

# A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site

Lucy Stols,\* Minyi Gu,† Lynda Dieckman,† Rosemarie Raffen,† Frank R. Collart,† and Mark I. Donnelly\*,<sup>1</sup>

*\*Environmental Research Division and †Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439*

Received October 8, 2001

To establish high-throughput methods for protein crystallography, all aspects of the production and analysis of protein crystals must be accelerated. Automated, plate-based methods for cloning, expression, and evaluation of target proteins will help researchers investigate the vast numbers of proteins now available from sequenced genomes. Ligation-independent cloning (LIC) is well suited to robotic cloning and expression, but few LIC vectors are available commercially. We have developed a new LIC vector, pMCSG7, that incorporates the tobacco etch virus (TEV) protease cleavage site into the leader sequence. This protease is highly specific and functions under a wide range of conditions. The new vector incorporates an N-terminal his-tag followed by the TEV protease recognition site and a *SspI* restriction site used for LIC. The vector functioned as expected, giving high cloning efficiencies and strong expression of proteins. Purification and cleavage of a target protein showed that the his-tag and the TEV cleavage site function properly. The protein was purified and cleaved under different conditions to simulate both plate-based screening methods and large-scale purifications for crystal production. The vector also includes a pair of adjacent, unique restriction sites that will allow insertion of additional modules between the his-tag and the cleavage site of the leader sequence to generate a family of vectors suitable for high-throughput production of proteins.

*Key Words:* high throughput; structural genomics;

ligation-independent cloning; TEV protease; affinity purification.

The emerging field of structural genomics embraces many diverse objectives, but in general depends on the generation of protein structural information at a much faster rate than occurs currently (1–3). To achieve this objective, every step involved in determining a protein's structure must be accelerated, including the cloning, expression, and purification of target proteins. Automation of labor-intensive steps, when possible, can dramatically increase throughput and reduce costs (4). The Midwest Center for Structural Genomics<sup>2</sup> (MCSG, [www.mcsg.anl.gov](http://www.mcsg.anl.gov)) has instituted robotic, microtiter plate-based protocols for high-throughput cloning of target proteins and analysis of their expression and solubility. The protocols incorporate ligation-independent cloning (LIC) of target genes (5, 6). LIC possesses several advantages for automated cloning. It eliminates the use of restriction endonuclease digestion and ligation of PCR products, allowing any gene to be cloned into the vector regardless of its sequence. In LIC, PCR primers are designed to append sequences that, after treatment with T4 DNA polymerase in the presence of a single deoxyribonucleotide triphosphate, generate 12- to 15-base-pair overhangs that are complementary to overhangs generated in the vector. These overhangs anneal sufficiently strongly to allow the transformation

<sup>1</sup> To whom correspondence should be addressed. Fax: (630) 252-7709. E-mail: [mdonnelly@anl.gov](mailto:mdonnelly@anl.gov).

<sup>2</sup> Abbreviations used: MCSG, Midwest Center for Structural Genomics; LIC, ligation-independent cloning; TEV, tobacco etch virus; DTT, dithiothreitol; LB, Luria–Bertani medium; IPTG, isopropyl  $\beta$ -galactose.

of hosts without ligation of the fragments; host repair enzymes ligate the introduced plasmid. Thus, LIC allows the consistent design of PCR primers. LIC also generates very high cloning efficiencies, increasing greatly the probability of successfully cloning target genes without labor-intensive screening. All procedures from the initial PCR through the analysis of expression can be automated and carried out in microtiter plates.

A serious bottleneck in high-throughput protein structure determination is the purification of expressed proteins. Typically, this effort is accelerated by attachment of affinity tags to the N- or C-terminus of expressed proteins (7). Some tags also improve the solubility of expressed proteins (8) or reduce their toxicity (9). Ideally, these tags are cleaved from the purified protein by specific proteases at recognition sites incorporated into the tags. Commercially available LIC vectors currently offer limited options with respect to affinity tags and cleavage sites. The pET-based LIC vectors from Novagen are dual tagged (polyhistidine plus S-peptide) and require cleavage with either enterokinase or factor Xa to remove both tags. A thrombin site allows removal of the his-tag only. Stratagene provides a LIC vector that combines the calmodulin binding peptide tag and an enterokinase cleavage site. Other options would be desirable.

The tobacco etch virus (TEV) protease is a highly specific protease that offers several advantages over the mammalian proteases currently used in LIC vectors. It is highly specific, cleaving a seven-amino-acid recognition sequence, and is active under a wide range of conditions including low temperature and high ionic strength (10–12). It is also insensitive or only mildly sensitive to many of the protease inhibitors commonly used to prevent spurious degradation of expressed proteins by host proteases (13). Its optimal recognition sequence is ENLYFQ followed by either S or G. Cleavage occurs after the Q residue. Some residues may be altered, but changes in the positions occupied by E, Y, or Q drastically reduce the efficiency of cleavage.

We have modified the commercial vector pET-30 Xa/LIC to generate a LIC vector that incorporates the TEV protease site. The modified vector encodes a leader sequence that consists of a polyhistidine affinity tag followed by an eight-amino-acid spacer and the TEV protease recognition sequence. An *SspI* restriction site downstream from the TEV recognition sequence allows LIC manipulations of the vector. In the path to the preparation of this vector, we also created a variant of pET-30 Xa/LIC lacking the S-peptide tag but retaining the factor Xa site. The leader sequence encoding regions of both vectors were transferred into the closely related vector pET-21a to give ampicillin-resistant vectors. Experiments with the new LIC-TEV vector showed that it functioned effectively in all aspects—cloning, expression, purification, and cleavage.

## MATERIALS AND METHODS

### *Construction of pMCSG3*

Vectors pMCSG3 and pMCSG7 were constructed by replacing portions of the leader sequence-encoding region of pET-30 Xa/LIC with nucleotide linkers that encoded the new leader sequences (Fig. 1). For pMCSG3, the entire region encoding the his-tag, thrombin site, and S-tag of pET-30 Xa/LIC was excised by treatment with *NdeI* and *BgIII*. The vector was dephosphorylated with calf intestinal phosphatase and purified by electrophoresis in an agarose gel and extracted using the Qi-aEx gel extraction kit (Qiagen). The synthetic oligonucleotides TATGCACCATCATCATCATCTTCTGTGTA and GATCTACACCAGAAGAATGATGATGATGATGGTGCA, encoding a six-histidine his-tag, a small spacer sequence, and the appropriate overhangs, were hybridized and phosphorylated with T4 DNA kinase. The phosphorylated linker was ligated into the modified pET-30 Xa/LIC with T4 DNA ligase and transformed into library competent DH5 $\alpha$  cells (Gibco). Clones were screened by restriction endonucleases for the presence of the shorter leader sequence-encoding region. To create an ampicillin-resistant vector, the region between *SphI* and *XhoI*, which contains the leader sequence and is nearly identical in pET-30 Xa/LIC and pET-21a, was transferred into pET-21a using standard methods. The resulting vector was named pMCSG3.

### *Construction of pMCSG7*

**Elimination of *SspI* sites.** The vector pET-21a contains two *SspI* sites, one located in the f1 replication origin and the other between the f1 origin and the ampicillin resistance-encoding gene. These sites were eliminated by cutting pMCSG3 with *SspI* and *DraIII* (located within the f1 origin), treating the linearized vector with T4 polymerase in the absence of dNTP's to generate a blunt end from the *DraIII* cleavage overhang, and circularizing the vector with T4 DNA ligase. The resulting vector was cloned and screened for resistance to ampicillin and the absence of *SspI* and *DraIII* sites. To insure that the modification had not damaged the vector's ability to express proteins efficiently, the gene encoding APC236 was transferred from pET-30 Xa/LIC into the modified vector and induced. Comparably strong expression occurred in both vectors.

**Introduction of the TEV protease recognition sequence and the *SspI* LIC site.** The region of the resulting vector between *KpnI* and *BamHI*, which encodes the factor Xa recognition site and the *BseRI* LIC site of pET-30 Xa/LIC, was excised and the vector was dephosphorylated and purified by agarose gel electrophoresis as described above. The synthetic oligonucleotides CGA-GAACCTGTA CTTCCAATCCAATATTGGAAGTGGATAACG and GATCCGTTATCCACTTCCAATATTGGA-TTGAAGTACAGGTTCTCGGTAC, encoding the TEV

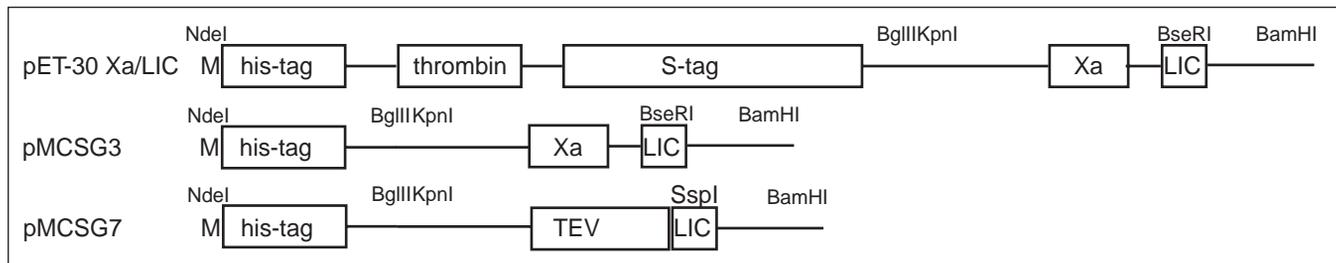


FIG. 1. Schematic illustration of the leader sequence encoding regions of LIC vectors. The commercial vector pET-30 Xa/LIC was modified by replacing the sequence between the *NdeI* and *BglII* sites to generate pMCSG3. This vector was subsequently modified by replacement of the sequence between the *KpnI* and *BamHI* sites to give pMCSG7. Functional modules, indicated by boxes, are his-tag, six histidine purification tag; thrombin, thrombin recognition site; S-tag, S-peptide detection tag; Xa, factor Xa recognition site; TEV, TEV protease recognition site; LIC, ligation-independent cloning restriction site.

recognition sequence, an *SspI* restriction site, and the appropriate overhangs, were hybridized, phosphorylated, ligated into the linearized pMCSG3, and transformed into DH5 $\alpha$  cells as described above.

**Analysis.** The final vector, pMCSG7, was sequenced with a T7 terminator primer (GCTAGTTATTGCT-CAGCGG) using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit and an ABI Prism 377 DNA Sequencer (PE Applied Biosystems, Foster City, CA).

### LIC Processing and Cloning Efficiencies

**Processing of pMCSG3, pMCSG7, and PCR products.** Preparation of pMCSG3 was the same as that described for the commercial vector pET-30 Xa/LIC (Novagen). Briefly, the vector was hydrolyzed with *BseRI* and treated with T4 DNA polymerase in the presence of only dCTP to generate the LIC overhang. LIC quality vectors (low background of clones lacking inserted genes) were prepared by two cycles of digestion with a five-fold excess of *BseRI* (5 units/ $\mu$ g DNA) for 1 h followed by ethanol precipitation. The linearized vectors were treated with T4 DNA polymerase (2 unit/ $\mu$ g DNA) in a reaction buffer containing 0.33 M Tris acetate, pH 8.0, 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM dithiothreitol (DTT), and 2.5 mM dCTP. The final vector stock was diluted to 50  $\mu$ g/ml and stored at  $-70$  or  $-180^{\circ}\text{C}$ . The LIC extensions on PCR products targeted for introduction into pET-30 Xa/LIC were compatible with pMCSG3. Treatment of these PCR products with T4 DNA polymerase (as described above) in the presence of 2.5 mM dGTP generated the complementary LIC overhangs.

Because of the DNA sequence required by the TEV protease site, the preparation of pMCSG7 was different. This vector was first cut with *SspI* and then treated with T4 DNA polymerase in the presence of only dGTP (as opposed to dCTP). To attain low backgrounds in LIC cloning, pMCSG7 was either digested with a high concentration of *SspI*, 20 units of enzyme/ $\mu$ g DNA, for

3 h (60-fold greater than the standard conditions of 1 unit/ $\mu$ g for 1 h) followed by ethanol precipitation, or purified by electrophoresis in agarose gels after linearization with 10 units of enzyme/ $\mu$ g DNA for 4 h (40-fold greater than standard conditions).

Because the LIC overhang of pMCSG7 is different, different cloning handles were incorporated into the design of the PCR primers. The sense primers began with the sequence TACTTCCAATCCAATGCX followed by the nucleotides encoding the target protein. Anti-sense primers began with the sequence TTATCCAATTCCAATG followed by the complement of a stop codon and the C-terminus of the target protein. Products were then treated with T4 DNA polymerase in the presence of dCTP to generate complementary overhangs.

The cloning efficiencies of the vector were evaluated by determining the frequency of expression of target proteins in individual clones generated by LIC. For these experiments we used a set of four target proteins of various size that spanned a range of cloning efficiency when introduced into commercially available LIC vectors (15). Good correspondence was observed between cloning efficiencies measured using these control fragments and results obtained with plates of many diverse targets proteins. Details of protocols for robotic cloning in 96-well microtiter plates are described elsewhere (15). Cultures were grown in 2 ml of Luria-Bertani (LB)/ampicillin medium at  $37^{\circ}\text{C}$  to an  $\text{OD}_{600} = 0.5$  and then induced with 1 mM isopropyl  $\beta$ -galactoside (IPTG). Cells were harvested after 3 h and lysed with a mixture of 2% SDS, 50 mM Tris (pH 6.8), 100 mM DTT, 10% glycerol, and 0.1% bromphenol blue dye. Expression was evaluated by denaturing polyacrylamide gel electrophoresis.

### Expression, Purification, and Cleavage of Proteins Cloned in pMCSG7

The functionality of the leader sequence was evaluated by purifying and cleaving one of the target proteins



A

--CTGTACTTCCAATCCAAT      ATTGGAAGTGGATAACGG--  
 --GACATGAAGGTTAGGTTA      TAACCTTCACCTATTGCC--

--CTG      ATTGGAAGTGGATAACGG--  
 --GACATGAAGGTTAGGTTA      GCC--

B

TACTTCCAATCCAATGC---CATTGGAAGTGGATAA  
 ATGAAGGTTAGGTTACG---GTAACCTTCACCTATT

TACTTCCAATCCAATGC---C  
 CG---GTAACCTTCACCTATT

C

- L Y F Q S N A - - - - -  
 ---CTGTACTTCCAATCCAATGCX-----  
 ---GACATGAAGGTTAGGTTACGX-----

FIG. 3. LIC processing of pMCSG7 and PCR products. (A) Cleavage of pMCSG7 with *SspI* followed by treatment with T4 polymerase in the presence of dGTP generates 15-base LIC overhangs. (B) Treatment of PCR products with complementary termini (incorporated into the primers used to amplify target genes) with T4 polymerase in the presence of dCTP generates complementary LIC overhangs. (C) Annealing of the overhangs generates a product encoding a leader sequence with the TEV protease recognition site (residues LYFQS of the seven-residue recognition site are shown) followed by an asparagine residue (N) derived from the *SspI* site and an alanine residue (A) determined by the PCR product. Subsequent amino acids are the target protein. Dashes indicate continuation of the nucleotide or amino acid sequences. Cleavage with TEV protease occurs between residues Q and S to generate a protein with the N-terminal sequence SNA appended to the target protein.

glutamine and serine residues of the recognition site and will generate a protein starting with ser-asn followed by amino acids determined by the PCR product.

Plasmid pMCSG3, a precursor of pMCSG7, was sequenced and verified to encode the same his-tag/spacer sequence, followed by the unaltered factor Xa cleavage site of pET-30 Xa/LIC (data not shown). Both vectors include the *BglII* and *KpnI* sites originally present in Figs. 1 and 2 which can allow insertion of additional modules into the leader sequence.

### Cloning and Expression

Primers were designed to amplify four target proteins (Table 1) and append the appropriate sequences to the ends of the PCR products to allow their introduction by LIC into pET-30 Xa/LIC, pMCSG3, and pMCSG7. The cloning efficiency of each vector, defined as the

TABLE 1

Genes Expressed in LIC Vectors

Protein <sup>a</sup>	Source	Identity
APC236	<i>B. subtilis</i>	iolI protein
APC327	<i>H. influenzae</i>	Ribosomal large subunit pseudouridine synthase D
APC331	<i>H. influenzae</i>	nifS protein homolog
APC385	<i>H. influenzae</i>	Hypothetical protein HI0033

<sup>a</sup> Additional information on APC target proteins can be obtained at the Midwest Center for Structural Genomics web site, www.mcsg.anl.gov.

percentage of colonies arising from LIC that expressed the target protein, was evaluated. For the two *BseRI*-based vectors, pET-30 Xa/LIC and pMCSG3, cleavage followed by ethanol precipitation prior to T4 DNA polymerase treatment gave efficiencies of 90 and 75% respectively (Table 2). For pMCSG7, which uses *SspI* as the LIC restriction site, poor efficiency was obtained by this procedure, and extensive digestion with high levels of *SspI* or purification by agarose gel electrophoresis was required to reduce the background sufficiently. The final efficiencies for all three vectors were at least 75% sufficient to give over 90% accuracy in screening expression of target proteins using a pair of replicate microtiter plates. Consistent expression efficiencies for each vector have been observed for different vector lots (results not shown). As expected for vectors based on T7 polymerase, pMCSG7 generated very strong overexpression of cloned proteins (Fig. 4). Similar strong overexpression of these proteins was obtained with pET-30 Xa/LIC and pMCSG3 (data not shown).

### Functionality of the Leader Sequence of pMCSG7

The functionality of the his-tag and TEV cleavage site of pMCSG7 was demonstrated by purifying one of

TABLE 2

Expression Efficiencies in the LIC Vectors

Vector	Processing <sup>a</sup>	Efficiency <sup>b</sup>
pET-30 Xa/LIC <sup>c</sup>	5-fold digestion + precipitation	0.90
pMCSG3	5-fold digestion + precipitation	0.75
pMCSG7	17-fold digestion + precipitation	0.17
pMCSG7	60-fold digestion + precipitation	0.75
pMCSG7	40-fold digestion + gel purification	0.75

<sup>a</sup> Vectors pET-30 Xa/LIC and pMCSG3 were treated by a 5-fold excess digestion with *BseRI* (fold refers to the product of units/ $\mu$ g DNA times hours, relative to a standard digestion with 1 unit/ $\mu$ g for 1 h) and purified by ethanol precipitation prior to treating with T4 DNA polymerase to create LIC overhangs. Vector pMCSG7 was treated by various fold excess digestions with *SspI* and purified by ethanol precipitation or agarose gel electrophoresis.

<sup>b</sup> Efficiency equals the fraction of colonies that expressed the target protein.

<sup>c</sup> Commercial LIC vector from Novagen.

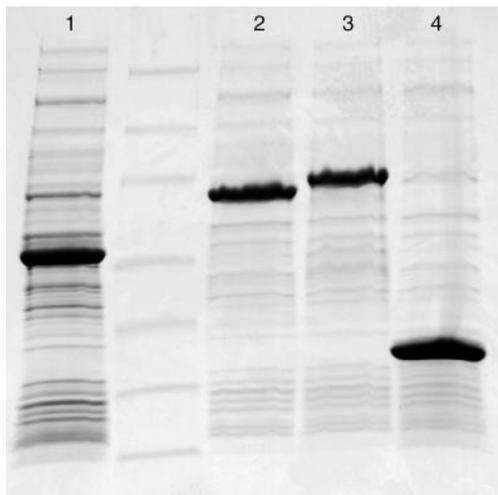


FIG. 4. Expression of target proteins cloned into pMCSG7. Cultures of colonies generated by LIC into pMCSG7 were grown at 37°C, induced with 0.5 mM IPTG at OD<sub>600</sub> = 0.5, harvested after 3 h, and lysed. Lane 1, APC236; lane 2, APC327; lane 3, APC331; lane 4, APC385. Identities of the proteins are given in Table 1. Prestained molecular weight markers (New England Biolabs) run with apparent masses of 175, 83, 62, 47.5, 32.5, 25, and 16.5 kDa.

the proteins, APC331. The protein was processed under two contrasting protocols, one to assess the potential to extend the robotic microtiter plate analyses to include binding, release, and cleavage of target proteins and the other to mimic large-scale purification for crystallization. In the first case, APC331 was purified using Pierce's Swell-Gel nickel IMAC disks in a process that replicated in a microfuge tube the manufacturer's recommended steps for purification in microtiter plates. Highly purified protein was easily obtained in less than 1 h. Dialysis of the purified protein followed by treatment with commercial TEV protease (Gibco) at room temperature resulted in nearly complete hydrolysis in 4 h (Fig. 5).

For evaluation of conditions used for large-scale processing of his-tagged proteins, APC331 was purified on a Ni-NTA column (Qiagen) and eluted stepwise with imidazole. The nearly homogeneous protein was

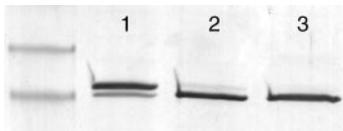


FIG. 5. Cleavage of APC331 purified with Swell-Gel disks. Lysates prepared from 2 ml of induced cells were incubated with Swell-Gel matrix removed from commercial 96-well microtiter plates (Pierce). The protein was eluted with 250 mM imidazole, dialyzed into 50 mM Tris, pH 8.0, containing 5 mM EDTA, and 1 mM DTT and treated with 10 units of commercial TEV protease (Gibco) at room temperature for: lane 1, 6 min; lane 2, 1 h; and lane 3, 4 h. Prestained molecular weight markers (New England Biolabs) shown run with apparent masses of 62 and 47.5 kDa.

cleaved under two conditions. In the first case, purified protein was cleaved with commercial TEV protease at 4°C directly in the elution buffer, which contained 250 mM imidazole and 300 mM NaCl (Fig. 6A). Cleavage was nearly complete in 18 h, reflecting the ability of TEV protease to function under a wide range of conditions. In the second test, his-tagged TEV protease that had been expressed and purified from the vector pRK508 (8) was added to the purified protein and the mixture was dialyzed overnight into optimal proteolysis buffer free of imidazole and NaCl. The his-tagged protease (noncleavable) was removed by batchwise incubation with Ni-NTA agarose (Fig. 6B).

## DISCUSSION

Successful automation of LIC-based cloning promises to provide a steady stream of proteins for crystallographic studies, but could lead to a serious bottleneck in the purification of the proteins. Affinity tags associated with proteins help greatly in automating their purification, but ideally one should be able to remove these tags quickly under mild conditions with no hydrolysis of the target protein. Commercially available LIC vectors incorporate cleavage sites for factor Xa, enterokinase, and thrombin. While reasonably effective, none of these proteases is ideal, often requiring extended incubation at 30–37°C to obtain complete hydrolysis of the tag. Some degradation of the target protein frequently occurs, rendering the previously pure protein impure. The capsid protease of the tobacco etch virus (TEV protease)

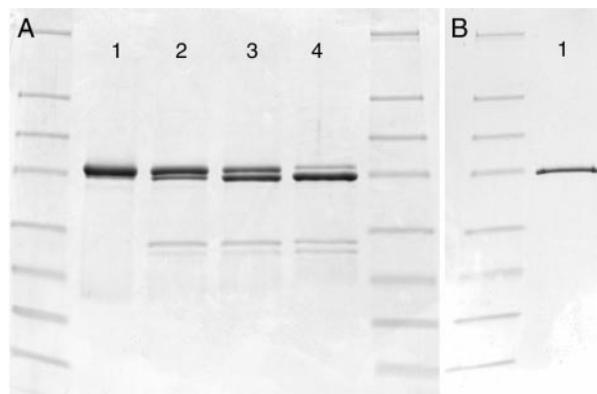


FIG. 6. Cleavage of APC331 purified by column chromatography. (A) APC331 was purified on Ni-NTA agarose (Qiagen), eluted with buffer containing 250 mM imidazole, and treated with purified his-tagged TEV protease at 4°C. Lane 1, untreated; lane 2, 1 h; lane 3, 4 h; lane 4, 18 h. The bands around 28 kDa in lanes 2 through 4 are the TEV protease. (B) TEV protease was added to the same sample and the mixture was dialyzed overnight into phosphate buffer free of imidazole. Batch treatment of the dialysate with Ni-NTA agarose removed the his-tagged protease. Prestained molecular weight markers (New England Biolabs) run with apparent masses of 175, 83, 62, 47.5, 32.5, 25, 16.5, and 6.5 kDa.

is highly specific because of its seven-amino-acid recognition sequence (12). Degradation of the target protein occurs only rarely. The enzyme is also active under a wide range of conditions, allowing efficient cleavage of appended tags even at 4°C without dialysis into an optimal buffer (14). Currently, no LIC-based vector incorporates the TEV protease site. This report describes the construction and use of such a vector.

The mechanism of preparation of LIC vectors and compatible PCR product inserts for cloning imposes considerable limitations on vector design. Both the linearized vector and the PCR product containing the gene to be cloned are treated with T4 polymerase in the presence of a single dNTP to generate complimentary overhangs. The 3'-endonuclease activity of the polymerase removes bases from the 3'-end of each duplex DNA until it encounters the nucleotide matching the single added dNTP. At that point, polymerase activity dominates, stopping the endonuclease reaction and leaving a 5' single-stranded overhang. To create a 12- to 15-base-pair overhang in the linearized vector, the 3'-sequence of both ends must lack one of the four bases, for example, G, as occurs in pMCSG7. In the presence of dGTP (only) the 3'-5'-exonuclease activity of T4 polymerase hydrolyzes 3'-bases until it encounters a G, thereby generating the desired overhang (Fig. 3A). To be compatible, PCR products must have complementary termini so that treatment with polymerase and the complementary dNTP will generate compatible overhangs (Fig. 3B). Because of these restraints, only a limited number of LIC vectors are available commercially.

Additional factors dictated the design of the TEV/LIC vector. To place the protease cleavage site adjacent to the target protein, its sequence must be included in the N-terminal overhang (Figs. 2 and 3). The optimal TEV recognition sequence is ENLYFQ followed by either S or G, with cleavage occurring after the Q residue (11). The E, Y, and Q residues are required for efficient hydrolysis. Because glutamine (Q) is encoded only by codons beginning CA and tyrosine (Y) only by codons beginning TA, the only option for the nucleotide missing from the overhang is G. Consequences of this restriction were that (1) T4 polymerase processing of linearized vector must be done in the presence of dGTP, (2) glycine (encoded by GGX) cannot be the +1 amino acid of the recognition site, and (3) an alternative LIC restriction site to *Bse*RI must be used. We were able to design such a sequence by use of *Ssp*I as the LIC restriction site and serine as the +1 residue of the recognition site (Fig. 2). *Ssp*I cleaves its recognition site AATTTA in the middle to generate a blunt end terminating in AAT, which encodes asparagine. Treatment of linearized pMCSG7 with T4 polymerase in the presence of dGTP will generate 15-base overhangs (Fig. 3A). Complementary PCR products must be processed in the presence of the complementary dNTP, in this case dCTP. The

required use of dCTP imposes an additional restriction, that the first base after the overhang in the PCR product must be G, complementary to the C that must be present to terminate the action of the polymerase (Fig. 3B). This requirement places a G at the start of the next codon and limits the possible amino acids to glycine, alanine, valine, aspartate, or glutamate. For cloning into pMCSG7, alanine (GCX) is recommended because of its small size and lack of charge. Glycine is not recommended because of the instability of the asparagine-glycine peptide bond it would generate. The annealed product (Fig. 3C) encodes a leader sequence with the TEV protease site followed by the amino acids asparagine-alanine. Processing of expressed proteins generates a product beginning with the sequence serine-asparagine-alanine followed by the amplified protein sequence.

The results reported here show that the new TEV/LIC vector functions as intended. The vector allows the high cloning efficiency needed for high-throughput applications. We found, however, that in contrast to the *Bse*RI-based LIC vectors, preparation of pMCSG7 for LIC requires additional treatment. Apparently, traces of uncut vector remain after treatment with *Ssp*I and T4 polymerase that generate an unacceptably high background of transformants lacking an inserted gene. Intensive digestion or gel purification of the processed plasmid reduced the background sufficiently. Expression of proteins cloned into pMCSG7 was strong, as expected for T7 promoter based expression systems, and comparable to that obtained in pET-30 Xa/LIC and in pMCSG3, a derivative of pET-30 Xa/LIC that, like pMCSG7, only encodes a his-tag.

Purification trials demonstrated that the his-tag and the TEV protease cleavage site function properly. A representative protein, APC331, was purified by standard, vendor prescribed protocols both on small Swell-Gel disks containing chelated Ni (simulating purification in microtiter plates, for which the disks were designed) and on a Ni-NTA agarose resin (simulating large-scale purification for crystallization of target proteins). The purified protein was cleaved efficiently with TEV protease under various conditions using two different forms of the protease. In one experiment, commercial TEV protease cleaved the target protein completely in fewer than 4 h at room temperature. In the second case, his-tagged TEV protease purified locally cleaved the target efficiently at 4°C overnight either in the presence of 250 mM imidazole or during dialysis into buffer lacking imidazole. Treatment of the dialyzed, cleaved target with Ni-NTA agarose effectively removed the his-tagged protease.

The results presented show that pMCSG7 provides an alternative LIC vector suitable for high-throughput cloning. The incorporation of the TEV protease cleavage

site—which should permit cleavage reactions to be carried out under a wider range of conditions—will allow greater flexibility in the design of purification protocols, potentially allowing greater automation of those steps and eliminating possible bottlenecks. Because of the variability of biological materials, no one vector is ideal for all proteins. pMCSG7 is intended to serve as a platform for the development of a series of vectors. The adjacent *Bgl*III and *Kpn*I sites, located between the his-tag and cleavage site, should allow easy introduction of additional modules into the leader sequence of expressed proteins to improve expression, solubility, or purification. Because the LIC overhangs would be unaffected by such additions, a single PCR product could be introduced by LIC into a series of vectors. Robotic screening could then include evaluation of the preferred vector for each target, determining which would be most likely to generate sufficient soluble material for purification and crystallization.

## ACKNOWLEDGMENTS

We thank Andrzej Joachimiak, Irina Dementieva, and Tom Zarembinski for helpful discussion, Cindy Sanville Millard for technical support, Ruiying Xu for purified his-tagged TEV protease, and Paul Gardner of Howard Hughes Medical Institute of the University of Chicago for sequence analysis. This work was supported by National Institute of Health Grant GM62414-01, A. Joachimiak, PI, and by the U.S. Department of Energy, Office of Health and Environmental Research, Under Contract W-31-109-Eng-38.

## REFERENCES

1. Terwilliger, T. C. (2000) Structural genomics in North America. *Nat. Struct. Biol.* 7 (Suppl.), 935–993.
2. Stevens, R. C., and Wilson, I. A. (2001) Tech.Sight. Industrializing structural biology. *Science* 293, 519–520.
3. Hol, W. G. (2000) Structural genomics for science and society. *Nat. Struct. Biol.* 7 (Suppl.), 964–966.
4. Lesley, S. A. (2001) High-throughput proteomics: Protein expression and purification in the postgenomic world. *Protein Expr. Purif.* 22, 159–164.
5. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18, 6069–6074.
6. Haun, R. S., Serventi, I. M., and Moss, J. (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques* 1992, 515–518.
7. Sheibani, N. (1999) Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins. *Prep. Biochem. Biotechnol.* 29, 77–90.
8. Kapust, R. B., and Waugh, D. S. (2000) Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* 19, 312–318.
9. Donnelly, M. I., Stevens, P. W., Stols, L., Xiaoyin Su, S., Tollaksen, S., Giometti, C., and Joachimiak, A. (2001) Expression of a highly toxic protein, Bax, in *Escherichia coli* by attachment of a leader peptide derived from the GroES cochaperone. *Protein Expr. Purif.* 22, 422–429.
10. Dougherty, W. G., Carrington, J. C., Cary, S. M., and Parks, T. D. (1988) Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* 7, 1281–1287.
11. Dougherty, W. G., and Parks, T. D. (1989) Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology* 172, 145–155.
12. Dougherty, W. G., Cary, S. M., and Parks, T. D. (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: A model. *Virology* 171, 356–364.
13. Dougherty, W. G., Parks, T. D., Cary, S. M., Bazan, J. F., and Fletterick, R. J. (1989) Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* 172, 302–310.
14. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* 216, 413–417.
15. Dieckman, L., Gu, M., Stols, L., Donnelly, M. I., and Collart, F. R. (2002) High throughput methods for gene cloning and expression. *Protein Expr. Purif.* 25, 1–7.