

Positive Selection Vectors to Generate Fused Genes for the Expression of His-Tagged Proteins

T. Van Reeth, P.-L. Drèze, J. Szpirer, C. Szpirer and P. Gabant
Université Libre de Bruxelles, Rhode-Saint-Genèse, Belgium

BioTechniques 25:898-904 (November 1998)

ABSTRACT

Epitope tagging simplifies detection, characterization and purification of proteins. Gene fusion to combine the coding region of a well-characterized epitope with the coding region for a protein of interest generally requires several subcloning steps. Alternatively, a PCR strategy can be used to generate such a chimeric gene. In addition to its simplicity, this approach allows one to limit the size of the multiple cloning sites present in conventional expression vectors, thus reducing the introduction of artifactual amino-acid sequences into the fused protein. In this communication, we describe new vectors that allow PCR cloning and selection of chimeric genes coding for N- or C-terminal His-tagged proteins. These vectors are based on the control of cell death CcdB direct selection technology and are well adapted to the cloning of blunt-ended PCR products that were generated by using thermostable polymerases that provide proof-reading activity.

INTRODUCTION

In vitro analysis of protein function (i.e., enzymatic assays, protein-protein and protein-DNA interactions) requires large amounts of purified polypeptide. Often, this is hindered by difficulties in purifying the expressed protein and/or the lack of specific antibodies. To circumvent these problems, a technique known as epitope tagging has been developed. This approach involves the construction of a recombinant fusion protein containing a Tag epitope attached to either the N or the C terminus of the protein of interest. The chimeric protein can then be easily identified in whole cell extracts and purified by

immunoprecipitation using antibodies raised against the Tag epitope. Epitope tagging has been widely used for the study of newly isolated gene products in biological systems as diverse as *Escherichia coli*, yeast, insect cells and mammalian cells.

Several Tag epitopes have been used including: β -galactosidase, glutathione-S-transferase (GST) and peptides like c-myc or poly-histidine peptide (His-Tag). Some chimeric expression systems include a specific endopeptidase cleavage site (e.g., Enterokinase [EK], Thrombin or Xa factor) that allows one to clear the purified protein from its Tag epitope. Different genetic systems are now available to express tagged proteins (8,12). The His-Tag is a short region consisting generally of six histidine residues [6x(His)-Tag] with a high affinity for Cu_2^+ , Ni_2^+ , Co_2^+ or Zn_2^+ , allowing the fusion product to be separated from a complex protein mixture by metal chelate affinity chromatography (6). The fusion proteins that bind to the affinity matrix are eluted with a solution containing either ammonium chloride, glycine, histidine, imidazol and EDTA or with high concentrations of competing metal ions. In contrast with long, fused Tag sequences that can interfere with the activity of the fused protein of interest, it can be expected that in most cases the short His-Tag epitope will not affect protein activities.

Complex strategies have to be used to fuse the Tag epitope coding sequence in-frame with the gene of interest. To simplify the generation of such fusions, expression systems containing three vectors presenting the His-Tag coding sequence in the different frames were developed. However, this approach is limited by the possibility that compatible restriction sites, both in the studied gene and/or in the multiple cloning site

(MCS) of the vector, are present.

The fused proteins produced by these vectors generally contain additional amino acids (aa) because of the translation of vector sequences that might introduce artifacts in the final protein to be analyzed.

Alternatively, polymerase chain reaction (PCR) could be used to generate the fusion between the Tag coding region and the sequence of interest. However, current expression vectors are not well adapted to the cloning of PCR products. We have recently developed new direct-selection vectors adapted to the cloning of PCR-generated inserts (4). They are based on the insertional inactivation of the control of cell death *ccdB* gene (3) and were shown to be particularly efficient for the cloning of blunt-ended PCR products generated by thermostable DNA polymerases, that also provide proof-reading activity (e.g., *Pfu* and Vent®). Commercial versions of these vectors (e.g., pCR®-Blunt) are now available in the Zero Blunt™ PCR Cloning Kit (Invitrogen, Carlsbad, CA, USA). We generated new vectors to use the CcdB technology for the expression of His-tagged proteins in *E. coli*. Their advantages are presented and illustrated in this communication.

Construction of pKILHIS-1

The fragment containing pT7-RBS-ATG-6×(His)-EK from pRSET-B (Invitrogen) was amplified by PCR and cloned (*Hind*III-*Eco*RV) into the *Hind*III-*Eco*RV fragment of pZERO-2 (Invitrogen). The terminator of pRSET-B (fragment *Xho*I-*Bgl*II) was cloned into the *Xho*I-*Bgl*II sites of this first construction. The fragment containing *lacZa-ccdB* (fragment *Eco*RV-*Eco*RI of pZERO-2) was then cloned in the *Eco*RV-*Eco*RI sites of the construction to generate the pKILHIS-1 plasmid. The sequence of pKILHIS-1 appears in the GenBank® database (Accession No. AF028837).

Construction of pKILHIS-2

The *Nhe*I-*Hind*III fragment from pKILHIS-1, containing EK*-*lacZa-ccdB* was cloned into the *Nhe*I-*Hind*III sites of pTrcHisB (Invitrogen), and the fragment *Ssp*I-*Ssp*I of this new construction (pKILHIS-B) was cloned in the *A*fIII(blunted)-*Stu*I sites of pZERO-2 to generate the pKILHIS-2 plasmid. The sequence of pKILHIS-2 appears in the GenBank database (Accession No. AF028838).

N-Terminal His Tagging

Direct selection vectors pKILHIS-1 and pKILHIS-2 were genetically engineered to generate and express N-terminal His-tagged proteins.

Figure 1 shows the structure of the pKILHIS-1 and -2 vectors. These vectors differ from the original pKILPCR-2 plasmid (4) in two ways: (i) they contain a 6×(His)-Tag coding sequence introduced in the *ccdB* gene, and (ii) they contain strong prokaryotic promoters for the expression of the fused protein. No difference was observed between the killing activity of these derivatives and that of the parental pKILPCR-2 plasmid (5), showing that the 6×(His)-CcdB poison produced by these vectors is not less toxic than the selective marker of the parental construct. Amplified fragments can be cloned into the blunt-end *Eco*RV site of the plasmids. The two vectors have different promoters (Figure 1A; *Plac* and *PT7* for pKILHIS-1; *P_{trc}* for pKILHIS-2) that are used to regulate the

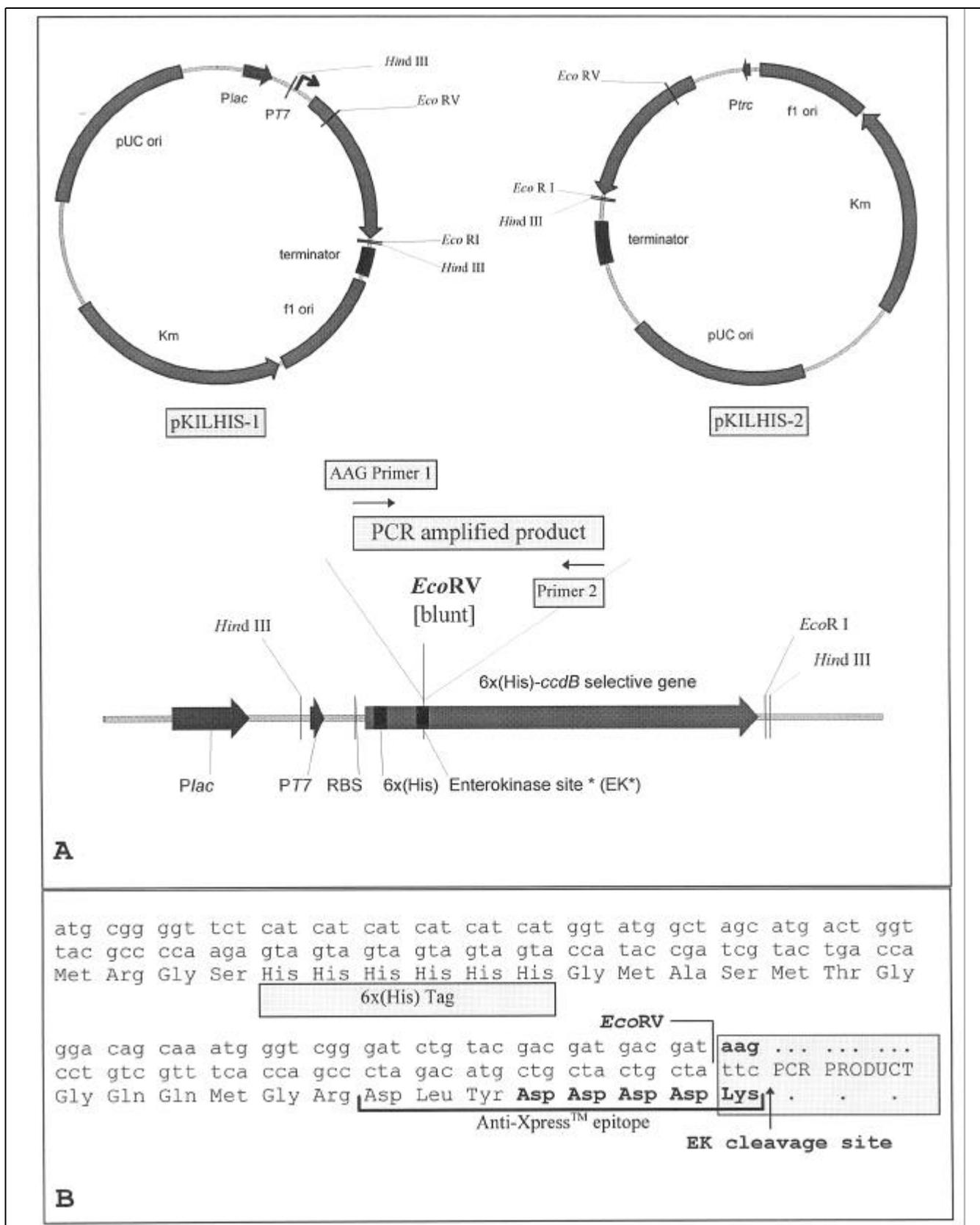


Figure 1. Structure of pKILHIS-1 and pKILHIS-2. The two plasmids contain a 6x(His)-*ccdB* fusion under the control of strong promoters (*Plac* and *PT7* for pKILHIS-1 and *P_{trc}* for pKILHIS-2). (A) General map of pKILHIS-1 and pKILHIS-2, details of the different elements of the fused gene of pKILHIS-1. The *EcoRV* (blunt) site that can be used to clone the PCR-generated insert, the partial Enterokinase site (EK*) overlapping the ribosome binding site (RBS) are indicated. Primers (primer 1 and 2) represent the oligonucleotides designed to produce the insert. A lysine (AAG) codon is added at the 5' end of the primer 1 to generate a complete EK recognition coding sequence (see text). (B) Schematic representation of the insertion of a PCR-generated insert in the *EcoRV* cloning site of pKILHIS-1 vector, showing the EK cleavage site and the Anti-Xpress™ epitope generated by this insertion.

expression of the fused protein in *E. coli*. The 5' primer (primer 1) is chosen in the gene of interest and designed in such a way that the amplified fragment is inserted in-frame with the 6×(His)-Tag coding sequence of both vectors. Because the reading frame is imposed by the 5' primer, a single vector can be used to generate any N-terminal fusion with the 6×(His)-Tag epitope of these vectors. In its C-terminal end, the fusion is limited by either the natural stop codon present in the amplified gene or by a stop codon inserted in the 3' primer.

It may be important for certain applications to remove the Tag epitope from the fused protein. Most systems are designed in such a way that an endopeptidase cleavage site allows clearance of the epitope from the fused protein, generally leaving artifactual aa at the N terminal of the cleared protein that, consequently, do not correspond to the natural one. This drawback can be overcome by using the vectors described and a PCR strategy that generates a peptidase site at the junction between the epitope and the polypeptide of interest. To limit the length of sequence added in the 5' primer, we

searched for restriction sites that generate blunt-ends and contain at least part of the coding sequence for a peptidase recognition site. EK and Xa factor are the most useful peptidases because they cleave at the carboxy-terminal side of their respective recognition sequence. The *EcoRV* site of pKILHIS-1 or -2 is part of the coding sequence of a partial EK cleavage site (Figure 1A, EK*), and a complete site is generated if a lysine codon (AAG) is included at the 5' end of the 5' primer (Figure 1B, primer 1).

To analyze the potential of fusion strategies using these vectors, the *E. coli* chloramphenicol acetyltransferase (*cat*) gene present in the pBLCAT2 plasmid (9) was amplified and cloned into pKILHIS-2. This amplification was performed using the *Pfu* DNA Polymerase (Catalog No. 600153; Stratagene, La Jolla, CA, USA) and the following two primers: 5'-AAGATGGAGAAAAAATCACTGG-3' (primer 1), in which the 5' nucleotides (AAG) are added to generate the EK cleavage site after ligation, and 5'-TTACGCCCCGCCCTGC-CACTC-3' (primer 2), in which the three 5' nucleotides (TTA)

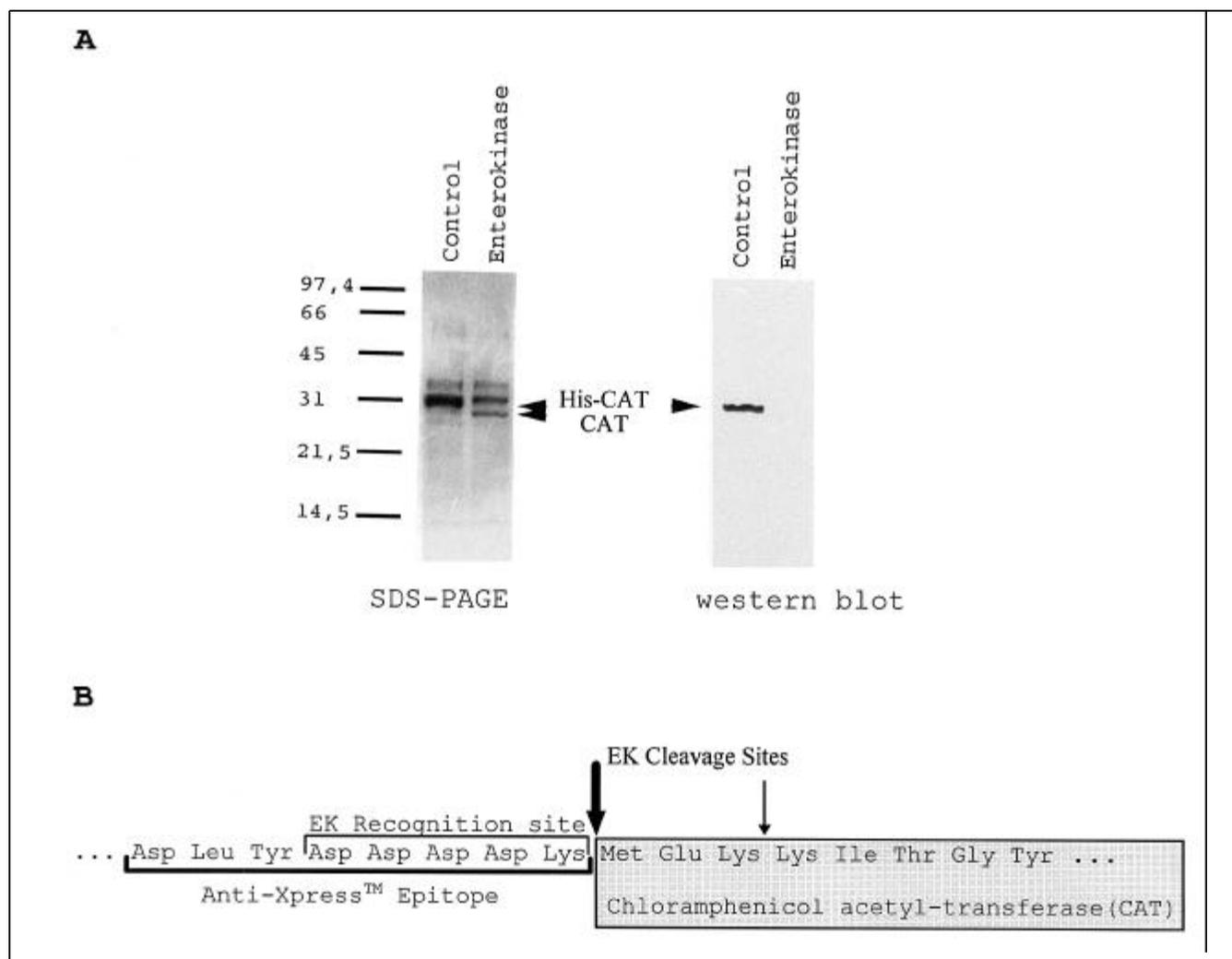


Figure 2. Analysis of the removal by EK digestion of the 6×(His)-Tag from the CAT protein expressed by pKILHIS-2. (A) SDS-polyacrylamide gel electrophoresis (PAGE) (left) and Western blot (right) of the fused CAT protein, before (control) and after EK treatment of the protein extract. As shown on the left, a 30-kDa polypeptide corresponding to the 6×(His)-CAT fused protein was detected in the control extract, and a smaller band (28 kDa) corresponding to the natural CAT protein was detected after EK treatment. As shown on the right, Anti-Xpress antibody gave a single band in the control and no signal after EK treatment of the extract, indicating that the epitope has been cleaved from the fusion protein. (B) To determine whether the cleavage had occurred at the EK site, the CAT protein obtained after epitope cleavage was extracted from the gel, and the N-terminal end of the protein was sequenced. As indicated by a bold arrow, 80% of the product was shown to correspond to a cleavage at the expected site, with 20% corresponding to a CAT internal secondary cleavage site (indicated by an arrow).

constitute the *cat* stop codon. The resulting 663-bp amplification product was directly cloned into *EcoRV*-digested pKILHIS-2. Recombinants were selected on kanamycin plates after transformation of the CcdB sensitive TOP10F' Competent Cells (Invitrogen). One recombinant containing the *cat* gene fused to the 6x(His)-Tag of the pKILHIS-2 vector was chosen to over-express the fusion protein (see Figure 2).

The selected *E. coli* TOP10F' strain (Invitrogen) harboring pKILHIS-2 containing the *cat* coding sequence was grown in 30 mL LB medium at 37°C to an optical density (OD) of 0.6. Expression was then induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The culture was grown for 2 h under induced conditions. After centrifugation at 4500× g, the bacterial pellet was

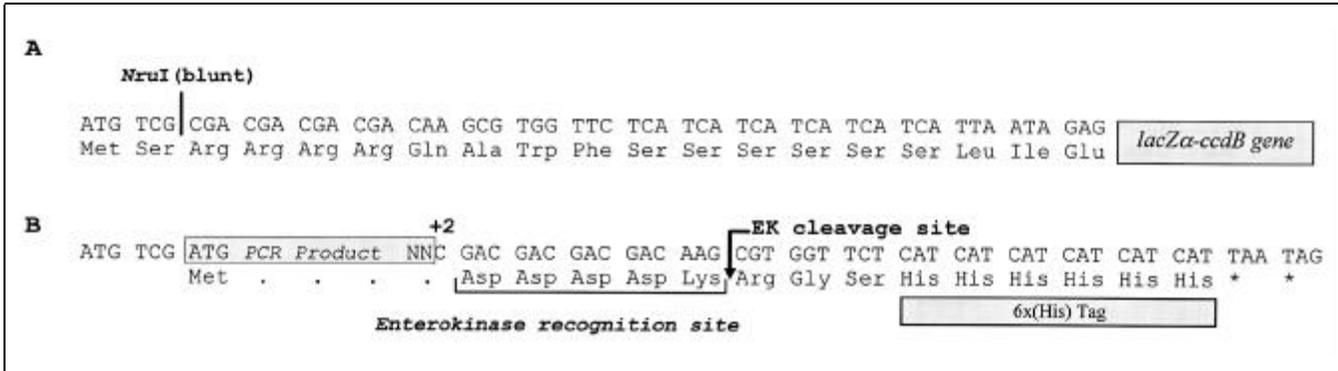


Figure 3. Structure of the pKILHIS-3 cloning site. pKILHIS-3 was generated by insertion in the *NcoI*-*Bam*HI sites of pKILHIS-B (for description see text) of the following sequence: 5'-CATGTCGCGACGACGACGACAAGCGTGGTTCTCATCATCATCATCATTAATAGAGG-3'. The sequence of pKILHIS-3 appears in the GenBank database (Accession No. AF0500464). (A) Sequence of the pKILHIS-3 cloning site. The 5' end of the *lacZα-ccdB* sequence is indicated with its corresponding translation. The *NruI* restriction site that can be used to clone the PCR-generated insert is indicated. (B) Schematic representation of the structure obtained after insertion of a PCR product in the *NruI* site. The translation of the C-terminal peptide obtained after introduction of a +2 frame-shift is indicated by the two additional bases added at the 5' end of primer 2 are represented (NN). The EK recognition site, the EK cleavage site, the 6x(His)-Tag and the stop codons (*) generated by the insertion are also shown.

suspended in 50 mM Tris-HCl, pH 7.5. The protein lysate was obtained by two freeze-thaw cycles. This crude protein extract was cleared by centrifugation and divided in two aliquots. One aliquot was treated with 3 µg of EK (Boehringer Mannheim, Mannheim, Germany) overnight at 37°C, and the other aliquot was used as a control (placed overnight at 37°C without treatment). For analysis, 5 µL of each sample were electrophoresed on a 12% sodium dodecyl sulfate (SDS) polyacrylamide gel, and the gel was stained with Coomassie® Blue. The protein samples were also blotted on a nitrocellulose membrane and revealed with the Xpress™ Antibody (Invitrogen), raised against the epitope containing the EK recognition motif.

The 30-kDa 6×(His)-CAT fusion protein we obtained was treated with an excess of EK to generate the wild type CAT protein (28 kDa). A Western blotting experiment using Anti Xpress Antibody (Invitrogen) raised against the Tag polypeptide was performed to confirm that the epitope had been cleaved from the fused CAT protein. To assess that the N-terminal end of CAT was restored in the cleaved polypeptide, the product obtained after EK digestion was sequenced. The correct N terminus was found (Figure 2B).

C-Terminal His Tagging

Direct selection expression vectors aimed at generating and expressing C-terminal, His-tagged proteins were constructed because the insertion of an epitope at the N terminus of the protein of interest could alter its function. Another vector, pKILHIS-3, has been designed and allows the generation of a fusion protein with an epitope at its C terminus. Insertion of the sequence of interest in the blunt-ended *NruI* site of pKILHIS-3 generates a fusion protein with the 6×(His) epitope. As described in Figure 3, the cloning site of pKILHIS-3 is designed so that the insertion of a PCR product introducing a +2 frame shift not only disrupts *ccdB* but also imposes the translation of (i) an EK cleavage site, (ii) a 6×(His)-Tag and (iii) two stop codons limiting the tagged protein at its C terminus. Because EK has a recognition site in the C terminus, cleavage by EK of any fused protein expressed by this vector will leave 6 additional aa at the C terminus of the expressed protein.

We observed that the pKILHIS-3 plasmid is toxic for the CcdB-sensitive TOP10F' strain even without IPTG induction of the *P_{trc}* promoter (which controls the *ccdB* gene). This unexpected observation opens the possibility of selecting recombinants in the absence of *P_{trc}* induction. This is a real advantage, especially if high-level expression of the studied protein is anticipated to be toxic.

To test the usefulness of pKILHIS-3 as a 6×(His)-Tag fusion expression vector, new primers to the *E. coli cat* gene were designed as follows and then used for amplification: 5'-ATGGAGAAAAAATCACTGG-3' (primer 1) and 5'-TC-CGCCCCGCCCTGCCACTC-3' (primer 2). The 659-bp PCR product was inserted into *NruI*-digested pKILHIS-3. The ligation reaction was transformed in TOP10F'. Selection in the presence or absence of IPTG led to more than 90% of recombinants. However, in recombinants selected without induction, inserts were found in both orientations; whereas in recombinants selected in the presence of the inducer (IPTG), only *cat* genes inserted in the opposite orientation to the *P_{trc}* promoter were recovered. This observation suggests that high-level production of C-terminal-tagged CAT protein is toxic to the bacterial host. Moreover, recombinant plasmids

selected without induction and containing the fused *cat* gene under the control of *P_{trc}* were killed when grown in the presence of IPTG. The toxic effect of the C-terminal 6×(His)-CAT protein was surprising because no toxic effect was observed by induction of the N-terminal-tagged *cat* gene in pKILHIS-2. Because the toxicity could potentially be due to residual activity of *ccdB* in the pKILHIS-3-derived construct, the CcdB resistant strain B462 (1) was transformed either with pKILHIS-3 or with a recombinant containing the C-terminal 6×(His)-tagged *cat* coding sequence. In presence of IPTG, pKILHIS-3 carrying the fused *cat* gene was lethal, while no killing was observed with the vector. This strongly suggests that the C-terminal 6×(His)-CAT protein is toxic for *E. coli* when over-expressed.

To definitively show that the killing was caused by the 6×(His) C-terminal tagged CAT, the corresponding coding sequence was cloned in another expression vector, namely Xpress Vector pTrcHis B (Invitrogen). Bacteria harboring recombinant plasmids obtained from pTrcHis B were also killed by induction conditions. These observations demonstrate that high-level expression of the C-terminal tagged CAT protein is toxic for *E. coli* and illustrate the potential value of selecting for recombinants in non-induced conditions.

DISCUSSION

The *ccdB* gene of the F plasmid *ccd* operon has been shown to be an efficient marker for the selection of recombinant clones. The original vectors pKIL18–19 were designed to clone inserts generated by conventional restriction enzyme digestion (3). The positive selection system could be seen as a technology that allows one to avoid the vector dephosphorylation step and/or the cost of the 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) reagent used in *lacZ* inactivation screening. To our knowledge, the major advantage of the CcdB technology over other selection methods developed so far, is that this approach is based on a small gene whose function is known (*ccdB* encodes a natural gyrase poison) and that a resistant strain has been selected (2). The function of the CcdB gyrase poison and the conservation of its target in different bacterial species opens the door to a broader use of the *ccdB* gene in bacterial genetics (5).

Cloning of PCR-generated fragments is currently very popular; however, this approach precludes vector dephosphorylation and is rather inefficient. The possibility of selecting vectors with insertion singularly improves the power of PCR cloning, especially with the new generation of thermostable polymerases that generate blunt-ended products (4). Unlike the original *Taq* DNA polymerases, these proofreading polymerases generate blunt-ended products, thereby preventing the TA cloning used to limit vector self-ligation and to increase cloning efficiency of *Taq*-generated inserts (7,10,11).

In this communication, we describe new expression vectors, combining the CcdB selection systems and His tagging. The major advantage of these new vectors is that they allow, in one cloning step, the generation of a chimeric gene generated by PCR. In addition, His tagging either at the N or C terminus is possible. The combination of PCR flexibility and selection of recombinants allowed us to propose the possibility of removing the N-terminus epitope from the chimeric protein to release the “authentic” polypeptide.

This new application demonstrates the versatility of the

positive selection using *ccdB* gene derivatives.

The potential of this selection technique to generate fusion proteins opens the door to new eukaryotic expression vectors containing coding sequences for signal peptides to target the expressed protein to the extracellular medium or to specific cell compartments.

Positive selection relying on insertional inactivation of *ccdB* genes is filed for patent (2) protection and is licensed by the Université Libre de Bruxelles to Invitrogen Corporation.

ACKNOWLEDGMENTS

We are grateful to Michel Faelen and Laurence Van Melder for critical review of the manuscript. This work was supported by CGER-Assurances, the Association contre le Cancer, the Communauté française de Belgique (ARC) and the Fund for Medical Scientific Research (FRSM). T.V.R. is a fellow of the Fonds pour la Recherche et l'Industrie Agroalimentaire (FRIA). C.S. is a Research Director of the Fonds National pour la Recherche Scientifique (FNRS).

REFERENCES

1. **Bernard, P.** 1996. Positive selection of recombinant DNA by CcdB. *BioTechniques* 21:320-323.
2. **Bernard, P. and P. Gabant**, inventors. Université Libre de Bruxelles, assignee. Cloning and/or sequencing vector. PCT patent PCT/BE93/00051 and Belgian patent 09200696.
3. **Bernard, P., P. Gabant, El.M. Bahassi and M. Couturier.** 1994. Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene* 148:71-74.
4. **Gabant, P., P.-L. Drèze, T. Van Reeth, J. Szpirer and C. Szpirer.** 1997. Bifunctional *lacZα-ccdB* genes for selective cloning of PCR products. *BioTechniques* 23:938-941.
5. **Gabant, P., C.Y. Szpirer, M. Couturier and M. Faelen.** 1998. Direct selection cloning vectors adapted to the genetic analysis of Gram-negative bacteria and their plasmids. *Gene* 207:87-92.
6. **Hochuli, E., W. Bannwarth, H. Döbeli, R. Gentz and D. Stüber.** 1988. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology* 6:1321-1325.
7. **Invitrogen Corporation**, inventor. T/A cloning. US patent 5,487,993 and European patent 0550693.
8. **LaVallie, E.R., D.B. Smith and P. Riggs.** 1994. Introduction to expression by fusion protein vectors, p. 16.4.1-16.4.17. *In* F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (Eds.), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
9. **Luckow, B. and G. Schütz.** 1987. CAT constructions with unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res.* 15:5490.
10. **Marchuck, D., M. Drumm, A. Saulino and F.S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* 19:1154.
11. **Mead, D.A., N.K. Pey, C. Herrnstadt, R.A. Marzil and L.M. Smith.** 1991. Universal method for the direct cloning of PCR amplified nucleic acid. *Bio/Technology* 9:657-663.
12. **Zhang, N. and J.-L. Chen.** 1998. Purification of recombinant proteins and study of protein interaction by epitope tagging, p. 10.15.1-10.15.9. *In* F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (Eds.), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Address correspondence to Dr. Philippe Gabant, Université Libre de Bruxelles, Département de Biologie Moléculaire, Laboratoire de Biologie du Développement, Rue des Chevaux, 67, B-1640 Rhode-Saint-Genèse, Belgium. Internet: pgabant@dbm.ulb.ac.be