Benchmarks

Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*

Bettina Rohweder^{1,*}, Florian Semmelmann^{1,*}, Christiane Endres¹, and Reinhard Sterner¹

Institute of Biophysics and Physical Biochemistry, University of Regensburg, Regensburg, Germany

*B.R. and F.S. contributed equally to this work.

BioTechniques 64:24–26 (January 2018) doi 10.2144/000114628

Keywords: *Bsal*; *Escherichia coli*; Golden Gate cloning; type II restriction enzymes; protein production; gene library

Supplementary material for this article is available at www.BioTechniques.com/article/114628.

Here, we modified the multiple cloning sites from commonly used expression vectors to create a new suite of cloning plasmids that simplify and speed up cloning procedures in *Escherichia coli*. Each of our standardized plasmids contains two *Bsal* restriction sites, allowing for highly efficient cloning of genes and bringing their expression under control of either a T7 (pET21a_*Bsal*, pET28a_*Bsal*, and pMAL-c5T_*Bsal*) or T5 promoter (pUR22 and pUR23). Another plasmid in our suite (pTNA_*Bsal*) allows for generation of large gene libraries containing $>10^8$ variants, which can be constitutively expressed in selection experiments using metabolic complementation of auxotrophic *E. coli* strains. Coupling restriction and ligation with the *Bsal* restriction enzyme minimizes hands-on time, while the need for only three different primers to clone a target gene into the six different vectors keeps overall cloning costs low.

The advancement of restriction enzymes as tools for the digestion of target genes and the invention of PCR for their specific amplification have simplified the construction of vectors for protein production (1) and the generation of large plasmid-based gene libraries for selection or screening

experiments (2) in *Escherichia coli*. Even though a variety of vectors for these tasks are now commercially available, there is still room for improvement when it comes to cost and efficiency. First, commonly used cloning vectors lack versatility and standardization: Most vectors differ in their restriction

sites; thus, re-cloning of a gene from one plasmid into another one before expression often requires additional PCR steps using different primer sets. Moreover, use of two restriction enzymes always necessitates digestion of both the gene of interest and the cloning vector followed by gel purification before ligation. Second, in the pET expression vectors, which are commonly used for protein production in *E. coli* (1), *NdeI* is usually the choice for the 5' cloning site, since all other restriction site sequences included in the multiple cloning site add additional amino acids between the N-terminal methionine residue and the remainder of the protein of interest. However, *NdeI* digestion results in only a 2-bp overhang, significantly lowering ligation efficiency and increasing the risk of generating false-positive clones through non-digested or re-ligated vector. Third, some expression vectors encode unwanted linkers between purification tags and the coding sequence, which could impair function, oligomerization, or crystallization of target proteins. An example is the 10-amino acid linker between the N-terminal hexa-histidine tag and the protein of interest in pET28a vectors, which is introduced upon cloning with the restriction enzymes *NdeI* and *XhoI*. These drawbacks often render cloning in commonly used expression vectors inefficient, tedious, and time consuming. This is especially disadvantageous in directed evolution experiments for which large gene libraries with $>10^8$ variants of the randomized gene of interest are sought because beneficial mutations can be very rare, and the chances of success increase with library size (3).

The Golden Gate Cloning method (4) offers an elegant solution to the ligation efficiency problem through the utilization of type II restriction enzymes such as *Bsal*. *Bsal* cuts outside its recognition sequence, leaving 4-bp overhangs that allow for efficient restriction enzyme digestion

METHOD SUMMARY

A set of six expression vectors that simplifies cloning procedures for both protein production and the generation of large gene libraries in *Escherichia coli* is described. These vectors contain standardized *Bsal* restriction sites for highly cost-efficient cloning.

pET28a_ *Bsa*I

T7 promoter lac operator
 AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT

His-Tag *Bsa*I *Bsa*I
 ATACCATGCATCATCATCATCATCTCGACATGAGAGACCGCGGATCCGAATTCGGGTCTCATTAAGTCGAGCACCACCACCACCACCCTGAGATCCGGC
 MetHisHisHisHisHisHisLeuAspMet — protein of interest — End

T7 terminator
 TGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTT

pMAL-c5T_ *Bsa*I

maltose binding protein thrombin
 TCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAATAACAATAACAACAACCCCGGGCTGGTGCCGCGCGCAGCCAATGA
 GlySerHisMet—

*Bsa*I *Bsa*I
 GAGACCGCGGATCCGAATTCGGGTCTCAATAGTCGAGTGATAAGCTTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTATGAAAGGCT
 — protein of interest — End

pET21a_ *Bsa*I

T7 promoter lac operator
 AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT

*Bsa*I *Bsa*I His-Tag
 ATACACATGAGAGACCGCGGATAATTCGGGTCTCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG
 Met — protein of interest — LeuGluHisHisHisHisHisHisEnd

T7 terminator
 CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTT

pUR22

T5 promoter
 AAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTACACAGAATTCATTAAAGAGGAGAAA

*Bsa*I *Bsa*I His-Tag
 TTAAGATGCGAGACCGAGGATCCTAAGGTCTCACTGAGCACCACCACCACCACCCTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGAC
 Met — protein of interest — LeuGluHisHisHisHisHisHisEnd

pUR23

T5 promoter
 AAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTACACAGAATTCATTAAAGAGGAGAAA

His-Tag *Bsa*I *Bsa*I
 TTAAGATGCATCATCATCATCATCTCGACATGCGAGACCGAGGATCCTAAGGTCTCATTAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCA
 MetHisHisHisHisHisHisLeuAspMet — protein of interest — End

pTNA_ *Bsa*I

E. coli tna operon operator *Bsa*I *Bsa*I
 TATCGACAATGAAAGCAATTTTCGTACTGTAAGGTTGAATTCATTAAAGAGGAGAAATTAAGATGCGAGACCGAGGATCCTAAGGTCTCATTAAGCTTAATT
 Met — protein of interest — End

Figure 1. The multiple cloning sites and promoter regions of the vectors pET21a_ *Bsa*I, pET28a_ *Bsa*I, pMAL-c5T_ *Bsa*I, pUR22, pUR23, and pTNA_ *Bsa*I. The standardized 4-bp overhangs after digestion with *Bsa*I are highlighted in orange. The resulting amino acid sequence is indicated in gray using the three-letter code.

Table 1. Overview of the protein production and gene library vectors constructed or modified in this work.

Vector name	Function	Tag	Promoter
pET21_ <i>Bsal</i>	protein production	C-terminal hexa-histidine	phage T7
pET28_ <i>Bsal</i>	protein production	N-terminal hexa-histidine	phage T7
pMAL-c5T_ <i>Bsal</i>	protein production	N-terminal MBP	phage T7
pUR22	protein production	C-terminal hexa-histidine	phage T5
pUR23	protein production	N-terminal hexa-histidine	phage T5
pTNA_ <i>Bsal</i>	gene library generation	—	<i>E. coli</i> tryptophanase operon

with simultaneous and seamless ligation, thereby speeding up the cloning process. We have generated a versatile set of six novel plasmids (Supplementary Material: Vector construction) that exploits the benefits of the Golden Gate method while simultaneously addressing the shortcomings of commonly used vectors: non-standardization of restriction sites and the addition of unnecessary amino acids to the protein coding sequence. We simplified the multiple cloning sites of 6 commercially available and commonly used vectors by introducing 2 *Bsal* restriction sites that leave distinct 4-bp 5' and 3' overhangs that are identical in all our plasmids (Figure 1). These standardized restriction sites allow for cloning of a gene of interest in *E. coli* for (i) protein production with different purification tags using only three different primers (Supplementary Material: Primer design and cloning protocol) or (ii) gene library generation (Supplementary Material: Construction of large gene libraries protocol).

The vectors pET21_*Bsal*, pET28_*Bsal*, and pMal-c5T_*Bsal* were constructed from pET21a (Novagen/Merck KGaA, Darmstadt, Germany), pET28a (Novagen/Merck KGaA), and pMAL-c5x (New England Biolabs, Ipswich, MA), respectively, allowing for protein production with C- or N-terminal hexa-histidine tags or maltose binding protein (MBP) fusion in *E. coli* strains containing a chromosomal copy of the T7 polymerase gene (Table 1). In pET28_*Bsal*, three superfluous amino acids between the N-terminal tag and the start codon were removed, along with the often unnecessary thrombin-cleavage site between the tag and the gene of interest. This strategy left only two amino acids for increased flexibility between the N-terminal tag and the protein, analogous to the link between the protein and the C-terminal tag in pET21a and pET21_*Bsal*. In pMal-c5T_*Bsal*,

the protein of interest is fused at the N terminus to the solubility-enhancer MBP, which can be cleaved off using thrombin. Vectors pUR22 and pUR23 were constructed from a shortened pQE70 vector (Qiagen, Venlo, The Netherlands) and the *lacI*^q fragment from pCA24N (5), allowing for protein production with the newly introduced N- and C-terminal hexa-histidine tags in *E. coli* strains containing no T7 polymerase gene (Table 1).

Vector pTNA_*Bsal* (Table 1) was constructed from pTNA (6,7) to generate large gene libraries using the *Bsal*-based Golden Gate cloning method (4) and standard laboratory equipment. Using pTNA_*Bsal*, we prepared four gene libraries from four different *E. coli* gene templates. These libraries consistently contained $>10^8$ gene variants (ranging 1.4×10^8 – 2.0×10^9 gene variants) with ligation efficiencies $>95\%$.

These libraries are larger by 2 to 3 orders of magnitude compared with libraries obtained with pTNA and a combination of the *Bam*HI and *Hind*III restriction sites (8), which significantly increases the chances of identifying rare, beneficial combinations of mutations. Use of the promoter region of the tryptophanase operon (P_{TNA}) leads to low constitutive expression in all *E. coli* strains, allowing for in vivo selection or screening under stringent conditions.

There are a number of benefits provided by our vector set. First, the use of the same set of primers and only a single restriction enzyme for multiple vectors significantly lowers the cost of a cloning experiment. Second, the high efficiency of Golden Gate cloning, where restriction digestion and ligation are combined in a one-pot reaction, eliminates the need for gel purification of the PCR reaction. Finally, a ligation efficiency $>95\%$ makes colony PCR before sequencing dispensable, again lowering the overall cloning time and cost.

Author contributions

F.S. and B.R. designed the experiments. F.S., B.R., and C.E. performed the experiments. F.S. and B.R. analyzed the data. F.S., B.R., and R.S. wrote the manuscript.

Acknowledgments

F.S. was supported by a doctoral fellowship from the Konrad-Adenauer-Stiftung. We thank Alexandra Holinski, Josef Sperl, and Andrea Kneuttinger for helpful suggestions.

Competing interests

The authors declare no competing interests.

References

1. Rosano, G.L. and E.A. Ceccarelli. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 5:172.
2. Bershtein, S. and D. Tawfik. 2008. Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol.* 12:151-158.
3. Jäkel, C., P. Kast, and D. Hilvert. 2008. Protein design by directed evolution. *Annu. Rev. Biophys.* 37:153-173.
4. Engler, C., R. Kandzia, and S. Marillonne. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3:e3647.
5. Kitagawa, M., T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, F. Inamoto, H. Toyonaga, and H. Mori. 2005. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 12:291-299.
6. Merz, A., M.C. Yee, H. Szadkowski, G. Pappenberger, A. Cramer, W.P. Stemmer, C. Yanofsky, and K. Kirschner. 2000. Improving the low catalytic activity of a thermophilic enzyme at low temperature. *Biochemistry* 39:880-889.
7. Beismann-Driemeyer, S. and R. Sterner. 2001. Imidazole glycerol phosphate synthase from *Thermotoga maritima*: Quaternary structure, steady-state kinetics and reaction mechanism of the bi-enzyme complex. *J. Biol. Chem.* 276:20387-20396.
8. Claren, J., C. Malisi, B. Höcker, and R. Sterner. 2009. Establishing wild-type levels of catalytic activity on natural and artificial ($\beta\alpha$) 8-barrel protein scaffolds. *Proc. Natl. Acad. Sci. USA* 106:3704-3709.

Received 26 September 2017; accepted 29 November 2017.

Address correspondence to Reinhard Sterner, Institute of Biophysics and Physical Biochemistry, University of Regensburg, D-93040 Regensburg, Germany. E-mail: Reinhard.Sterner@ur.de

To purchase reprints of this article, contact: biotechniques@fosterprinting.com