ORIGINAL ARTICLE

Optimization of a 2A self-cleaving peptide-based multigene expression system for efficient expression of upstream and downstream genes in silkworm

Yuancheng Wang[1,2] · Feng Wang[1,2,3] · Sheng Xu[1,2] · Riyuan Wang[1,2] · Wenjing Chen[1,2] · Kai Hou[1,2] · Chi Tian[1,2] · Fan Wang[3] · Ping Zhao[1,2] · Qingyou Xia[1,2]

Abstract

The multigene expression system is highly attractive to co-express multiple genes or multi-subunit complex-based genes for their functional studies, and in gene therapy and visual tracking of expressed proteins. However, the current multiple gene co-expression strategies usually suffer from severe inefficiency and unbalanced expression of multiple genes. Here, we report on an improved 2A self-cleaving peptide (2A)-based multigene expression system (2A-MGES), by introducing an optimized Kozak region (Ck) and altering the gene arrangement, both of which contributed to the efficient expression of two fluorescent protein genes in silkworm. By co-expressing *DsRed* and *EGFP* genes in insect cells and silkworms, the potent Ck was first found to improve the translation efficiency of downstream genes, and the expression of the flanking genes of 2A were improved by altering the gene arrangement in 2A-MGES. Moreover, we showed that combining Ck and an optimized gene arrangement in 2A-MGES could synergistically improve the expression of genes in the cell. Further, these two flanking genes, regulated by modified 2A-MGES, were further co-expressed in the middle silk gland and secreted into the cocoon, and both achieved efficient expression in the transgenic silkworms and their cocoons. These results suggested that the modified Ck-2A-MGES will be a potent tool for multiple gene expression, for studies of their functions, and their applications in insect species.

Keywords 2A self-cleaving peptide · Kozak sequence · Multiple-gene co-expression · Silkworm

Yuancheng Wang and Feng Wang contribute equally to this study.

✉ Qingyou Xia
xiaqy@swu.edu.cn

[1] Biological Science Research Center, Southwest University, Chongqing 400715, People's Republic of China

[2] Chongqing Key Laboratory of Sericultural Science, Southwest University, Chongqing 400715, People's Republic of China

[3] Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing 400715, People's Republic of China

Introduction

The efficient co-expression of multiple genes is a highly attractive system for studying their function and interaction relationships in gene function research, co-expression of multi-subunit protein complexes in gene therapy, and visual tracking target proteins of interest by fluorescent protein fusion co-expression in cells and tissues (Foti et al. 2009; Tang et al. 2009; Kim et al. 2013). However, an efficient strategy for multiple gene expression in cells and individuals is lacking. Commonly, in insect and mammalian cells and individuals, the strategy for multigene co-expression includes genetic hybridization (Iizuka et al. 2009), using multiple promoters in a single vector (Baron et al. 1995), introducing a post-translational proteolysis site between two genes (Kunes et al. 2009), and using a multicistronic expression vector, mediated by internal ribosomal entry sites (Ye et al. 2013) or a 2A self-cleaving peptide (2A) (Bieniossek et al. 2012). Up to now, many multi-subunit proteins

or multiple target proteins including the recombinant mouse mAb light-chain plus heavy-chain proteins, human type IV collagen plus *Bombyx mori* prolyl 4-hydroxylase a subunit proteins (Adachi et al. 2005) and human anti-CD21 antibody (Tada et al. 2015b) were successfully expressed in silkworm and cocoon. Moreover, many studies attempted to express more than two target proteins at the same time (AL et al. 2004; Deng et al. 2011a; Hurh et al. 2013). Comparatively, 2A has a shorter length (usually 19–22 amino acids), higher cleavage efficiency, and, theoretically, equivalently co-expresses its two flanking genes in the same cassette via a non-proteolytic self-cleavage incident, known as the putative STOP–GO mechanism, during the translation process (Ryan et al. 1991; Donnelly et al. 2001; Atkins et al. 2007). Thus, the 2A-based multigene expression system (2A-MGES) was widely studied and developed in different cells (Szymczak et al. 2004; Fang et al. 2005; Simmons et al. 2008), insects (Daniels et al. 2014; Wang et al. 2015c), plants (Ha et al. 2010; Mikkelsen et al. 2010), and animals (Trichas et al. 2008; Deng et al. 2011b; Dempsey et al. 2012; Tian et al. 2013). The cleavage site of 2A is highly conserved, and is located between the last glycine of its C-terminal and the first proline of the downstream protein (–EXNPG↓P–) (Donnelly et al. 2001; Atkins et al. 2007).

Until now, although 2A-MGES (F2A, T2A, P2A, etc.) has been widely applied to co-express multiple genes (Szymczak et al. 2004; Luke et al. 2008), it still suffers from the expression imbalance of both flanking genes to some degree, which severely hinders the application of 2A-MGES. For example, Hurh et al. (Hurh et al. 2013) found that the expression of the downstream *EGFP* gene of T2A was gradually and dramatically decreased when the expression efficiency of its upstream genes was reduced (*HO*1, *hTBM*, and *hCD*46) in porcine fibroblasts. Moreover, expression of the anti-rat CD4 scFv protein, upstream of F2A, was significantly higher than its downstream gene in the culture supernatant of HEK-293A cells (Appleby et al. 2013). In addition, the imbalanced expression phenomenon of 2A was also confirmed in the lentiviral shuttle plasmid and other expression vectors in HeLa cells (Appleby et al. 2013; Minskaia et al. 2013; Liu et al. 2017).

The Kozak region is a consensus nucleotide sequence (GCCA/GCCAUGG) near the start codon AUG, which was first identified in eukaryotic mRNA in 1981 (Kozak 1981) and vertebrate mRNA in 1987 (Kozak 1987), and the most efficient context contains purine in position −3 (three nucleotides upstream from the AUG codon) and G in the +4 position (Kozak 1986). Up to now, a large number of studies confirm that the Kozak region efficiently improves the translation initiation efficiency and increases gene expression. In 2014, a study found that WNT16 expression was increased 3.7-fold when substituting the native regulatory sequence (GCACCC) (Hendrickx et al. 2014) with the optimized Kozak sequence

(GCCACC), and GATA4 expression was severely decreased by a mutation in its Kozak sequence (G to C in position −6) (Mohan et al. 2014). Since the reason for the low expression level of the gene downstream of 2A is the relatively low translation efficiency according to the putative "STOP–GO" 2A cleavage mechanism (Donnelly et al. 2001; Atkins et al. 2007), we predict that introducing the Kozak sequence into 2A-MGES may efficiently increase the downstream gene expression of 2A. Moreover, according to previous studies, gene arrangement may also be an important factor for efficient co-expression of both flanking genes in 2A-MGES. For example, Ho et al. (2013) reported that monoclonal antibody expression level and quality, by the L-chain+F2A+H-chain order vector, was obviously better than that by the H-chain+F2A+L-chain order vector in CHO cells. Recently, Liu et al. (2017) also showed that downstream 2A gene expression efficiency in bi-, tri-, or quad-cistronic constructs was obviously affected by varying the gene position and 2A.

Since simultaneous expression of multiple genes in specific tissues of silkworm will also accelerate both fundamental and applied research in silkworms, we focused upon multiple-gene expression strategies in silkworms. Commonly, there are two strategies to express two or more genes in silkworms. For example, when expressing multiple genes in the silk gland of silkworms, the first strategy is to express them as fusion protein, such as the human type-III procollagen and EGFP fusion protein in posterior silk gland (Tomita et al. 2003). The other strategy is to genetically cross two or more transgenic target gene silkworms, such as the combination of the human type-III collagen gene and *Bombyx mori* prolyl-hydroxylase α-subunit gene in posterior silk gland (Adachi et al. 2006), mouse IgG monoclonal antibody (Iizuka et al. 2009), and anti-CD20 monoclonal antibody in the middle silk gland (Tada et al. 2015a). Recently, two fluorescent protein genes, *DsRed* and *EGFP*, were simply co-expressed using our previously constructed 2A-MGES in the middle silk gland of silkworms (Wang et al. 2015c); however, upstream *DsRed* gene expression was more efficient than downstream *EGFP* gene expression in silkworms.

In this study, we aimed to construct a modified 2A-MGES by combining the Kozak sequence and an optimized gene arrangement for efficient expression of multiple genes in silkworms. Our results suggested that the modified Ck-2A-MGES will be a potent multiple-gene expression tool for functional and application studies in insect species.

Materials and methods

Cell lines

The Sf9 cell line (Invitrogen, USA), derived from the ovary of *Spodoptera frugiperda*, and BmE cell line, derived from

the embryo of *Bombyx mori*, were all maintained in our laboratory and cultured in Grace's medium (Gibco, USA) supplemented with 10% fetal bovine serum (Gibco, USA).

Vector construction

The method of constructing the expression vector was according to our previous reports (Wang et al. 2013, 2015c). The DsRed and EGFP genes were commercially synthesized by Genscript. The DsRed-BmCPV2A(30)-EGFP, DsRed-BmCPV2A(30)-CkEGFP, DsRed-BmCPV2A(30)-SkEGFP, DsRed-GSG_P2A-EGFP, EGFP-GSG_P2A-CkDsRed, EGFP-GSG_P2A-CkDsRed, and CkEGFP-GSG_P2A-CkDsRed fragments with *Bam*HI or *Not*I restriction endonuclease sites were amplified and assembled by overlapping PCR, with the primers detailed in Supplementary Table 1. Then all PCR fragments were cloned into the commercial Peasy-T5 Zero Vector (TransGen Biotech, China) and sequenced. Next, the appropriate gene fragments digested with *Bam*HI and *Not*I from the respective cloning vector were subcloned into the transient expression vector pSL1180[hr3A4DsRedSer1PA], which was under the control of the nuclear polyhedrosis virus enhancer (hr3), the *actin4* gene promoter from silkworm genomic DNA (A4), and the 3′-UTR of the *sericin1* gene (Ser1PA) to gain transient expression vectors in cells.

The EGFP-GSG-P2A-DsRed (G-R) and CkEGFP-GSG-P2A-CkDsRed (CkG-CkR) gene fragments digested with *Bam*HI and *Not*I were subcloned into the pSL1180[hr3Ser1DsRedSer1PA] vector, which consisted of hr3, silkworm *Ser1* gene promoter (Ser1), and Ser1PA. Finally, the whole transcriptional regulatory cassette was subcloned into the pBac(3xp3DsRedaf) transgenic vector (Wang et al. 2015c), and were, respectively, designated as pG-R and pCkG-CkR.

Transient expression in cells

The transfection method in BmE cells was performed according to our previous report (Wang et al. 2015a). After BmE, cells were seeded in six-well plates overnight; plasmids that were prepared using Plasmid Miniprep Kits (Qiagen, Germany) were mixed with X-tremeGENE HP DNA Transfection Reagent (Roche, Switzerland) at a ratio of 1 µg:3 µL in 500 µL antibiotic-free serum-free medium. After 6 h of culture, the antibiotic-free serum-free medium was replaced with the DEME medium with 10% FBS for 3 days.

Generation of transgenic silkworm

The method to generate transgenic silkworm was according to our previous reports (Wang et al. 2015c). The pG-R and pCkG-CkR transgenic expression vectors, prepared with Plasmid Miniprep Kits (Qiagen, Germany), were separately mixed with the hsp70-PIG helper vector at a ratio of 1:1, and microinjected into the D9L non-diapause eggs. The injected embryos were incubated at 25 °C and 90% relative humidity for 11 days to hatch, and the hatched silkworm larvae were carefully fed with fresh mulberry leaves. The G0 moths were sibling crossed and screened when the embryos were at 6 days post-fertilization, in the G1 generation, using the DsRed maker gene expressed in the eyes and nerve, and the positive broods were named pG-R and pCkG-CkR.

Real-time reverse transcription PCR

The method to measure target gene expression by quantification of mRNA levels was also according to our previous report (Wang et al. 2015c). First, total RNA was carefully extracted from the middle silk gland (MSG) of transgenic silkworms on the seventh day of the fifth instar, using Total RNA Extraction Kit (Promega, USA). Then 3 µg of each RNA was used to synthesize cDNA by M-MLV reverse transcriptase (Promega, USA), and all cDNAs were diluted with sterile water twofold, for use as RT-PCR templates. The relative expression based on mRNA levels of the DsRed-BmCPV2A(30)-EGFP gene, DsRed-BmCPV2A(30)-CkEGFP gene, DsRed-BmCPV2A(30)-SkEGFP gene, DsRed-GSG_P2A-CkEGFP gene, EGFP-GSG_P2A-DsRed gene, CkEGFP-GSG_P2A-CkDsRed gene, endogenous *Ser1* gene, and SW22934 gene was, respectively, quantified on an ABI Prism 7000 Sequence Detection System (Applied Biosystems, USA) using the primers detailed in Supplementary Table 2, and an SYBR Premix exTaq Kit (Takara, Japan). The silkworm *Ser*1 and SW22934 genes were used as controls.

Protein analysis

Recombinant RFP and EGFP proteins in BmE cells and transgenic cocoons were analyzed using the methods described in our previous report (Wang et al. 2015c). Total protein in BmE cells was extracted with RIPA lysis buffer (Beyotime, China) containing EDTA-free protease inhibitor cocktail (Sigma, USA), according to the instruction manual, and total cocoon protein was extracted from 20 mg transgenic cocoon powder using 1 mL Tris–HCl extraction buffer (135 mM NaCl, 2.7 mM KCl, 1.5 mM KH2PO4, and 8 mM K2HPO4, pH 7.2) at 80 °C (Wang et al. 2015c). Prior to SDS-PAGE, the protein content of all samples was measured using the Enhanced BCA Protein Assay Kit (Beyotime, China) and an equal amount of protein was analyzed by 12% SDS-PAGE with Coomassie Brilliant Blue staining and Western blotting. For Western blotting, the primary antibodies used were anti-EGFP (Genscript, China), anti-RFP (Biovisi, USA), and β-tubulin (Genscript, China), and

the secondary antibody was rabbit IgG (Beyotime, China). Bands on membranes were visualized using an ECL Western Blotting Detection System (Amersham Biosciences, USA). Images were recorded by autoexposure for a few seconds using a Chemiscope Series (Clinx science instruments, China). For quantitative analysis of *DsRed* and *EGFP* gene expression, the band intensities from Western blots were measured using ImageJ software. The relative expression of *DsRed* and *EGFP* genes in BmE cells was the ratio of their expression to β-tubulin expression. *p* values were determined using two-tailed Student's *t* test ($n = 3$).

Image

The detailed method for imaging the fluorescence signals of BmE cells, MSGs, and cocoons of silkworms was according to our previous report (Wang et al. 2015c). The cultured BmE cells, the middle silk gland on the 7th day in the fifth stage larva of silkworms, and its cocoon were directly photographed with an Olympus SZX12 fluorescent stereomicroscope (Olympus, Japan). For imaging BmE cells, the exposure times were 15 ms for DsRed and 100 ms for EGFP following excitation; and for imaging the middle silk gland and cocoon of silkworms, the exposure times were 30 ms for DsRed and 80 ms for EGFP following excitation.

Results

Kozak improved the EGFP expression downstream of 2A by enhancing its translation efficiency

To examine whether the Kozak sequence could efficiently increase the 2A downstream gene expression, the native 2A self-cleaving peptide from the silkworm cytoplasmic polyhedrosis virus BmCPV2A(30) was used to link both RFP and EGFP genes. This sequence was chosen as BmCPV2A(30) does not allow an efficient expression of the downstream EGFP gene in the DsRed-BmCPV2A(30)-EGFP fusion gene expression vector (Wang et al. 2015c). To potentially discover a better Kozak sequence, both the classic Kozak sequence (Ck), which was designed according to the Kozak principle (Kozak 1981), and the endogenous *ser1* Kozak region of silkworm (Sk) were separately fused to the N terminus of the downstream *EGFP* gene in the DsRed-BmCPV2A(30)-EGFP fusion gene expression vector (Fig. 1a). Transfection of these vectors into the BmE cells showed that no obvious difference in the transcription levels of *DsRed* and *EGFP* genes was observed among the BmCPV2A(30), BmCPV2A(30) + Ck, and BmCPV2A(30) + Sk groups (Fig. 1b). Fluorescence intensities and protein production in western blotting also showed that there was no difference in the upstream *DsRed* gene expression among the BmCPV2A(30), BmCPV2A(30) + Ck, and BmCPV2A(30) + Sk groups. By contrast, a stronger green fluorescence intensity and improved expression of the downstream *EGFP* gene were detected in BmCPV2A(30) + Ck and BmCPV2A(30) + Sk (Fig. 1c, d), suggesting both Ck and Sk significantly improved the protein yields of the downstream *EGFP* gene by improving the mRNA translation efficiency of downstream *EGFP* gene. Furthermore, band intensity analysis of EGFP protein showed that EGFP expression was, respectively, improved 4.3-fold by Ck and 1.9-fold by Sk, respectively (Fig. 1e, f), which indicated that Ck was more powerful than Sk for the efficient expression of downstream genes in 2A-MGES. In addition, the transgenic DsRed-2A-CkEGFP silkworm (pR-CkG) was also generated to confirm that the Ck modified 2A-MGES (Ck-2A-MGES) could efficiently increase downstream 2A *EGFP* gene expression compared to the huge difference between expression of *DsRed* and *EGFP* genes (Supplementary Fig. 1) in transgenic DsRed-2A-EGFP fusion gene silkworm (pR-G) (Wang et al. 2015c). These results showed that 2A downstream *EGFP* gene expression was significantly improved, and that there was no statistically significant difference between the expressed RFP and EGFP protein contents ($0.18 \pm 0.03\%$, $0.23 \pm 0.04\%$) in the middle silk gland of pR-CkG silkworms (Supplementary Fig. 2 and Table 1). Thus, we concluded that the Ck-optimized 2A-MGES could efficiently co-express multiple genes in silkworms.

Gene expression was efficiently improved by altering their arrangement in 2A-MGES

To investigate the effects of gene arrangement in the optimized GSG-P2A expression system on their expressions, two vectors R-G and G-R, with different arrangements of *DsRed* and *EGFP* genes, were constructed and transfected into BmE cells (Fig. 2a). The results showed that there was an obvious expression difference between *DsRed* and *EGFP* genes according to protein levels, although both DsRed-2A-EGFP and EGFP-2A-DsRed fusion genes were normally transcribed because DsRed mRNA and EGFP mRNA were statistically identical in the cells of R-G and G-R groups (Fig. 2b, c). By analyzing the fluorescence intensity, we found that a stronger red fluorescent signal in the G-R group was observed, compared to that in the R-G group (Fig. 2d). In addition, the expressed RFP content in the G-R group was also significantly improved (Fig. 2e), which showed that the efficient expression of *DsRed* gene was dependent on its location, flanking GSG-P2A. Band intensity analysis showed that the expressed RFP content was improved 5.3-fold, while EGFP expression was not significantly improved in the G-R group (Fig. 2f, g). Furthermore, transgenic EGFP-2A-DsRed silkworm (pG-R)
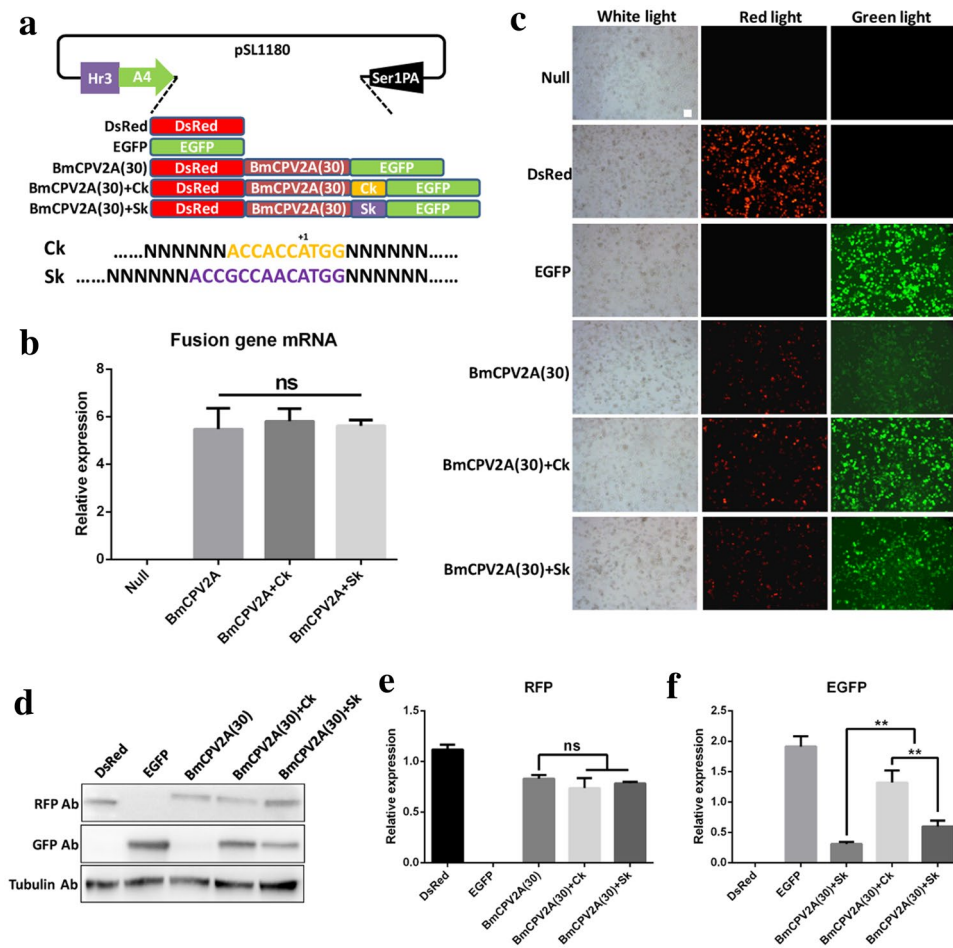
Fig. 1 Kozak sequence improved the expression of 2A downstream gene by its enhanced translation efficiency. **a** Schematic diagram of BmCPV2A(30), BmCPV2A(30)+Ck and BmCPV2A(30)+Sk fusion gene expression vectors. The amino sequence of BmCPV2A(30) is TAFDFQQDVFRSNYDLLKLCGDIESNPGP. Ck was designed according to classic Kozak sequence principle and Sk was cloned from the endogenous Ser1 gene Kozak region of silkworm. The Hr3, A4 and Ser1PA, respectively, indicated the enhancer from the nuclear polyhedrosis virus, the promoter of the silkworm *actin 4* gene and the 3′-UTR of the silkworm *sericin 1* gene. **b** The transcription level of the DsRed and EGFP fusion genes. **c** The fluorescent signal results after transfection in BmE cell. The white light, red light and green light, respectively, indicated the results in the white, red and green lights (the same below). Scale bar, 400 μm. **d** Protein analysis of samples from BmE cells. In western blotting, three primary antibodies (anti-RFP antibody, anti-GFP antibody and anti-β-tubulin antibody) and two secondary antibodies (anti-mouse IgG antibody and anti-rabbit IgG antibody) were used (the same below). **e, f** The relative RFP and EGFP expression level in BmE cells

Table 1 Microinjection analysis of the pR-CkG, pG-R and pCkG-CkR silkworms

Strain	Injected eggs	Hatched eggs (%)	G1 brood	Positive brood (%)
pR-CkG	400	35 (8.8)	5	2 (40)
pG-R	400	28 (7)	12	3 (25)
pCkG-CkR	400	115 (28.7)	35	15 (42.9)

was also generated to confirm whether the optimized gene arrangement of *EGFP* gene located upstream and *DsRed* gene located downstream of GSG_P2A could efficiently co-express these two genes compared to pR-G silkworms (Supplementary Fig. 1), and the results showed that there was no obvious difference in the expressed RFP and EGFP contents ($0.26 \pm 0.02\%$, $0.19 \pm 0.05\%$) in pG-R silkworm cocoon (Supplementary Fig. 3 and Table 1). These results strongly suggested that the expression of the analyzed genes could be efficiently improved by optimizing their arrangement in 2A-MGES.

Fig. 2 Gene efficient expression was affected by their arrangement in 2A-MGES. **a** Schematic diagram of R-G and G-R fusion gene expression vectors. The GSG-P2A was an optimized 2A, which is derived from the porcine teschovirus-1 2A by adding a glycine–serine–glycine spacer (GSG) at its N terminus and its amino sequence is GSGATNFSLLKQAGD-VEENPGP. **b, c** The transcription levels of DsRed gene and EGFP gene in the R-G and G-R groups, respectively. **d** The fluorescent signal results after transfection in BmE cell. Scale bar, 400 μm. **e** Protein analysis of samples from BmE cells. **f, g** The relative RFP and EGFP expression levels in BmE cells, respectively
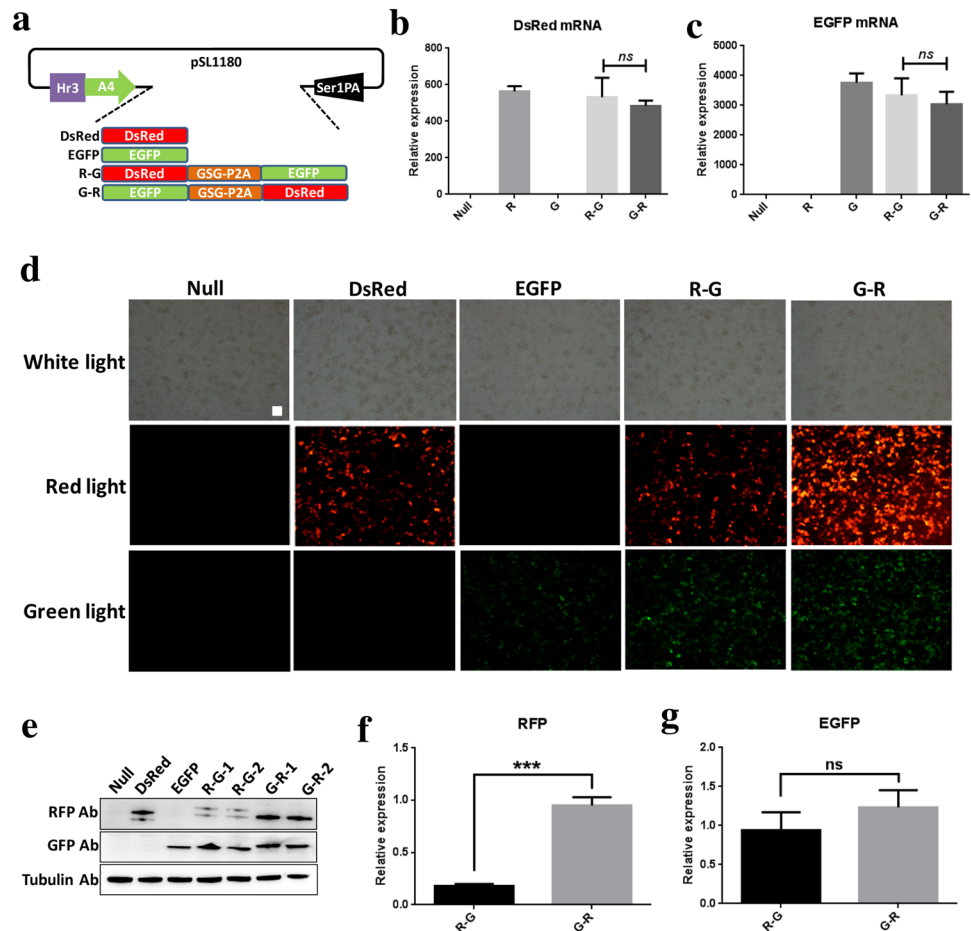
Combining Ck and optimized gene arrangement further improved the gene expression efficiency in 2A-MGES

Since both Ck and an optimized gene arrangement in 2A-MGES could efficiently improve the expression of genes flanking 2A, two transient expression vectors, G-CkR and CkG-CkR (Fig. 3a), were also constructed and transfected into BmE cells to examine whether combining Ck and a modified gene arrangement in 2A-MGES could further synergistically improve gene expression. The results showed that the expression of downstream *DsRed* gene in G-CkR was only slightly improved compared to that in G-R (Fig. 3b, c), most probably because the optimized gene arrangement, with the *EGFP* gene located upstream and the *DsRed* gene located downstream of GSG_P2A, could already efficiently co-express the *DsRed* and *EGFP* genes (Fig. 2 and Supplementary Fig. 3). However, when Ck was introduced to both upstream *EGFP* gene and downstream *DsRed* gene in CkG-CkR, both red and green fluorescent signals in the CkG-CkR group were significantly improved and stronger than those in the G-R and G-CkR groups (Fig. 3b). Furthermore, the improved expressions of DsRed and EGFP genes in the

CkG-CkR group, when considering protein levels, were also confirmed by Western blot (Fig. 3c). Grey intensity calculation indicated that expression of *DsRed* and *EGFP* genes in the CkG-CkR group were 2.1- and 1.6-fold higher, respectively, than in the G-R group (Fig. 3d, e). The results suggested that gene expression could be systematically improved by combining Ck and optimized gene arrangement in 2A-MGES.

Modified 2A-MGES achieved efficient and consistent expression of DsRed and EGFP genes in the silk gland and cocoon of transgenic silkworms

Finally, the optimized CkEGFP-GSG_P2A-CkDsRed fusion gene (CkG-CkR), which theoretically has the best expression efficiency for both *DsRed* and *EGFP* genes, was used to construct a transgenic expression vector and generate the transgenic silkworm, pCkG-CkR (Fig. 4a, b and Table 1). Transcription levels of the CkG-CkR fusion gene in the pCkG-CkR transgenic silkworm was ~ 56% of the endogenous *Ser1* gene (Fig. 4c, d). The silk gland and cocoon from pCkG-CkR transgenic silkworm exhibited both strong red and green fluorescent signals

Fig. 3 Combination of Kozak sequence and optimized gene arrangement further improved the expressions of upstream and downstream genes in 2A-MGES. **a** Schematic diagram of G-R, G-CkR and CkG-CkR fusion gene expression vectors. **b** The fluorescent signal results after transfection in BmE cell. Scale bar, 400 μm. **c** Protein analysis of samples from transfected BmE cells. **d, e** The relative RFP and EGFP expression level in the transfected BmE cells
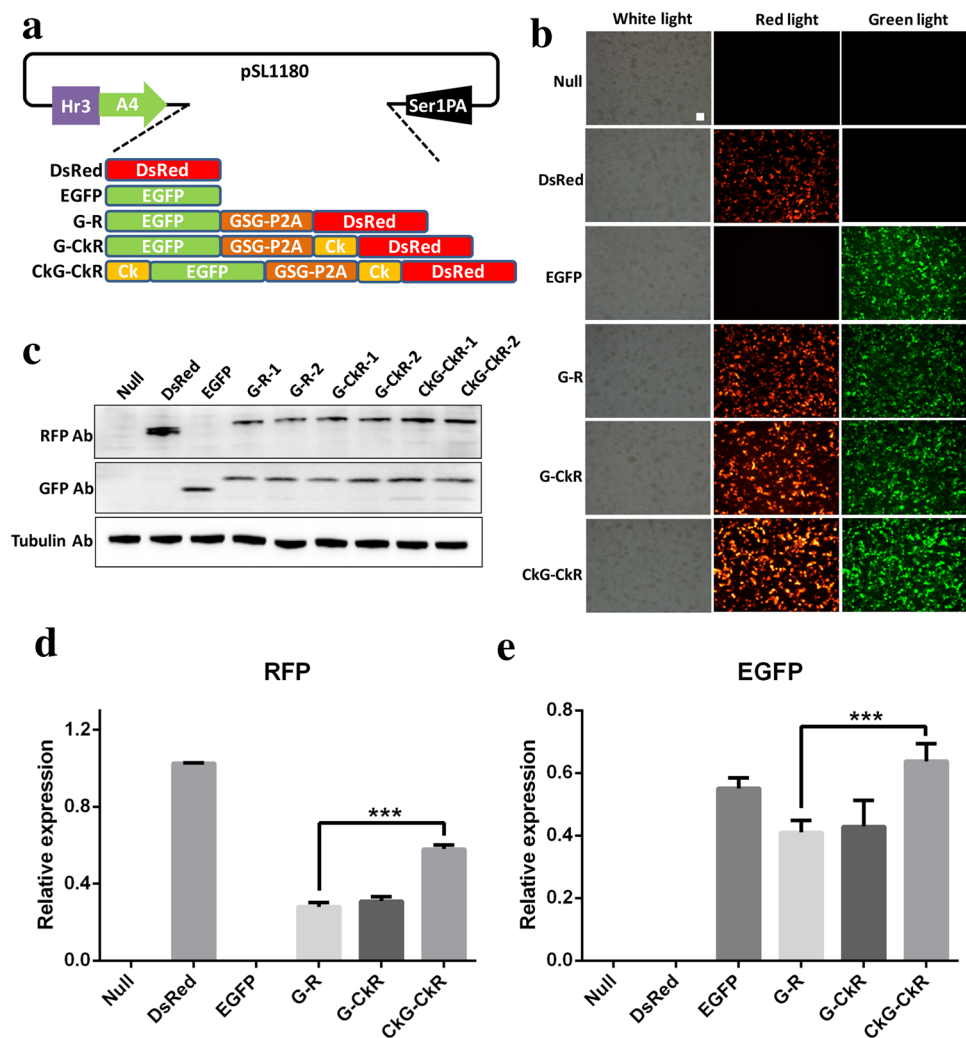
(Fig. 4e), and Western blot showed the specified immunoblot bands of cleaved EGFP-P2A (~ 36 kDa) and RFP (~ 27 kDa) from the total pCkG-CkR cocoon proteins (Fig. 4f and Supplementary Fig. 4), which indicated that *DsRed* and *EGFP* genes were successfully co-expressed in the pCkG-CkR silkworm, and that the cleaved RFP and EGFP proteins were secreted into the silk thread used to make the cocoon. Most importantly, there was no significant difference between RFP and EGFP protein contents in the pCkG-CkR silkworm ($0.36 \pm 0.03\%$, $0.28 \pm 0.07\%$) (Fig. 4g). Moreover, both stronger red and green fluorescent signals and higher RFP and EGFP protein contents (improved 1.4- and 1.5-fold, respectively) were observed in the pCkG-CkR silkworm, compared to that in the pG-R silkworm. All in all, we conclude that the modified 2A-MGES with Ck and optimized gene arrangement could efficiently co-express multiple genes in silkworms.

Discussion

Simultaneous co-expression of multiple genes is a good strategy to study the relationship among more than two target genes, express complete multi-subunit protein complexes by, respectively, co-expressing their multiple subunit units, visually track target protein distribution and movement in tissues using a fluorescent protein fusion, and even study the continuous and programmed processes of some biological phenomena with multiple targets. However, the common strategy of co-expressing multiple genes in individuals is by genetic hybridization, which is a complicated process (van den Akker et al. 2002; Adachi et al. 2006; Iizuka et al. 2009; Tada et al. 2015a). In the silkworm, *B. mori*, there is also an increasing need for co-expression of multiple genes for functional and
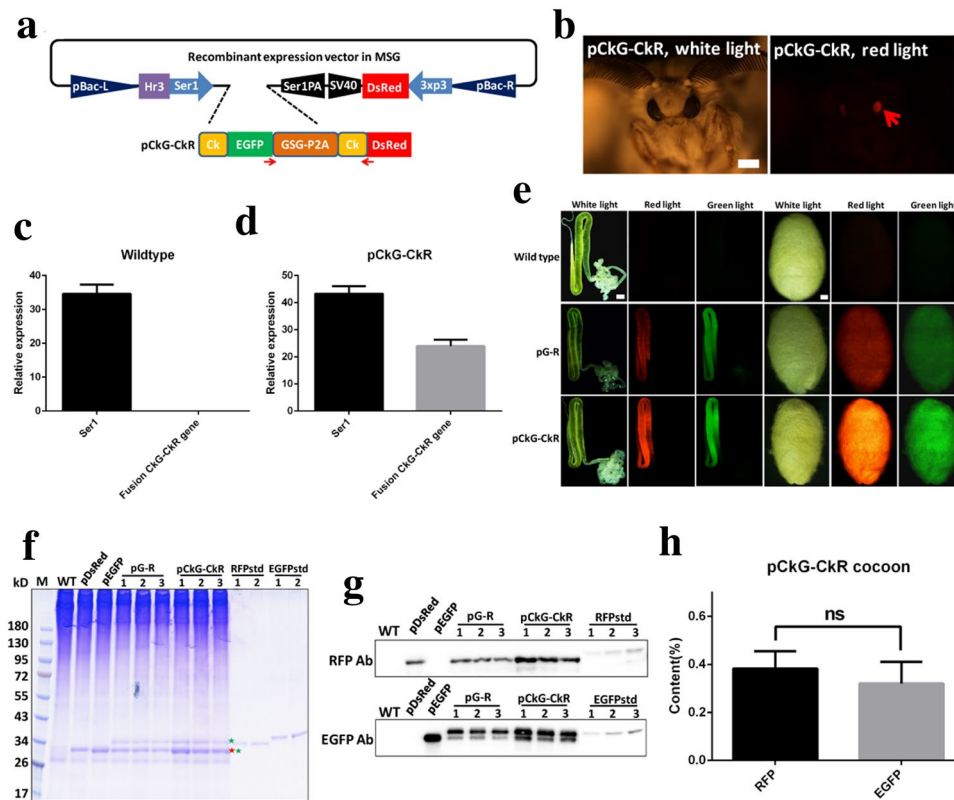
Fig. 4 The modified 2A-MGES achieved the efficient co-expressions of DsRed and EGFP genes in the transgenic silkworms. **a** Schematic diagram of the transgenic CkG-CkR fusion gene expression vectors. 3xp3DsRedSV40, pBacL and pBacR, respectively, indicate the fluorescent maker gene transcription regulatory cassette, the left and right arms of piggyBac transposon (Wang et al. 2015a). The *DsRed* and *EGFP* genes were linked by the GSG-P2A, in which a furin cleavage site RAKR was added at the front of GSG-P2A and it can remove the GSG-P2A peptide. The red arrows represented the RT-PCR primers. **b** Screen for positive transgenic silkworm. **c, d** The transcription of the CkG-CkR fusion gene in the wild type and pCkG-CkR silkworms by RT-PCR. Ser1 is the endogenous gene, as a control. **e** Fluorescent images of the middle silk glands and cocoons of the wild type, pG-R and pCkG-CkR silkworms. Scale bar, 2 mm. **f, g** Protein analysis of samples from the pG-R and pCkG-CkR cocoons by Coomassie blue brilliant blue staining and western blotting. The pG-R-1,2,3 represented the transgenic cocoons of the pG-R silkworm without Ck optimization and pCkG-CkR-1,2,3 represented the transgenic cocoons of the pCkG-CkR silkworm. Red star and green star meant the cleaved RFP and cleaved EGFP proteins. In western blotting, the content of RFP standard proteins was 25 ng, 50 ng, 100 ng and the content of EGFP standard proteins was 25 ng, 50 ng, 100 ng. **h** RFP and EGFP contents in the pCkG-CkR cocoon

applied research (Xia et al. 2004, 2009, 2014). Thus, we previously constructed 2A-MGES, which achieved the co-expression of multiple genes in silk glands and cocoons of transgenic silkworms (Wang et al. 2015c), while these two genes can not be expressed in the same amount in silkworms.

In this study, we further improved 2A-MGES to achieve the efficient co-expression of *DsRed* and *EGFP* genes in the silk gland and cocoon of transgenic silkworm, by introducing the Kozak sequence and altering the arrangement of genes flanking 2A. To the best of our knowledge, the Kozak sequence was, for the first time, introduced into 2A-MGES to improve the expression of genes flanking 2A, in particular, the downstream gene expression. We found that the designed Ck was more efficient than the endogenous Sk in improving BmCPV2A downstream *EGFP* gene

expression by enhancing the *EGFP* translation efficiency, in vitro and in vivo. Moreover, our results also confirmed that gene arrangement flanking 2A would affect their efficient expression, which was the same with previous reports (Appleby et al. 2013; Hurh et al. 2013; Minskaia et al. 2013; Liu et al. 2017). With the *EGFP* gene located upstream and the *DsRed* gene located downstream of 2A, there was an obvious improvement in their co-expression by 2A-MGES, suggesting it was a more optimized gene arrangement. These observations were also confirmed in pCkG-CkR silk gland and cocoon, and most importantly, compared to that in the pR-G silkworm (Wang et al. 2015c), their expressions in this modified 2A-MGES with Ck and optimized gene arrangement were more efficient and were statistically equal, which suggested that our modified 2A-MGES efficiently and consistently co-expressed *DsRed* and *EGFP* genes in silkworm.

According to the modified version of the scanning mechanism of the ribosome (Kozak 1981), when a migrating 40S ribosomal subunit scans the AUG triplet, it will stop and initiate the downstream gene translation irrespective of the flanking sequences at the AUG triplet, since it is like a "stop signal". Meanwhile, if the Kozak region bordering the AUG codon is optimal, all 40S subunits will stop at that AUG, so the non-self-cleaving sequence (NCS) linked to the downstream *EGFP* gene, in which the Kozak region is optimal in the DsRed-NCS-CkEGFP or CkDsRed-NCS-CkEGFP fusion gene expression vectors, should also be expressed in the transfected BmE cells. However, unfortunately, there was no any ~ 27 kDa EGFP protein of NCS downstream to be detected, except the RFP-EGFP fusion protein (data not shown); thus, it is still unclear how the 2A downstream gene expression was significantly improved in the modified 2A-MGES. Based on the putative STOP–GO cleavage mechanism of 2A (Donnelly et al. 2001; Atkins et al. 2007), gene expression imbalance in the translation products linked by 2A was most probably caused by leaky scanning of the ribosome, because this translational complex is prone to disintegration, terminating the downstream mRNA translation, since it suffered an abnormal course of protein synthesis when the upstream nascent peptide was hydrolyzed and released out of the ribosome complex. Considering there was no obvious difference between RFP and EGFP protein expression in the pCkG-CkR silkworm by the modified 2A-MGES, we thought that, when Ck was inserted at the N terminus of the target gene, most of these disintegrating and separating ingredients from ribosome complexes were re-recruited and reformed new ribosome complexes, because there were a large number of translation initiation factors in the Ck region. Then the newly formed ribosome complexes continued to translate the downstream gene, efficiently avoiding the leaky scanning of the 2A downstream gene; thus, genes upstream and downstream of 2A were co-expressed in equal amounts. In addition, both the previous studies (Ho et al. 2013; Liu et al. 2017) and our results showed that expression of genes flanking 2A was highly dependent on their arrangement. We guessed that the downstream gene expression efficiency of 2A would be severely limited by the interaction degree among the upstream nascent peptide, 2A peptide, and ribosome complexes, or between their formed mRNA structure and ribosome complexes. Altogether, our results suggested that gene expression flanking 2A was more efficient when EGFP gene locates the upstream of 2A rather than DsRed gene.

From the report in 2000, that piggyBac transposon was used to efficiently create a transgenic silkworm (Tamura et al. 2000), lots of transgenic overexpression systems in silkworm have been built, including in the middle and posterior silk gland of silkworms (Tomita et al. 2003; Adachi et al. 2006). Previously, by overexpressing the human acidic

fibroblast growth factor (aFGF) into the outer layer of silk using our constructed middle silk gland expression system, we successfully created aFGF-containing silk, which significantly promoted NIH3T3 cell proliferation and wound healing (Wang et al. 2014, 2015b). Potentially, if multiple functional genes (promoting wound healing plus antibacterial or anti-inflammatory activities, silkworm nuclear polyhedrosis virus resistant gene plus cytoplasmic polyhedrosis virus resistant gene (Jiang et al. 2012, 2017)) can be efficiently co-expressed by modified 2A-MGES, multi-functional silk, or multi-resistant silkworms could be created.

Conclusion

In summary, we successfully improved the 2A self-cleaving peptide-based multigene expression system by optimizing the Kozak sequence and altering gene arrangement, which resulted in efficient co-expression of two flanking genes in the middle silk gland of silkworms and cocoons. This system is a novel and useful tool to simultaneously express multiple genes in silkworms for research into their function and in functional silk material innovation. In addition, these results will also guide the construction of more efficient 2A-MGES for other *Lepidopteran* insects and species in the future.

References

Adachi T, Tomita M, Yoshizato K (2005) Synthesis of prolyl 4-hydroxylase α subunit and type IV collagen in hemocytic granular cells of silkworm, *Bombyx mori*: involvement of type IV collagen in self-defense reaction and metamorphosis. Matrix Biol J Int Soc Matrix Biol 24:136–154

Adachi T, Tomita M, Shimizu K, Ogawa S, Yoshizato K (2006) Generation of hybrid transgenic silkworms that express *Bombyx mori* prolyl-hydroxylase alpha-subunits and human collagens in posterior silk glands: Production of cocoons that contained collagens with hydroxylated proline residues. J Biotechnol 126:205–219

AL Szymczak et al. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol 22:589

Appleby SL, Irani Y, Mortimer LA, Brereton HM, Klebe S, Keane MC, Cowan PJ, Williams KA (2013) Co-expression of a scFv antibody fragment and a reporter protein using lentiviral shuttle plasmid containing a self-processing furin-2A sequence. J Immunol Methods 397:61–65

Atkins JF, Wills NM, Loughran G, Wu CY, Parsawar K, Ryan MD, Wang CH, Nelson CC (2007) A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). RNA 13:803–810

Baron U, Freundlieb S, Gossen M, Bujard H (1995) Co-regulation of two gene activities by tetracycline via a bidirectional promoter. Nucleic Acids Res 23:3605–3606

Bieniossek C, Imasaki T, Takagi Y, Berger I (2012) MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem Sci 37:49–57

Daniels RW, Rossano AJ, Macleod GT, Ganetzky B (2014) Expression of multiple transgenes from a single construct using viral 2A peptides in Drosophila. PLoS One 9:e100637

Dempsey WP, Fraser SE, Pantazis P (2012) PhOTO zebrafish: a transgenic resource for in vivo lineage tracing during development and regeneration. Plos One 7:e32888

Deng W, Yang D, Zhao B, Ouyang Z, Song J, Fan N, Liu Z, Zhao Y, Wu Q, Nashun B (2011a) Use of the 2A peptide for generation of multi-transgenic pigs through a single round of nuclear transfer. Plos One 6:e19986

Deng W, Yang D, Zhao B, Ouyang Z, Song J, Fan N, Liu Z, Zhao Y, Wu Q, Nashun B, Tang J, Wu Z, Gu W, Lai L (2011b) Use of the 2A peptide for generation of multi-transgenic pigs through a single round of nuclear transfer. PLoS One 6:e19986

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J Gen Virol 82:1013–1025

Fang J, Qian JJ, Yi S, Harding TC, Tu GH, VanRoey M, Jooss K (2005) Stable antibody expression at therapeutic levels using the 2A peptide. Nat Biotechnol 23:584–590

Foti SB, Samulski RJ, McCown TJ (2009) Delivering multiple gene products in the brain from a single adeno-associated virus vector. Gene Ther 16:1314–1319

Ha SH, Liang YS, Jung H, Ahn MJ, Suh SC, Kweon SJ, Kim DH, Kim YM, Kim JK (2010) Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. Plant Biotechnol J 8:928–938

Hendrickx G, Boudin E, Fijalkowski I, Nielsen TL, Andersen M, Brixen K, Van Hul W (2014) Variation in the Kozak sequence of WNT16 results in an increased translation and is associated with osteoporosis related parameters. Bone 59:57–65

Ho SC, Bardor M, Li B, Lee JJ, Song Z, Tong YW, Goh LT, Yang Y (2013) Comparison of internal ribosome entry site (IRES) and Furin-2A (F2A) for monoclonal antibody expression level and quality in CHO cells. PLoS One 8:e63247

Hurh S, Cho B, You DJ, Kim H, Lee EM, Lee SH, Park SJ, Park HC, Koo OJ, Yang J, Oh KH, Lee BC, Hwang JI, Ahn C (2013) Expression analysis of combinatorial genes using a bi-cistronic T2A expression system in porcine fibroblasts. PLoS One 8:e70486

Iizuka M, Ogawa S, Takeuchi A, Nakakita S, Kubo Y, Miyawaki Y, Hirabayashi J, Tomita M (2009) Production of a recombinant mouse monoclonal antibody in transgenic silkworm cocoons. FEBS J 276:5806–5820

Jiang L, Cheng T, Zhao P, Yang Q, Wang G, Jin S, Lin P, Xiao Y, Xia Q (2012) Resistance to BmNPV via overexpression of an exogenous gene controlled by an inducible promoter and enhancer in transgenic silkworm, Bombyx mori. PLoS One 7:e41838

Jiang L, Peng Z, Guo H, Sun J, Sun Q, Xia F, Huang C, Xu G, Xia Q (2017) Enhancement of antiviral capacity of transgenic silkworms against cytoplasmic polyhedrosis virus via knockdown of multiple viral genes. Dev Comp Immunol 77:138–140

Kim C, Wilson T, Fischer KF, Williams MA (2013) Sustained interactions between T cell receptors and antigens promote the differentiation of CD4(+) memory T cells. Immunity 39:508–520

Kozak M (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res 9:5233–5252

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kunes YZ, Gion WR, Fung E, Salfeld JG, Zhu RR, Sakorafas P, Carson GR (2009) Expression of antibodies using single-open reading frame vector design and polyprotein processing from mammalian cells. Biotechnol Prog 25:735–744

Liu Z, Chen O, Wall JBJ, Zheng M, Zhou Y, Wang L, Ruth Vaseghi H, Qian L, Liu J (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7:2193

Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD (2008) Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. J Gen Virol 89:1036–1042

Mikkelsen MD, Olsen CE, Halkier BA (2010) Production of the cancer-preventive glucoraphanin in tobacco. Mol Plant 3:751–759

Minskaia E, Nicholson J, Ryan MD (2013) Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. BMC Biotechnol 13:67

Mohan RA, van Engelen K, Stefanovic S, Barnett P, Ilgun A, Baars MJ, Bouma BJ, Mulder BJ, Christoffels VM, Postma AV (2014) A mutation in the Kozak sequence of GATA4 hampers translation in a family with atrial septal defects. Am J Med Genet A 164A:2732–2738

Ryan MD, King AM, Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J Gen Virol 72(Pt 11):2727–2732

Simmons AD, Moskalenko M, Creson J, Fang J, Yi S, VanRoey MJ, Allison JP, Jooss K (2008) Local secretion of anti-CTLA-4 enhances the therapeutic efficacy of a cancer immunotherapy with reduced evidence of systemic autoimmunity. Cancer Immunol Immunother 57:1263–1270

Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, Vignali DA (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol 22:589–594

Tada M, Tatematsu K, Ishii-Watabe A, Harazono A, Takakura D, Hashii N, Sezutsu H, Kawasaki N (2015a) Characterization of anti-CD20 monoclonal antibody produced by transgenic silkworms (Bombyx mori). MAbs 7:1138–1150

Tada M, Tatematsu KI, Ishiiwatabe A, Harazono A, Takakura D, Hashii N, Sezutsu H, Kawasaki N (2015b) Characterization of anti-CD20 monoclonal antibody produced by transgenic silkworms (Bombyx mori). Mabs 7:1138–1150

Tamura T, Thibert C, Royer C, Kanda T, Abraham E, Kamba M, Komoto N, Thomas JL, Mauchamp B, Chavancy G, Shirk P, Fraser M, Prudhomme JC, Couble P (2000) Germline transformation of the silkworm Bombyx mori L. using a piggyBac transposon-derived vector. Nat Biotechnol 18:81–84

Tang W, Ehrlich I, Wolff SB, Michalski AM, Wolfl S, Hasan MT, Luthi A, Sprengel R (2009) Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. J Neurosci 29:8621–8629

Tian Y, Li W, Wang L, Liu C, Lin J, Zhang X, Zhang N, He S, Huang J, Jia B, Liu M (2013) Expression of 2A peptide mediated tri-fluorescent protein genes were regulated by epigenetics in transgenic sheep. Biochem Biophys Res Commun 434:681–687

Tomita M, Munetsuna H, Sato T, Adachi T, Hino R, Hayashi M, Shimizu K, Nakamura N, Tamura T, Yoshizato K (2003) Transgenic silkworms produce recombinant human type III procollagen in cocoons. Nat Biotechnol 21:52–56

Trichas G, Begbie J, Srinivas S (2008) Use of the viral 2A peptide for bicistronic expression in transgenic mice. BMC Biol 6:40

van den Akker E, Forlani S, Chawengsaksophak K, de Graaff W, Beck F, Meyer BI, Deschamps J (2002) Cdx1 and Cdx2 have

overlapping functions in anteroposterior patterning and posterior axis elongation. Development 129:2181–2193

Wang F, Xu H, Yuan L, Ma S, Wang Y, Duan X, Duan J, Xiang Z, Xia Q (2013) An optimized sericin-1 expression system for mass-producing recombinant proteins in the middle silk glands of transgenic silkworms. Transgenic Res 22:925–938

Wang F, Xu H, Wang Y, Wang R, Yuan L, Ding H, Song C, Ma S, Peng Z, Peng Z, Zhao P, Xia Q (2014) Advanced silk material spun by a transgenic silkworm promotes cell proliferation for biomedical application. Acta Biomater 10:4947–4955

Wang F, Wang R, Wang Y, Xu H, Yuan L, Ding H, Ma S, Zhou Y, Zhao P, Xia Q (2015a) Remobilizing deleted piggyBac vector post-integration for transgene stability in silkworm. Mol Genet Genomics 290:1181–1189

Wang F, Wang R, Wang Y, Zhao P, Xia Q (2015b) Large-scale production of bioactive recombinant human acidic fibroblast growth factor in transgenic silkworm cocoons. Sci Rep 5:16323

Wang Y, Wang F, Wang R, Zhao P, Xia Q (2015c) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. Sci Rep 5:16273

Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Yu J, Wang J, Li R, Shi J, Li H, Li G, Su J, Wang X, Li G, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Xu J, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li C, Li D, Sun Y, Zhang Z, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Li J, Ye J, Chen H, Zhou Y, Liu B, Wang J, Ye J, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B, Wang W, Ye C, Li S, Wang J, Wong GK, Yang H, Biology Analysis G (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*) Science 306:1937–1940

Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu S, He N, Zhou Y, Liu H, Zhao J, Ye C, Du Z, Pan G, Zhao A, Shao H, Zeng W, Wu P, Li C, Pan M, Li J, Yin X, Li D, Wang J, Zheng H, Wang W, Zhang X, Li S, Yang H, Lu C, Nielsen R, Zhou Z, Wang J, Xiang Z, Wang J (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). Science 326:433–436

Xia Q, Li S, Feng Q (2014) Advances in silkworm studies accelerated by the genome sequencing of *Bombyx mori*. Annu Rev Entomol 59:513–536

Ye L, Zhuang L, Li J, You Z, Liang J, Wei H, Lin J, Zhong B (2013) Analysis of the activity of virus internal ribosome entry site in silkworm *Bombyx mori*. Acta Biochim Biophys Sin 45:534–539