Published in final edited form as:
Methods Mol Biol. 2009 ; 498: 105–115. doi:10.1007/978-1-59745-196-3_7.

A Family of LIC Vectors for High-Throughput Cloning and Purification of Proteins[1]

William H. Eschenfeldt, **Lucy Stols**, **Cynthia Sanville Millard**, **Andrzej Joachimiak**, and **Mark I. Donnelly**

Summary

Fifteen related ligation-independent cloning vectors were constructed for high-throughput cloning and purification of proteins. The vectors encode a TEV protease site for removal of tags that facilitate protein purification (his-tag) or improve solubility (MBP, GST). Specialized vectors allow coexpression and copurification of interacting proteins, or *in vivo* removal of MBP by TVMV protease to improve screening and purification. All target genes and vectors are processed by the same protocols, which we describe here.

Keywords

Structural genomics; High throughput; Protein purification; Ligation-independent cloning; Coexpression; *In vivo* proteolysis; Maltose-binding protein; TEV protease; TVMV protease

7.1. Introduction

A family of compatible ligation-independent cloning (LIC) vectors (1) was created to enable effective high-throughput cloning and purification of recombinant proteins for structural studies. All the vectors contain a sequence encoding the tobacco etch virus (TEV) protease cleavage site (2) next to an *Ssp*I site used for LIC (Fig. 7.1). The base vector, pMCSG7 (3), appends an N-terminal hexahistidine tag to proteins that is followed by the protease recognition sequence. Derivatives of the base vector (Table 7.1) add maltose-binding protein (4,5), glutathione-*S*-transferase (6,7), or a loop of GroES (8–10) to the leader, replace the his-tag with the S-tag (7,11), incorporate a second protease cleavage site for *in vivo* tag removal (5, 12), or move the entire LIC region into different, compatible vectors to allow coexpression of proteins (5,7). In all cases, expression is driven by T7 polymerase under control of the lac promoter in specific host strains (13). To introduce genes into the vector, LIC-compatible extensions are added through the use of specific primers during amplification by PCR (1). Following appropriate processing, the PCR product can be introduced into any member of the family by a standard LIC protocol. This chapter describes the manual, nonhigh-throughput LIC of genes into the pMCSG vectors. **Chapter 8** describes plate-based methods for high-throughput applications.

The LIC process is identical for all members of the family (Fig. 7.2). The vector is first linearized by cleavage with *Ssp*I then treated with T4 polymerase in the presence of dGTP only. Exo-nuclease activity of the polymerase hydrolyzes nucleotides from the 3′ ends of the

vector until it reaches a G residue, creating a 15-base, single-stranded 5′ overhang. Conversely, treatment of appropriate PCR products with polymerase in the presence of only dCTP generates a complementary overhang. The two treated fragments are combined, allowed to anneal, and introduced into a suitable host. The host's native enzymes ligate then propagate the plasmid. In order to place nucleotides encoding a TEV cleavage site as close as possible to the introduced gene, restraints were placed on the design of the LIC region (3). The use of *Ssp*I at the center of the LIC region allows the resulting proteins to carry only three additional amino acids on their N termini after proteolysis – serine from the TEV site, asparagine from the *Ssp*I site, and alanine, the preferred choice for the limited number of amino acids possible due to the G used to stop the action of T4 polymerase in creating the overhang.

In addition to having the exact 16 base pairs at each end needed to create the complementary 15-base pair overhangs and stop the T4 polymerase, PCR primers must carry additional nucleotides (Fig. 7.2). For the sense primer, an additional two nucleotides are needed to maintain the proper reading frame and complete the codon begun with the required G. Because alanine is the most benevolent of the amino acids whose codons begin with G, the bases CC (or CA, CT, or CG) are added to give TACTTCCAATCCAATGCC. For the anti-sense primer, the required 16 nucleotides should be followed by the complement of a stop codon to prevent readthrough into vector sequences, giving TTATCCACTTCCAATGTTA. Following these sequences, primers terminate with nucleotides complementary to the gene being amplified, according to the requirements of the PCR conditions to be used. Primers may be designed manually or with software, such as the Express Primer tool (http://tools.bio.anl.gov/bioJAVA/jsp/ExpressPrimer-Tool/) (14).

Use of highly purified vectors is not essential to successful LIC, but improves efficiencies. However, complete cutting of the vectors with *Ssp*I is critical and problematic. Simple restriction digests with commercial, normal-strength *Ssp*I resulted in high backgrounds – clones containing unmodified vector free of an inserted gene (15). To circumvent this problem, vectors are cleaved with excess *Ssp*I using high concentration, LIC-qualified commercial enzyme, then purified on agarose gels to remove traces of uncut vector. When processed in this fashion, we routinely achieve efficiencies between 75 and 90%.

7.2. Materials

7.2.1. PCR Amplification of Genes and Processing for LIC

7.2.1.1. PCR

1. Sense primer: 5′-TACTTCCAATCCAATGCC– – – and Antisense primer: 5′-TTATCCACTTCCAATGTTA– – – obtained from oligonucleotide production service of choice. The dashes denote a series of nucleotides identical to those of the target gene (*see* **Section 7.1** and (14))
2. Platinum *Pf*x DNA polymerase (Invitrogen, kit cat. no. 11708-013)
3. Platinum *Pf*x 10× buffer (included in kit)
4. $MgSO_4$ (50 mM) (included in kit)

7.2.1.2. Purification

1. Qiagen spin column (QIAquick PCR Purification Kit #28104)
2. Qiagen buffers (included in kit)

7.2.1.3. T4 Treatment

1. dCTP (100 mM) (Promega cat. no. U1221)

2. Dithiothreitol (DTT, 100 mM), molecular biology grade (Sigma cat. no. D-9779)
3. T4 DNA polymerase, LIC-qualified (Novagen cat. no. 70099)
4. 10× T4 polymerase buffer (included with polymerase)

7.2.2. Preparation of Vector for LIC

7.2.2.1. Vector Prep

1. LB Broth, Miller or other rich medium
2. LB Agar, Miller
3. Sterile polystyrene Falcon tube (14 ml)
4. QIAGEN Plasmid Midi Kit (Qiagen, Inc. cat. no. 12143 or 12145)
5. Ampicillin
6. Chloramphenicol
7. Spectinomycin
8. Oakridge centrifuge tubes (Nalgene 3118-0050)

7.2.2.2. SspI Treatment

1. *Ssp*I, Genome-Qualified, High Concentration, ca. 50 U/μl (Promega R4604) and *Ssp*I buffer (included with enzyme)
2. 50× TAE Buffer concentrate for gel: 2 M Tris/Acetate, pH 8.0, 50 mM EDTA
3. Ultrapure agarose for gel
4. Ethidium bromide, 50 μg/ml
5. Tracking dye: 0.25% bromophenol blue plus 0.25% xylene cyanol FF in 30% glycerol
6. Gel electrophoresis apparatus – Biorad Mini Sub Cell or other comparable equip
7. DNA molecular weight markers, 1-kb ladder
8. QIAEX II gel extraction kit (Qiagen)

7.2.2.3. T4 Treatment

1. dGTP (100 mM)
2. Dithiothreitol (DTT, 100 mM), molecular biology grade
3. T4 DNA polymerase, LIC-qualified (Novagen cat. no. 70099)
4. 10× T4 polymerase buffer (included with enzyme)

7.2.3. LIC Annealing and Transformation

1. Sterile polystyrene Falcon tube (14 ml)
2. DH5α cells (Invitrogen Library Competent, cat. no. 18263-012), or equivalent
3. SOC (included with DH5α cells)

7.3. Methods

7.3.1. PCR and Preparation of PCR Product for LIC

7.3.1.1. PCR

1. In a thin-walled PCR tube combine 5 μl 10x reaction buffer and primers at 1 μM with approximately 1 ng template DNA in a total volume of 48.5 μl.
2. Add 0.5 μl polymerase (1.25 U) and 0.5–1.5 μl 50 mM $MgSO_4$ (as needed).
3. Perform PCR at appropriate temperatures for amplifying specific targets. Typical reactions comprise an initial 3-min denaturation at 94°C, amplification through 35 cycles of 30 s at 94°C, 45 s at 55–60°C, and 1 min at 68°C, with a final extension for10 min at 68°C (*see* Note 1).

7.3.1.2. Purification of PCR Product

1. Apply PCR to Qiagen Spin column from QIAquick PCR Purification Kit.
2. Follow the protocol detailed in the kit.
3. Determine concentration spectrophotometrically (e.g., NanoDrop Technologies' ND-1000 spectrophotometer) or by another suitable method.

7.3.1.3. T4 Polymerase Treatment of PCR Product

1. To a 0.4-ml Eppendorf tube on ice add

	Volume
PCR product 20 ng	x μl ($x + y = 32$ μl)
T4 polymerase 10× reaction buffer	4 μl
dCTP (100 mM)	1 μl
DTT (100 mM)	2 μl
Sterile water	y μl ($x + y = 32$ μl)
T4 DNA polymerase	1 μl
	Total 40 μl

2. Incubate the reaction mix at room T for 30 min
3. Inactivate the T4 DNA polymerase by heating at 75°C for 20min

7.3.2. Preparation of Vector for LIC

7.3.2.1. Isolation and Purification of Vector

1. Inoculate 2 ml of LB broth containing the appropriate anti-biotic in a 14-ml Falcon tube from a glycerol stock of *E. coli* DH5α containing the desired vector. Incubate for several hours at 37°C, agitating at 250 rpm. Antibiotic concentrations: ampicillin, 100 μg/ml; chloramphenicol, 30 μg/ml; spectinomycin, 50 μg/ml. (*see* Note 2).
2. Inoculate 50 ml of LB plus antibiotic in a sterile, notched Erlenmeyer flask with the entire 2-ml culture and incubate overnight at 37°C at 250 rpm.
3. Transfer the culture to two 50-ml Oakridge centrifuge tubes and centrifuge at 6,000 × *g* for 15 min at 4°C. Pour off supernatant fluid and invert the tubes to drain.

[1]Other proof reading DNA polymerases and PCR conditions may be used for amplification of genes. For high-throughput cloning into the MCSG vectors, PCR is routinely performed in 96-well plates using 1 U of KOD polymerase, 10 ng genomic target DNA, and primers at 0.2 μM in a 50-μl reaction.

[2]If cultures are started in the morning, they should be slightly turbid by afternoon.

4. Lyse cells and purify vector using QIAGEN Plasmid Midi Kit according to the instructions detailed in the QIAGEN Plasmid Purification Handbook that accompanies the product.

5. Dissolve the pellet resulting from the purification in 200–500 μl water or 10 mM Tris (pH 8.0). The DNA concentration can be estimated by running an aliquot on an agarose gel or spectrophotometrically with a Nanodrop instrument (*see* Note 3).

7.3.2.2. SspI Digestion

1. To a 1.5-ml Eppendorf tube on ice add

	Volume
Vector DNA 15 μg instrument	*x* μl (*x* + *y* = 52 μl)
*Ssp*I 10× reaction buffer	6 μl
Sterile water	*y* μl (*x* + *y* = 52 μl)
*Ssp*I	2 μl
	Total 60 μl

2. Incubate for 2 h at 37°C

3. Prepare a 50-ml, 0.8% TAE agarose gel containing ethidium bromide at 0.5 μg/ml in a 6.5 × 10-cm flat bed electrophoresis tray fitted with an 8-well comb

4. Add 10 μl of 6× tracking dye to the digestion mixture

5. Load 10–15 μl per well

6. Run the gel at 75 V for approximately 2.5 h or until the molecular weight ladder is well separated

7. Visualize plasmid with a UV light box and excise band containing cut vector and transfer to tared 1.5-ml sterile Eppendorf tubes (*see* Note 4)

8. Extract DNA from gel slices using a QIAEX II Gel Extraction kit following the instructions detailed in the product's manual (*see* Note 5)

7.3.2.3. T4 Polymerase Treatment

1. To a 0.4-ml Eppendorf tube on ice add

	Volume
Vector DNA 200 ng	*x* μl (*x* + *y* = 32 μl)
T4 polymerase 10× reaction buffer	4 μl
dGTP (100 mM)	1 μl
DTT (100 mM)	2 μl
Sterile water	*y* μl (*x* + *y* = 32 μl)
Novagen T4 DNA polymerase	1 μl
	Total 40 μl

2. Incubate the reaction mix at room T for 30′

3. Inactivate the T4 DNA polymerase by heating at 75°C for 20′ (*see* Note 6)

[3]The DNA pellet usually is quite clear and difficult to see. Incubating the water/buffer over the position where the pellet is expected for several minutes to allow time for it to dissolve can improve yields moderately. Absorbance at 260 nm tends to overestimate the DNA concentration.

[4]Compared to normal gel purifications, the gel is grossly overloaded and the DNA band often appears deformed. Run the gel out until the tracking dye approaches the end of the gel, so that the 5- to 6-kb vector band is well separated.

[5]The agarose gel fragments should be cut into small pieces before extraction and can be stored overnight at 4°C prior to extraction if necessary. The extracted DNA can be frozen at −20°C if necessary.

7.3.2.4. Large-Scale Preparation of Vector for Microtiter Plate Experiments—For stocks for several 96-well plates, all steps are scaled up 5- to 10-fold (*see* Note 7).

7.3.3. LIC Annealing and Transformation of Cells

1. In a 14-ml Falcon tube on ice mix

15 ng. vector DNA	(ca. 3 μl)
30–45 ng PCR product DNA	(ca. 2–3 μl)

7.3.3.1. LIC Annealing—Place the two aliquots together in a small droplet in the bottom of the tube in a total volume less than 8 μl.

2. Incubate for 30 min on ice

7.3.3.2. Transformation

1. To the tube, add 50 μl of Invitrogen Library Competent DH5α cells.
2. Incubate on ice for 30 min.
3. Heat shock 45 s at 42°C.
4. Chill on ice 2 min.
5. Add 0.45 ml SOC.
6. Incubate at 37°C, 250 rpm, 1 h.
7. Plate on LB agar containing the appropriate antibiotic(s). Antibiotic concentrations: ampicillin, 100 μg/ml; chloramphenicol, 30 μg/ml; spectinomycin, 50 μg/ml. Plate 100 μl of the culture, then centrifuge the remaining 400 μl gently in an Eppendorf tube to concentrate the cells. Decant the supernatant, resuspend the pellet in the small volume of medium remaining (ca. 50 μl), and plate on a second LB agar plate.

Acknowledgments

This work was supported by grants from the NIH (GM62414, GM074942), A. Joachimiak, PI, and the U.S. Department of Energy, Office of Biological and Environmental Research under Contract W-31-109-ENG-38.

References

1. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 1990;18:6069–74. [PubMed: 2235490]
2. Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. Anal Biochem 1994;216:413–7. [PubMed: 8179197]

[6]Can store frozen, but precipitation will occur. Be sure to bring the solution to room temperature after storage and wait for cloudiness to clear.

[7]To scale up preparation for cloning in microtiter plates (**Chapter 8**), start two 5-ml cultures of LB in the morning (inoculate heavily from glycerol stocks) and subculture into 500-ml medium for overnight incubation. Harvest cells in 250-ml centrifuge bottles and purify the vector using the QIAGEN Plasmid Maxi Kit, resuspending the final pellet in 0.5-ml water or buffer. This preparation typically yields about 500 μg of vector. For SspI digestion, incubate 100–200 μg vector with 200-U SspI for 2 h at 37°C in a volume of 600 μl. Add 100-μl tracking dye and purify the cut vector on a 150-ml agarose gel in a 11 × 14-cm tray fitted with a 14-tooth comb, loading 50 μl per well. Combine gel fragments and elute the extract the DNA using the QIAEX II Gel Extraction kit according to the instructions provided by the vendor.

3. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif 2002;25:8–15. [PubMed: 12071693]
4. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999;8:1668–74. [PubMed: 10452611]
5. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Expr Purif 2006;47:446–54. [PubMed: 16497515]
6. Nygren PA, Stahl S, Uhlen M. Engineering proteins to facilitate bioprocessing. Trends Biotechnol 1994;12:184–8. [PubMed: 7764901]
7. Stols L, Zhou M, Eschenfeldt WH, Millard CS, Abdullah J, Collart FR, Kim Y, Donnelly MI. New vectors for coexpression of proteins: structure of *Bacillus subtilis* ScoAB obtained by high-throughput protocols. Protein Expr Purif 2007;53:396–403. [PubMed: 17363272]
8. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 1997;388:741–50. [PubMed: 9285585]
9. Dieckman LJ, Zhang W, Rodi DJ, Donnelly MI, Collart FR. Bacterial expression strategies for human angio-genesis proteins. J Struct Funct Genomics 2006;7:23–30. [PubMed: 16688392]
10. Donnelly MI, Stevens PW, Stols L, Su SX, Tollaksen S, Giometti C, Joachimiak A. Expression of a highly toxic protein, Bax, in *Escherichia coli* by attachment of a leader peptide derived from the GroES cochaperone. Protein Expr Purif 2001;22:422–9. [PubMed: 11483004]
11. Raines RT, McCormick M, Van Oosbree TR, Mierendorf RC. The S. Tag fusion system for protein purification. Methods Enzymol 2000;326:362–76. [PubMed: 11036653]
12. Nallamsetty S, Kapust RB, Tozser J, Cherry S, Tropea JE, Copeland TD, Waugh DS. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. Protein Expr Purif 2004;38:108–15. [PubMed: 15477088]
13. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 1990;185:60–89. [PubMed: 2199796]
14. Yoon HY, Hwang DC, Choi KY, Song BD. Proteolytic processing of oligopeptides containing the target sequences by the recombinant tobacco vein mottling virus NIa proteinase. Mol Cells 2000;10:213–9. [PubMed: 10850664]
15. Dieckman L, Gu M, Stols L, Donnelly MI, Collart FR. High throughput methods for gene cloning and expression. Protein Expr Purif 2002;25:1–7. [PubMed: 12071692]

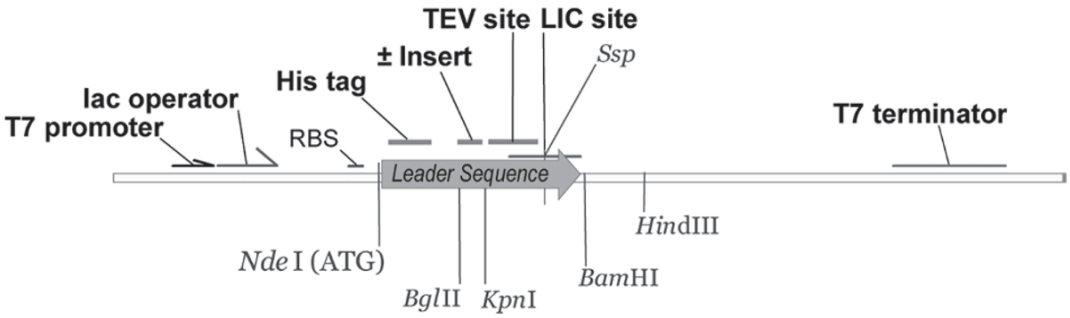

Fig. 7.1.

Generalized organization of MCSG vectors. MCSG vectors encode an N-terminal leader sequence (*arrow*) that terminates in an LIC region centered on an *Ssp*I site. Restriction sites in and around the coding/cloning region allow insertion of protein or peptide modules into the leader (at *Bgl*II and/or *KpnI*), replacement of the his-tag (*NdeI* to *Bgl*II or *Kpn*I), or transfer of the entire region to different backgrounds (*Nde*I to *Bam*HI, *Hin*DIII or *Xho*I). Derivatives (**Table 7.1**) add MBP, GST, or a loop of GroES (Sloop), replace the His-tag with the S-tag, and move these expression regions to pACY-CDuet-1 and pCDFDuet-1, allowing cotransformation with two or three vectors and coexpression of multiple proteins. Another modification, pMCSG19, positions untagged MBP followed by the TVMV protease recognition sequence (12) ahead of the His-tag, allowing *in vivo* removal of MBP by coexpressed TVMV protease.

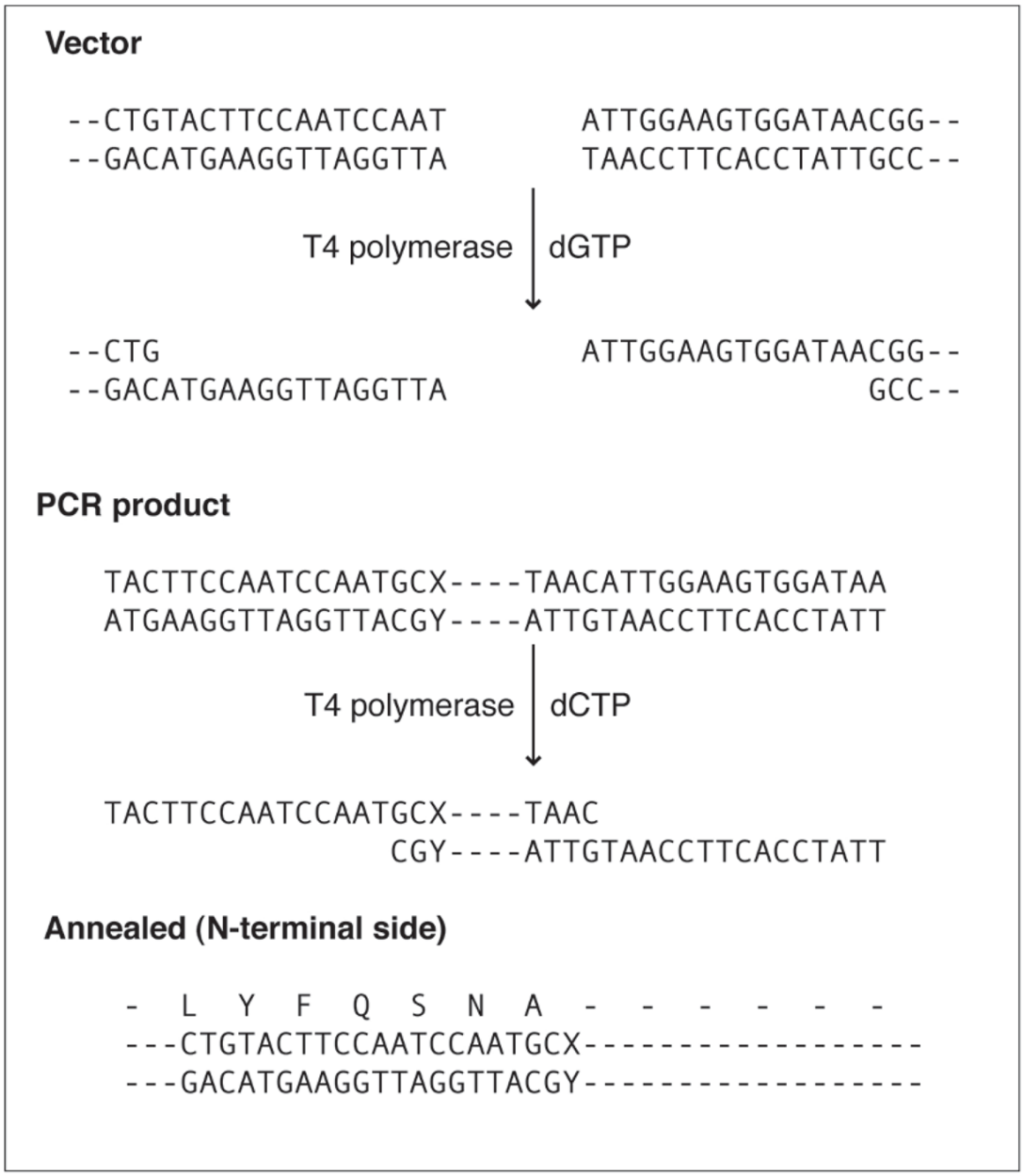

Fig. 7.2.

LIC procedure using pMCSG vectors. All MCSG vectors contain an *Ssp*I site (AATATT) positioned immediately after the sequence encoding the TEV protease recognition site. Cleavage with *Ssp*I (a blunt cutter) followed by treatment with T4 DNA polymerase in the presence of only dGTP generates 15-base pair overhangs on both ends of the vector. PCR products must be generated using primers that begin with the complements of these overhangs followed by nucleotides required for LIC processing and proper expression. The sense primer must begin, TACTTCCAATCCAATGCC– – –, where *dashes* indicate nucleotides identical to the target gene, and requires the nucleotides GCC to stop the action of the polymerase and encode alanine in the correct reading frame. The antisense primer, TTATCCACTTCCAATGTTA– – –, requires a G, complement of the C that stops the endonuclease activity of T4 polymerase, and TTA, the complement of a stop codon. Annealing

to treated vector positions the TEV protease site adjacent to the gene in the correct reading frame. The resulting protein has the residues SNA appended to its N terminus after TEV cleavage. Primers conforming to these restrictions can be designed manually or using commercial programs or online tools such as the Express Primer Tool (14), http://tools.bio.anl.gov/bioJAVA/jsp/ExpressPrimerTool/.

Table 7.1

MSCG vectors for high-throughput ligation-independent cloning

Vector	Parental vector	Antibiotic	Leader sequence	MW (leader)[a]	kb	Purpose
pMCSG7	pET-21a[b]	Amp	N-His-TEV-LIC	2,755	5,286	production
pMCSG8	pMCSG7	Amp	N-His-Sloop-TEV-LIC	4,399	5,341	toxicity
pMCSG9	pMCSG7	Amp	N-His-MBP-TEV-LIC	43,713	6,147	solubility
pMCSG10	pMCSG7	Amp	N-His-GST-TEV-LIC	29,046	5,961	solubility
pMCSG11	pACYC-Duet-1[c]	Cam	N-His-TEV-LIC	2,755	4,079	coexpression
pMCSG12	pACYC-Duet-1	Cam	N-His-Sloop-TEV-LIC	4,399	4,144	coexpression
pMCSG13	pACYC-Duet-1	Cam	N-His-MBP-TEV-LIC	43,713	4,940	coexpression
pMCSG14	pACYC-Duet-1	Cam	N-His-GST-TEV-LIC	29,046	4,754	coexpression
pMCSG17	pMCSG7	Amp	N-Stag-TEV-LIC	3,760	5,316	coexpression
pMCSG19	pMCSG7	Amp	N-MBP-TVMV-His-TEV-LIC	45,050/2,711[d]	6,441	production
pMCSG20	pMCSG17	Amp	N-Stag-GST-TEV-LIC	30,051	5,991	coexpression
pMCSG21	pCDFDuet-1[c]	Spec	N-His-TEV-LIC	2,755	3,852	coexpression
pMCSG22	pCDF-Duet-1	Spec	N-His-Sloop-TEV-LIC	4,399	3,906	coexpression
pMCSG23	pCDF-Duet-1	Spec	N-His-MBP-TEV-LIC	43,713	4,971	coexpression
pMCSG24	pCDF-Duet-1	Spec	N-His-GST-TEV-LIC	29,046	4,527	coexpression

All 15 vectors are processed for LIC the same way and accept the same properly prepared PCR products.

Abbreviations: Amp ampicillin; *Cam* chloramphenicol; *Spec* spectinomycin; *His* hexahistidine tag; *TEV* tobacco etch virus protease recognition sequence; *LIC* ligation-independent cloning site centered on an *Ssp*I site; *Sloop* GroEL-binding loop of GroES; *MBP* maltose-binding protein; *GST* glutathione-*S*-transferase; *Stag* S-tag fragment of ribonuclease; *TVMV* tobacco vein mottling virus protease recognition sequence

[a] Molecular weight removed by cleavage with TEV protease

[b] Vector pET-21a is a product of Novagen, Inc. (Madison, WI)

[c] Vectors pACYCDuet-1 and pCDFDuet-1 are products of Novagen, Inc. (Madison, WI)

[d] First value is molecular weight removed by cleavage by TEV protease without prior treatment with TVMV protease. Second value is that removed after prior treatment with TVMV protease. TVMV pro-tease removes 42,339 Da comprising untagged MBP and flanking amino acids