Accepted Manuscript

High-level expression of LMW-GS and α-gliadin genes promoted by the expressed tag sequence of 5’ end in *Escherichia coli*

Peihong Jiang, Ke Wang, Jiansheng Gao, Xiaoran Zheng, Jianing Feng, Wujun Ma, Yueming Yan, Xiaohui Li

PII: S1046-5928(14)00213-7
DOI: http://dx.doi.org/10.1016/j.pep.2014.09.016
Reference: YPREP 4574

To appear in: *Protein Expression and Purification*

Received Date: 19 June 2014
Revised Date: 12 September 2014

Please cite this article as: P. Jiang, K. Wang, J. Gao, Xiaoran. Zheng, J. Feng, W. Ma, Y. Yan, X. Li, High-level expression of LMW-GS and α-gliadin genes promoted by the expressed tag sequence of 5’ end in *Escherichia coli*, *Protein Expression and Purification* (2014), doi: http://dx.doi.org/10.1016/j.pep.2014.09.016

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
High-level expression of LMW-GS and α-gliadin genes promoted by the expressed tag sequence of 5’ end in Escherichia coli

Peihong Jiang¹, Ke Wang¹, Jiansheng Gao¹, Xiaoran Zheng¹, Jianing Feng¹, Wujun Ma², Yueming Yan¹, Xiaohui Li¹*

¹Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing, China.
²School of Veterinary & Life Sciences, Murdoch University and Australian Export Grains Innovation Centre, Perth, Australia.

*Corresponding author
X.H. Li, Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing 100048, China.
Fax: +8601068902777.
E-mail address: lixiaohui@cnu.edu.cn(X. Li).

Abbreviations: HMW-GS, high molecular weight glutenin subunit; LMW-GS, low molecular weight glutenin subunit; Gli-α, α-type gliadin; qRT-PCR, quantitative real time PCR; MCS, multiple cloning sites; SNP, single nucleotide polymorphism; LCCS, low compositional complexity segment.

Abstract

Wheat storage protein genes, especially low molecular weight glutenin subunit (LMW-GS) and gliadin genes are difficult to be expressed in E. coli, mainly due to the presence of highly repetitive sequences. In order to establish a high efficiency expression system for these genes, five different expression plasmids combining with
9 genes, viz. 6 LMW-GS and 3 α-gliadin genes isolated from common wheat and related species, were studied for heterologous expression in *E. coli*. In this study, when an expressed tag sequence encoding signal peptide, His-S or GST-tag was fused to the 5′end of LMW-GS or gliadin gene as the leading sequence, all recombination genes could be stably expressed at a high level. On the contrast, as expected, the inserted genes encoding mature protein failed without an expressed tag sequence. This result indicated that using expressed tag sequences as leading sequences could promote LMW-GS and gliadin genes to be well expressed in *E. coli*. Further transcriptional analysis by quantitative real-time PCR (qRT-PCR) showed transcription levels of recombination genes (e.g. GST-Glutenin, His-S-Glutenin and SP*-His-Glutenin) were 4-fold to 33-fold higher than those of the LMW-GS genes, which suggested these expressed tag sequences might play an important role in stimulating transcription. The possible molecular mechanism under this phenomenon was discussed.

**Keywords:** Low molecular weight subunit; α-gliadin; heterologous expression; repetitive sequence; expressed tag sequence; *Escherichia coli*. 
Background and introduction

The wheat storage proteins consisting of gliadins (including α-, β-, γ- and ω-type) and glutenins (including high and low molecular weight glutenin subunits or respective HMW-GS and LMW-GS) are vital in bread making due to its ability in conferring dough elasticity and extensibility [1-3]. Heterologous expression is particularly important for storage protein structure-function studies. The expressed proteins can be used to determine the functional properties of single subunit by small-scale testing [4-8], and to identify the corresponding native subunit in the wheat endosperm [9-14]. In the past several decades, three main types of expression systems have been widely used for heterologous protein production, viz. *Escherichia coli*, yeasts and cultured insect cells [15], among which the most popular for wheat storage protein gene expression is the *E. coli* because of its flexibility, ease of use and low cost [15-17].

However, heterologous expression of wheat storage protein genes had largely remained as a challenge. The special structural characteristics, highly repetitive sequences, present in glutenin and gliadin genes probably affect their heterologous expression, which may result in unfavorable mRNA secondary structures and inhibit ribosome processivity through mRNA stem-loops [18, 19]. These highly repetitive domains contain tripeptide motif GQQ, hexapeptide motif PGQGQQ, and nonapeptide motif GYYPTSLQQ in HMW-GSs [20], the motif P_{1-2}FSQ_{2-6} in LMW-GSs [21], and the motif PQPQPFP or PQQPY in gliadins [22]. Even though all of the wheat storage protein genes bear repetitive sequences, to some extent, the expression of LMW-GS and gliadin genes are more difficult than that of the HMW-GS gene in *E. coli*, which resulted in this phenomenon had been unclear.

In this work, we constructed various expression plasmids of LMW-GS and gliadin
genes to explore the biochemical mechanism underlying the suppression of expression and ultimately to establish an effectively heterologous expression system for wheat storage protein genes.

**Materials and methods**

**Construction of expression plasmids**

On the base of two vectors pET-30a (Novagen) and pGEX-4T-2 (Amersham), five types of expression plasmids were constructed with different expressed tag sequences located at the site adjacent to downstream of the initiation codon, designated as plasmid pET30a-Glutenin/Gliadin, pGEX4T2-GST-Glutenin/Gliadin, pET30a-His-S-Glutelin, pET30a-SP-Glutelin and pET30a-SP*-His-Glutenin. The 9 genes and their corresponding primers designed for plasmid construction were listed in Table 1 and Table S1, respectively. The first type of plasmid pET30a-Glutenin/Gliadin was constructed with the pET-30a vector, in which the upstream His- plus S-tag sequence of the multiple cloning sites (MCS) was replaced by each of the 9 genes removed signal peptides sequence. Similarly, the pGEX-4T-2 vector was used to construct the second type of plasmid pGEX4T2-GST-Glutelin/Gliadin, and the 9 genes encoding mature protein were inserted the site downstream of the GST-tag sequence (GST-tag, about 26 kDa), respectively. The third type of plasmid pET30a-His-S-Glutelin was constructed using the pET-30a vector, and 2 LMW-GS genes excluded signal sequence inserted into the site downstream of the His- and S-tag, respectively. In the fourth type of plasmid pET30a-SP-Glutelin, the His- plus S-tag sequence of pET-30a was replaced with 3 complete open reading frames of LMW-GS gene in this study, and the His-tag was located at 3’end of the MCS to allow the detection of induced proteins. The fifth type of plasmid pET30a-SP*-His-Glutenin was a modified pET-30a, in which the
primordial His- plus S-tag sequence was replaced by a synthesized nucleotide sequence encoding a His-tag, the signal peptide of LMW-GS AnLMW-m1 (GenBank accession no. EF536031) and a short peptide recognized by the thrombin. By this construction, a fusion protein consisted of a His-tag, the signal peptide of AnLMW-m1 and the mature glutenin would be acquired.

**Table 1.** Wheat LMW-GS and gliadin genes used for heterologous expression in *E. coli.*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Type</th>
<th>Specie</th>
<th>Cultivar/accession</th>
<th>Length (bp)</th>
<th>GenBank no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnLMW-m1</td>
<td>LMW-m</td>
<td><em>Ae. neglecta</em></td>
<td>PI298897</td>
<td>903</td>
<td>EF536031</td>
</tr>
<tr>
<td>AnLMW-m2</td>
<td>LMW-m</td>
<td><em>Ae. neglecta</em></td>
<td>PI298897</td>
<td>906</td>
<td>EF536032</td>
</tr>
<tr>
<td>AnLMW-m3</td>
<td>LMW-m</td>
<td><em>Ae. neglecta</em></td>
<td>PI298897</td>
<td>903</td>
<td>EF536033</td>
</tr>
<tr>
<td>AnLMW-m1</td>
<td>LMW-m</td>
<td><em>Ae. markgrafii</em></td>
<td>PI254863</td>
<td>900</td>
<td>EU329425</td>
</tr>
<tr>
<td>ZyLMW-m1</td>
<td>LMW-m</td>
<td><em>T. aestivum</em></td>
<td>Zhongyou9507</td>
<td>915</td>
<td>EU329426</td>
</tr>
<tr>
<td>ZyLMW-m2</td>
<td>LMW-m</td>
<td><em>T. aestivum</em></td>
<td>Zhongyou9507</td>
<td>912</td>
<td>EU329427</td>
</tr>
<tr>
<td>Gli-A4</td>
<td>Gli-α</td>
<td><em>Ae. tauschii</em></td>
<td>T26</td>
<td>855</td>
<td>EF561273</td>
</tr>
<tr>
<td>Gli-G4</td>
<td>Gli-α</td>
<td><em>T. aestivum</em></td>
<td>Gaocheng8901</td>
<td>870</td>
<td>EF561277</td>
</tr>
<tr>
<td>Gli-Z4</td>
<td>Gli-α</td>
<td><em>T. aestivum</em></td>
<td>Zhongyou9507</td>
<td>864</td>
<td>EF561284</td>
</tr>
</tbody>
</table>

**Protein expression and extraction**

Plasmids transformation and bacterial culture as well as the identification of fusion proteins by SDS-PAGE and Western blotting were performed according to Li et al [13]. After being induced for four hours by isopropyl-β-D-thiogalactopyranoside (IPTG), the 1.8 ml cooled bacterial culture was transferred into a 2.0 ml microcentrifuge tube. The total proteins were collected by being centrifuged at 10,000×g, 4°C for 5 min, and the supernatant was removed. In the following steps, the sample was treated by one of the two methods. 1) 100 µl deionized distilled water (ddH₂O) was added immediately. After boiling for 10 min, 100 µl 2×Loading buffer containing 100 mM Tris-HCl(PH 6.8), 4%(w/v) SDS, 20%(v/v) Glycerine, 0.1%(w/v) Bromophenol Blue and 1% (w/v) DTT was added and incubated for 45 min at 65°C.
The sample was centrifuged for 5 min at 10,000 g at room temperature and the supernatants were used for SDS-PAGE analysis. 2) 60 µl of 55% isoamyl alcohol containing 1% (w/v) DTT was added immediately and incubated for 45 min at 65°C and centrifuged for 5 min at 10,000 g at room temperature. Equal volume of 2×Loading buffer and the supernatant were mixed in a clean microcentrifuge. After being incubated for 10 min at 65°C, the sample was centrifuged for 5 min at 10,000 g at room temperature. The supernatants were used for Western blotting analysis.

**mRNA extraction, purification and cDNA synthesis**

The *E. coli* BL21Star (DE3) plysS competent cells transformed with five types of expression plasmids (pET30a-Glutelin/Gliadin, pGEX4T2-GST-Glutelin/Gliadin, pET30a-His-S-Glutelin, pET30a-SP-Glutelin and pET30a-SP*-His-Glutelin) plus the control pET-30a and pGEX-4T-2 vectors were shaken at 220 rpm and 37°C until the cultured bacteria reached to a density of 0.6 (OD₆₀₀). Exogenous genes were induced with 0.6 mM IPTG for 2.5 h at 220 rpm and 37°C. The 1.8 ml of cooled bacterial culture was centrifuged for 1 min at 10,000 g and 4°C. After discarding the supernatant, 1 ml of Trizol reagent was used to re-suspend the bacterial pellet. The follow steps were carried out according to Li et al [13]. High quality RNA was used to synthesize cDNA with Random primers from approximately 1 µg total RNA by a SuperScript III First-Strand Synthesis System (Invitrogen, catalog no. 18080-051).

**Quantitative real-time PCR (qRT-PCR)**

1) Specific primer designed: the 16s rRNA was used as internal control (GenBank accession no. [V00348]), and the specific primer sequences used in this study were listed in Table S2. 2) qRT-PCR reaction and running program: qRT-PCR was performed on a Rotor Gene 3000 machine (Corbett Research, Australia). The reaction was performed in 20 µl volume including 10 µl RealMasterMix (Tiangen, catalog no.
FP202-02), 0.2 μM of each primer and 2 μl cDNA. Running programs were as follows: 94°C for 3 min, and 40 cycles of 94°C for 20 s, 58°C for 30 s and 72°C for 20 s. The products were heat-denatured over a temperature gradient of 0.1°C/s from 72°C to 95°C. The fluorescence signal was collected at the end of extension in every cycle. 3) Construction of double standard curves of internal control and LMW-GS genes: double standard curves were constructed in each run of PCR by diluting the cDNA to a tenfold concentration gradient. Every concentration was performed in three independent replicates. Double standard curves were applied to accumulate the relative quantification of the internal control and LMW-GS genes expression.

Results

Construction of plasmids and expression in *E. coli*

Six LMW-GS and three gliadin genes excluded signal sequences were subcloned into the expression vector pET-30a (Table 1, Fig.1a), in which the His- plus S-tag sequence had been removed. The constructed plasmids pET30a-Glutelin/Gliadin were transformed into the BL21Star (DE3) plysS *E. coli* cells. No induced protein was separated by SDS-PAGE under optimized induction conditions (Fig.2a). It had been reported that AT-content of codons immediately downstream of the initiation codon could affect the heterologous gene expression in *E. coli* [23]. The 9 nucleotide sequences encoding mature protein were ligated into the MCS downstream of GST sequence in expression vector pGEX-4T-2, in which the GST-tag sequence was used as a leading sequence for the glutenin and gliadin genes (Fig. 1b). The total proteins treated with the method 1) (see Materials and methods) were separated by SDS-PAGE, and the induced proteins including mature proteins of glutenin (gliadin) and a 26 kDa GST-tag could be identified (Fig. 2b). The fusion proteins were purified according to the method 2) and verified further by Western blotting conducted with anti-GST-tag
antibody (Fig. 2b). Under optimized induction conditions, the products of 9 genes reached 0.4-0.6 g/l. These suggested that the GST as a leading sequence had a positive effect on the expression of the glutenin and gliadin genes in E. coli.

**Fig. 1. Structural models of vectors and constructed expression plasmids.** Diagrams showed the structural models of pET-30a, pGEX-4T-2 expression vectors and the constructed plasmids of glutenin and gliadin genes (a, b, c, d and e). Promoters, initiation codons, stop codons, MCSs and expressed tag sequences were indicated. MCS, multiple cloning site; SP, signal sequence; SP*, the signal sequence of AnLMW-m1.

With the characteristics of high efficiency, conciseness and well expressing, the His-tag is widely applied to affinity purification of heterologous proteins. Therefore, the His- plus S-tag sequence located at the site upstream of MCS in pET-30a vector was used as the leading sequence for 2 LMW-GS genes removed signal sequences (Fig. 1c). Fusion proteins related to LMW-GS genes ZyLMW-m1 or ZyLMW-m2 were separated successfully by SDS-PAGE and purified fusion proteins were also identified by Western blotting with anti-His-tag antibody (Fig. 2c), which indicated that the His-plus S-tag sequence could also improve the expression of LMW-GS genes.
Fig. 2. Identification of induced proteins. a: The products of nine genes without any leading sequence were not detected in *E. coli*. Lane 1-9, AnLMW-m1, AnLMW-m2, AnLMW-m3, AmLMW-m1, Gli-At4, Gli-G4, Gli-Z4, ZyLMW-m1 and ZyLMW-m2 respectively. M, protein marker; NC, bacteria containing the pET-30a vector. b: Nine genes with a GST-tag sequence at 5’end were expressed in *E. coli*. Lane 1-9, AnLMW-m1, AnLMW-m2, AnLMW-m3, AmLMW-m1, Gli-At4, Gli-G4, Gli-Z4, ZyLMW-m1 and ZyLMW-m2 respectively. The purified fusion proteins indicated by upward arrow heads were detected by Western blotting with anti-GST antibody. M, protein marker; NC, bacteria containing the pGEX-4T-2 vector. c: ZyLMW-m1 and ZyLMW-m2 with a His-tag sequence at 5’end were expressed in *E. coli*. Lane 1-2, ZyLMW-m1 and ZyLMW-m2 respectively; the purified fusion proteins indicated by upward arrow heads were detected by Western blotting with anti-His-tag antibody. M, protein marker; NC, bacteria containing the pET-30a vector. d: Three genes carrying their native signal sequences were expressed in *E. coli*. Lane 1-3, AnLMW-m1, AnLMW-m2 and AnLMW-m3, respectively. The purified fusion proteins indicated by upward arrow heads were detected by Western blotting with anti-His-tag antibody. M, protein marker; NC, bacteria containing the pET-30a vector.

On the base of both transcription and translation processes starting at the signal sequence in eukaryotic cell, we speculated the signal sequence should be well expressed and function as promoting the expression of its subsequent sequences (encoding mature proteins). To confirm this hypothesis, the plasmids pET30a-"SP-Glutelin" were constructed with pET-30a vector and the nucleotides encoding both signal peptide and mature protein of 3 LMW-GS genes (*AnLMW-m1-3*).
The His- plus S-tag sequences between \( T_7 \) promoter and MCS were removed as shown in Fig. 1d. Simultaneously, the stop codons of these exogenous genes were excluded so that the His-tag sequence in the 3’end of MCS of the vector pET-30a could be fused with the inserted genes for the purpose of detecting the fusion proteins. The expression results were shown in Fig. 2d. The expressed proteins of the 3 genes were identified by both SDS-PAGE and Western blotting, indicating that the signal sequence was able to generate a high-level expression, similarly to those exogenous genes used the GST or His+S as leading sequences.

Since some SNPs (single nucleotide polymorphism) were present in the signal sequences of glutenin genes, it is interesting to understand whether a highly-conserved signal sequence was essential for glutenin expression in \( E. coli \). Alignment of nucleotide sequences showed that the signal sequence was different among \( AnLMW-m1 \), \( ZyLMW-m1 \), \( ZyLMW-m2 \), and \( AmLMW-m1 \). Compared with \( AnLMW-m1 \), there were two nucleotides substitutions present in the signal sequences of \( ZyLMW-m1 \) and -2, and three in \( AmLMW-m1 \) genes (Fig. 3a). A synthesized nucleotide sequence encoding the signal peptide of AnLMW-m1, a His-tag and a thrombin recognized peptide was used as the leading sequence for \( ZyLMW-m1 \), \( ZyLMW-m2 \), and \( AmLMW-m1 \) genes. The induced proteins of the three LMW-GS genes were separated by SDS-PAGE as shown in Fig. 3b, and the corresponding fusion proteins were further verified by Western blotting with anti-His-tag antibody (Fig. 3b). Based on analysis with software AlphaEaseFC version 4.0.0, the purified proteins reached 0.25-40 mg/L. These implied that the specific signal sequence was not essential for the LMW-GS gene expression.
Fig. 3. An exogenous signal sequence using as the leading sequence for 3 LMW-GS genes.

a: Alignment of signal sequences among 4 LMW-GS genes. * indicated the sequence was used as the leading sequence for AmLMW-m1, ZyLMW-m1 and ZyLMW-m2. b: Three genes with an exogenous signal sequence were expressed in E. coli. Lane 1, AmLMW-m1; lane 2, ZyLMW-m1; lane 3, ZyLMW-m2. The downward arrow heads indicated the fusion proteins separated by SDS-PAGE, and the upward indicated the corresponding proteins detected by Western blotting with His-tag antibody. M, protein marker; NC, bacteria containing pET-30a vector.

Identification of transcriptional levels of LMW-GS genes from different expression plasmids

Transcription analysis was conducted for two LMW-GS genes, ZyLMW-m1 and ZyLMW-m2. The mRNA of the two genes were extracted from the bacteria containing expression plasmid pET30a-Glutelin, pGEX4T2-GST-Glutelin, pET30a-His-S-Glutelin and pET30a-SP*-His-Glutelin after being induced for 2.5 hours. The gDNA was digested and the purified RNA was reversely transcribed into cDNA for using as templates of qRT-PCR. The specificity of primers was firstly verified as shown in Fig. 4. The unique DNA fragments (146 bp) of two LMW-GS genes were amplified by both ZyLMWm1F+R and ZyLMWm2F+R primer pairs when the cDNA prepared from plasmids pGEX4T2-GST-Glutelin, pET30a-His-S-Glutelin and pET30a-SP*-His-Glutelin as templates (Lane 4-6, Lane 11-13), but no any product by the same primers from the negative control vectors pGEX-4T-2 and pET-30a (Lane 1 and 8, Lane 2 and 9, respectively). As for internal
control 16s rRNA, a 164-bp DNA fragment was also uniquely amplified by 16srRNAF1+2 primer pair from the E. coli BL21 Star (DE3) plysS (Lane 7).

![Fig. 4. Specificity analysis of primers designed for qRT-PCR.](image)

**Fig. 4. Specificity analysis of primers designed for qRT-PCR.** M, 1 kb plus maker; lane 1-6, lane 7 and lane 8-13, the productions were amplified with the primer pairs ZyLMWm1F+R, 16srRNAF+R and ZyLMWm2F+R, respectively. Total RNA was extracted from BL21 containing one of different expression plasmids. Lane 7, no plasmid; lane 1 and 8, the pGEX-4T-2 vector; lane 2 and 9, the pET-30a vector; lane 3-6, the pET30a-ZyLMWm1, pGEX4T2-GST-ZyLMWm1, pET30a-His-S-ZyLMWm1, and pET30a-SP*-His-ZyLMWm1 plasmids, respectively; lane 10-13, the pET30a-ZyLMWm2, pGEX4T2-GST-ZyLMWm2, pET30a-His-S-ZyLMWm2, and pET30a-SP*-His-ZyLMWm2 plasmids, correspondingly.

The relative quantification method was used to detect the relative transcriptional level of ZyLMW-m1 and -m2 from different expression plasmid. Both ZyLMW-m1 and ZyLMW-m2 genes could be transcribed from pET30a-Glutelin, pGEX4T2-GST-Glutelin, pET30a-His-S-Glutelin and pET30a-SP*-His-Glutelin plasmids, but their transcriptional levels were significantly different (Fig. 5). The mRNA quantities of both two genes from pGEX4T2-GST-Glutelin, pET30a-His-S-Glutelin and pET30a-SP*-His-Glutelin were 4-fold to 33-fold higher than those from the pET30a-Glutelin. This result showed expressed tag sequences as leading sequences could promote the transcription of LMW-GS genes.
Fig. 5. Analysis of transcription levels of gene \(\text{ZyLMW-}m1\) (a) and \(\text{ZyLMW-}m2\) (b) with different expressed tag sequences as leading sequences in \(E.\ coli\).

Since both expression plasmid pET30a-\(\text{His-S-Glutelin}\) and pET-\(\text{SP*-His-Glutelin}\) were constructed with the vector pET-30a, the relative translational levels of \(\text{ZyLMW-}m1\) and \(\text{ZyLMW-}m2\) from the two types of expression plasmids were also analyzed. The total proteins were separated by 12% SDS-PAGE (Fig. 6) and the quantities of induced proteins were analyzed with software AlphaEaseFC version 4.0.0. The bacterial proteins indicated by upward arrow heads in Fig. 6 were used as internal controls to normalize the loading samples. The induced protein of \(\text{ZyLMW-}m1\) and \(\text{ZyLMW-}m2\) from pET30a-\(\text{His-S-Glutelin}\) were two-fold higher than those from pET-\(\text{SP*-His-Glutelin}\) plasmid. In general, the translational level of genes is well consistent with their transcription level (Fig. 5).

Fig. 6. Quantity analysis of the induced glutenins expressed by
pET30a-His-S-Glutenin and pET30a-SP*-His-Glutenin plasmids. The total proteins were separated by 12% SDS-PAGE. Induced proteins were indicated by downward arrow heads, and bacterial proteins serving as internal controls by upward arrow heads. M, protein marker; lane 1 and 2, the pET30a-His-S-ZyLMWm1 and -ZyLMWm2 plasmids; lane 3 and 4, the pET30a-SP*-His-ZyLMWm1 and -ZyLMWm2, respectively.

Discussion

Possible reasons for expression suppression of wheat storage genes in E. coli

Three types of expression systems are widely used in heterologous protein production, among which pET series of vectors and E. coli were considered to be suitable for the heterologous expression of wheat storage genes [15]. Despite of many improvements in expression vectors including combining strong promoters with suitable length, stringent regulators, transcription start sites, alternative translation start codons and multiple cloning site and transcription terminators, it is still difficult to express LMW-GS and gliadin genes in these vectors.

Previous investigations demonstrated that the usage of codon following the initiation codon could affect gene expression at translation level in E. coli [24-27]. In this study, only a small amount of transcriptional products were detected by qRT-PCR without any leading sequence adjacent to the 5’end of LMW-GS genes, which indicated that transcriptional suppression should be responsible for LMW-GS gene failing to express in E. coli. Incorporation of AT-rich codons immediately downstream of the initiation codon was able to enhance the expression efficiency of a recombinant gene [23, 28]. We found the AT percentages of signal peptide sequences were between 41.7% and 46.7%, while those of the N-terminal region sequences ranged from 46.2% to 48.7% among AmLMW-m1, ZyLMW-m1 and ZyLMW-m2 genes. According to the AT-rich hypothesis, it could be deduced that the nucleotide encoding
N-terminal regions should be more beneficial than signal peptide sequences in heterologous expression of LMW-GS genes, and the reservation of signal peptide sequence would decrease the AT percentage of 5’end region and suppress gene expression. However, the reservation of signal peptide sequence significantly enhanced the expression level of LMW-GS gene in our study. This may imply that the AT-rich hypothesis may be only applicable in a certain range of AT content, and it was not the key factor in determining the expressions of LMW-GS genes in the current study. In addition, two types of methylases, DNA adenine methylase and DNA cytosine methylase recognizing GATC and CC(A/T)GG, respectively, have been found to regulate gene expression in E. coli [29]. In this work, the sites recognized by methylases were not found in the ZyLMW-m2 gene but it still failed to express on the context of pET30a-Glutelin. It suggested that there was no direct relation between DNA methylations and suppression of LMW-GS genes expression in E. coli.

An interesting finding had been reported that the low compositional complexity segment (LCCS) in C-terminal domain of ICK could result in the silence of full-length ick in E. coli [30]. The structures of fusion and mature proteins of ZyLMW-m1 and ZyLMW-m2 were predicted by the SMART [31], and the results were shown in the Table 2. The LCCSs were present in the near N-terminal domain of both fusion and mature proteins of two LMW-GSs, which were well corresponding to their repetitive domains. For example, in the mature LMW-m1, the LCCS located at 14-123 residues and repetitive domain at 14-112 residues. On the other hand, the position of LCCS encoded by the same gene was different between fusion and mature proteins, the LCCS at 14-123 residues in mature LMW-m1 but at 58-167 residues in the HIS-S-LMW-m1. This indicated that the inserted leading sequence made the repetitive region shift away from the initiation codon. Combined with quantitative
analysis of mRNA (Fig. 5) and protein (Fig. 6), we thought there were two possibilities about the sequence between the initiation codon and repetitive region resulted in the different expression level of LMW-GS gene in *E. coli*. One was the length of sequence. Our previous results showed that the nucleotide sequences encoding mature protein of 13 HMW-GS genes, inserted into the MCS downstream of *T*$_7$ promoter in pET-30a/28a vector, could be expressed successfully in *E. coli* whether the *His, His-S* or *His-T*$_7$ tag sequence exist (Table 3). Then we thought the leading sequence was unessential for the heterologous expression of HMW-GS genes removed signal sequence because they possessed sufficient length N-terminal sequences. As summarized in Fig. 7, the major difference among four types of storage protein genes was the length of 5’non-repetitive region. Comparing with 39 bp in m-type LMW-GS genes and 15 bp in α-type gliadins, the 5’non-repetitive region of y-type HMW-GS genes contained 312 bp and those of x-type HMW-GS genes ranged from 243 to 267 bp. Some others’ researches also supported this hypothesis. The synthetic spr gene encoding repeat motif PQQPY, directly ligating with the *T*$_7$ promoter, only expressed in a very low level of polypeptides [37]. However, being fused with thioredoxin gene without low compositional complexity segment, spr and the nucleotide sequence encoding the N-terminal repetitive domain of γ-gliadin could be over-expressed as fusion proteins in *E. coli* [38, 39]. It was also reported that using the *pelB, ompA* or *CSP* signal sequence as the leading sequence, both insoluble and soluble proteins of GM-CSF and scFv-phOx were achieved at high level, but no product was detected when the signal peptide was absent [40]. The other was the identity of sequence. Sletta et al (2007) analyzed the effect of *pelB, ompA* and *CSP* signal sequences on the expression of GM-CSF in *E. coli*. They found the *ompA* caused two-fold higher expression than *pelB*, and only 22% DNA sequence similarity
was among the three signal sequences [40]. But in our study, for AmLMW-m1, ZyLMW-m1 and ZyLMW-m2, exchange of native signal sequence with the exogenous did not lead to significant difference in expression level. We thought high sequence similarity might be responsible for this phenomenon (Fig. 3a).

Table 2. Low compositional complexity segments in non-recombinant and recombinant proteins of ZyLMW-m1 and ZyLMW-m2 predicted by the SMART.

<table>
<thead>
<tr>
<th>Expressed tag</th>
<th>LMW-GS</th>
<th>LCCS</th>
<th>Repetitive domain</th>
<th>mRNA</th>
<th>Induced protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Start-stop</td>
<td>Start-stop</td>
<td>SDS</td>
</tr>
<tr>
<td>Null</td>
<td>ZyLMW-m1</td>
<td>14-123</td>
<td>14-112</td>
<td>#</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>ZyLMW-m2</td>
<td>14-110</td>
<td>14-115</td>
<td>#</td>
<td>Null</td>
</tr>
<tr>
<td>His-tag</td>
<td>ZyLMW-m1</td>
<td>58-167</td>
<td>58-156</td>
<td>###</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>ZyLMW-m2</td>
<td>58-154</td>
<td>58-159</td>
<td>###</td>
<td>F</td>
</tr>
<tr>
<td>Signal peptide*</td>
<td>ZyLMW-m1</td>
<td>47-156</td>
<td>47-145</td>
<td>#</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>ZyLMW-m2</td>
<td>47-143</td>
<td>47-148</td>
<td>#</td>
<td>F</td>
</tr>
</tbody>
</table>

*, signal peptide of AnLMW-m1; LCCS, low compositional complexity segment; #, indicated the relative quantity of mRNA or induced protein; SDS, SDS-PAGE; WB, Western blotting; F, fusion protein; -, no identification.
Table 3. Heterologous expression of HMW-GS genes in *E. coli.*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Type</th>
<th>Specie</th>
<th>Cultivar/accession</th>
<th>Vector</th>
<th>Leading sequence</th>
<th>Length (bp)</th>
<th>GenBank no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dx3</td>
<td>HMW-x</td>
<td><em>Ae. tauschii</em></td>
<td>T67</td>
<td>pET-28a</td>
<td>Null</td>
<td>2538</td>
<td>HM347447</td>
<td>[32]</td>
</tr>
<tr>
<td>Dx4</td>
<td>HMW-x</td>
<td><em>Ae. tauschii</em></td>
<td>T132</td>
<td>pET-30a</td>
<td>Null</td>
<td>2511</td>
<td>DQ307383</td>
<td>[32]</td>
</tr>
<tr>
<td>Dx5</td>
<td>HMW-x</td>
<td><em>Ae. tauschii</em></td>
<td>TD81</td>
<td>pET-30a</td>
<td>Null</td>
<td>2481</td>
<td>DQ681076</td>
<td>[33]</td>
</tr>
<tr>
<td>Dx5.1</td>
<td>HMW-x</td>
<td><em>Ae. tauschii</em></td>
<td>TD130</td>
<td>pET-30a</td>
<td>Null</td>
<td>2526</td>
<td>DQ681077</td>
<td>[33]</td>
</tr>
<tr>
<td>By16</td>
<td>HMW-y</td>
<td><em>T. turgidum</em></td>
<td>KU1952</td>
<td>pET-28a</td>
<td>Null</td>
<td>2190</td>
<td>-</td>
<td>[34]</td>
</tr>
<tr>
<td>Akx2.3</td>
<td>HMW-x</td>
<td><em>Ae. speltoides</em></td>
<td>PI487280</td>
<td>pET-28a</td>
<td>His</td>
<td>2907</td>
<td>-</td>
<td>[14]</td>
</tr>
<tr>
<td>Akx1</td>
<td>HMW-x</td>
<td><em>Ae. kotschyi</em></td>
<td>PI393493-1</td>
<td>pET-28a</td>
<td>His</td>
<td>2538</td>
<td>-</td>
<td>[14]</td>
</tr>
<tr>
<td>Akx3</td>
<td>HMW-x</td>
<td><em>Ae. kotschyi</em></td>
<td>PI554298-1</td>
<td>pET-28a</td>
<td>His</td>
<td>2583</td>
<td>-</td>
<td>[14]</td>
</tr>
<tr>
<td>Dy12.2</td>
<td>HMW-y</td>
<td><em>Ae. tauschii</em></td>
<td>TD87</td>
<td>pET-28a</td>
<td>His</td>
<td>1947</td>
<td>DQ307385</td>
<td>[35]</td>
</tr>
<tr>
<td>Aky20</td>
<td>HMW-y</td>
<td><em>Ae. speltoides</em></td>
<td>PI487280</td>
<td>pET-28a</td>
<td>His</td>
<td>1974</td>
<td>-</td>
<td>[14]</td>
</tr>
<tr>
<td>Aky8</td>
<td>HMW-y</td>
<td><em>Ae. speltoides</em></td>
<td>PI487280</td>
<td>pET-28a</td>
<td>His</td>
<td>1911</td>
<td>-</td>
<td>[14]</td>
</tr>
<tr>
<td>By8</td>
<td>HMW-y</td>
<td><em>T. durum</em></td>
<td>Simeto</td>
<td>pET-28a</td>
<td>His+T7</td>
<td>2163</td>
<td>AY245797</td>
<td>[36]</td>
</tr>
<tr>
<td>Gy7</td>
<td>HMW-y</td>
<td><em>T. timophevi</em></td>
<td>CWI17006</td>
<td>pET-30a</td>
<td>His+S</td>
<td>2202</td>
<td>EF151424</td>
<td>[13]</td>
</tr>
</tbody>
</table>
Fig. 7. The general structural models of different wheat storage protein genes. The signal sequence, 5’ non-repetitive region, repetitive region and 3’ non-repetitive region were indicated by white, black, gridding and splashing frames, respectively.

**Effect of signal sequence on the transcription level of exogenous genes**

The signal peptides were once regarded as a negative factor for heterologous expression of storage protein genes, which could be attached into the membrane and block regular cellular transform system of *E. coli* [41]. For this reason, the signal sequences of these genes were generally removed before being transformed into *E. coli* [2, 8, 10, 13, 14, 32, 34, 36, 42, 43]. However, being removed signal sequence, the mRNA of HMW-GS gene *Dx* failed to be detected by Northern blotting [11]. In this study, our results showed that the signal sequence, similar to the 5’ non-repetitive region of HMW-GS gene and GST-tag sequence as well as His-S tag sequence, functioned as improving expression of LMW-GS genes in *E. coli*. It was reasonable that the signal sequence should facilitate transcription and translation of exogenous gene in *E. coli*, especially for those genes containing complex structures such as glutenin and gliadin genes.

In this work, we found an expressed tag sequence located at the region between the start codon and repetitive region, such as encoding GST-tag, His-tag or signal peptide, was essential for the expression of LMW-GS and gliadin genes in *E. coli*. Integrating an expressed tag or reserving intact signal sequence could improve
transcription and then acquire a high-level expression of LMW-GS and gliadin genes in *E. coli*.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (31271703, 31101145), the Natural Science Foundation of Beijing (6122002) and the National Key Projects for Transgenic Crops of China (2011ZX08009-003-004, 2011ZX08002-004).

**References**


Highlights

1. Expressed tag sequences as leading sequences can promote the expression of glutenin.
2. Expressed tag sequences can stimulate the transcription of glutenin.
3. Transcriptional level of glutenin was consisted with its translational level in *E. coli*.