Recombinant protein expression of *Magnaporthe oryzae* MGG06069 and its polysaccharide monooxygenase domain in *Escherichia coli*

Bui Cong Chinh, Ho Ta Giap, Vu Van Van

NTT Hi-tech Institute, Nguyen Tat Thanh University
chinhbc@ntt.edu.vn

Abstract

MGG06069 is a polysaccharide monooxygenase protein that originated from the fungus that causes rice blast, *Magnaporthe oryzae*. It has the potential to speed up the cellulolysis process, which can be beneficial for industry. In this study, the full-length MGG06069 protein and its polysaccharide monooxygenase domain were expressed in *Escherichia coli* in order to produce recombinant proteins on a large scale for further enzymatic experiment. It was shown that the amount of the recombinant polysaccharide monooxygenase domain protein was much higher compared to that of the full-length protein. In addition, only one band of expected size was obtained for the recombinant PMO domain protein. These results suggest the purified polysaccharide monooxygenase domain of MGG06069 that can be used for enzymatic assays in the future.

Received 21/03/2022
Accepted 20/08/2022
Published 12/09/2022

Keywords
AA9, MGG06069, *Escherichia coli*, Polysaccharide Monooxygenase, PMOs

1 Introduction

Polysaccharide monooxygenase (PMO) is a group of proteins that have the ability to oxidize the glycosidic linkage of the polysaccharide thanks to its monocopper active site. This ability allows PMO to break down or enhance other enzymes to hydrolyse long and complex polysaccharide chains (cellulose, chitin, etc) [1]. In the CAZy and Uniprot databases, MGG06069 is a PMO belonging to the AA9 group. In nature, this protein is secreted by *Magnaporthe oryzae*, (*M. oryzae*) a rice blast fungus [2]. The MGG06069 protein consists of a signal peptide, a glyco-hydro-61 domain (PMO domain), and a chitin-binding domain [3]. In one study, it has been shown that during the appressorium morphogenesis, one of the first steps in the *M. oryzae*'s infection process, the expression of the *mgg06069* gene increases dramatically [4]. This suggests that MGG06069 may have an essential role in the infection process. It is possible that MGG06069 protein assists *M. oryzae* in breaking down the cellulose wall, thus making way for the pathogen to enter the host. Through studying MGG06069, we can have a better understanding of how *M. oryzae* infects rice and from that there is a chance that MGG06069 can be utilized as an enzyme to speed up the cellulolysis. Therefore, it is necessary to produce MGG06069 on a large scale by recombinant expression.

Nowadays, many hosts including bacteria, insect, fungal, plant, and animal cell can be used for recombinant protein expression [5-7]. Among them, *Escherichia coli* (*E. coli*) is one of the most common hosts. Despite many drawbacks, *E. coli* is still widely used because of its well-studied genome, fast growth rate, high efficiency, simple requirements, and cost-effectiveness. Moreover, in one study, a protein belonging to the AA9 group has been successfully

expressed in *E. coli* [8]. Therefore, *E. coli* can be a suitable host to express and produce MGG06069 protein. Since both the full-length and the PMO domain of the protein can exhibit PMO activity, we expressed and purified recombinant MGG06069 and its PMO domain in *E. coli* [9]. The obtained recombinant proteins can be used for further enzymatic tests.

2 Materials and methods

2.1 Preparation of expression plasmid for the full-length and PMO domain of MGG06069

In this study, the commercial pGEX-4t-3 plasmid was used as expression vector on the *E. coli* BL-21 strain.The *mgg06069* sequence was obtained from the Uniprot database. The *mgg06069* gene encodes a signal peptide, a glyco-hydro-61 domain (PMO domain), and a chitin-binding domain (Fig. 1). Based on that structure, some modifications had been made to the gene before it was integrated into the plasmid.

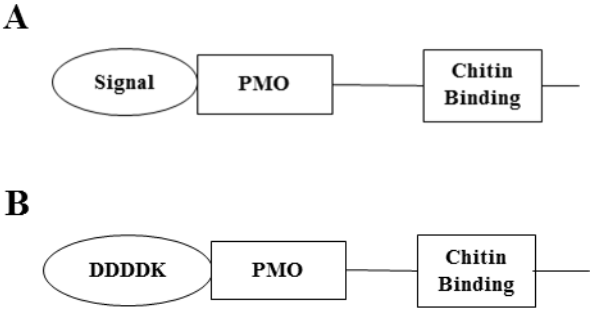

Fig. 1 A diagram showing the structure of the native MGG06069 protein (A) and the modified protein (B). Signal: the signal peptide, PMO: the polysaccharide monooxygenase domain, Chitin Binding: the chitin-binding domain (Chitin Binding). DDDDK: the peptide sequence recognized and cleaved by enterokinase.

The signal peptide-coding sequence was replaced by the DDDDK-coding sequence. DDDDK is a peptide that can be recognized and cleaved by the enterokinase enzyme. The sequence was optimized by Eurofin to be suitable for *E. coli*. Then, the sequence was synthesized and inserted into the pGEX-4t-3 plasmid. The plasmid was transformed into the *E. coli* BL-21 via heat shock. To clone the sequence encoding the PMO domain into the pGEX-4t-3 plasmid, this sequence, together with the DDDDK-coding sequence, was amplified by PCR using synthesized *mgg06069* gene as the template. The stop codon was added at the end of the PMO-coding sequence and restriction sites (RS) were added on both ends (BamHI at the 5' end and XhoI at the 3' end). The forward primer was 5' TATATAGGATCCGATGACGACGACAAG 3' and the reverse primer was 5' TTAATTCTCGAGTGAATAAGGAAGGTCTAACG 3'. The amplified sequence was inserted into the pGEX-4t-3 plasmid using restriction enzymes and ligase enzyme. The cloned plasmid was then transformed into the BL-21 *E. coli* via heat shock. The transformed bacteria cells were selected on the LB plate containing ampicillin.

2.2 Purification of the full-length MGG06069 and its PMO domain

Six Erlenmeyer flasks containing 100 mL of LB medium were prepared for bacteria inoculation. Five of the flasks were numbered from 1 to 5 and the other flask was labeled as "N" (negative control). These flasks were inoculated with 10 mL of overnight culture of transformed *E. coli*. Ampicillin was also added to the flasks. All of the flasks were incubated at 37 ^{0}C until the OD reached 0.4-0.6. After that, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the flask. The concentration of IPTG in the flasks 1 to 5 was 0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM and 1 mM respectively. IPTG was not added to the "N" flask. The flasks were then incubated at 25 ^{0}C with shaking for 16 hours. After that, bacteria cells in each flask were collected by centrifugation and resuspended in phosphate buffered saline (PBS). Then the cells were broken down by sonification. The lysate was centrifuged at 6,000 rpm for 15 minutes. The pellet and the supernatant from each flask were electrophoresed by SDS-PAGE to analyze the protein induction and the optimal IPTG concentration.

The transformed *E. coli* was then expressed at a larger scale (0.5-1) L using the procedure described above and the optimal concentration of IPTG. The non-transformed *E. coli* was also expressed to serve as the negative control. The supernatant of the lysate was added to the Glutathione Agarose Resin column. The procedure of protein purification was conducted following the manual from GoldBio which was based on the Sandra Harper and David W. Speicher's research [10]. After the cell was broken down by sonification, the lysate was centrifuged at 6,000 rpm for 15 minutes to remove the cell debris. The

lysate supernatant was loaded into the Glutathione Agarose Resin column. After the lysate had flown out of the column completely, the column was washed twice with PBS (pH = 7.4). Then, 1 mL of elution buffer (50 mM Tris-HCl pH = 8.0; 10 mM L-Glutathione reduced) was added to the column to elute the protein. The elution step was repeated 5 times. All of the wash and eluted fractions were analyzed by SDS-PAGE.

To purify the PMO domain, each colony selected on the ampicillin-containing LB plate was picked and bacterial cells were broken down by sonification. The lysate of each colony was analyzed by SDS-PAGE. Bacteria cells from the colony overexpressing the protein at the right size (~ 50 kDa) were kept and their plasmids were collected. The plasmid was re-transformed into *E. coli*. The newly transformed bacteria cells were expressed at a larger scale (0.5-1) L. The procedure used for the purification of the PMO domain was similar to that used for the full-length MGG06069. All of the wash and elution fractions were analyzed by SDS-PAGE.

2.3 SDS-PAGE

SDS-PAGE, a discontinuous electrophoretic system separates the proteins based on their molecular masses. In this study, 12.5 % acrylamide gel was used for all SDS-PAGE runs. To prepare the sample, 30 µL of the sample was mixed with 10 µL of loading dye. The mixture was heated to 100 ^{0}C for 10 minutes. After that, 10 µL of the mixture was loaded into the gel. The gel was run at the voltage of 50 V for 30 minutes and then at 100 V for 90 minutes. The gel was soaked into the comassie blue dye overnight. Finally, the gel was washed with a solution containing 10 % of acid acetic and 10 % of ethanol to reveal the protein band.

2.4 Measurement of protein concentration

In this study, the protein concentration was determined via measuring the absorbance at 280 nm. After measuring, the number was put into an online tool (https://www.aatbio.com/) to calculate the protein concentration.

3 Results and discussion

3.1 The expression of the full-length MGG06069 in *E. coli*

In Fig. 2, the SDS-PAGE result showed that two overexpressed protein bands were observed at the expected size (~ 90 kDa). Based on the boldness of the band, the optimal IPTG concentration was 0.1 mM (P1). The result also showed that a substantial amount of the overexpressed proteins was in the pellet of the lysate, suggesting that the protein might not be well-soluble.

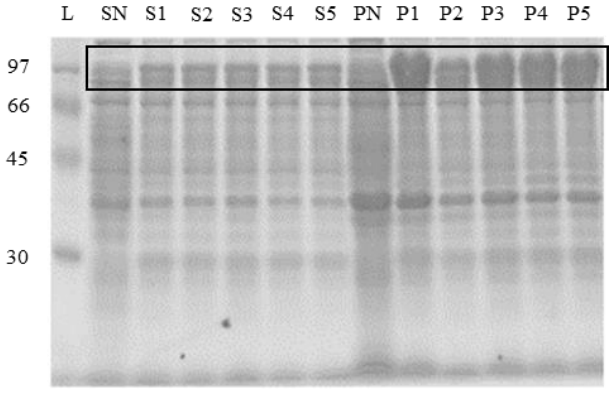

Fig. 2 The expression of the full-length MGG06069 in *E. coli*. S: supernatant; P: pellet; L: ladder; N: negative control; (1, 2, 3, 4, and 5: 0.1, 0.25, 0.5, 0.7 5 and 1) mM of IPTG, respectively.

In Fig. 3, in the E1 and E2 columns, two protein bands were observed at the approximate position of 90 kDa. The protein concentrations of the E1 and E2 were 35 ng/mL and 41 ng/mL, respectively. The protein concentrations in E3, E4, and E5 were undetectable. With this amount of purified protein, it would be inconvenient to use it for further enzymatic experiments. Furthermore, a second overexpressed band at the expected location (~ 90 kDa) indicates that some errors could have occurred. The large size of the expressed protein could be the reason that caused the problem. Therefore, it was expected that the PMO domain with smaller protein size could eliminate the error and improve the purified protein yield.

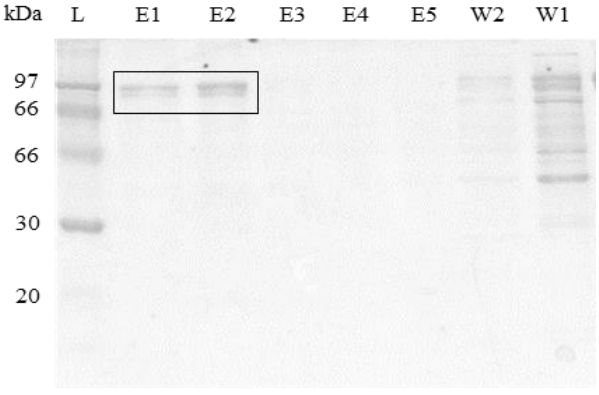

Fig. 3 The wash and elution fractions after the full-length MGG06069 protein was purified. L: protein ladder; E1-E5: elution fractions; W1-W2: wash fractions.

3.2 The expression of the PMO domain of MGG06069 in *E. coli*

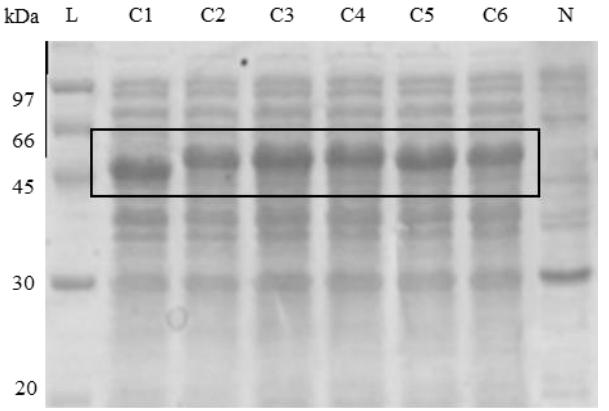

Fig. 4 The expression of the PMO domain of MGG06069 in *E. coli.* L: protein ladder; C1-C6: colony 1-6; N: negative control

In Fig. 4, the SDS-PAGE result showed that there was only one overexpressed band at the expected size (~ 50 kDa). The overexpressed band of colony 1 was slightly lower than that of the other colonies. Therefore, the strain from colony 1 was discarded. Other colonies were kept to collect their plasmids.

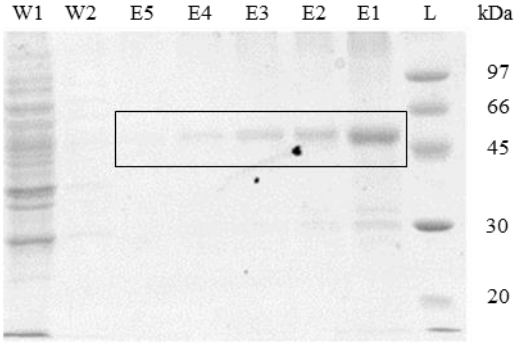

Fig. 5 The wash and elution fractions after the PMO domain protein was purified. L: protein ladder; E1-E5: elution fractions; W1-W2: wash fractions.

In Fig. 5, the purification result also showed that there was only one purified protein at the expected size (~ 50 kDa). The protein concentrations from E1 to E5 were (105, 83, 41, 21 and 9) ng/mL, respectively. This amount of purified protein was much higher than that of the purified full-length MGG06069 protein and it is sufficient for downstream experiments. This indicates that the expression of the PMO domain has improved the efficiency and eliminated the two-band problem that occurred in the expression of the full-length MGG06069. Therefore, it is possible to apply the PMO domain for further enzymatic assays.

4 Conclusion

It is possible to express both the full-length MGG06069 and its PMO domain in *E. coli.* However, the amount of the purified full-length MGG06069 was rather low. The presence of two bands, which might be due to errors in the protein processing, causes difficulties in downstream experiments. Meanwhile, the amount of the purified PMO domain was much higher and only one band of the expected size was obtained. This suggests that the reduction of the protein size in this study can improve both the quality and quantity of the purified protein. Thus, the purified PMO domain would be suitable for further enzymatic assays.

Acknowledgement

This research is funded by NTTU Foundation for Science and Technology Development under grant number 2020.01.160

References
1. V. V. Vu, W. T. Beeson, C. M. Phillips, J. H. D. Cate, and M. A. Marletta (2014). Determinants of Regioselective Hydroxylation in the Fungal Polysaccharide Monooxygenases. *J. Am. Chem. Soc., Vol. 136, no. 2, pp. 562-565,* DOI: 10.1021/ja409384b.
2. Te Beest (2007). Rice Blast. *PHI,* DOI: 10.1094/PHI-I-2007-0313-07.
3. Y. Li et al. (2020). Magnaporthe oryzae Auxiliary Activity Protein MoAa91 Functions as Chitin-Binding Protein to Induce Appressorium Formation on Artificial Inductive Surfaces and Suppress Plant Immunity. *mBio, Vol. 11, no. 2,* DOI: 10.1128/mBio.03304-19.
4. D. M. Soanes, A. Chakrabarti, K. H. Paszkiewicz, A. L. Dawe, and N. J. Talbot (2012). Genome-wide Transcriptional Profiling of Appressorium Development by the Rice Blast Fungus Magnaporthe oryzae. *PLoS Pathog, Vol. 8, no. 2, p. e1002514,* DOI: 10.1371/journal.ppat.1002514.
5. E. A. McKenzie and W. M. Abbott (2018). Expression of recombinant proteins in insect and mammalian cells. *Methods, Vol. 147, pp. 40-49,* DOI: 10.1016/j.ymeth.2018.05.013.
6. B. Owczarek, A. Gerszberg, and K. Hnatuszko-Konka (2019). A Brief Reminder of Systems of Production and Chromatography-Based Recovery of Recombinant Protein Biopharmaceuticals. *BioMed Research International, Vol. 2019, pp. 1-13,* DOI: 10.1155/2019/4216060.
7. J. Puetz and F. M. Wurm (2019). Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes, Vol. 7, no. 8, p. 476,* DOI: 10.3390/pr7080476.
8. I. J. Kim et al. (2015). Optimization of synergism of a recombinant auxiliary activity 9 from Chaetomium globosum with cellulase in cellulose hydrolysis. *Appl Microbiol Biotechnol, Vol. 99, no. 20, pp. 8537-8547,* DOI: 10.1007/s00253-015-6592-3.
9. Z. Forsberg et al. (2014). Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U.S.A., Vol. 111, no. 23, pp. 8446-8451,* DOI: 10.1073/pnas.1402771111.
10. Harper, S., & Speicher, D. W. (2011). Purification of Proteins Fused to Glutathione S-Transferase. In D. Walls & S. T. Loughran (Eds.), Protein Chromatography (Vol. 681, pp. 259-280). Humana Press. https://doi.org/10.1007/978-1-60761-913-0_14

Biểu hiện protein tái tổ hợp MGG06069 của loài *Magnaporthe oryzae* và phân đoạn PMO của protein trên *Escherichia coli*

Bùi Công Chính, Hồ Tá Giáp, Vũ Văn Vân
Viện Kĩ thuật Công nghệ cao, Đại học Nguyễn Tất Thành
chinhbc@ntt.edu.vn

Tóm tắt MGG06069 là một polysaccharide monooxygenase (PMO) protein có nguồn gốc từ loài nấm *Magnaporthe oryzae* gây bệnh đạo ôn trên cây lúa. Đây là một protein có tiềm năng trong việc đẩy nhanh quá trình phân hủy xenlulo, do đó protein này có thể có giá trị trong công nghiệp. Trong nghiên cứu này, protein MGG06069 với độ dài hoàn chỉnh cũng như phân đoạn PMO của protein MGG06069 được biểu hiện trong *E. coli* nhằm sản xuất ra protein tái tổ hợp ở quy mô lớn để phục vụ cho các nghiên cứu về hoạt tính enzym sau này. Kết quả cho thấy lượng protein tái tổ hợp thu được của phân đoạn PMO cao hơn so với lượng protein tái tổ hợp thu được của protein MGG06069 có độ dài hoàn chỉnh. Thêm vào đó, việc biểu hiện phân đoạn PMO đã thu được một dải protein duy nhất tại vị trí kích thước dự đoán. Các kết quả trên cho thấy phân đoạn PMO của protein MGG06069 được tinh sạch có thể được sử dụng cho các thí nghiệm về hoạt tính enzym trong tương lai.

Từ khóa AA9, MGG06069, Escherichia coli, Polysaccharide Monooxygenase, PMOs.