or the *Sm*LPMO10A signal sequence (SP), and the target LPMO gene downstream of the *Pm* promoter, as well as the reading frame are shown in the detail in the grey box. Additional elements relevant for stable maintenance as well as transcriptional terminators are not shown (see ref. 22 for further information).

Table 2. Amino acid sequence of the native signal peptides for the selected LPMOs and an OmpA-based signal peptide. Gram classification and order of the source organism are

included. Text formatting indicates the different features of the sequences as follows. **Bold**: N-domain, regular: H-domain, underlined: C-domain, *italics*: helix-breaking residues Gly and Pro, ***bold italics:*** interspersed positively charged residue.

Source	Amino acid sequence	Gram classification	Order
*Sm*LPMO10A	**MNKTSR**TLLSL*G*LLSAAMF*G*VSQQANA	Gram-negative	Enterobacteriales
*Jd*LPMO10A	**MKKRKLR**ASAAIAVLL*GA*G*L*V*P*ALSAT*P*AAA	Gram-positive	Actinomycetales
*Sc*LPMO10C	**MVRRTR**LLTLAAVLATLL*G*SL*G*VTLLL*GQG*RAEA	Gram-positive	Actinomycetales
*Cj*LPMO10A	**MFNTRH**LLAGVSQLV***K****P*ASMMILAMASTLAIHEASA	Gram-negative	Pseudomonales
*Ma*LPMO10A	**MSTPYRR***P*L*P*LAAAIL*G*VCAVVAALLTTAFS*GP*ASA	Gram-positive	Actinomycetales
OmpA-like (*E. coli*)	**MKK**TAIAIAVALA*G*FATVAQAAAMA	Gram-negative	Enterobacteriales

Overall, all signal sequences used have the standard signal peptide architecture consisting of the N-region (composed of 1-5 residues), the hydrophobic H-region (7-15 residues) and the C-region (3-7 residues), which is a determinant of specificity[35]. In particular, the end of the C-region has a characteristic Ala-X-Ala motif[36], where the cleavage site is located. However, they differ in typical features such as the number of positively charged amino acids in the N-region, the number and position of helix-breaking Pro, Gly or Ser residues[37] as well as their length. Secondary structure analysis revealed that two of the signal peptides (*Cj*LPMO10A and OmpA-like) most likely form two consecutive helices instead of a typical single alpha helix in the H-domain, and that the signal sequence from *Sm*LPMO10A seems to be in best agreement with this "consensus" signal sequence.

3.2 The pJB_SP_*Sm* and pJB_SP_*Bl* vectors can be used for efficient production and translocation of LPMOs

Controlled protein production is a prerequisite for correct processing of the signal peptide and translocation of LPMOs to the periplasmic space. A hallmark of the XylS/*Pm* expression

cassette is the dependency on the inducer dosage[22], allowing control of the recombinant protein production levels. The constructed expression vectors were evaluated by cultivating *E. coli* strains RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*), expressing *Sm*LPMO10A and *Bl*LPMO10A, respectively (see Table 1), in 100 mL shake flasks, and protein production was controlled by using a low inducer concentration (0.1 mM *m*-toluic acid) and a low cultivation temperature (16 °C). The semiquantitative amount of translocated protein (Table 3) was determined by using SDS-PAGE from the intensity of the bands at 21 kDa and 19.4 kDa (*Sm*LPMO10A and *Bl*LPMO10A, respectively) in the periplasmic extract fraction, and a ½ dilution series (Figure 2). The absence of visible bands at 21 kDa and 19.4 kDa in the cell pellet after periplasmic extraction (Figure 2) indicates low accumulation of recombinant protein in the cytoplasm. These results show that, under the chosen cultivation conditions, the XylS/*Pm* expression cassette efficiently produced LPMOs, which were then successfully translocated to the periplasm. Moreover, the expression system showed minimal levels of leak expression (in absence of inducer), a feature consistent with the XylS/*Pm* system[21].

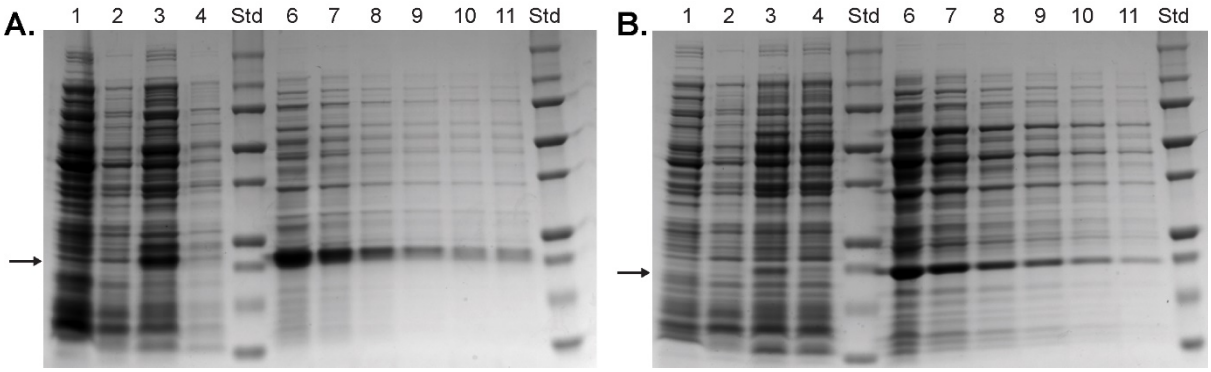


Figure 2: SDS-PAGE analysis of recombinant strains producing *Sm*LPMO10A and *Bl*LPMO10A. Panel A corresponds to pJB_SP_*Sm*, Panel B corresponds to pJB_SP_*Bl*. For both panels the lanes were as follows: (1) induced negative control vector (pAT64), (2) non-induced cells, (3) induced cells, (4) cell pellet after periplasmic extraction, (6-11) serial ½ dilutions of the extract, and (Std) Bio-Rad Precision Plus Protein™ Dual Color Marker. The

arrows mark bands corresponding to the molecular weights of *Sm*LPMO10A (21 kDa, Panel A) and *Bl*LPMO10A (19.4 kDa, Panel B).

3.3 The native signal sequence of *Sm*LPMO10A can be used for efficient translocation of heterologous LPMOs

A signal peptide is vital for the correct translocation of the protein to the periplasmic space. Since the native signal sequences of the target LPMOs (Table 2) showed large translocation variability, it was of interest to test the effects of substituting with alternative signal sequences. Thus, using the XylS/*Pm* cassette design, LPMO genes were inserted with their native signal sequences (nSP) or with the signal sequence of *Sm*LPMO10A (SP), which has been shown to successfully produce mature (correctly processed) LPMOs[24].

Table 3 shows a summary of the test expressions of the six target LPMOs. Where applicable, a comparison of the production level of mature protein for each signal peptide is given. The protein levels were assessed semiquantitatively using serial ½ dilutions of the periplasmic extracts and SDS-PAGE, as described in the previous section. The results show that *E. coli* consistently produced higher amounts of mature LPMOs with SP as the fusion partner compared to when using nSP. In addition, an *E. coli* OmpA-like signal peptide was tested together with *Sc*LPMO10C, but using this signal sequence did not result in higher protein production levels than SP.

Table 3. Relative expression levels of the different mature LPMOs using three different signal peptides are indicated with plus (+) signs, where +++ corresponds to the highest observed level of expression. N/A means that data for that LPMO/signal peptide combination is not available. In the case of *Sm*LPMO10A, nSP and SP are the same native signal peptide.

LPMO	Source	Native signal peptide (nSP)	*Sm*LPMO10A signal peptide (SP)
*Sm*LPMO10A	*S. marcescens*	+++	
*Jd*LPMO10A	*J. denitrificans*	+	+++
*Sc*LPMO10C	*S. coelicolor*	+	+++
*Cj*LPMO10A	*C. japonicus*	++	+++
*Ma*LPMO10A	*M. aurantiaca*	+	+++
*Bl*LPMO10A	*B. licheniformis*	N/A	+++

It is known that combinations of host strain, signal sequences and target proteins can influence translocation, solubility and expression levels[26,38]. It is likely to assume that signal sequences from more closely related bacteria would work better in a certain expression host (Table 2). Indeed, the SP from *S. marcescens,* which is a Gram-negative Enterobacterium commonly found in the gastrointestinal tract just like *E. coli*, was here demonstrated to be the best alternative from among the LPMOs that were screened. The signal sequence from *C. japonicus* LPMO10A led to intermediate expression levels in *E. coli*. *C. japonicus* was originally classified as *Pseudomonas fluorescens* ssp. *cellulosa,* which is also a mesophilic Gram-negative bacterium. The remaining LPMOs originate from Gram-positive bacteria, which are characterized by marked differences in the signal sequences, such as a longer H-

region than signal sequences from Gram-negative organisms[39]. Therefore, it would appear that the choice of signal sequence is primarily dependent on the origin of the target protein.

3.4 The XylS/*Pm* expression cassette is useful for high-level production of isotopically enriched LPMOs

NMR spectroscopy is a versatile and powerful technique for the study of protein structure and function. NMR has provided insights into various aspects of LPMOs including mobility, metal binding, and substrate interactions[40,41]. However, NMR studies require ~0.5-1 mM of LPMO containing the NMR active ^{15}N and ^{13}C isotopes, which have natural abundances of 0.36 % and 1.1 %, respectively. This means that LPMOs must be produced in isotopically enriched media with defined C and N sources such as ^{13}C-glucose and $(^{15}NH_4)_2SO_4$[42]. As glucose represses the *lac* operon, the production of LPMOs in defined media with T7 or T7lac promoters containing vectors (such as pRSET or pET) in *E. coli* strains harboring the T7 polymerase gene (typically under control of the *lac*UV5 promoter) represents a bottleneck for NMR investigations.

Regulation of the XylS/*Pm* system is independent from the carbon source used in *E. coli*. Therefore, the LPMO expression cassette was further tested to evaluate its performance when producing isotopically enriched LPMOs with a defined carbon source (i.e. glucose). Recombinant strains RV308 (pJB_SP_*Sm*), RV308 (pJB_SP_*Bl*) and T7 Express (pJB_SP_*Cj*), were cultivated in defined ^{15}N-enriched M9 medium using 500 mL shake flasks. These three LPMOs were selected based on the simplicity of their purification protocols (see Materials and Methods) as well as previous knowledge of their ^{15}N HSQC fingerprint spectra[24,43]. Following subsequent purification, the three isotope labeled proteins were analyzed in terms of protein yields and quality using SDS-PAGE, UV/Vis and NMR spectroscopy. The SDS-PAGE results (Figure 3, top row) show that the elution fractions after

single step purification (Table 4) contained more than 95% pure protein. Although the data show the presence of weak double bands, the absence of strong, overlapping signals in the center of NMR spectra indicates that this is not a result of incorrect processing of the signal peptide[42,44]. The recombinant protein yields range from 7-26 mg per L of culture (Table 4), which were sufficiently high for NMR analyses, resulting in ^{15}N-HSQC spectra characteristic of pure, correctly processed and folded proteins (Figure 3, bottom row). Moreover, the product yields were two- to three-fold higher than previously reported for LPMOs produced in LB media[16,17]. This finding also highlights the positive effect of the signal sequence from *Sm*LPMO10A on expression and translocation of the three tested LPMOs.

Table 4: Amount of recombinant LPMO in *E. coli* after incubation in defined ^{15}N-enriched M9 medium and purification.

Plasmid	**Purification method**	**Yield (mg per L of culture)**	
pJB_SP_*Sm*	Chitin affinity column	2xLB	26
		M9	22
pJB_SP_*Bl*	Anion exchange DEAE FF	M9	7
pJB_SP_*Cj*	Anion exchange DEAE FF and SEC	M9	10

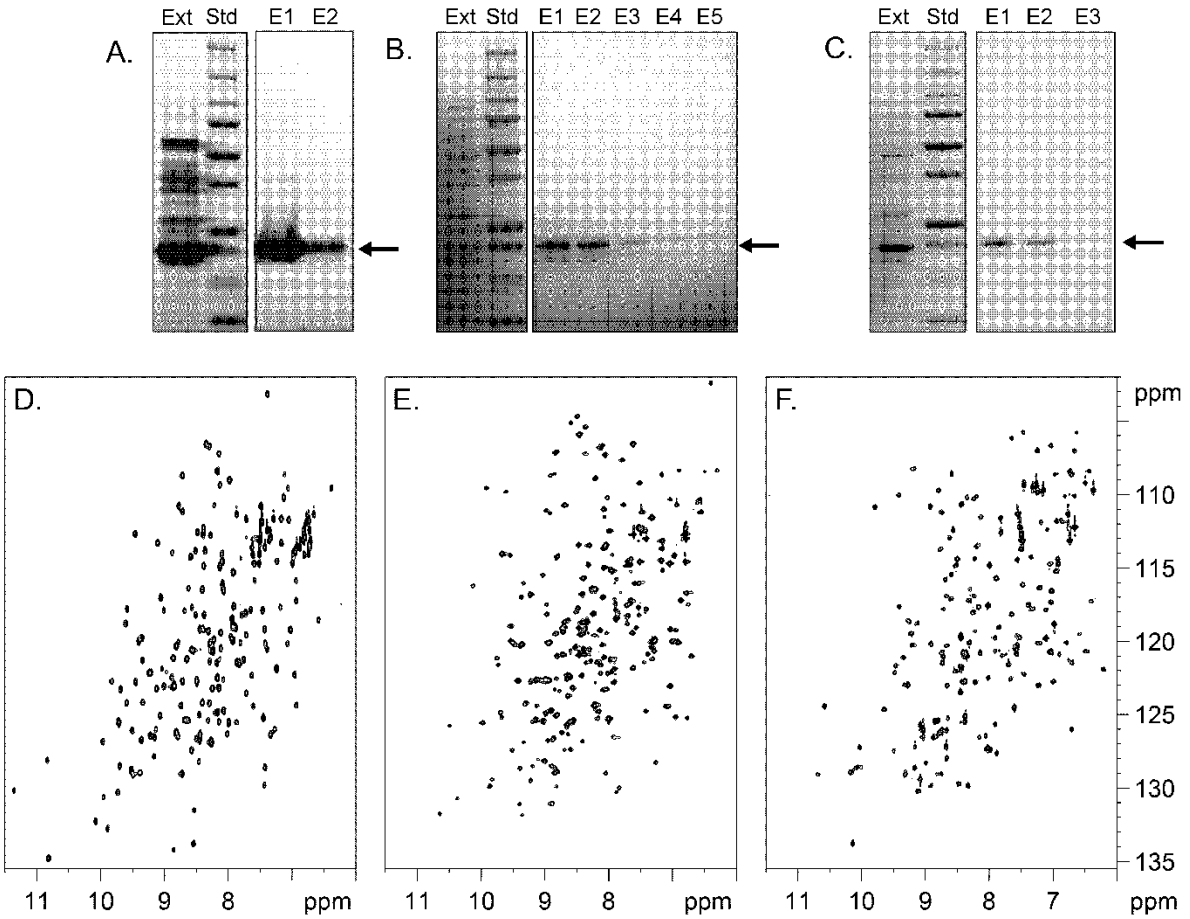


Figure 3: Top row: Purification of LPMOs from defined ^{15}N-enriched M9 medium. (Std) Bio-Rad Precision Plus Protein™ Dual Color Marker, (Ext) periplasmic extract. Bottom row: ^{15}N HSQC spectra of the LPMO samples, showing the characteristics of pure and correctly processed proteins[42]. A&D: *Sm*LPMO10A, B&E: *Bl*LPMO10A, C&F: *Cj*LPMO10A. . The arrows mark bands corresponding to the molecular weights of each LPMO.

3.5 The XylS/*Pm* expression cassette is applicable to high cell-density cultivation for production of LPMOs

Fermentation in a bioreactor allows precise control of the specific growth rate (μ), ensures that the bacterial population is sampled homogenously, and provides valuable information for further scale-up and optimization of the cultivation process.

The performance of the XylS/*Pm* at HCDC was evaluated by cultivating the recombinant *E. coli* strains RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*) in bioreactors run in batch-mode and 3xLB medium or fed-batch mode and Hf defined medium. The fermentation courses (Figure 4) show that the production of LPMOs started immediately after XylS/*Pm* induction and, for the fermentations in defined medium, the maximum amount of LPMO was produced within 5 h. Production of LPMOs in 3xLB medium followed cell-growth. Overall, LPMO production remained stable at high cell-density, as observed previously for other XylS/*Pm* containing vectors[22]. Whereas low cell-densities resulted in (protein) yields of 42 and 78 mg per L of culture (3.8 and 11.6 mg per OD unit) for 3xLB in batch-mode (Table 5), which were similar to the yields obtained with shaking flasks (Table 4), cultures in defined medium in fed-batch mode reached higher cell-densities and yields at 514 and 1298 mg per L of culture (3.9 and 9.2 mg per OD_{600nm} unit). No signs of plasmid instability were observed during the fermentation, as the protein yields were maintained or increased, which is expected for the XylS/*Pm* system[22]. The tunable induction of XylS/*Pm* also enabled a controlled and efficient translocation of the target proteins even at high-cell densities, as observed for the shaking flask cultivations. Even at high levels of recombinant LPMO production, no indication of host toxicity was observed during the fermentation. This indicates the scale-up potential of LPMO production using this LPMO expression cassette, even though no further optimization of the cultivation conditions was performed.

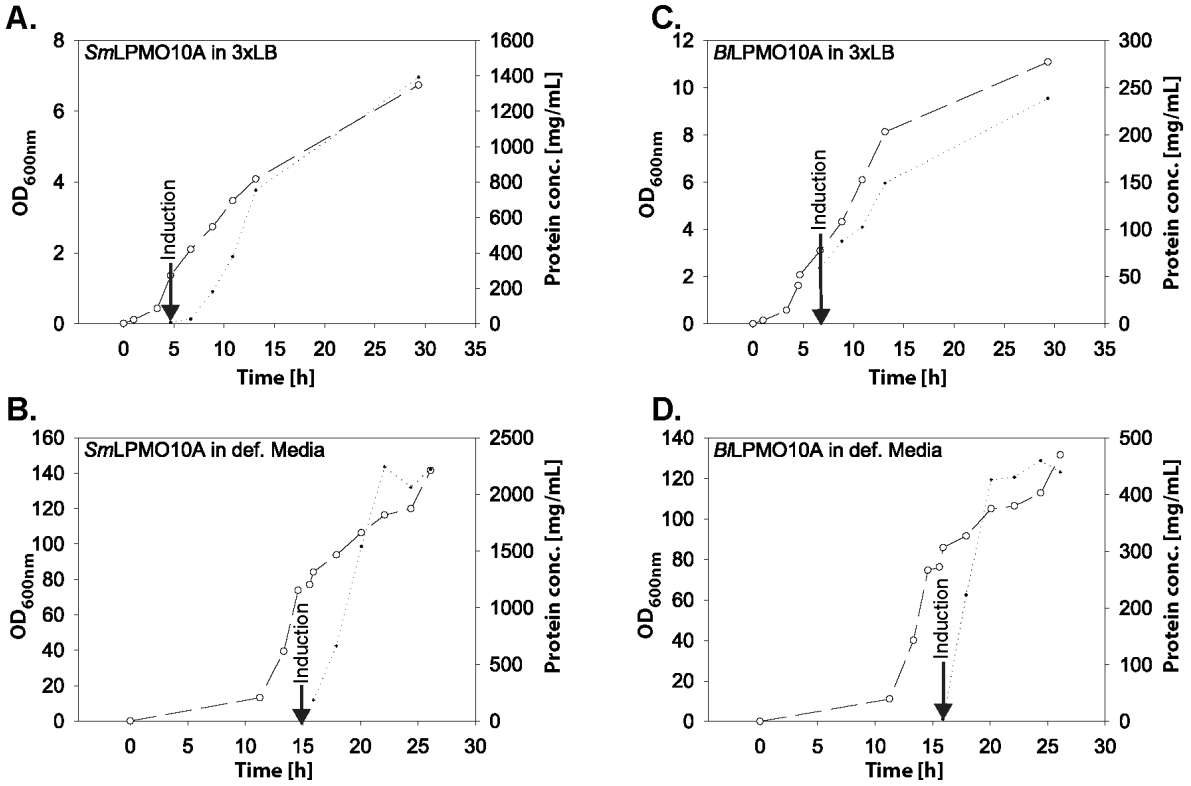


Figure 4: Fermentation course of *E. coli* RV308 containing pJB_SP_*Sm* (A and B) or pJB_SP_*Bl* (C and D) in 3xLB and Hf defined medium. Time courses of bacterial growth (OD_{600nm}) are shown in dashed lines and LPMO production is shown in dotted lines. An arrow indicates the time of induction with 0.5 mM *m*-toluic acid.

Table 5: Amount of pure, recombinant LPMO at harvest times in *E. coli* RV308 after HCDC in 3xLB or Hf defined medium**.** The yields in mg per L of culture were estimated from the SDS-PAGE bands of the elution fractions using the same band densitometry analysis described in Materials and Methods. To facilitate comparison between the different cultures, yields are also given as mg per OD_{600nm} unit.

Plasmid	**Media**	**Yield (mg per L of culture)**	**Yield (mg per OD_{600nm} unit)**
pJB_SP_*Sm*	Defined medium	1298	9.2

	3xLB	78	11.6
pJB_SP_*Bl*	Defined medium	514	3.9
	3xLB	42	3.8

4. Conclusions

The need for the development of a reliable microbial system for LPMO production, particularly in isotopically enriched media, led to the construction of a new LPMO expression cassette based on the XylS/*Pm* system. The combination of the XylS/*Pm* promoter/activator system together with the signal sequence from *Sm*LPMO10A was shown to give highest levels of mature LPMOs. The LPMO expression cassette also performed well under HCDC conditions, as well as when producing isotopically enriched LPMOs in defined media with glucose as the sole carbon source. Altogether, these outcomes show that the LPMO expression cassette has potential use for the production of LPMOs for scientific and commercial applications. Further studies would comprise optimization of other features of the vector (e.g. XylS/*Pm* variant, 5′-UTR, plasmid copy number), cultivation parameters (such as composition of the cultivation media, temperature and inducer dosage), and host strains.

Acknowledgements

This work was financed by strategic funds from the Norwegian University of Science and Technology, the MARPOL project, and NNP - Norwegian NMR Platform (grants 221576 and 226244 from the Research Council of Norway, respectively). We thank our partners at Vectron Biosolutions AS and the Protein Engineering and Proteomics group at the Norwegian University of Life Sciences who provided insight and expertise that greatly assisted the research. Anne Tøndervik, Per Odin Hansen and Randi Aune from SINTEF are thanked for excellent technical assistance.

References

1. Horn, S.J.; Vaaje-Kolstad, G.; Westereng, B.; Eijsink, V.G.H. *Biotechnol. Biofuels*. 2012, **5**, 45.

2. Vaaje-Kolstad, G.; Westereng, B.; Horn, S.J.; Liu, Z.; Zhai, H.; Sørlie, M.; et al. *Science*. 2010, **330**, 219–222.

3. Quinlan, R.J.; Sweeney, M.D.; Lo Leggio, L.; Otten, H.; Poulsen, J.-C.N.; Johansen, K.S.; et al. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**, 15079–15084.

4. Phillips, C.M.; Beeson, W.T.; Cate, J.H.; Marletta, M.A. *ACS Chem. Biol.* 2011, **6**, 1399–1406.

5. Kim, S.; Ståhlberg, J.; Sandgren, M.; Paton, R.S.; Beckham, G.T. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**, 149–154.

6. Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P.M.; Henrissat, B. *Biotechnol. Biofuels*. 2013, **6**, 41–54.

7. Hemsworth, G.R.; Henrissat, B.; Davies, G.J.; Walton, P.H. *Nat. Chem. Biol.* 2014, **10**, 122–126.

8. Vu, V. V; Beeson, W.T.; Span, E.A.; Farquhar, E.R.; Marletta, M.A. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**, 13822–13827.

9. Lo Leggio, L.; Simmons, T.J.; Poulsen, J.N.; Frandsen, K.E.H.; Hemsworth, G.R.; Stringer, M.A.; et al. *Nat. Commun.* 2015, **6**, 1–9.

10. Johansen, K.S. *Biochem. Soc. Trans.* 2016, **44**, 143–149.

11. Moser, F.; Irwin, D.; Chen, S.; Wilson, D.B. *Biotechnol. Bioeng.* 2008, **100**, 1066–

1077.

12. Forsberg, Z.; Mackenzie, A.K.; Sørlie, M.; Røhr, Å.K.; Helland, R.; Arvai, A.S.; et al. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**, 8446–8451.

13. Vaaje-Kolstad, G.; Houston, D.R.; Riemen, A.H.K.; Eijsink, V.G.H.; van Aalten, D.M.F. *J. Biol. Chem.* 2005, **280**, 11313–11319.

14. Kennell, D.; Riezman, H. *J. Mol. Biol.* 1977, **114**, 1–21.

15. Rosano, G.L.; Ceccarelli, E.A. *Front. Microbiol.* 2014, **5**, 1–17.

16. Hemsworth, G.R.; Taylor, E.J.; Kim, R.Q.; Gregory, R.C.; Lewis, S.J.; Turkenburg, J.P.; et al. *J. Am. Chem. Soc.* 2013, **135**, 6069–6077.

17. Gregory, R.C.; Hemsworth, G.R.; Turkenburg, J.P.; Hart, S.; Walton, P.H.; Davies, G.J. *Dalton Trans.* 2016, **45**, 16904–16912.

18. Griffiths, A.J.; Miller, J.H.; Suzuki, D.T.; Lewontin, R.C.; Gelbart, W.M. Catabolite repression of the lac operon: positive control. In *An Introduction to Genetic Analysis*, 7th ed; W.H. Freeman: New York, 2000; pp. 342–343.

19. Blatny, J.M.; Brautaset, T.; Haugan, K.; Blatny, J.M.; Brautaset, T.; Winther-Larsen, H.C. *Appl. Environ. Microbiol.* 1997, **63**, 370–379.

20. Brautaset, T.; Petersen, S.B.; Valla, S. *Metab. Eng.* 2000, **2**, 104–114.

21. Balzer, S.; Kucharova, V.; Megerle, J.; Lale, R.; Brautaset, T.; Valla, S. *Microb. Cell Fact.* 2013, **12**, 1–14.

22. Sletta, H.; Nedal, A.; Aune, T.E. V; Hellebust, H.; Hakvåg, S.; Aune, R.; et al. *Appl. Environ. Microbiol.* 2004, **70**, 7033–7039.

23. Berg, L.; Lale, R.; Bakke, I.; Burroughs, N.; Valla, S. *Microb. Biotechnol.* 2009, **2**, 379–389.

24. Courtade, G.; Balzer, S.; Forsberg, Z.; Vaaje-Kolstad, G.; Eijsink, V.G.H.; Aachmann, F.L. *Biomol. NMR Assign.* 2015, **9**, 207–210.

25. Mergulhão, F.J.M.; Summers, D.K.; Monteiro, G.A. *Biotechnol. Adv.* 2005, **23**, 177–202.

26. Sletta, H.; Tøndervik, A.; Hakvåg, S.; Vee Aune, T.E.; Nedal, A.; Aune, R.; et al. *Appl. Environ. Microbiol.* 2007, **73**, 906–912.

27. Sambrook, J.; Russell, D.W. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2001.

28. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. *Nat. Methods*. 2011, **8**, 785–786.

29. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; et al. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handook*; J.M. Walker, Ed.; Springer Publishing: New York, 2005; pp. 571–607.

30. Jeong, J.Y.; Yim, H.S.; Ryu, J.Y.; Lee, H.S.; Lee, J.H.; Seen, D.S.; et al. *Appl. Environ. Microbiol.* 2012, **78**, 5440–5443.

31. Forsberg, Z.; Røhr, A.K.; Mekasha, S.; Andersson, K.K.; Eijsink, V.G.H.; Vaaje-Kolstad, G.; et al. *Biochemistry*. 2014, **53**, 1647–1656.

32. Forsberg, Z.; Nelson, C.E.; Dalhus, B.; Mekasha, S.; Loose, J.S.M.; Crouch, L.I.; et al. *J. Biol. Chem.* 2016, **291**, 7300–12.

33. Mekasha, S.; Forsberg, Z.; Dalhus, B.; Bacik, J.-P.; Choudhary, S.; Schmidt-Dannert, C.; et al. *FEBS Lett.* 2016, **590**, 34–42.

34. Horn, U.; Strittmatter, W.; Krebber, A.; Knüpfer, U.; Kujau, M.; Wenderoth, R.; et al. *Appl. Microbiol. Biotechnol.* 1996, **46**, 524–32.

35. Paetzel, M.; Strynadka, N.C.J. *CSBMCB/SCBBMC Bull.* 2001, 60–65.

36. von Heijne, G. *Eur. J. Biochem.* 1983, **133**, 17–21.

37. von Heijne, G. *J. Membr. Biol.* 1990, **115**, 195–201.

38. Retallack, D.M.; Schneider, J.C.; Mitchell, J.; Chew, L.; Liu, H. *Biotechnol. Lett.* 2007, **29**, 1483–1491.

39. von Heijne, G.; Abrahmsèn, L. *FEBS Lett.* 1989, **244**, 439–446.

40. Aachmann, F.L.; Sørlie, M.; Skjåk-Bræk, G.; Eijsink, V.G.H.; Vaaje-Kolstad, G. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**, 18779–18784.

41. Courtade, G.; Wimmer, R.; Røhr, Å.K.; Preims, M.; Felice, A.K.G.; Dimarogona, M.; et al. *Proc. Natl. Acad. Sci.* 2016, **113**, 5922–5927.

42. Kwan, A.H.; Mobli, M.; Gooley, P.R.; King, G.F.; MacKay, J.P. *FEBS J.* 2011, **278**, 687–703.

43. Aachmann, F.L.; Eijsink, V.G.H.; Vaaje-Kolstad, G. *Biomol. NMR Assign.* 2011, **5**, 117–119.

44. Chupin, V.; Killian, J.A.; Breg, J.; de Jongh, H.H.; Boelens, R.; Kaptein, R.; et al. *Biochemistry*. 1995, **34**, 11617–11624.

45. Pechsrichuang, P.; Songsiriritthigul, C.; Haltrich, D.; Roytrakul, S.; Namvijtr, P.;

Bonaparte, N.; et al. *Springerplus*. 2016, **5**, 1–10.