Combinatorial engineering for heterologous gene expression

Friederike Zwick, Rahmi Lale, and Svein Valla*
Department of Biotechnology; Norwegian University of Science and Technology; Trondheim, Norway

Keywords: combinatorial engineering, Escherichia coli, heterologous gene expression, XylS/Pm, recombinant protein production, synthetic biology, metabolic engineering

Submitted: 03/15/13
Revised: 04/12/13
Accepted: 04/15/13
http://dx.doi.org/10.4161/bioe.24703
*Correspondence to: Svein Valla;
Email: svein.valla@ntnu.no

Addendum to: Zwick F, Lale R, Valla S. Strong stimulation of recombinant protein production in Escherichia coli by combining stimulatory control elements in an expression cassette. Microb Cell Fact 2012; 11:33; PMID:23031552; http://dx.doi.org/10.1186/1475-2859-11-33

Tools for strain engineering with predictable outcome are of crucial importance for the nascent field of synthetic biology. The success of combining different DNA biological parts is often restricted by poorly understood factors deriving from the complexity of the systems. We have previously identified variants for different regulatory elements of the expression cassette XylS/Pm. When such elements are combined, they act in a manner consistent with their individual behavior, as long as they affect different functions, such as transcription and translation. Interestingly, sequence context does not seem to influence the final outcome significantly. Expression of reporter gene bla could be increased up to 75 times at the protein level by combining three variants in one cassette. For other tested reporter genes similar results were obtained, except that the stimulatory effect was quantitatively less. Combination of individually characterized DNA parts thus stands as suitable method to achieve a desired phenotype.

Engineering approaches are gaining momentum within the field of molecular biology. Several registries have been established which serve as repository for DNA parts with characterized behaviors.1,2 However, despite its immediate appeal, design and construction of biological systems with predictable outcome by combining DNA parts or elements is challenging due to the complexity of living organisms. Often the combined DNA parts are not acting independently and their function is influenced by their genetic and cellular context and the environment.

Previously, we have studied variants of the DNA elements that are part of the expression system XylS/Pm: the positively regulated transcription factor XylS; the promoter Pm; and the DNA sequence corresponding to the 5' untranslated region of Pm (5'-UTR). These variants were isolated by screening procedures from large libraries, constructed either by directed mutagenesis (for xylS) or by doped oligonucleotide synthesis (for Pm and 5’-UTR). Both variants with increased (hereafter high-expression variants) and decreased (hereafter low-expression variants) expression levels relative to the wild type were identified in a context of surrounding wild type sequences. We became interested in assessing whether expression levels could be further increased by engineering the system by means of combining some of these variants with known phenotype. Here we use the term “engineered” in the sense of building a system that behaves as expected, given the characteristics of its components. The outcome obviously would have an impact on rational engineering approaches, which are crucial within the nascent field of synthetic biology.

Variants for the promoter region have been shown to mainly influence transcription, as both transcript and final protein production were stimulated to a similar degree by these variants. Most likely it is the step of transcription initiation which is influenced. The best high-expression variant (ML2–5,3 here termed comP), increased protein expression of the reporter gene bla (coding for β-lactamase, conferring ampicillin resistance to the host cell) about 5 times. Also a high-expression variant chosen for the activator protein gene (StEP-13,4 termed comX) seemed to mainly act at the level of transcription and led to a similar final protein expression as...
the variant comP (Fig. 1A). Based on a model of the N-terminal domain of XylS it can be assumed that the mutations in this variant influence different functional aspects, as for example inducer-binding and dimerization, which both lead to a higher rate of transcription initiation from $P_m$.

A combination of the two high-expression variants in the same expression cassette (comXP) did not significantly stimulate expression levels further (Fig. 1B), and one explanation could be that both variant elements act optimally in combination with the wild type elements, which were present during their identification. Alternatively, the capacity of the expression cassette or host might be saturated.

However, higher protein amounts could be obtained, when a high-expression 5'-UTR variant was used. For this DNA element, several variants have been identified that seem to primarily act on translation, which is reflected by a higher increase at the protein than at the transcript level, even if the putative Shine-Dalgarno-site was not mutated. Translation initiation efficiency is known to be influenced by mRNA secondary structure formation via the accessibility of the translation initiation region, explaining the influence of mutations in the 5'-UTR on this step. The chosen variant (comU) stimulated translation of $\beta$-lactamase about 15 times, while transcript levels were similar to those for comP and comX (Fig. 1A). Interestingly, when comU was combined either with comX or comP, translation was further stimulated, up to 50 times (comXU), compared with the levels achieved with the wild-type system (Fig. 1B). This clearly demonstrates that higher expression levels than those observed for the single variants are achievable as long as two single variants are combined which individually influence different steps of expression (here: transcription and translation). Thus the absence of an additional stimulatory effect for comXP compared with the single variants comX and comP seems to be caused rather by restrictions deriving from the fact that they both influence the same step of expression than by cell-born limitations.

Combination of all three high-expression variants led to a yet further improvement, and $\beta$-lactamase enzyme activity about 75 times higher than that of the wild type cassette was reached (Fig. 1C).

The stimulatory effect achieved by the combination of high-expression variants clearly demonstrated that the individual elements can be combined very efficiently. Despite this promising finding, some degree of context dependency cannot be excluded and therefore it was decided to test whether a screening, in which one high-expression variant was used as starting point to identify high-expression variants of another regulatory element, would result in even stronger stimulation.

This was assessed by three different procedures: (1) screening for $P_m$-variants in a comX background seemed to be most promising, because no stimulatory effect had been observed from the direct combination of comP and comX, which might be caused by an adaptation of the identified variants to the functionally related wild type element. However, no variant combination with increased expression levels compared with the levels achieved by comX could be identified from the resulting $P_m$ library (size of ~220,000 clones), indicating that maximum transcription initiation improvement was reached already by the single variants. In other words the $P_m$ mutations did not
have any additive effect on the observed phenotype of comX.

(2) Also in the second screening comX was used as a starting point and combined with a mutated 5′-UTR region. The most promising variant combination, which was identified by this screening (library size of ~100 000 clones), tolerated slightly less ampicillin than comXU on agar medium (10 mg/mL vs. 12 mg/mL). In principle one might expect to at least re-identify the comU variant.

(3) Neither a screening for improved 5′-UTR variants in a comP background led to variant combinations with higher protein expression levels than levels obtained by the direct combination comP/ comU (library size of ~800 000 clones). Even if it cannot be excluded that more promising variant combinations could be identified by more extensive screenings, it was concluded that the direct combination of independently identified variants was the better procedure since all elements could be identified in parallel and the final outcome was at least as good as from the more laborious screening procedure. This conclusion is certainly not valid for all biological systems and contexts, but it is encouraging that this approach can be used for applications that are relevant both for protein production and synthetic biology.

Initially we had focused mainly on combining high-expression variants with the aim to further improve the final protein production levels. However, we also wanted to investigate how combination of low-expression variants with high-expression variants would perform. Several 5′-UTR variants with low-expression phenotypes have previously been reported,7 and one of them (DI-6) was combined with comP. Like comU, DI-6 also acts mainly at the level of translation. For the resulting construct similar transcript amounts as for comP were observed, while β-lactamase enzyme activity was about 5 times less than that of the wild type system (data not shown). This might be useful for reduction of transcript levels for very high-expression elements, as for example the bacteriophage T7 promoter, which is known to produce vast amounts of transcripts, which often cannot be handled by the translational machinery.8 It also demonstrates that the combination approach is not limited to maximization of expression, but could for example also be applied for metabolic engineering purposes.

For rational engineering it is desirable to have a defined system that results in similar output regardless of the gene to be expressed. Therefore we also tested whether combination of stimulatory variants would lead to a similar increase in expression levels with other reporter genes (celB, luc, gm-csf, and scFv-phOx). In general the strategy was found to stimulate expression of all tested reporter genes, but to varying extents (Table 1). Variations might be caused by effects such as gene- and host-dependent limitations for increases in expression and context-dependency between gene sequences and the variant elements.

With celB as the reporter gene (coding for phosphoglucomutase), which is known to be expressed at high levels from the wild type XylS/Pm system,9 a higher increase than the threshold level could be observed for the different combination cassettes, indicating translational limitations. Also in the case of the host-toxic antibody fragment scFv-phOx, gene-born restrictions might play a role, as the high background expression of the combination plasmids probably leads to cell death prior to production, which was reflected by low growth rates of cells with the corresponding plasmids.

Expression of all reporter genes in the test might have been influenced by the gene dependency of the variants. ComP and comX have been shown to act relatively gene-independent and led to increased expression also of other genes than celB.8,9 For variants of the 5′-UTR there seems to exist a certain degree of gene-dependency, especially for those variants that mainly act on translation.7 This might also have an influence on the outcome of combination of variants, and thus it might be a better approach to directly use 5′-UTR variants that have been identified as stimulating for the gene of interest.

Generally the studies show that the output of the combination approach correlates with the properties of the individual elements that are combined, independent of what gene is used. However, the quantitative effects vary among genes, presumably reflecting a variety of gene-specific parameters that to different degrees mask the effects of the expression elements. Strong additive effects of the combined elements also appear to be more likely if the variants influence different functional steps in expression, such as transcription and translation.

Disclosure of Potential Conflicts of Interest
No potential conflict of interest was disclosed.

Table 1. Stimulation of expression by comKPU or comP/ lac at the transcript and final protein production levels for different reporter genes

| Reporter gene | Transcript level | Protein level |
|---------------|------------------|---------------|
| wild type a   | 1                | 1             |
| bce           | 63               | 75            |
| luc           | Not tested       | 4             |
| gm-csf        | 5                | 8             |
| ScFv-phOx b   | Not tested       | a             |
| celB          | 15               | 8.5           |

a: coding for luciferase; gm-csf, coding for granulocyte-macrophage colony-stimulating factor; scFv-phOx: coding for a single-chain antibody variable fragment; celB: coding for a phosphoglucomutase. All expressed by comKPU, except for luc, which was only tested with comP. For each construct with different genes expression levels are relative to its cognate wild type. Wild-type values were arbitrarily set to 1. b: Higher increase might be restricted by production of the protein in insoluble form at high expression levels. c: Strong band on western blot, while protein is not detectable when expressed from the wild-type system. d: This gene is known to be host toxic. e: Weak band on western blot, while protein is not detectable when expressed from the wild-type system.
References

1. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Artuk AP, et al. BglBricks: A flexible standard for biological part assembly. J Biol Eng 2010; 4:1-12; PMID:20205762; http://dx.doi.org/10.1186/1754-1611-4-11.

2. Sherry RD, Eady E, Knight TF Jr. Engineering RedBlue vectors from RedBlue parts. J Biol Eng 2004; 2:5-16; PMID:14818088; http://dx.doi.org/10.1186/1754-1611-2-5.

3. Bakke I, Berg L, Aune TE, Brænset T, Stein H, Tøndervik A, et al. Random mutagenesis of the EM promoter as a powerful strategy for improvement of recombinant-gene expression. Appl Environ Microbiol 2009; 75:2862-11; PMID:19501973; http://dx.doi.org/10.1128/AEM.02395-08.

4. Vee Aune TE, Bakke I, Darbøs F, Lale R, Brænset T, Valla S. Directed evolution of the transcription factor XylR for development of improved expression systems. Microb Biotechnol 2010; 3:50-67; PMID:21195188; http://dx.doi.org/10.1111/j.1751-7915.2009.00126.x.

5. Berg L, Lale R, Bakke I, Burroughs N, Valla S. The expression of recombinant genes in Escherichia coli can be strongly stimulated at the transcript production level by mutating the DNA region corresponding to the 5′ untranslated part of mRNA. Microb Biotechnol 2009; 2:378-99; PMID:19212492; http://dx.doi.org/10.1111/j.1751-7915.2009.00101.x.

6. Schal JP, Broun NE. An array secondary structure analysis of mRNA from rifampicin-induced strains. J Mol Biol 1990; 211:427-45; PMID:2146585; http://dx.doi.org/10.1016/0022-2836(90)90363-Q.

7. Lale R, Berg L, Stången F, Næss R, Stønnes M, Brænset T, et al. Continuous control of the flow in biochemical pathways through 5′ untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter Pm. Appl Environ Microbiol 2011; 77:7648-55; PMID:21935887; http://dx.doi.org/10.1128/AEM.02991-10.

8. Balzer S, Kachona Y, Mogenk J, Lale R, Brænset T, Valla S. A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. Microb Cell Fact 2013; 12:26-39; PMID:23506076; http://dx.doi.org/10.1186/1475-2859-12-26.

9. Bruny JM, Brænset T, Wüster-Leonard BC, Karunakaran P, Valla S. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid 1997; 38:35-51; PMID:9281494; http://dx.doi.org/10.1006/plas.1997.1294.

10. Slehta H, Nedal A, Aune TE, Hellebust H, Hakvåg S, Aune R, et al. Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in Escherichia coli. Appl Environ Microbiol 2004; 70:793-9; PMID:15574897; http://dx.doi.org/10.1128/AEM.70.12.793-799.2004.

References