Nucleic Acids Research, 1998, Vol. 26, No. 22 **5109–5115**

Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step

Shaorong Chong*, Geoffrey E. Montello, Aihua Zhang, Eric J. Cantor, Wei Liao, Ming-Qun Xu and Jack Benner

New England Biolabs Inc., 32 Tozer Road, Beverly, MA 01915, USA

Received August 10, 1998; Revised and Accepted October 1, 1998

ABSTRACT

A conventional affinity protein purification system often requires a separate protease to separate the target protein from the affinity tag. This paper describes a unique protein purification system in which the target protein is fused to the C-terminus of a modified protein splicing element (intein). A small affinity tag is inserted in a loop region of the endonuclease domain of the intein to allow affinity purification. Specific mutations at the C-terminal splice junction of the intein allow controllable C-terminal peptide bond cleavage. The cleavage is triggered by addition of thiols such as dithiothreitol or free cysteine, resulting in elution of the target protein while the affinity-tagged intein remains immobilized on the affinity column. This system eliminates the need for a separate protease and allows purification of a target protein without the N-terminal methionine. We have constructed general cloning vectors and demonstrated single-column purification of several proteins. In addition, we discuss several factors that may affect the C-terminal peptide bond cleavage activity.

INTRODUCTION

A widely used technology for recombinant protein expression and purification is to express a target protein as a fusion to an affinity tag protein, such as *Schistosoma* glutathione *S*-transferase (GST), *Escherichia coli* maltose-binding protein (MBP), etc. (1). The affinity tag, often highly expressed in *E.coli*, can elevate the expression level of the fusion protein and thus the yield of the target protein (1). One potential limitation of this technology is the use of a separate protease to cleave the target protein from its affinity tag. Treatment with proteases adds an extra step to the purification, and proteases are sometimes non-specific and inefficient. Protein splicing elements, named inteins (2), capable of catalyzing specific splicing reactions, offer a unique alternative to cleave a peptide bond without the use of a separate protease (3,4). Previously, we investigated an inducible peptide bond cleavage reaction at the N-terminus of an intein from *Saccharomyces cerevisiae*, the *Sce* VMA intein (3,5,6). Our study led to the development of the first intein-based affinity protein purification system (5). The target protein was fused to the N-terminus of the intein whose C-terminus was linked to a chitin-binding domain (CBD) from *Bacillus circulans* as an affinity tag (5). Thiols such as dithiothreitol (DTT), β-mercaptoethanol and free cysteine induced peptide bond cleavage at the intein N-terminus, thereby releasing the target protein from the rest of the fusion protein immobilized on the column (5). Though this N-terminal cleavage system eliminates the need for a separate protease, it loses the advantage of having a highly expressed tag protein as the N-terminal fusion partner, which, in many cases, is necessary for achieving a high level of protein expression. The target protein (as the N-terminal domain) often determines the expression level of the fusion protein.

In this paper, we describe an affinity purification system in which the target protein was fused to the C-terminus of a modified *Sce* VMA intein capable of peptide bond cleavage at its C-terminus. The modified intein contained a double substitution of His453 and Cys455 at the intein C-terminal splice junction. The cleavage activities of the intein containing the double substitution, His453Gln/Cys455Ala, were previously studied in an MYT4 fusion system in which the intein was fused between MBP (as the N-extein) and phage T4 DNA ligase (as the C-extein) (6). We have found that induction of the intein N-terminal cleavage in MYT4 by thiols such as DTT or free cysteine could trigger specific cleavage at the intein C-terminus (at Asn 454) (6). To utilize this inducible C-terminal cleavage activity of the intein for protein purification, we have constructed general cloning vectors from pMYT4. A sequence consisting of the first 10 residues of MBP was used to replace the MBP sequence in MYT4 to initiate the translation of the fusion protein. The CBD from *B.circulans* was then inserted in a loop region of the intein endonuclease domain to allow affinity purification. Finally, the T4 DNA ligase sequence was replaced by multiple cloning sites. The resulting vectors, named pIMC vectors, were used for the expression and purification of several recombinant proteins. Since a target protein may require a residue other than methionine at its N-terminus, we also investigated the effect of substituting the first C-extein residue (residue 455, equivalent to

*To whom correspondence should be addressed. Tel: +1 978 927 5054; Fax: +1 978 921 1350; Email: chong@neb.com

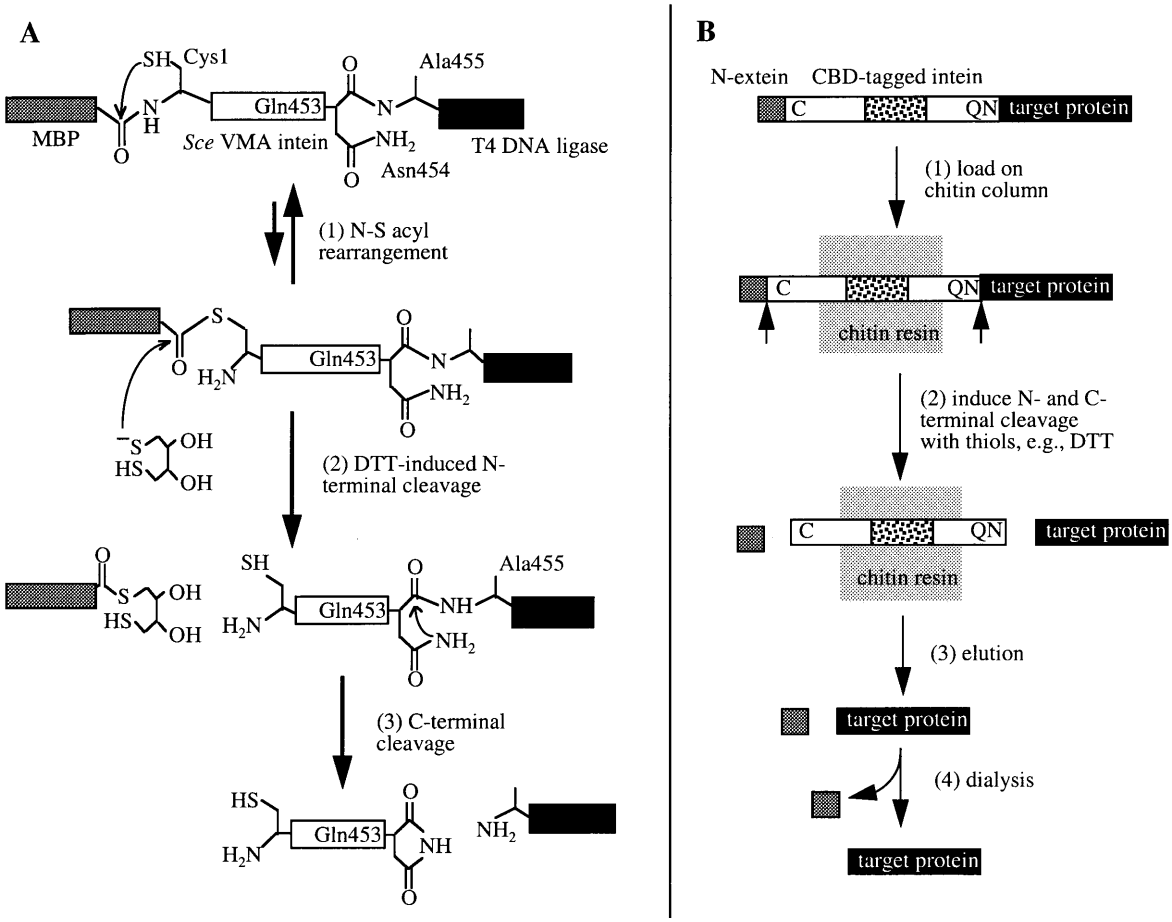

Figure 1. (**A**) A proposed mechanism for thiol-induced cleavage reactions at the N- and C-termini of the *Sce* VMA intein. The MYT4 system consists of the modified intein (white box) fused between MBP (grey box) as the N-extein and T4 DNA ligase (black box) as the C-extein. The double substitution, His453Gln/Cys455Ala, allows DTT-induced N-terminal cleavage [steps (1) and (2)] and *in vitro* succinimide formation [i.e. C-terminal cleavage, step (3)]. Similarly, cysteine can also induce cleavage and consequently attach to the N-extein through a peptide bond (for more details see 5,6). (**B**) Schematic diagram of the protein purification procedures utilizing inducible cleavage activities of a modified intein. The target protein (black box) is fused to the C-terminal residue Asn (N) of the intein containing a single substitution, His453Gln (Q), and the CBD insertion (dotted box). After induction of cleavage [step (2)], the target protein along with the N-extein sequence (grey box) are eluted from the column while the intein is immobilized on the chitin resin (shaded area) [step (3)]. The target protein can usually be purified away from the N-extein sequence by dialysis [step (4)].

the N-terminal residue of a target protein) on the intein C-terminal cleavage in the MYT4 fusion system.

MATERIALS AND METHODS

Construction of pIMC vectors for expression and affinity purification of target proteins fused to the C-terminus of the intein

The first step was to insert the CBD in the intein. Using the single-stranded pLitYP (3) as the template for mutagenesis by the method of Kunkel (7), the mutagenic primer 5′-AAGATTATTGC-GAATACCACCATGGTGACCACCGCTAGCACCTCTGACA-ACTTTAGAGTA-3′ created two unique restriction sites, *Nhe*I and *Nco*I, between residues 272 and 275 of the intein, to yield pLitYP (*Nhe–Nco*). Amino acid numbers refer to the position in the *S.cerevisiae* VMA intein with the first residue being Cys1 and last Asn454 (Fig. 1) (3). The CBD gene was amplified by the polymerase chain reaction (PCR) from pMYB129 (5) using the primers 5′-GGTGGTGCTAGCACAAATCCTGGTGTATCCGC-TTGG-3′ and 5′-GGTGGTACCATGGCCACCTTGAAGCTG-CCACAAGGCAGGAAC-3′ and following a previously described protocol (6). The product was digested with *Nhe*I and *Nco*I and cloned into the same sites in pLitYP (*Nhe–Nco*) to generate a CBD-tagged intein. The *Xho*I–*Bam*HI fragment containing the CBD-tagged intein was then used to replace the corresponding segment in pMYT4 (H453Q/C455A) (6) to yield pMY(B)T4 (H453Q/C455A). Through linker insertion, the MBP sequence in pMY(B)T4 (H453Q/C455A) was replaced by a sequence encoding the first 10 amino acid residues of MBP to create pCT4. The complete N-extein amino acid sequence (the coding sequence before Cys 1 of the intein) is as follows: MKIEEGKLVIGSLEG. Finally, the T4 ligase sequence in pCT4 was replaced with multiple cloning sites (Fig. 2) to create pIMC vectors. Unless otherwise stated, all enzymes and plasmids used are the products of New England Biolabs, Inc.

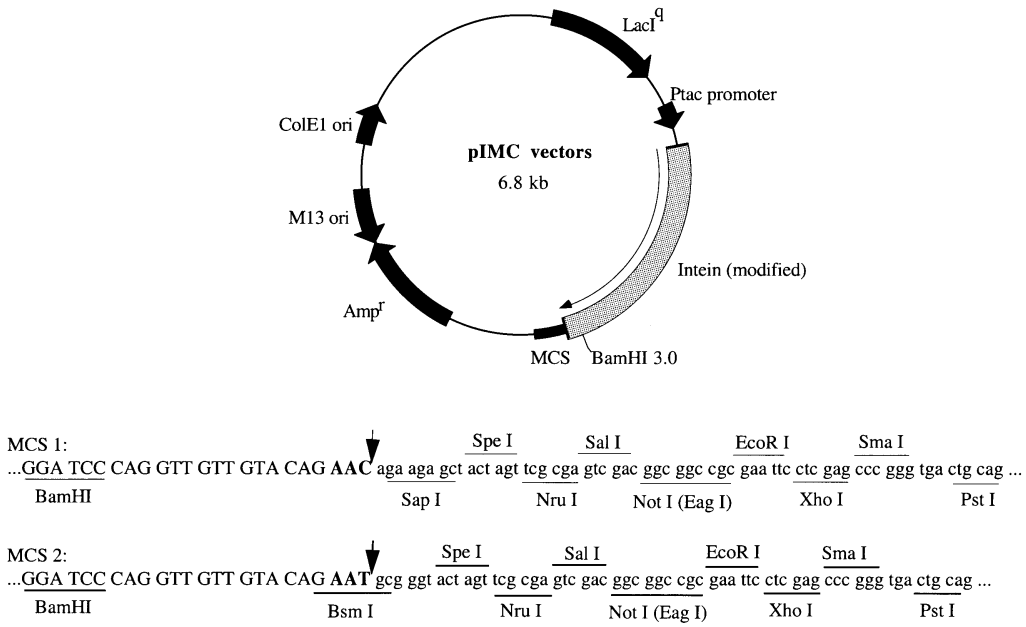

Figure 2. The pIMC *E.coli* expression vectors. The pIMC vectors have a *Col*E1 replication origin (*Col*E1 ori), an M13 origin (M13 ori) for generating single-stranded DNA and an ampicillin resistance gene (Amp^r). The modified intein contains the mutation for inducible C-terminal cleavage and the CBD insertion. It is flanked by a short N-extein sequence encoding the first 10 amino acid residues of MBP and the multiple cloning site (MCS). Expression of the fusion proteins is under control of a *tac* promoter and regulated by the *lacI*^{q} repressor. Two multiple cloning sites are shown (MCS 1 and MCS 2) with the recognition sequences of the restriction enzymes underlined. The arrows indicate the cleavage site at the intein C-terminal residue, Asn (codon AAT or AAC).

Expression and purification of recombinant proteins in pIMC vectors

The genes for the target proteins (Table 2) were amplified by PCR to include suitable restriction sites for cloning into one of the pIMC vectors (Fig. 2). The PCR protocol was essentially the same as previously described (5). For the T4 gene 32 product (8), the forward primer, 5′-GGTGGTTGCTCTTCCAACGCTG-GT**TTTAAACGTAAATCTACTGCTGAACTC**-3′, contains a *Sap*I site (underlined) and two extra amino acid codons (GCT, GGT) added to the first nine codons of the T4 gene 32 (bold). The reverse primer, 5′-GGTGGTCTGCAGTCA**AAGGTCATT-CAAAAGGTCATCCAGGTC**-3′, contained a *Pst*I site (underlined), a stop codon (TCA) and the sequence encoding the last nine amino acid residues of the T4 gene 32 product. The PCR product of the T4 gene 32 was cloned into the *Sap*I and *Pst*I sites in MCS1 of a pIMC vector (Fig. 2). Similarly, the genes for T4 endo VII (9) and rabbit protein phosphatase-1 (PP1) (10) were cloned in the *Spe*I, *Xho*I sites of MCS2; the endonuclease *Fse*I (11) and green fluorescent protein (GFP) (12) in the *Spe*I, *Pst*I sites of MCS2; a calmodulin-dependent protein kinase (CamKII) (13) and the *S.cerevisiae* Invertase (14) in the *Bsm*I, *Pst*I sites of MCS2. T4 DNA ligase (15) was expressed from pCT4. All target genes in pIMC vectors were expressed in *E.coli* strain ER2683 (New England Biolabs, Inc.). The culture was induced with 0.3 mM isopropyl-β-D-thiogalactoside (IPTG) at 15°C for 16 h. Low induction temperatures (15–20°C) can often improve protein expression. The purification procedures were conducted at 0–4°C and are shown schematically in Figure 3. The cell pellet (3–5 g) from each 1 l culture was resuspended in 50 ml of column buffer (20 mM HEPES, pH 8.0; 0.5 M NaCl) and broken by sonication. After centrifugation at 25 000 *g* for 30 min, the supernatant was passed through an Econo-Column (2.5 cm × 10 cm, Bio-Rad, Hercules, CA) packed with 30 ml of chitin beads (New England Biolabs, Inc.) at a flow rate of 0.5 ml/min and the column was then washed with the column buffer at a flow rate of 2 ml/min until the protein content of the eluate reached a minimum (OD_{595nm} < 0.05 as measured by Bradford assay). Column buffer (50 ml) containing 50 mM DTT was quickly passed through the column in order to distribute DTT evenly throughout the resin and the flow was stopped. The column was incubated at 4 or 23°C for 16–40 h. Fractions (3 ml each) containing the target protein were obtained by eluting the column with the column buffer. After elution, 50 ml of 2% SDS in the column buffer was passed through the column at room temperature to strip the remaining bound proteins, and fractions (4 ml each) were collected. The samples were analyzed by SDS–PAGE using 12% Tris-glycine gels (Novex, San Diego, CA) (Fig. 3). The target genes for T4 DNA ligase, *Fse*I, GFP, CamKII and PP1 (Table 2) were also cloned into pCYB vectors (5) and the expression and purification were conducted following essentially the same protocol as described above. Protein concentrations were estimated by the method of Bradford (16).

Construction of MYT4 mutants for the study of the effect of residue 455 on the intein C-terminal cleavage

Unique *Bam*HI and *Age*I sites flanking the C-terminus of the intein in pMYT4 (6) allowed convenient substitutions of His453 and Cys455 through linker insertion. pMYT4 was digested with *Bam*HI and *Age*I and then ligated with the complementary oligomers, 5′-GATCCCAGTTGTAGTACAAACNNNGGTG-GCCTGA-3′ and 5′-CCGGTCAGCCACCNNNGTTCTGTACT-ACAACCTGG-3′, that encoded each of the 20 amino acids (NNN) at position 455 as well as a His453Gln substitution. The resulting

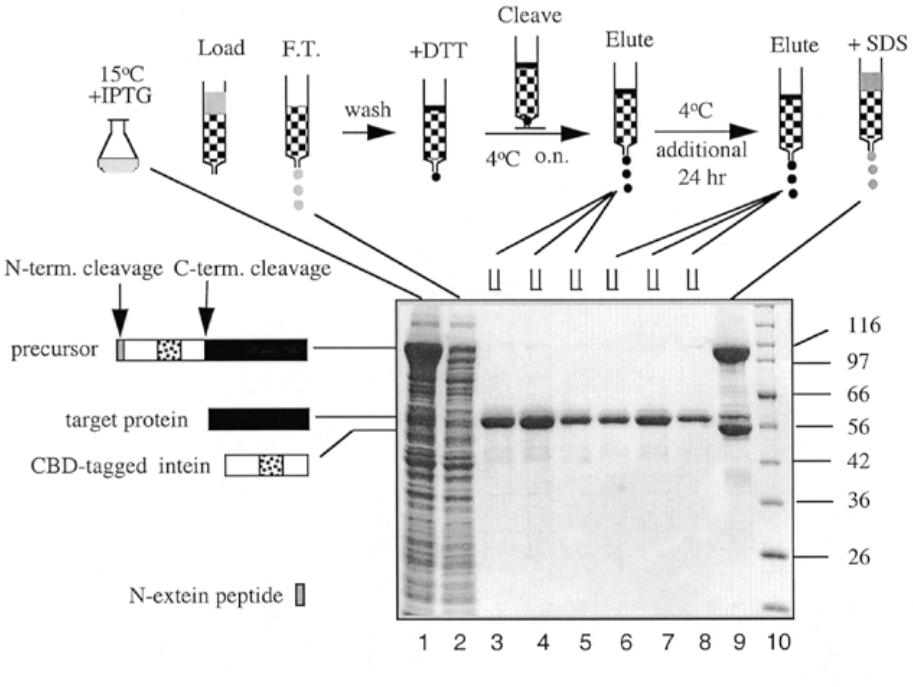

Figure 3. Expression and purification of T4 DNA ligase using a pIMC vector. Crude cell extract from 1 l culture (lane 1) was loaded onto a chitin column. The expressed precursor protein bound to the chitin resin as seen in the flow-through (F.T., lane 2). After thoroughly washing the column, 50 ml column buffer containing 50 mM DTT were passed through the column. The cleavage reaction was then allowed to occur at 4 °C overnight. The first three fractions from the elution are shown in lanes 3, 4 and 5, respectively. After the first elution, the column flow was stopped and the cleavage reaction was allowed to continue for additional 24 h. The first three fractions from the second elution are shown in lanes 6, 7 and 8, respectively. The 2% SDS elution is shown in lane 9. Protein molecular weight standards (kDa, lane 10) are indicated on the right. Schematic representations of the structures of the fusion precursor and cleavage products, using the same shading scheme as Figure 1B, are shown on the left. The N-terminal cleavage product (N-extein peptide, ~1 kDa) is not detectable by SDS–PAGE.

constructs were named pMYT4 (H453Q/C455X), in which X refers to any one of the 20 amino acid residues. After protein expression and purification from each construct following a previously described protocol (6), the fusion proteins (0.5–1.0 mg/ml) were subjected to DTT-induced cleavage. The reactions were conducted at three different temperatures, 4, 16 and 23 °C for up to 40 h. The rest of the cleavage reaction conditions were essentially the same as for the fusion protein from pMYT4 (H453Q/C455A) (6). The cleavage products were resolved on SDS–PAGE followed by staining with Coomassie Blue. The stained gels were digitized with a Microtec Scanmaker 600 ZS and the scanned images were analyzed with NIH Image 1.47 software. The percentage of cleavage was determined by comparing cleavage of the precursor from the DTT-treated samples with that from samples without DTT treatment.

RESULTS AND DISCUSSION

Modification of the intein and the N-terminal fusion protein sequence to allow expression and affinity purification of target proteins

The wild-type *Sce* VMA intein catalyzes efficient protein splicing reactions *in vivo* resulting in rapid peptide bond cleavage and ligation at both termini of the intein (3,17). In order for the intein to be useful for protein purification, the cleavage activity of the intein has to be attenuated *in vivo* but remain inducible *in vitro*. Previously, we have investigated the modulation of protein splicing reactions of the *Sce* VMA intein and found that a double substitution, His453Gln/Cys455Ala, allowed inducible cleavage at the intein C-terminus when the N-terminal cleavage was induced by thiols (Fig. 1A) (6). In this study, we used this intein mutation as the basis for constructing a protein expression and purification system. To allow affinity purification, the CBD from *B.circulans* was inserted in the intein at a site previously identified as protease-sensitive (data not shown). The insertion site was confirmed by the crystal structure data (18) to be in a loop region of the endonuclease domain. The intein containing the CBD underwent efficient splicing and cleavage reactions similar to the intein without the insertion (data not shown).

It is known that many important factors that affect the expression level of a protein are located around the 5′ region of an open reading frame (19), e.g. the promoter region, the 5′ untranslated region (UTR), the Shine–Dalgarno sequence and the translational initiation region including the downstream box (20). Fusion of a target protein to the C-terminus of the intein allows the modification of the N-terminal sequence of the fusion protein to improve protein expression. Here we found that the first 10 residues of MBP as the N-terminal sequence of the fusion protein (or N-extein, Fig. 1B) was sufficient for high level expression of many target proteins (Table 1). For instance, 8.4 mg/l culture of T4 DNA ligase was obtained from the pCT4 construct (Table 1). When the N-terminal MBP sequence was deleted from pCT4, the yield was <1.0 mg (data not shown). It is likely that sequences from other proteins (e.g. GST) that are known to have a high

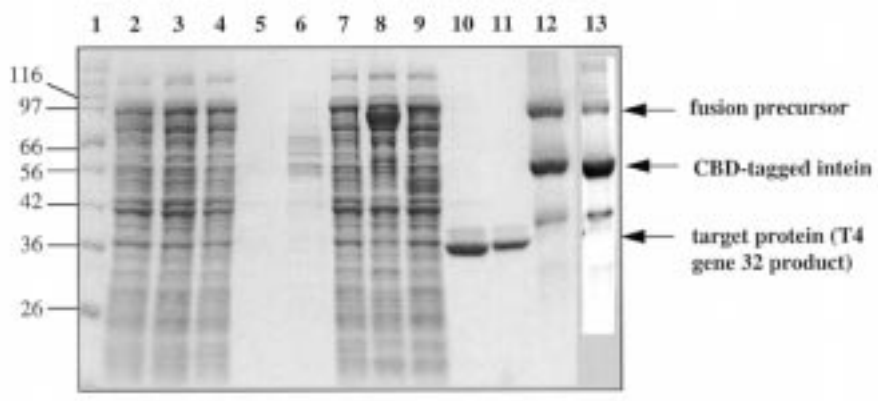

Figure 4. SDS–PAGE gel showing the expression and purification of T4 gene 32 product fused to either N-terminus (in pCYB, lanes 2–6) or C-terminus (in pIMC, lanes 7–13) of the modified *Sce* VMA intein. Lane 1, protein molecular weight standards (kDa); lanes 2–6, T4 gene 32 product expressed in pCYB. Lane 2, uninduced cell extract; lane 3, induced cell extract; lane 4, flow through; lane 5, a fraction of the elution after the 4°C overnight cleavage reaction in the presence of DTT; lane 6, a fraction from the SDS elution. Lanes 7–13, T4 gene 32 product expressed in pIMC. Lane 7, uninduced cell extract; lane 8, induced cell extract; lane 9, flow through; lanes 10 and 11, fractions of the elution after the 4°C overnight cleavage reaction in the presence of DTT; lane 12, a fraction from the SDS elution of the resin after the 4°C overnight cleavage reaction; lane 13, a fraction from the SDS elution of the resin after the 23°C overnight cleavage reaction.

expression level in a particular host can also be used as the N-terminal sequence (N-extein) of the fusion protein.

Table 1. Recombinant proteins expressed and purified in pIMC and pCYB vectors

Target proteins[a]	Yields in pIMC (mg/l culture)	Yields in pCYB (mg/l culture)	Activities[b]
T4 DNA ligase	8.4	8.0	+
T4 gene 32 product	6.0	<0.1	+
*Fse*I	2.0	<0.5	+
GFP	1.9	<0.5	+
PPI	1.9	0.6	+
CamKII	2.2	0.8	n.d.
Invertase	1.7	n.d.	n.d.
T4 endo VII	4.6	n.d.	n.d.

[a]Abbreviations are described in Materials and Methods.
[b]The specific activity of the purified T4 DNA ligase was determined to be within 90% of the specific activity of the enzyme purified by the conventional method. The purified T4 gene 32 product showed specific binding of the single-stranded substrate. The purified *Fse*I showed activity similar to the conventionally purified enzyme. The purified GFP displayed UV light-induced fluorescence. PP1 activity was assayed on Myelin Basic Protein using the PSP Assay system (10; T.Barshevsky, unpublished data).

Using the pIMC vectors to express and purify target proteins

The pIMC vectors are *E.coli* expression vectors which have a *Col*E1 replication origin, an ampicillin resistance gene and an IPTG-inducible *tac* promoter. Two multiple cloning sites, MCS1 and MCS2, were constructed (Fig. 2). Using the *Sap*I site (MCS1, Fig. 2) for cloning a target protein allows fusion of the N-terminus of the target protein to the intein C-terminal cleavage site (Asn 454, codon AAC, Fig. 2), and consequently, the target protein retains its native N-terminus after the cleavage. Cloning a target protein into *Bsm*I or *Spe*I sites in MCS1 (Fig. 2), on the other hand, results in an extra Ala or Ala-Gly-Thr-Ser sequence at the N-terminus of the target protein, respectively. The extra Ala or Ala-Gly-Thr-Ser sequence allows more efficient cleavage at the intein C-terminus (discussed below). Several proteins have been successfully expressed and purified in pIMC vectors and most retained their activities (Table 1). The purification procedures are shown in Figure 3 and are essentially the same as those for the pCYB vectors (5). As the cleavage at the intein N-terminus in pIMC vectors is required for the cleavage at the intein C-terminus, the small N-extein peptide (containing the first 10 residues of MBP) co-eluted with the target protein (Fig. 3, the N-extein peptide is not detectable on SDS–PAGE). The N-extein peptide (~1 kDa) was normally separated from the target protein by dialysis (data not shown). The presence of the N-extein peptide was confirmed by HPLC and mass spectrometric analyses (data not shown). Due to the specificity of the intein cleavage reaction and high affinity of CBD for chitin resin, pIMC vectors usually resulted in highly purified target proteins. However, small amounts of other proteins were sometimes co-purified (e.g. some smaller molecular weight bands in Fig. 3, lanes 3 and 4, and higher molecular weight bands in Fig. 4, lane 10). These impurities were probably the result of insufficient washing of the column, protease degradation or unspecific binding to chitin resin and/or the target protein. Expression of a target protein in a different *E.coli* strain (e.g. a protease-deficient strain) sometimes improves the purity of the target protein (data not shown).

The level of protein expression in pIMC vectors could be affected by induction temperature. Low induction temperatures (15–20°C) sometimes increased expression of soluble fusion precursors and final yields of the target proteins (e.g. T4 gene 32 product, *Fse*I, GFP); however, in other cases (e.g. MBP, T4 DNA ligase), the induction temperature made no difference (data not shown). Thus, different induction temperatures should be tested for each target protein in order to achieve optimal protein expression. The yields of most target proteins can also be elevated by increasing the duration and/or temperature of the on-column cleavage reaction. For instance, after induction of on-column

cleavage at 4°C overnight, 5.0 mg of T4 DNA ligase was eluted (Fig. 3). Incubation of the column for an additional 24 h resulted in further C-terminal cleavage and elution of T4 DNA ligase (3.4 mg). As indicated by the SDS elution (Fig. 3, lane 9), there was a significant amount of the precursor that remained bound to the resin. Increasing the temperature of the cleavage reaction decreased the amount of the uncleaved precursor resulting in a higher yield of T4 DNA ligase (data not shown). In the case of the T4 gene 32 product, induction of on-column cleavage at 4°C resulted in 6.0 mg of the protein (Fig. 4); however, 14 mg of T4 gene 32 product was obtained at 23°C. The increase in the yield was due to a higher efficiency of the cleavage reaction at higher temperatures as suggested by a decrease in the amount of the fusion precursor at 23 versus 4°C (Fig. 4, lanes 12 and 13). We have shown that pH also affects the cleavage at the intein C-terminus (6). The on-column cleavage reaction was usually conducted at pH 8.0–8.5. Lowering the pH to 6.0 inhibited the cleavage reaction and the final yield of a target protein.

Effect of the N-terminal residue of a target protein on the intein C-terminal cleavage

Cloning a target protein in the pIMC vectors using *Sap*I in the multiple cloning site (Fig. 2) places the N-terminal residue of the target protein immediately adjacent to the intein C-terminal cleavage site (Asn 454). To investigate the effect of the N-terminal residue of a target protein on the intein C-terminal cleavage, we substituted the first C-extein residue Cys455 in MYT4 with all 19 amino acids. The MYT4 sytem allows quantification of both N-terminal and C-terminal cleavage products by SDS–PAGE (6). As shown in Table 2, most of the 20 amino acid residues at position 455 allowed >50% cleavage at 16°C after 40 h, or 23°C after 16 h. The cleavage was most efficient at 4°C for Met, Ala and Gln, but less efficient for other residues, especially Val, Ile, Asp, Glu, Lys, Arg and His (Table 2). Pro inhibited the intein C-terminal cleavage, whereas Ser and Cys resulted in only the splicing product (i.e. ligation of T4 and M, data not shown). Some splicing products were also observed in the case of Thr (data not shown).

It is conceivable that the N-terminal structure of a target protein also affects the cleavage at the intein C-terminus. For some target proteins, adding a few extra amino acid residues (e.g. cloning into *Bsm*I or *Spe*I sites in MCS1, Fig. 2) at their N-termini may increase the cleavage efficiency and thus the yield. For instance, we have obtained 6.0 mg/l culture of T4 gene 32 product with two amino acid residues, Ala-Gly, added to its N-terminus (Table 1). Removing these two residues resulted in 3-fold decrease in the yield of the purified T4 gene 32 product (data not shown). Examination of the SDS elution fractions suggested that the decrease in the amount of the purified T4 gene 32 product was caused by a decrease in the cleavage efficiency of the fusion precursor (data not shown).

Comparison of target proteins as an N-terminal fusion versus C-terminal fusion to the modified intein

A number of target proteins have been cloned in both pIMC and pCYB vectors and the expression and purification were conducted under the same conditions. For many target proteins, fusion to the C-terminus of the intein (in pIMC vectors) resulted in higher yields than fusion to the N-terminus of the intein (in pCYB vectors) (Table 2). For instance, ~2 mg/l culture of *Fse*I and GFP were obtained in pIMC vectors compared to <0.5 mg/l culture in pCYB vectors (Table 2). For CamKII, the yield was 0.8 mg/l culture in pCYB and 2.2 mg/l culture in pIMC. Similarly, PP1 resulted in only 0.4 mg/l culture in pCYB but 1.9 mg/l culture in pIMC. In all cases, the higher yields were due to a higher level expression of the fusion precursors in the pIMC vectors. The expression and purification of T4 gene 32 product in both pCYB and pIMC vectors are shown in Figure 4. The fusion precursor can be easily identified in the crude extract of the induced cells of the pIMC expression (Fig. 4, lanes 7 and 8) but not of the pCYB expression (Fig. 4, lanes 2 and 3). As a result, 6.0 mg/l culture of T4 gene 32 product was obtained from the pIMC vector but <0.1 mg/l culture was obtained from the pCYB construct (not detectable by SDS–PAGE, Fig. 4, lane 5). The SDS elution fraction (Fig. 4, lane 6) suggests that only trace amounts of the fusion protein were produced from the pCYB construct. It is not clear why no significant amount of the fusion precursor was expressed when T4 gene 32 product was fused to the N-terminus of the intein. One explanation could be that proper folding of the fusion precursor was only possible when the N-terminus of T4 gene 32 product was fused to the intein. Not all target proteins exhibited significant differences in expression when fused to different termini of the intein. For T4 DNA ligase, MBP, expression in either the pCYB or pIMC vector resulted in similar yields (data not shown).

Table 2. Effect of the first C-extein residue (the first N-terminal reside of a target protein) on DTT-induced cleavage at the C-terminus of the *Sce* VMA intein

	cleavage $t_{1/2}$ (hr)*			% cleavage after 16 hr*			% cleavage after 40 hr*		
the first C-extein Residue	4°C	16°C	23°C	4°C	16°C	23°C	4°C	16°C	23°C
Met Ala Gln	**14-24**	5-10	2-6	40-60	>80	>95	60-90	>90	>95
Gly Leu Asn Trp Phe Tyr	30-60	10-20	<10	10-40	50-80	75-95	40-60	>90	>90
Val Ile Asp Glu Lys Arg His	>60	20-30	10-20	<10	30-50	50-80	10-20	70-90	70-95
Pro	n.d.	n.d.	n.d.	<10	<10	<10	<10	<10	<10
Thr Ser Cys	100 n.d. n.d.	20 n.d. n.d.	10 n.d. n.d.	7 n.d. n.d.	40 n.d. n.d.	80 n.d. n.d.	20 n.d. n.d.	80 n.d. n.d.	>90 n.d. n.d.

*Cleavage reactions were conducted by treatment of the purified MYT4 fusion proteins with 40 mM DTT in 30 mM HEPES, pH 8.0, 0.5 M NaCl for 16 or 40 h at 4, 16 and 23°C. The percentage of cleavage was determined by comparing the MYT4 precursors from the DTT-treated samples with those from the samples without the DTT-treatment in scanned images of Coomassie Blue stained SDS–PAGE gels. The half time (h) of the cleavage reaction (cleavage $t_{1/2}$) is the incubation time required for DTT to cleave 50% of the MYT4 precursors.
n.d., not determined.

There are other advantages of using the C-terminal fusion (pIMC) system. For instance, the system allows the purification of a target protein without an N-terminal methionine. Most of the 20 amino acids (with the exception of Ser, Cys and Pro, Table 2) can be placed at the N-terminus of a target protein. We have shown that free cysteine, instead of reducing agent DTT, can induce peptide bond cleavage at both termini of the intein (3,6). In the N-terminal fusion (pCYB) system, the cysteine-induced cleavage results in a modified target protein as cysteine is covalently attached to the target protein (5). In the C-terminal fusion (pIMC) system, on the other hand, the target protein is unmodifed since cysteine is attached to the N-extein sequence (Fig. 1A).

Conclusions

In this paper, we describe a unique protein expression and purification system which utilizes a modified *Sce* VMA intein capable of inducible peptide bond cleavage at its C-terminus. An affinity tag, the CBD, was inserted in the intein without affecting the cleavage activities and a small N-extein sequence consisting of the first 10 residues of MBP allowed a favorable translational start for protein expression. The intein-catalyzed peptide bond cleavage eliminates the need for a protease normally required by other affinity fusion systems such as MBP, GST or His-tag systems. The cleavage at the C-terminus of the intein is mostly affected by three factors: pH, temperature and the first residue of the target protein. Fusions to the C-terminus of the intein can increase the yield of some target proteins compared to the N-terminal fusion, and in some cases, allow the purification of target proteins that are not feasible as N-terminal fusions due to either protein misfolding or unfavorable C-terminal residues (e.g. Asp; 5). In addition, the C-terminal fusion opens the way for varying the N-terminal sequence of the fusion protein according to the expression hosts thereby allowing the sequence from a known highly expressed protein to be chosen as the N-extein sequence. This facilitates the application of the C-terminal cleavage system to other expression hosts such as yeast, insect cells and mammalian cells.

ACKNOWLEDGEMENTS

We would like to thank Donald G. Comb, Richard Roberts, Francine Perler, Maurice Southworth and Fana B. Mersha for valuable discussions and reading of the manuscript, and Tanya Barshevsky, Melissa Scott and Richard D. Whitaker for technical assistance. This work is supported by NIH grant GM 57734-01 and New England Biolabs, Inc.

REFERENCES

1 LaVallie,E.R. and McCoy,J.M. (1995) *Curr. Opin. Biotechnol.*, **6**, 501–506.
2 Perler,F.B., Davis,E.O., Dean,G.E., Gimble,F.S., Jack,W.E., Neff,N., Noren,C.J., Thorner,J. and Belfort,M. (1994) *Nucleic Acids Res.*, **22**, 1125–1127.
3 Chong,S., Yang,S., Paulus,H., Benner,J., Perler,F.,B. and Xu,M.-Q. (1996) *J. Biol. Chem.*, **271**, 22159–22168.
4 Xu,M.-Q. and Perler,F.B. (1996) *EMBO J.*, **15**, 5146–5153.
5 Chong,S., Mersha,F.B., Comb,D.G., Scott,M.E., Landry,D., Vence,L.M., Perler,F.B., Benner,J., Kucera,R.B., Hirvonen,C.A., Pelletier,J.J., Paulus,H. and Xu,M.-Q. (1997) *Gene*, **192**, 271–281.
6 Chong,S., Williams,K.S., Wotkowicz,C. and Xu,M.-Q. (1998) *J. Biol. Chem.*, **273**, 10567–10577.
7 Kunkel,T.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 488–492.
8 Krisch,H.M. and Allet,B. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 4937–4941.
9 Tomaschewski,J. and Ruger,W. (1987) *Nucleic Acids Res.*, **15**, 3632–3633.
10 Zhang,A.J., Bai,G., Deans-Zirettu,S., Browner,M.F. and Lee,E.Y. (1992) *J. Biol. Chem.*, **267**, 1484–1490.
11 Nelson,J.M., Miceli,S.M., Lechevalier,M.P. and Roberts,R.J. (1990) *Nucleic Acids Res.*, **18**, 2061–2064.
12 Shimomura,O., Johnson,F.H. and Saiga,Y. (1962) *J. Cell. Comp. Physiol.*, **59**, 223–227.
13 Lin,C.R., Kapiloff,M.S., Durgerian,S., Tatemoto,K., Russo,A.F., Hanson,P., Schulman,H. and Rosenfield,M.G. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 5962–5966.
14 Taussig,R. and Carlson,M. (1983) *Nucleic Acids Res.*, **11**, 1943–1954.
15 Armstrong,A., Brown,R.S. and Tsugita,A. (1983) *Nucleic Acids Res.*, **11**, 7145–7156.
16 Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
17 Cooper,A.A., Chen,Y., Lindorfer,M.A. and Stevens,T.H. (1993) *EMBO J.*, **12**, 2575–2583.
18 Duan,X., Gimble,F.S. and Quiocho,F.A. (1997) *Cell*, **89**, 555–564.
19 Makrides,S.C. (1996) *Microbiol. Rev.*, **60**, 512–538.
20 Sprengart,M.L., Fuchs,E. and Porter,A.G. (1996) *EMBO J.*, **15**, 665–674.