# IMPACT<sup>™</sup>-CN

# Protein Purification System Now Featuring Fusion to C- or N- Terminus of the Target Protein

For additional information, including vector sequences and frequently asked questions, visit: www.neb.com

#### Instruction Manual

Catalog #E6900S Store at –20°C Note: Store chitin beads at 4°C



Version 1.9 7/01

# Table of Contents:

Introduction	3
System Components	7
Overview of the Expression and Purification Protocol	13
Cloning the Target Gene	15
Cell Culture and Fusion Protein Expression	24
Affinity Purification and On-column Cleavage	25
Simplified Purification Protocol for Test Experiment	29
Media and Solutions	31
Frequently Asked Questions	32
References	38
IMPACT Usage Publications	39
Appendices	
Appendix I: Comparison Between C-terminal and N-terminal Fusions	42 44 46
Appendix V: Research Use Assurance Statement	
Kit Components Sold Separately/Companion Products	

Notice to Buyer/User: The buyer and user have a non-exclusive license to use this system or any components thereof for RESEARCH PURPOSES ONLY. See RESEARCH USE ASSURANCE STATEMENT attached hereto as Appendix V for detail on terms of the license granted hereunder.

Information presented herein is accurate and reliable to the best of our knowledge and belief, but is not guaranteed to be so. Nothing herein is to be construed as recommending any practice or any product in violation of any patent or violation of any law or regulation. It is the user's responsibility to determine for himself or herself the suitability of any material and/or procedure for a specific purpose and to adopt such safety precautions as may be necessary.

1

# The System Includes:

- Vector DNA (4 vectors) 10 μg of each (50 μl)
- Sequencing Primers (4 primers) 200 picomoles of each
- pMYB5 Control Plasmid 10 µg
- E. coli Strain ER2566 0.2 ml stock cells
- Anti-Chitin Binding Domain Serum (rabbit) 50 بلا
- Chitin Beads (store at 4°C) 20 ml
- 1, 4-Dithiothreitol (DTT), 1M 5 ml
- 3X SDS Sample Buffer 1 ml
- Instruction Manual

IMPACT<sup>™</sup>-CN System components are described in detail on page 7. A list of components sold separately and companion products can be found on page 48.

# Introduction:

IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) is a novel protein purification system which utilizes the inducible self-cleavage activity of a protein splicing element (termed intein) to separate the target protein from the affinity tag. It distinguishes itself from all other purification systems by its ability to purify, in a single chromagraphic step, a native recombinant protein without the use of a protease. The concept of IMPACT purification has evolved from our studies of the

protease. The concept of MirACT purification protein splicing mechanism [Appendix II & IV (1–6)]. The IMPACT-CN System utilizes an intein (454 amino acid residues) from the *Saccharomyces cerevisiae* VMA1 gene. A target protein is fused to a self-cleavable intein tag in which a chitin binding domain allows affinity purification of the fusion precursor on a chitin column. In the presence of thiols such as DTT,  $\beta$ mercaptoethanol or cysteine, the intein undergoes specific self-cleavage which releases the target protein from the chitinbound intein tag resulting in a singlecolumn purification of the target protein (Figure 1, 2).

The IMPACT-CN System contains expression vectors (pTYB vectors) which allow the fusion of the cleavable intein tag to either the C-terminus (pTYB1 and pTYB2, C-terminal fusion) or N-terminus (pTYB11 and pTYB12, N-terminal fusion) of a target





protein. This flexibility in fusion protein construction maximizes the probability of successful expression and purification of a target protein. To allow the cloning of the same amplified target gene in either fusion construction, the same or compatible restriction sites are designed in the multiple cloning region of pTYB2 and pTYB12 vectors. pTYB1 and pTYB11 vectors, on the other hand, allow the cloning of a target gene immediately adjacent to the intein cleavage site. This results in the purification of a native target protein without any vector-derived extra residues after the cleavage. A detailed comparison between the N-terminal and C-terminal fusion with the intein tag is described in Appendix I. The pTYB vectors use a T7 promoter-driven system (11) to achieve high levels of expression and tight transcriptional control in *E. coli*.

The IMPACT-CN System provides 4 major advances over other expression systems: (i) rapid and simple purification of a native protein without extra vector-derived residues; (ii) high affinity of the chitin binding domain to allow more stringent washing conditions and reduce non-specific binding; (iii) easy separation of the target protein from the affinity tag without the use of a separate, expensive protease; and (iv) the only protein expression and purification system to allow specific C-terminal labeling of a target protein and the Intein-mediated Protein Ligation (IPL) (see Appendix IV) (8,9).

 Table 1: Recombinant proteins expressed and purified by the IMPACT-CN System.

IMPACT-CN Vectors	Target Protein	Yield (mg/L culture)
	Maltose-Binding Protein	20
	McrB	6.0
	T4 DNA Ligase	5.0
C-terminal Fusion	Bst DNA Polymerase Large Fragment	2.0
	BamH I	1.0
	Bgl II	4.0
	CDK2	1.0
	<i>Cam</i> K II	0.8
	T4 DNA Ligase	8.4
	T4 Gene 32 Product	6.0
	Fse I	2.0
N-terminal Fusion	GFP	1.9
	<i>Cam</i> K II	2.2
	Invertase	1.7
	T4 Endo VII	4.6





Lane 1: Protein Marker, Broad Range (NEB #P7702).

Lane 2: Crude extract from uninduced cells.

Lane 3: Crude extract from cells, induced at 15°C for 16 hours.

Lane 4: Clarified crude extract from induced cells.

Lane 5: Chitin column flow through (F.T.).

Lane 6: Chitin column wash

Lane 7: Quick DTT wash to distribute DTT evenly throughout the chitin column.

Lanes 8-9: Fractions of eluted MBP after stopping column flow and inducing a self-cleavage reaction at 4°C overnight.

Lane 10: SDS stripping of remaining proteins bound to chitin column (mostly the cleaved intein-CBD fusion).

# System Components:

## **Cloning Vector DNA**

The pTYB vectors are used for cloning and expression of recombinant proteins in *E. coli* (Figure 2). pTYB1 (7,477 bp) and pTYB2 (7,474 bp) are C-terminal fusion vectors in which the C-terminus of the target protein is fused to the intein tag. pTYB11 and pTYB12 are N-terminal fusion vectors in which the N-terminus of the target protein is fused to the intein tag; when pTYB11 (7,414 bp) and pTYB12 (7,417 bp) are used, a small peptide (15 amino acid residues, 1.6 kDa) is also cleaved from the intein tag and co-eluted with the target protein. It cannot be detected on a regular SDS-PAGE gel but can be separated from the target protein by dialysis. pTYB1 and pTYB12 use ATG of the *Nde* I site in the multiple cloning region for translation initiation. For both pTYB1 and pTYB11 vectors only the *Sap* I site should be used to clone the 3' and 5' end, respectively, of the target gene. This strategy will result in the fusion of the target gene adjacent to the intein tag (and the cleavage site). The target protein can be purified without any extra non-native residues. The use of the *Sap* I site also allows for the addition of amino acid residues favorable for cleavage (by engineering them into the coding region of the primers). pTYB2 and pTYB12 contain the same restriction sites (except *Nhe* I and *Spe* I, which are compatible) in the multiple cloning region. This allows fusion of the intein tag to either termini of the same amplified target gene.

Use of pTYB2 or pTYB12 yields a target protein with extra residue(s) added to its C-terminus or Nterminus, respectively, after the cleavage of the intein tag. For example, cloning the 3' end of a target gene using the *Sma* I site in pTYB2 adds an extra glycine residue to the C-terminus of the target protein. Likewise, cloning the 5' end of a target gene using the *Nde* I site in pTYB12 adds an extra three residues (Ala-Gly-His) to the N-terminus the target protein. For details, see **Cloning the Target Gene** on page 15.

The pTYB vectors use a T7/*lac* promoter to provide stringent control of the fusion gene expression. The vectors carry their own copy of the *lac* I gene encoding the *lac* repressor. Binding of the *lac* repressor to the *lac* operator sequence immediately downstream of the T7 promoter suppresses basal expression of the fusion gene in the absence of IPTG induction. The vectors also contain the origin of DNA replication from bacteriophage M13, which allows for the production of single-stranded DNA by helper phage (M13K07 Helper Phage, NEB #N0315S) superinfection of cells bearing the plasmid. pTYB vectors carry the Amp' gene marker (the *bla* gene), which conveys ampicillin resistance to the host strain.

pTYB	1				T7 Universal	Primer $\rightarrow$			
5´	.CGG GGA	TCT CGA	TCC CGC	GAA AT <u>T</u>	AAT ACG ACT				AGC
					T7 Promo			<i>ac</i> operator	
	<u>GGA TAA</u>	CAA TTC	<u></u> <i>Xb</i>		AAT TTT GT	TAA CTT		<u>GGA G</u> AT Dalgarno	ATA _Intein →
					Gly Gly Arg				
					G <u>GC GGC CGC</u>				<u>IC</u> L IGL
	Nde I	Nhe I	Nru I	Sal I	Not	<i>Eco</i> R I	Xho I	Sap I	
	TTT GCC			GTT TTA	ATG GCG GA	GGG TCT	ATT GAA	TGT ATT	
		K	'pn l						
	GAA AAC				GTC ATG GG	3´			
			$\leftarrow$ Intein F	leverse Seq	uencing Primer				
pTYB:		112 500	Con Ing	Val Aco	Gly Gly Arc	Clu Dho	Lou Clu		Intein→
					GGC GGC CGC				
	Ndel	Nhel	Nrul	Sall	Not I	EcoR I	Xho I	Sma I	Tuc
*pTYI	811				←Intein				
Inteir	n Forward I			Va	1 Gln Asn A				
Inteir	n Forward I			Va					
Inteir (117 Gly	<b>Forward  </b> bp) 5′ Gly Arg (	GGA TC Glu Phe	C CAG GT Leu Glu	Va T GTT GTA Pro Gly	1 G1n Asn An A CAG AAC A <u>G</u>	A AGA GC <sup>®</sup> <i>Sap</i> I	T <u>ACT AG</u> Spe I	<u>TCG CGA</u> <i>Nru</i> I	GTC GAC Sal I
Inteir (117 Gly G <u>GC</u>	<b>Forward  </b> bp) 5′ Gly Arg (	GGA TC Glu Phe <u>GAA TTC</u>	C CAG GT Leu Glu <u>CTC GAG</u>	Va T GTT GTA Pro Gly	1 G1n Asn An A CAG AAC A <u>G</u>	A AGA GC <sup>®</sup> <i>Sap</i> I	T <u>ACT AG</u> Spe I	<u>TCG CGA</u> <i>Nru</i> I	GTC GAC
Inteir (117 Gly G <u>GC</u> Not	(Forward I bp) 5 <sup>-</sup> Gly Arg <u>GGC CGC</u> ( <i>Eag</i> I) a: Use the	GGA TC Glu Phe <u>GAA TTC</u> <i>Eco</i> RI	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone 1	Va T GTT GTA Pro Gly <u>CCC GGG</u> <i>Sma</i> I the 5'end o	I GIN ASN AI A CAG AAC A <u>G</u> TGA <u>CTG CAG</u> <i>Pst</i> I	<u>A AGA GC</u> <i>Sap</i> I 3´ (58	T <u>ACT AG</u> <i>Spe</i> ∣ bp)←T7	T <u>CGCGA</u> <i>Nru</i> I Terminator	GTC GAC Sal I
Inteir (117 Gly G <u>GC</u> Not *Note	Gly Arg ( GGC CGC ( ( <i>Cag</i> I) CE Use the the 3'en	GGA TC Glu Phe <u>GAA TTC</u> <i>Eco</i> RI <i>Sap</i> Isite	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone 1	Va T GTT GTA Pro Gly <u>CCC GGG</u> <i>Sma</i> I the 5'end o	I GIN ASN AI A CAG AAC A <u>G</u> TGA <u>CTG CAG</u> <i>Pst</i> I	<u>A AGA GC</u> <i>Sap</i> I 3´ (58	T <u>ACT AG</u> <i>Spe</i> ∣ bp)←T7	T <u>CGCGA</u> <i>Nru</i> I Terminator	<u>GTC_GAC</u> Sal I r Reverse Primer
Inteir (117 Gly G <u>GC</u> Not *Note	Gly Arg ( GGC CGC ) ( <i>Eag</i> I) : Use the the 3'en	GGA TC Glu Phe <u>GAA TTC</u> <i>Eco</i> RI <i>Sap</i> Isite Id of the t	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone <sup>-</sup> arget gen	Va T GTT GTA Pro Gly <u>CCC GGG</u> <i>Sma</i> I the 5'end o	I GIN ASN AI A CAG AAC A <u>G</u> TGA <u>CTG CAG</u> <i>Pst</i> I	<u>A AGA GC</u> <i>Sap</i> I 3´ (58	T <u>ACT AG</u> <i>Spe</i> ∣ bp)←T7	T <u>CGCGA</u> <i>Nru</i> I Terminator	<u>GTC_GAC</u> Sal I r Reverse Primer
Inteir (117 G1y G <u>GC</u> Not *Note PTYB: Inteir	Gly Arg ( GGC CGC ( ( <i>Eag</i> I) CE Use the the 3'en Corward I	GGA TCG Glu Phe GAA TTC EcoRI SapIsite Id of the to Primer →	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone s arget gene	Va T GTT GTA Pro Gly <u>CCC GGG</u> Smal the 5'end c 3.	I GIN ASN AI A CAG AAC A <u>G</u> TGA <u>CTG CAG</u> . <i>Pst</i> I of the target g ←Intein ↓ I GIN ASN AI	<u>A AGA GC</u> <i>Sap</i> I 3 <sup>*</sup> (58 ene. The o a Gly Hi:	T <u>ACT AG</u> Spel bp) $\leftarrow$ T7 ther sites a s Met Thu	<u>TCG CGA</u> Nrul Terminator are used O	GTC GAC Sal I r Reverse Primer NLY for cloning Arg
Inteir (117 G1y G <u>GC</u> Not *Note PTYB: Inteir	Gly Arg ( GGC CGC ( ( <i>Eag</i> I) CE Use the the 3'en Corward I	GGA TCG Glu Phe GAA TTC EcoRI SapIsite Id of the to Primer →	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone s arget gene	Va T GTT GTA Pro Gly <u>CCC GGG</u> Smal the 5'end c 3.	a Gin Asn Ai A CAG AAC A <u>G</u> TGA <u>CTG CAG</u> <i>Pst</i> I Dof the target g ←Intein $\checkmark$	<u>A AGA GC</u> <b>Sap I</b> 3 <sup>*</sup> (58 ene. The o a Gly Hi: I GGT <u>CA</u>	$f ACT AGT Spe I bp) \leftarrow T7 ther sites a s Met Thu LATG ACT$	<u>TCG CGA</u> Nru I Terminator are used O Ser Ser LAGT ICG	GTC GAC Sall r Reverse Primer NLY for cloning
Inteir (117 G1y G <u>GC</u> Not <b>*Note</b> <b>*Note</b> Inteir (117 Val	Forward I bp) 5 <sup>°</sup> Gly Arg ( <u>GGC CGC 1</u> ( <i>Eag</i> I) e: Use the the 3 <sup>°</sup> en forward I bp) 5 <sup>°</sup> Asp Gly (	GGA TCI Glu Phe GAA TTC EcoR I Sap I site d of the t Primer $\rightarrow$ GGA TCI Gly Arq	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone - arget genu C CAG GT Glu Phe	Va T GTT GTA Pro Gly <u>CCC GGG</u> Smal the 5'end G a. Val T GTT GTA	I GIN ASM AI A CAG AAC A <u>G</u> PstI of the target g ← Intein ↓ I GIN ASM AI A CA <u>G AAT GC</u> BSm I Pro GIY	<u>A AGA GC</u> <b>Sap I</b> 3' (58 ene. The o a Gly Hi: <u>I GGT CA</u> A	$f ACT AGTSpe Ibp) \leftarrow T7ther sites as Met ThuLATG ACTide I Sp$	T <u>CG CGA</u> Nru I Terminator are used O are used O Ser Ser LAGT TCG pe I Nra	GTC GAC Sall r Reverse Primer NLY for cloning Arg CGA
Inteir (117 G1y G <u>GC</u> Not <b>*Note</b> <b>*Note</b> Inteir (117 Val	Forward I           bp)         5 <sup>°</sup> Gly Arg (         GGC CGC I           ( <i>Eag</i> I)         3 <sup>°</sup> s: Use the the 3 <sup>°</sup> en         1           the 3 <sup>°</sup> en         1           bp)         5 <sup>°</sup> Asp Gly (         GGC CGC I           ( <i>Eag</i> I)         3 <sup>°</sup> s: Use the 3 <sup>°</sup> en         1           bp)         5 <sup>°</sup> Asp Gly (         GGC CGC I	GGA TCI Glu Phe GAA TTC EcoR I Sap I site d of the t Primer $\rightarrow$ GGA TCI Gly Arq	C CAG GT Leu Glu <u>CTC GAG</u> Xho I to clone - arget genu C CAG GT Glu Phe	Va T GTT GTA Pro Gly <u>CCC GGG</u> <b>Smal</b> the <b>5</b> end G Val T GTT GTA Leu Glu I <u>CTC GAG</u> G	I GIN ASM AI A CAG AAC A <u>G</u> PSTI of the target g ←Intein ↓ I GIN ASN AI A CA <u>G AAT GC</u> BSMI Pro GIy <u>CCC GGG</u> TGA	<u>A AGA GC</u> <b>Sap I</b> 3' (58 ene. The o a Gly Hi: <u>I GGT CA</u> A	$f ACT AGTSpe Ibp) \leftarrow T7ther sites as Met ThuLATG ACTide I Sp$	$\begin{array}{c c} \underline{\Gamma CG CGA} \\ \hline Nru 1 \\ \hline Terminator \\ are used 0 \\ \hline are used 0 \\ \hline are Ser Ser \\ \underline{\Gamma AGT ICG} \\ \underline{F CG} \\$	GTC GAC Sall r Reverse Primer NLY for cloning Arg CGA

Figure 3A: Multiple cloning sites in pTYB vectors. ▼ indicates intein cleavage site.



*Figure 3B:* pTYB1 vector. All unique sites are shown. Coordinates refer to the position of the 5'-base in each recognition sequence.





Additional vectors (pTYB3, pTYB4, pTXB1, pTXB3, pTWIN1, pTWIN2, pKYB1) are also available for cloning. See Frequently Asked Questions (page 32 and web site) and Companion Products (page 48).

# **Sequencing Primers**

Four primers (200 picomoles of each) are included for sequencing the target gene cloned in the multiple cloning region of the pTYB vectors (Figure 3A). The T7 Universal Primer (NEB #S1248S, 5'-TAA TAC GAC TCA CTA TAG GG-3') is complementary to the T7 promoter and yields a sequence in the direction of transcription. The Intein Reverse Primer (NEB #S1261S, 5'-ACC CAT GAC CTT ATT ACC AAC CTC-3') is complementary to the intein sequence 64–87 nucleotides downstream from the intein cleavage site and yields sequences opposite the direction of transcription. The T7 Universal Primer and Intein Reverse Primer are used for sequencing a target gene cloned in the C-terminal fusion vectors (pTYB1 and pTYB2). The Intein Forward Primer (NEB #S1263S, 5'-CCC GCC GCT GTT TTG GCA GGT GAG-3') anneals to the intein sequence 117–141 nucleotides upstream from the intein cleavage site and with the T7 Terminator Reverse Primer (NEB #S1271S, 5'-TAT GCT AGT ATT TGC TCA G-3') can be used for sequencing a target gene cloned in the N-terminal fusion vectors (pTYB11 and pTYB12).

# pMYB5 Control Plasmid

The pTYB vectors are not recommended as controls for testing expression. Accordingly, the kit provides pMYB5 Control Plasmid (8,602 bp), which carries the *E. coli malE* gene, encoding the maltose binding protein (MBP), fused in-frame to the coding region of the intein tag. It should be transformed into *E. coli* strain ER2566 and used to test cell culture, induction and purification procedures. After induction with 0.3 mM IPTG at 30°C for 3 hours (or 15°C for 12–16 hours), 100 ml of cells should yield 2–3 mg of a 97 kDa fusion protein. After chitin column purification and DTT-induced cleavage, about 1.0–1.5 mg of the MBP (42 kDa) is usually obtained.

# E. colistrain ER2566

ER2566 is provided as a host strain for the expression of a target gene cloned in the pTYB vectors. ER2566 cells carry a chromosomal copy of the T7 RNA polymerase gene inserted into the *lac*Z gene, and thus under the control of the *lac* promoter. In the absence of IPTG induction, expression of T7 RNA polymerase is suppressed by the binding of *lac* I repressor to the *lac* promoter. The strain is deficient in both *lon* and *omp*T proteases. ER2566 is supplied as a 0.2 ml 50% glycerol stock; these cells are not competent. Recommended long term storage at  $-70^{\circ}$ C.

ER2566 Genotype:  $F \sim fhuA2$  [lon] ompT lacZ::T7 gene1 gal sulA11  $\Delta$ (mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm].

## **Chitin Beads**

An affinity matrix used for the isolation of the fusion precursor containing the target protein. 20 ml of chitin beads (~ 50–100  $\mu$ m in size) are supplied as a 38 ml slurry in 20% ethanol. The binding capacity, which has been tested using the control vector pMYB5, is 2 mg of eluted MBP protein per ml of chitin beads. Note: Store Chitin Beads at 4°C. Temporary storage at –20°C will not affect the binding capacity.

### Anti-Chitin Binding Domain Serum (anti-CBD)

 $50 \ \mu$  of rabbit serum raised against a peptide derived from the *Bacillus circulans* chitin binding domain is provided for Western blot analysis. A 1:5000 dilution of anti-CBD serum may be used to analyze crude extracts or purified samples.

# 1, 4-Dithiothreitol (DTT)

5.0 ml of a 1.0 M solution is provided in the kit. Store at  $-20^{\circ}$ C. Since DTT is not particularly stable after dilution, Cleavage Buffer should be freshly prepared before use. DTT stock solutions should be aliquoted to minimize freeze/thaw cycles.

### **3X SDS Sample Buffer**

187.5 mM Tris-HCl (pH 6.8 @  $25^{\circ}$ C), 6% (w/v) SDS, 30% glycerol and 0.03% (w/v) bromphenol blue (store at room temperature). DTT should be added to the 3X SDS Sample Buffer, to a final concentration of 40 mM (see note in Table 2A).

# **Overview of the Expression and Purification Protocol:**

- 1. **Cloning:** Choose an appropriate pTYB vector to clone the target gene. Consider: (i) extra residue(s) vs native sequence at the target protein termini; (ii) C-terminal fusion (pTYB1 and pTYB2) vs N-terminal fusion (pTYB11 and pTYB12).
- 2. **Transformation:** Transform the plasmid bearing the target gene into an appropriate *E. coli* host strain that carries the T7 RNA Polymerase gene.
- 3. **Cell Culture:** Grow the cells at 37°C in LB medium containing 100  $\mu$ g/ml ampicillin. When the OD<sub>600</sub> of the culture reaches 0.5–0.8, induce protein expression at 15–20°C with IPTG at a final concentration of 0.3–0.5 mM.
- Column Preparation: Equilibrate a chitin column (20 ml for 1 liter culture) with 10 volumes of Column Buffer [20 mM Hepes or Tris-HCI (pH 6.0–8.5), 500 mM NaCl, 1 mM EDTA].
- 5. **Loading:** Break cells in Column Buffer and slowly load the clarified lysate onto the chitin column.
- 6. **Washing:** Wash the column with at least 10 bed volumes of Column Buffer to thoroughly remove the unbound proteins.
- Adding Thiols: Quickly wash the column with 3 bed volumes of Cleavage Buffer [20 mM Hepes or Tris-HCl (pH 7.5 or above for using pTYB11 or pTYB12; pH 8.0 for using pTYB1 or pTYB2), 500 mM NaCl, 1 mM EDTA] containing 50 mM DTT or cysteine.
- 8. **On-column Cleavage:** Stop the flow and leave the column at 4°C–23°C for 16–40 hours. The temperature and duration of the cleavage reaction are dependent on the on-column cleavage efficiency which can be checked by analyzing samples from the SDS elution of the chitin resin (see next page).

- 9. **Elution:** Elute the target protein by continuing the column flow with Cleavage Buffer without DTT or cysteine.
- 10. **Dialysis:** Dialyze the target protein into an appropriate storage buffer; this also removes the excessive DTT or cysteine used in the Cleavage Buffer and the coeluted small peptide (when using pTYB11 and pTYB12).
- 11. **Cleavage:** To examine cleavage efficiency remove 200  $\mu$ l of chitin resin and mix with 100  $\mu$ l of 3X SDS Sample Buffer. After boiling for 5 minutes, the supernatant is analyzed on SDS-PAGE to determine the cleavage efficiency.
- 12. Regeneration of Chitin Resin: To regenerate chitin resin, wash the column with 3 bed volumes of the 0.3 M NaOH (stripping solution). Allow resin to soak 30 minutes and wash the resin with an additional 7 bed volumes of 0.3 M NaOH. Wash with 20 bed volumes of water, followed by 5 bed volumes of column buffer.

# **Cloning the Target Gene:**

## **Choice of Vectors**

Two factors determine the choice of a pTYB vector for cloning the target gene: (i) C-terminal fusion or N-terminal fusion; (ii) a target protein with a native sequence or with extra vector-derived residues.

pTYB1 and pTYB2 (C-terminal fusion vectors) allow the fusion of the target protein C-terminus to the intein tag, whereas pTYB11 and pTYB12 (N-terminal fusion vectors) allow the fusion of the target protein N-terminus to the intein tag. Our studies suggest that some target proteins can achieve a higher expression level and final yield in either the C-terminal or N-terminal fusion vectors. To facilitate the choice of an appropriate fusion vector, pTYB2 and pTYB12 are designed to have the same restriction sites (except *Nhe* I in pTYB2 and *Spe* I in pTYB12, which are compatible for cloning) in the multiple cloning regions. This allows cloning of the same amplified target gene to both N- and C-terminal fusion vectors with the same restriction site(s) (e.g., *Nde* I and *Sma* I sites).

Cloning a target gene into pTYB2 or pTYB12 results in extra vector-derived residues added to the N- or C-terminus of the target protein, respectively. For instance, cloning the 3' end of a target gene into the *Sma* I site of pTYB2 will add an extra glycine residue to the C-terminus of the target protein as the cleavage occurs at the cysteine residue of the intein tag. Similarly, cloning the 5' end of a target gene into the *Nde* I site of pTYB12 will add an extra three residues (Ala-Gly-His) to the N-terminal methione of the target protein, as the cleavage occurs at the asparagine residue of the intein tag. These extra residues, which are immediately adjacent to the intein cleavage site, are sometimes necessary for efficient cleavage by the intein.

To obtain a target protein without any vector-derived residues, one can use the *Sap* I site to clone the 3' end of a target gene in pTYB1 or 5' end of a target gene in pTYB11.

C-terminal Residue of	In vivo	In vitro Cleavage	with DTT (40 mM)
the Target Protein	Cleavage	4°C	16 <sup>°</sup> C
Gly	_	+++	+++
Ala	-	+++	+++
lle*	-	+	+
Leu*	-	+	+++
Met*	-	+++	+++
Phe*	-	+++	+++
Val*	-	+	++
GIn*	-	+++	+++
Ser	-	++	+++
Trp*	-	+++	+++
Tyr*	-	+++	+++
Lys*	-	+++	+++
Thr*	25%	++	+++
Glu*	50%	++	+++
His*	50%	++	++
Arg*	75%	not determined	not determined
Asp	100%	not determined	not determined
Asn	-	-	-
Cys	-	-	-
Pro	-	-	-

**Table 2A:** Effect of the C-terminal residue of a target protein on DTT-induced cleavage of the intein tag when pTYB1 is used as the cloning vector.

Data is based on analysis of cleavage reactions using MBP as the target protein (MYB) with amino acid substitutions at the position (–1) immediately upstream of the cleavage site in the sequence  $(L_{a}E_{a}X_{a}/C_{n})$ .

(-) = Less than 10% cleavage; (+) = 30%-49% cleavage;

(++) = 50%-74% cleavage; (+++) = 75%-100% cleavage.

\* Boiling in DTT-containing SDS Sample Buffer can cause partial or complete cleavage with this amino acid at the -1 position, resulting in an overestimation of in vivo cleavage. If substantial in vivo cleavage is observed, the cell extract should be evaluated in a SDS Sample Buffer containing no DTT.

	% Cleavage After 16 Hours*			% Cleavage After 40 Hours*			
N-Terminal Residue of the Target Protein	4°C	16°C	23°C	4°C	16°C	23°C	
Met							
Ala	40–60	> 80	> 95	60–90	> 90	> 95	
GIn							
Gly							
Leu							
Asn	10–40	50–80	75–95	40–60	> 90	> 90	
Trp							
Phe							
Tyr							
Val							
lle							
Asp							
Glu	< 10	30–50	50–80	10–20	70–90	70–95	
Lys							
Arg							
His							
Pro	< 10	< 10	< 10	< 10	< 10	< 10	
Thr	7	40	80	20	80	> 90	
Ser	not determined	not determined	not determined	not determined	not determined	not determined	
Cys	not determined	not determined	not determined	not determined	not determined	not determined	

**Table 2B:** Effect of the first N-terminal residue of a target protein on the DTT-induced Cleavage of the intein tag when pTYB11 is used as the cloning vector.

\*Cleavage reactions were conducted in a model system [the MYT4 fusion system (1)], in which T4 DNA ligase is the target protein. The purified fusion protein was treated with 40 mM DTT in 30 mM Hepes, pH 8.0, 0.5 M NaCl for 16 or 40 hours at 4°C, 16°C and 23°C. The percentage of cleavage (% cleavage) was determined by comparing the fusion precursors from the DTT-treated samples with those from the samples without the DTT-17 treatment in scanned images of Coomassie Blue stained SDS-PAGE gels. Note: If a site other than *Sap* I is chosen for cloning the 5<sup>'</sup> end of a target gene, one should use pTYB12. The other sites in pTYB11 are used ONLY for cloning the 3<sup>'</sup> end of a target gene. These are illustrated in Figure 4. Cloning into the *Sap* I site in pTYB1 and pTYB12 places the N- or C-terminal residue of a target protein immediately adjacent to the cleavage site of the intein tag. The effects of the adjacent residues on the intein cleavage have been extensively studied (Table 2A and 2B). In cases of unfavorable residues (e.g., Pro, Cys. Asn, Asp, Arg for pTYB1; Pro, Ser, Cys for pTYB12), it may be necessary to add extra residues to the terminus of a target protein.

# Primer Design

Normally, a target gene is amplified by PCR\* before it is inserted in-frame in the multiple cloning region (MCS) of one of the pTYB vectors. Appropriate restriction sites, absent in the target gene, are incorporated in the forward and reverse primers. The choice of the restriction site in the primers determines whether any or which extra amino acid residues will be attached to the terminus of the target protein after the cleavage of the intein tag.

For example, to obtain a target protein with no extra vector derived residues, one can clone a target gene between the *Nde* I and *Sap* I sites in pTYB1 or the *Sap* I and *Pst* I sites in pTYB11. Cloning the 3' end of a target gene into *Sma* I of pTYB2 or 5' end of a target gene into *Bsm* I of pTYB12 adds an extra residue to the target protein (a glycine to the 3' end or an alanine to the 5' end, respectively). Table 3 illustrates some examples of designing forward and reverse primers containing different restriction sites for cloning into pTYB vectors.

\* PCR Technology is disclosed in certain patents owned by Hoffman-LaRoche and/ or Perkin-Elmer. The purchase of this product does NOT carry with it a license to practice PCR.

Restriction Site Used the Forwar	in	
Primer	Forward Primer <sup>†</sup>	Cloning Vector
Nde I	<b>5</b> ′-GGT GGT <u>CAT ATG</u> NNN NNN <b>3</b> ′	pTYB1, 12
Sap I*	5´-GGT GGT T <u>GC TCT TC</u> C AAC NNN NNN3´	pTYB11
<i>Bsm</i> I**	5´-GGT GGT GG <u>G AAT GC</u> T NNN NNN3´	pTYB12

Table 3A: Examples of forward primer design.

The target gene starts at "NNN NNN...". Restriction sites are underlined. The "GGT GGT" sequence at the 5' end of the primer is to ensure efficient DNA cleavage by the restriction enzyme when the restriction site is close to the 5' end.

- \* Sap I site is not regenerated after the cloning.
- \*\* Bsm I digestion is at 65°C.

#### Table 3B: Examples of reverse primer design.

Restriction Site Used the Revers	in	
Primer	Reverse Primer <sup>+</sup>	<b>Cloning Vector</b>
Sap I*	5'-GGT GGT T <u>GC TCT TC</u> C GCA NNN NNN3'	pTYB1
<i>Sma</i> I**	5'- NNN NNN3'	pTYB2
Pst I***	5'-GGT GGT <u>CTG CAG</u> TCA NNN NNN3'	pTYB11 or pTYB12

\* "NNN NNN..." is the C-terminal antisense strand sequence of a target gene. All restriction sites are underlined. The "GGT GGT" sequence at the 5' end of the primer is to ensure efficient DNA cleavage by the restriction enzyme when the restriction site is close to the 5' end.

- \* Sap I site is not recreated after the cloning.
- \*\* To clone into the Sma I site of pTYB2, the target gene needs not to have a Sma I site at its 3 end. PCR (with a proofreading polymerase) would generate a blunt 3 end which can be ligated with Sma I-digested (blunt end) pTYB2. This results in an extra glycine added to the C-terminus of the target protein. If a Sma I (recognition site: CCCGGG) is incorporated in the reverse primer, however, two extra residues (Pro-Gly) will be added to the target protein.

<sup>\*\*\*</sup> When using Pst I, a stop codon should be included in the primer.

# A. The putative target gene



# B. Creating sticky ends from a PCR fragment



**Figure 4A:** Cloning of PCR-amplified target gene into pTYB1. The sequence of the putative target gene is in lower case. The addition of extra bases (NNN NN) 5' to the restriction site are required for efficient cleavage by the corresponding enzyme. Following digestion with Nde I and Sap I, the insert and vector can be ligated to regenerate the Nde I site with the translation initiation codon (ATG), and the codon for Cys1 (TGC) at the intein N-terminus.



Sap I Site is Lost After Ligation

Figure 4B: An example of using Sap I (and Pst I) to clone a target gene into pTYB11.

### **Cloning into a pTYB Vector**

The following is a protocol we routinely use to clone an amplified target gene fragment using restriction enzymes that create different sticky ends. For blunt-end or single-site cloning, the vector may need to be treated with an alkaline phosphatase (CIP, NEB #M0290) and more units of T4 DNA Ligase (NEB #M0202) should be added in the ligation mixture. Users may follow their own protocols.

- Purification of the Amplified Gene Fragment: the reaction mixture containing the amplified target gene fragment (generally 100 μl) is directly loaded on a 1% low-melting gel. The correct fragment is cut out using a razor blade. The gel slice (< 100 μl, the smaller the volume, the better) is melted at 65°C for 7 minutes. After the gel slice is cooled at 42°C for 7 minutes, 2 μl of β-Agarase I (NEB #M0392) per 100 μl gel slice is added and the incubation at 42°C continues for 1 hour. The melted gel slice containing the target gene is directly used for restriction digestion; no DNA precipitation is necessary.
- 2. **Restriction Digestion and Ligation:** The above gel-purified gene fragment (use 80  $\mu$ l) is double-digested with appropriate restriction enzymes in a 100  $\mu$ l reaction mixture (Note: use the suggested 10X NEBuffer for double digestion, see the NEB catalog). At the same time, a pTYB vector (0.5  $\mu$ g) is digested with the same enzymes in a 50  $\mu$ l reaction mixture. After a 2 to 4 hour digestion, both reaction mixtures are loaded on a 1% low-melting gel. The gel slices containing the digested gene fragment and digested pTYB vector are combined in the same eppendorf tube and melted at 65°C for 10 minutes. After cooling at 42°C for 7 minutes, 2  $\mu$ l  $\beta$ -Agarase I (NEB #M0392) per 100  $\mu$ l gel slice of is added to digest the gel for 1 hour at 42°C. The ligation is then conducted in the same mixture. T4 DNA Ligase (NEB #M0202) (2  $\mu$ l) and an appropriate volume of 10X T4 DNA Ligase Buffer are added to the mixture and the ligation incubates at 16°C overnight.

- 3. Transformation: The above ligation mixture (15 μ) is used to transform 150 μl of competent cells. To reduce the background from self-ligation of the vector, one can digest the ligation sample prior to transformation with an enzyme whose site is deleted from the polylinker during cloning and is also absent in the insert. This linearizes any remaining parental vectors without insert. The fusion constructs should initially be established in a non-restricting, non-expression host. The vectors supplied in the kit are prepared from a restriction-deficient *E. coli* strain (r<sup>-m</sup>). When introduced into a strain with wild type *EcoK* (hsd<sup>+</sup>), DNA will be restricted. Therefore plasmid DNA or ligated DNA should initially be transformed into a restriction-deficient *E. coli* strain (r<sup>-m</sup>).
- 4. Screening for Inserts: Plasmids are isolated from each colony and those containing the correct target gene insert are identified by digesting the plasmid DNA with the same restriction enzymes used for cloning the target gene fragment. The only exception is when Sap I is used. The Sap I site is lost after the insert is ligated to the vector. Alternatively, one can use colony PCR or colony hybridization to screen a large number of transformants for the presence of the target gene inserts. Immunodetection can also be used with anti-CBD to detect the full-length fusion precursor in the total cell lysates. Clones should be further confirmed by DNA sequencing before proceeding to the cell culture and protein expression.

# Cell Culture and Fusion Protein Expression:

The expression of the fusion protein from a pTYB vector may be affected by the following factors: (i) *E. coli* strain; (ii) cell culture conditions (e.g., temperature, aeration, cell density, etc.); (iii) protein expression induction conditions (temperature, duration, the IPTG concentration, etc.). *E. coli* strain ER2566 is supplied in the kit as a host for expression of the fusion protein from a pTYB vector. Other commercially available strains [e.g. BL21(DE3) and derivatives] can also be used as a host for the pTYB vectors. The use of hosts carrying pLysS or pLysE plasmids encoding T7 lysozyme, a natural inhibitor of T7 RNA Polymerase, can reduce basal transcription in uninduced cells if more stringent control is needed. Different strains may be tested to achieve optimal expression of each fusion protein. Expression of any toxic target protein may require lowering the culture temperature to maintain the vector stability; induction of protein expression at 15–30°C can often help the folding and solubility of the fusion protein and increase the cleavage efficiency of the intein. The following protocol is provided as a guideline.

# **Cell Culture**

1 liter of LB medium containing 100  $\mu$ g/ml ampicillin is inoculated with a freshly grown colony or 10 ml freshly grown culture. The culture is incubated in an air shaker at 37°C until the OD<sub>600</sub> reaches 0.5–0.8.

#### Induction of Protein Expression

IPTG is added to a final concentration of 0.3–0.5 mM before the culture is transferred to a 15°C air shaker (one can achieve a 15°C incubation by installing a shaker in a cold room and adjusting the temperature of the shaker). This allows accumulation of sufficient T7 RNA Polymerase for protein expression). Induction at 15°C should be conducted overnight. Other induction conditions such as 30°C for 6 hours and 20–25°C overnight may also be tested.

### **Cell Harvest**

The cells from the above culture are spun down at 5000 X g for 10 minutes at  $4^{\circ}$ C. After discarding the supernatant, the cell pellet can be stored at  $-20^{\circ}$ C.

# Affinity Purification and On-column Cleavage:

### **Preparation of Chitin Column**

The chitin-binding domain (CBD) in the intein tag allows affinity purification of the fusion protein. The CBD has an extremely high affinity for the chitin beads, which allows efficient recovery of the fusion protein from the crude cell extract. In addition, stringent wash conditions (*e.g.* high salt concentration or use of detergents) can be employed to reduce nonspecific binding, thus increasing purity. Generally, 15–20 ml of chitin beads should be used for a one liter culture. The chitin beads should be equilibrated at 4°C with 10 volumes of the Column Buffer (see Media and Solutions) prior to the loading of the crude cell extracts.

#### **Preparation of Crude Cell Extract**

The cell pellet from a one liter culture is resuspended in 50 ml ice-cold Cell Lysis Buffer (see Table 4 and Media and Solutions). Some commonly used nonionic detergents have no adverse effect on the cleavage of the intein tag. 0.1% Triton X-100 or Tween 20 may be used in the Cell Lysis Buffer unless the target protein is known to be inactivated by these nonionic detergents. Furthermore, protease inhibitors such as phenylmethylsulfonyl fluoride (PMSF) show no inhibitory effects and can also be used. DTT should not be used in column buffer. The reducing agents TCEP [tris-(2-carboxyethyl)phosphine] or TCCP [tris-(2-cyanoethyl)phosphine] (0.1–1 mM) can be used to stabilize oxidation-sensitive proteins during purification. The cells can usually be broken by either sonication or a French press. Since egg white lysozyme is known to bind and digest chitin, it is not a preferred method for cell lysis. However, if a sonicator or a French press is not available, a low level of lysozyme (10–20  $\mu$ g/ml) can be used (incubate at 4°C for 1 hour). The increase in viscosity indicates when the cells are broken. If the mixture becomes extremely viscous, it may be necessary to dilute the cell lysate and/or add 10  $\mu$ g/ml of protease-free DNase plus MgCl<sub>2</sub> (to 5 mM final concentration) to reduce viscosity.

The clarified cell extract (supernatant) is obtained by centrifugation at 20,000 X g for 30 minutes. Samples (40  $\mu$ l) are taken from the total cell extract (before the centrifugation) and the supernatant (after the centrifugation), mixed with 20  $\mu$ l 3X SDS Sample Buffer, and analyzed by SDS-PAGE and/or Western blot. Boiling in DTT-containing SDS Sample Buffer can cause partial or complete cleavage, resulting in an overestimation of *in vivo* cleavage. If substantial *in vivo* cleavage is observed, the cell extract should be evaluated in SDS Sample Buffer containing no

DTT. If the fusion precursor is detected in the total cell extract but not in the supernatant, this may suggest that the fusion protein is expressed but not in a soluble form.

**Loading the Clarified Cell Extract:** The clarified extract is slowly loaded onto the chitin column at a flow rate no faster than 0.5–1 ml/minute. A sample (40  $\mu$ ) from the flow-through (undiluted) is mixed with 20  $\mu$  3X SDS Sample Buffer and analyzed by SDS-PAGE. Comparison between samples from the flow-through and the supernatant reveals the binding efficiency of the fusion precursor on the chitin column.

**Washing the Chitin Column:** Due to the high affinity of the CBD for the chitin beads, a higher flow rate (2 ml/min) and stringent wash conditions [high salt concentration (1 M NaCl), and nonionic detergents] can be used. Loading and washing at high salt with nonionic detergents may reduce nonspecific binding of other *E. coli* proteins. Normally, no less than 10 bed volumes of the Column Buffer is required to thoroughly wash the column. Be sure that all traces of the crude cell extract have been washed off the sides of the column.

Induction of On-column Cleavage: The target protein is released from the chitin column when the chitin-bound intein tag undergoes self-cleavage in the presence of DTT,  $\beta$ -mercaptoethanol or cysteine. Induction of the on-column cleavage is conducted by quickly flushing the column with 3 bed volumes of the Cleavage Buffer containing 30–50 mM of DTT,  $\beta$ -mercaptoethanol or cysteine. This evenly distributes thiols throughout the column. Though no significant cleavage should occur during this quick flush, it is recommended not to extend the process for more than 30 minutes to 1 hour. Both cysteine and DTT produce similar cleavage efficiency. For target proteins that are sensitive to high concentrations of DTT, a mixture of cysteine (50 mM) and DTT (1 mM) may be used. The composition of the Cleavage Buffer can be made similar to that of the final storage buffer for the target protein. If the target protein is stable only under certain buffer conditions, then that buffer plus 40–50 mM DTT or cysteine may be used for the cleavage reaction and subsequent elution. After the quick flush, the flow in the column is stopped, and the column left at 4–23°C for 16–40 hours.

Several factors affect the cleavage efficiency and thus the final yield: (i) amino acid residue(s) at the cleavage site; (ii) temperature of the cleavage reaction; (iii) duration of the cleavage reaction; (iv) pH of the Cleavage Buffer.

In most cases (see Table 2A & B), incubation of the column at 16–23°C for 16 hours (overnight) results in more than 50% cleavage of the fusion precursor. When the C-terminal fusion vectors (pTYB1 and pTYB2) are used, the on-column cleavage reaction can be conducted at 4°C overnight (except in the case of Leu, Ile or Val at the cleavage site where a higher temperature and/or longer reaction time are needed). When the N-terminal fusion vectors (pTYB11 and pTYB12) are used, higher temperatures (16–23°C) and longer cleavage reaction times (40 hours) are normally required. The data in Table 2A & B may provide a guideline for selecting an appropriate temperature and duration of the cleavage reaction. For instance, when Met, Ala, or Gln are at the cleavage site of pTYB11 and pTYB12, incubation can be conducted at 4°C for 40 hours. If the activity and/or structure of the target protein is sensitive to prolonged incubation, a shorter incubation time (8 hours) at 23°C may result in a sufficient yield. The cleavage efficiency can be determined by analyzing a chitin resin sample on SDS-PAGE following the cleavage or elution step (see below). If most of the precursor is not cleaved and/or the target protein remains stable, longer incubation time and higher temperature for the cleavage reaction are recommended.

Generally, higher pH favors the on-column cleavage reaction. For the C-terminal fusion vectors (pTYB1 and pTYB2), on-column cleavage can be conducted at pH 7–9; For the N-terminal fusion vectors (pTYB11 and pTYB12), however, on-column cleavage should be conducted between pH 7.5–9. Our studies show that the cleavage reaction with pTYB11 and pTYB12 is inhibited below pH 7.0.

**Elution of the Target Protein:** After the induction of the cleavage reaction, the target protein is released from the intein tag and can be eluted from the column using the Column Buffer or a specific storage buffer. The target protein is normally detected in the first few fractions. When pTYB11 and pTYB12 are used, a small peptide (1.6 kDa) is also cleaved from the intein tag and co-eluted with the target protein (for more details, see Appendix III). Due to its small molecular weight, the cleaved peptide can not be detected on a regular SDS-PAGE gel and can be separated from the target protein by dialysis.

The concentration of the target protein in each fraction can be determined by the Bradford assay. If the protein concentration is too low for SDS-PAGE, one can use acetone to precipitate the protein before electrophoretic analysis (add 0.6 ml acetone to a 0.2 ml aliquot, mix and store at

-20°C for 30 minutes). Centrifuge and carefully pour off the supernatant. Then add 40  $\mu l$  of 1X SDS Sample Buffer to the pellet (white precipitate) and boil before electrophoresis.

At this point, chitin resin can be removed to determine the cleavage efficiency. To check cleavage efficiency remove 200  $\mu$ l of resin and mix with 100  $\mu$ l 3X SDS Sample Buffer. After boiling for 5 minutes, the resin is spun down. The supernatant (3–10  $\mu$ l) is directly used for SDS-PAGE analyses. If a large amount of the precursor still remains uncleaved, the incubation of the column should continue for additional 24 hours before a second elution is conducted.

**Regeneration of Chitin Resin:** The chitin resin can be regenerated 4–5 times by the following protocol: To remove proteins from the resin wash the column with 3 bed volumes of the 0.3 M NaOH (stripping solution). Allow resin to soak for 30 minutes and wash with an additional 7 bed volumes of 0.3 M NaOH. Wash with 20 bed volumes of water followed by 5 bed volumes of column buffer. The resin can be stored at 4°C. For long term storage, 0.02% sodium azide should be added to the Column Buffer.

Standard Column Buffer	Substitutions
20 mM HEPES (pH 7–9)	Tris-HCI Na-Phosphate
500 mM NaCl	50–2000 mM NaCl
0.1% Triton X-100 (optional)	0.1–0.5% Triton X-100 0.1–0.2% Tween 20
1 mM EDTA (optional)	0.1–1 mM EDTA 20 μM PMSF 1 mM TCEP or TCCP <sup>1</sup>

Table 4: Buffer	conditions	for	bindina	and	on-columr	l cleavage
Table II Dunion	contantionio		onnanng	unu	on ooranni	olourugo

<sup>1</sup>TCEP [tris-(2-carboxyethyl)phosphine] and TCCP [tris-(2-cyanoethyl)phosphine] are reducing agents which specifically reduce disulfide bonds without affecting thioesters.

# Simplified Purification Protocol for Test Experiment:

The following is a protocol for a pilot experiment which can serve as a guideline for testing optimal conditions for the expression and purification of a target protein. Samples taken throughout the experiment can be analyzed on SDS-PAGE.

- 1. Culture Preparation: Inoculate 1 liter of LB supplemented with ampicillin to a final concentration of 100  $\mu$ g/ml with a freshly grown colony or 10 ml freshly grown culture. Incubate the culture at 37°C to an A<sub>600</sub> of 0.5–0.6.
- 2. Induction: The 1 liter culture should be split into 5 samples (200 ml each in 1 liter flasks) to test for optimal expression conditions. One sample should be used as a control for uninduced cells (no IPTG).

Note: The optimal incubation temperature and time for induction will vary depending on the target protein. We recommend testing several conditions such as 30°C for 3 hours, 22–25°C for 6 hours and 12–15°C overnight using 0.3 mM IPTG.

3. Cell Harvest: Spin down the 200 ml cultures at 5000 X g for 10 minutes at 4°C. Discard the supernatant. The pellet can be stored at -20°C.

Note: Steps 4–9 should be performed at 4°C.

4. Crude Cell Extracts: Resuspend each cell pellet in 5 ml ice-cold Column Buffer (see Media and Solutions) and lyse the cells by sonication on ice. Remove 40 μl of the crude cell extract and mix with 20 μl of 3X SDS Sample Buffer (Sample 1: crude extract/uninduced; Sample 2: crude extract/induced cells). Prepare clarified extracts by centrifugation at 20,000 X g for 30 minutes. Save the pellet at -20°C for future analysis. If the target protein is absent from the clarified lysate, this may indicate a solubility problem in which case the pellet should be tested for the presence of insoluble fusion protein. Remove 40 μl of the supernatant and mix with 20 μl of 3X SDS Sample Buffer (Sample 3: clarified extract). These samples can be analyzed by SDS-PAGE and/or by Western blot with the anti-CBD serum to detect the fusion protein.

- 5. Chitin Column Equilibration: The chitin beads have a binding capacity of about 2 mg/ml. Aliquot 6 ml of the chitin bead slurry in a separate column for each clarified extract sample prepared from the induced cell cultures. Equilibrate each column by passing 60 ml of Column Buffer through the column.
- 6. Loading: Slowly load the clarified extracts onto the chitin columns (at a rate no faster than 0.5 ml/min). Remove 40 μl from the flow-through and mix with 20 μl 3X SDS Sample Buffer (Sample 4). This sample, when compared with the clarified extract sample, illustrates the binding efficiency of the construct.
- 7. Washing the Column: Wash the column with 100 ml of Column Buffer at a flow rate of 1 ml/minute. Be sure that all traces of crude extract have been washed off the sides of the column.
- On-column Cleavage Reaction: Flush the column quickly with 18 ml (or about 3 column volumes) of 50 mM DTT (freshly diluted in Cleavage Buffer or other suitable buffer) to evenly distribute DTT throughout the column. Remove 40 µl of the flow through from the quick flush and mix with 20 µl 3X SDS Sample Buffer (Sample 5). Stop the column flow, and leave at 4°C overnight.
- 9. Target protein elution: Elute the target protein using additional Cleavage Buffer without DTT and collect 1 ml fractions for 3 column volumes. Monitor fractions by the Bradford assay to determine the concentration of the target protein in each fraction. Remove 40 μl from the fractions with the highest protein concentrations and mix with 20 μl 3X SDS Sample Buffer (Sample 6). If the protein concentration is too low for SDS-PAGE, add 0.6 ml acetone to a 0.2 ml aliquot, mix and store at 20°C for 30 minutes. Centrifuge and carefully pour off the supernatant. Then add 40 μl of 1X SDS Sample Buffer to the pellet (white precipitate) and boil 5 minutes before electrophoresis.
- 10. Cleavage Efficiency: Remove 200  $\mu$ l chitin resin and mix with 100  $\mu$ l 3X SDS Sample Buffer. After boiling for 5 minutes, the supernatant is analyzed on SDS-PAGE to determine the cleavage efficiency.

# Media and Solutions:

The following are suggested conditions for cell culture, cell lysis and protein purification. They can be modified according to the specific properties of the target protein.

LB broth (per liter) 10 g tryptone 5 g yeast extract 10 g NaCl Adjust the pH to 7.0 with NaOH

#### **Cell Lysis Buffer**

20 mM Na-HEPES or Tris-HCl or Na-Phosphate (pH 8.0 or if needed, pH 6.0–9.0) 500 mM NaCl (or 50–1,000 mM NaCl) 1 mM EDTA (optional) 0.1% Triton X-100 (optional) Other nonionic detergent (e.g., 0.1–0.2% Tween 20) and protease inhibitors (e.g., PMSF [20 μM]) can also be included. For a target protein sensitive to oxidation, 1 mM of TCEP [tris-(2carboxyethyl)phosphine] or TCCP [tris-(2-cyanoethyl)phosphine] may be used.

#### **Column Buffer**

20 mM Na-HEPES or Tris-HCl or Na-Phosphate (pH 8.0 or 6.0–9.0) 500 mM NaCl (or 50–1,000 mM NaCl) 1 mM EDTA (optional)

#### **Cleavage Buffer**

20 mM Na-HEPES or Tris-HCl or Na-Phosphate pH 7–9 for pTYB1 & pTYB2; pH 7.5 or above for pTYB11 & pTYB12 500 mM NaCl (or 50–1,000 mM NaCl) 1 mM EDTA (optional) 50 mM DTT or β-mercaptoethanol or cysteine\* (\*use 2-mercaptoethanesulfonic acid for IPL see Appendix IV, page 34)

#### **Stripping Solution**

0.3 M NaOH

# Frequently Asked Questions (FAQs):

## Cloning

■ Should I clone my target gene into the C-terminal fusion vectors (pTYB1 and pTYB2) or N-terminal fusion vectors (pTYB11 and pTYB12)?

pTYB1 and pTYB2 vectors allow the fusion of the C-terminus of a target protein to the intein tag whereas with the pTYB11 and pTYB12 vectors, the N-terminus of the target protein is fused to the intein tag. It is conceivable that different target proteins, due to certain structural constraints, may prefer either C-terminal or N-terminal fusion to allow proper folding of the fusion precursor and a high level of protein expression. pTYB2 and pTYB12 contain the same or compatible cloning sites which allow the cloning of the same amplified target gene. One may express the target gene in both vectors and determine which type of fusion results in a better expression and yield. Use of pTYB1 or pTYB11 allows expression and purification of a target protein without any extra vector-derived residues. One should also take into consideration the differences between the C-terminal and N-terminal fusions (see Appendix I). For instance, when the purified target protein is also intended for C-terminal labeling or peptide ligation, pTYB1 or pTYB2 should be used. If the target protein requires a defined N-terminal residue (other than Met) or has Pro, Cys, Asp or Arg as the C-terminal residue, pTYB11 or pTYB12 should be used.

# The target protein contains an aspartate or arginine at its C-terminus. Can I still use the Sap I site (in pTYB1) as the 3' cloning site?

The data in Table 2A is based on our studies using the maltose-binding protein (MBP) as the target protein. It may still be possible to use the *Sap* I site for 3' cloning in the pTYB1 vector since the aspartate residue at the C-terminus may not cause *in vivo* cleavage in the context of a different target protein. However, it is advisable to try both *Sma* I and *Sap* I sites for subcloning the target gene.

### ■ What are the features of other E. coli IMPACT vectors?

Two additional pTYB C-terminal fusion vectors, pTYB3 (NEB #N6703S) and pTYB4 (NEB #N6704S), are available for cloning a target gene in which the C-terminus of the target protein is fused to the intein-CBD tag. pTYB3 and pTYB4 contain an *Nco* I site, overlapping

the initiating methionine codon, in place of the *Nde* I site in pTYB1 and pTYB2, respectively. Digestion of the insert with *Bsp*H I, *Bsp*LU11 I and *Afl* III can also generate *Nco* I-compatible overhangs. pTXB1 (NEB #N6707S) and pTXB3 (NEB #N6708S) are IMPACT C-terminal fusion vectors using an engineered mini-intein (198 residues) from the gyrA gene of *Mycobacte-rium xenopi*. These vectors are designed for fusing the C-terminal thioester tagged proteins for labeling and protein ligation (8, see Appendix IV). Use of pTXB vectors may result in higher expression and reduce *in vivo* cleavage. The polylinker region of pTXB1 is identical to pTYB1 and the polylinker region of pTXB3 is identical to pTYB3.

Also available is the IMPACT-TWIN (Intein Mediated Purification with an Affinity Chitinbinding Tag-Two Intein) Kit (NEB #E6950S). The pTWIN vectors (pTWIN1 and pTWIN2) allow for the isolation of native recombinant proteins possessing an N-terminal cysteine and/or possessing a reactive C-terminal thioester in a single chromatographic step without the use of exogenous proteases. These reactive groups can be used in Intein-mediated Protein Ligation (IPL, Appendix IV) to specifically attach proteins, peptides, or labels to the N-and/or C-terminus of a target protein.

The pTWIN vectors allow a target protein to be sandwiched between two self-cleaving miniinteins. Chitin binding domains present on both inteins allow the affinity purification of the precursor protein on a chitin resin. Intein1 is a mini-intein derived from the *Synechocytis* sp *dnaB* gene engineered to undergo pH and temperature dependent cleavage at its C-terminus. Cleavage of this intein can liberate an N-terminal amino acid residue other than Met on a target protein. A protein with an N-terminal cysteine residue can be used in IPL reactions. Intein2 is either a mini-intein from the *Mycobacterium xenopi gyrA* gene (pTWIN1, NEB #N6951S) or from the *Methanobacterium thermoautotrophicum rir1* gene (pTWIN2, NEB #N6952S). These inteins have been modified to undergo thiol-induced cleavage at their Nterminus. The use of thiol reagents such as 2-mercaptoethanesulfonic acid (MESNA) releases a reactive thioester at the C-terminus of the target protein for use in IPL. Following cleavage of both inteins the target protein is eluted from the chitin resin while the inteins remain bound through the chitin binding domains. An intramolecular condensation generates circular protein species. (See www.neb.com for details)

### Expression

### What factors contribute to poor expression and low yield of some proteins?

The yield of a target protein is generally determined by (i) the expression level of the fusion precursor; (ii) the solubility of the fusion protein; (iii) the thiol-induced cleavage efficiency; (iv) the solubility of the target protein after cleavage. The expression level of the fusion protein is mostly influenced by the target protein. Poor codon usage, mRNA degradation or proteolysis due to protein misfolding may all contribute to poor expression. Different growth and induction conditions should be tested to optimize the expression of the fusion protein. Induction at lower temperatures may reduce the formation of inclusion bodies (i.e., improve the solubility of the fusion protein) as well as the level of proteolysis. Protease deficient hosts should also be tested to minimize proteolysis. Optimizing codon usage may be helpful in increasing the expression level. Another possibility for poor expression could be instability of the clone due to toxicity of the target protein to host cells. Accordingly, one should inoculate the medium with a freshly grown colony and induce the expression at lower temperatures with lower IPTG concentrations. Use of an eucarvotic expression system may ultimately help to increase the yields of eucaryotic proteins and achieve proper modification of the proteins. Some target proteins become insoluble after on-column cleavage and therefore are only eluted after incubation of the resin with SDS. In this case, increasing the salt concentration (0.5-2 M NaCl) or adding a nonionic detergent to the Cleavage Buffer may improve the solubility of the target protein.

#### What does it mean if the Western blot analysis with anti-CBD serum detects a 55 kDa product but not the fusion protein in the crude cell extract?

The data indicate that either the target protein is degraded or cleaved *in vivo*. The samples should be prepared in SDS Sample Buffer without DTT or  $\beta$ -mercaptoethanol, since boiling in DTT-containing Sample Buffer may cause cleavage of the fusion protein (see note in Table 2A). Perform a Western blot with anti-target protein serum to differentiate between proteolysis and intein-mediated cleavage. If proteolysis is evident, try different hosts. *In vivo* cleavage may be reduced by inducing cells at lower temperatures (for example, 12–15°C for overnight or room temperature for 3–6 hours). You may also use the *Sma* I site or *Xho* I site in pTYB2 to subclone the target gene. This places one or several residues between the target protein and the intein tag which may reduce *in vivo* cleavage. However, the purified target protein will contain extra residues after cleavage.

### Purification

### ■ Is there an alternative way to break cells other than sonication?

Cells can also be broken with a French press or by the addition of egg white lysozyme. However, since egg white lysozyme is known to bind and digest chitin, it is not recommended for cell lysis. If a sonicator or a French press is not available, try a low level of lysozyme (10–20  $\mu$ g/ml) and incubate at 4°C for 1 hour. The increase in viscosity indicates when the cells are broken. If the mixture becomes extremely viscous, it may be necessary to add 10  $\mu$ g/ml of protease-free DNase plus MgCl<sub>2</sub> (to 5 mM final concentration) to reduce viscosity before the clarified extract can be passed through the chitin column.

# Can I use a thiol reducing agent during purification to protect my protein from oxidation?

Yes, a low concentration (1 mM) of dithiothreitol (DTT) or  $\beta$ -mercaptoethanol can be used throughout the purification procedures as long as the purification (before the on-column cleavage step) can be finished in a short period of time. Alternatively, TCEP [tris-(2-carboxyethyl)phosphine] and TCCP [tris-(2-cyanoethyl)phosphine] can be used at 0.1–1 mM final concentration in the Cell Lysis and Column Buffers to stabilize oxidation-sensitive proteins during purification. These compounds specifically reduce disulfide bonds without affecting the intein-mediated cleavage reaction and thus can be used to stabilize proteins with essential thiols (Burns, J.A., et al. *J. Org. Chem.* 56, 2648–2650). Furthermore, TCEP or TCCP (3–5 mM) may also be used as a reducing agent in SDS Sample Buffer in place of DTT, which can cause partial or complete cleavage of the fusion protein when certain amino acids are at the position preceding the cleavage site (pTYB1 and pTYB2) (Table 2A).

#### Cleavage

What should I do if the fusion precursor is the major product on the chitin resin after target protein elution?

This means that the thiol-induced on-column cleavage is not efficient – invariably leading to a low yield of the target protein. The following options can be tried to increase the cleavage efficiency: (i) increase the temperature of the on-column cleavage; an 8 hour incubation at 16-23°C may result in more target proteins than a 16 or 40 hour incubation at 4°C; (ii)
increase the duration of the on-column cleavage; if the target protein is sensitive to high temperatures, a longer incubation time at a lower temperature may be needed; (iii) increase the pH of the on-column cleavage; the higher the pH (up to 9.0), the higher the cleavage efficiency; (iv) change the residue(s) adjacent to the intein cleavage site. Often the low cleavage efficiency is caused by an unfavorable terminal residue of the target protein which is placed adjacent to the intein cleavage site when *Sap* I is used for cloning in pTYB1 or pTYB11. Table 2 may serve as a guideline for placing a suitable residue adjacent to the cleavage site. Sometimes, it is necessary to add more than one residue to the terminus of a target protein in order to achieve a high cleavage efficiency. In this case, pTYB2 and pTYB12 should be used.

### What does it mean if the target protein is not eluted after on-column cleavage but is present on the chitin beads after target protein elution?

If both the target protein and the 55 kDa intein tag are present on the chitin beads after elution, it suggests that the target protein becomes insoluble after induced on-column cleavage. Increase the salt concentration (0.5-2 M NaCl) or add a nonionic detergent to the Cleavage Buffer to improve the solubility of the target protein. A number of nonionic detergents examined (0.1-0.5% Triton X-100 or 0.1-0.2% Tween 20) had little effect on binding or cleavage. If urea is used to elute the column, some intein tag may co-elute with the target protein. In this case, it may be necessary to repurify and refold the target protein.

### If my target protein is sensitive to DTT, are there alternative means to induce the on-column cleavage?

If the activity of the target protein is affected by high concentrations of DTT or  $\beta$ -mercaptoethanol, lower concentrations of DTT or  $\beta$ -mercaptoethanol (5–10 mM) may be used for on-column cleavage. However, longer incubation time or higher temperatures (up to room temperature) may be required for efficient cleavage. Alternatively, 50 mM of freshly prepared hydroxylamine (for pTYB1 and pTYB2) or cysteine solution (at pH 8–9) can be used to induce cleavage at 4–25°C. Be aware that when hydroxylamine or cysteine is used with pTYB1 or pTYB2, they form a stable covalent bond with the C-terminus of the target protein. One should determine whether a C-terminal hydroxamate or cysteine affects the activity of the target protein. When cysteine is used for cleavage with pTYB11 or pTYB12, the cysteine is not attached to the target protein (see Appendix III for details).

## The target protein has optimal activity at low salt concentration. How do I perform on-column cleavage?

The composition of the Cleavage Buffer is not critical for the cleavage reaction and can be made similar to the final storage buffer for the target protein. Wash the column extensively (10 volumes) with high salt Column Buffer (this removes proteins that nonspecifically bind to chitin). Wash the column with 3 volumes of the low salt Cleavage Buffer (without DTT or  $\beta$ -mercaptoethanol) and then flush with 3 volumes of DTT-containing Cleavage Buffer. This ensures that the target protein is eluted in the specified buffer.

# After the on-column cleavage, the SDS-PAGE analysis shows that the target protein is contaminated with other bands. What are these proteins?

These may be *E. coli* proteins that nonspecifically bind to chitin beads. Generally, washing the column extensively with the Column Buffer containing high salt (0.5–1 M NaCl ) and nonionic detergent will effectively reduce the background. If lysozyme is used for breaking cells, it may be present in the eluant since lysozyme can bind and digest chitin. It is also possible that the proteins have been eluted by virtue of their affinity for the target protein. For example, the *E. coli* host chaperone protein GroEL (~60 kDa) has been seen to co-purify with some eucaryotic proteins or mutant proteins which fold poorly in *E. coli* host cells. Addition of ATP to Column Buffer may allow separation of the GroEL protein from the target protein. Finally, in some cases, small amounts of the fusion protein or the intein tag (55 kDa) may appear in the eluate following the on-column cleavage reaction. The latter two possibilities can be checked by Western blot analysis using antibodies against the target protein and the CBD. Use of protease inhibitors, protease-deficient hosts and induction at low temperatures (15°C for 16 hours) may reduce proteolytic degradation.

### ■ How do I remove DTT after cleavage?

After elution of the target protein, free DTT can be removed from the sample by dialyzing at least twice against an appropriate buffer.

# <u>References:</u>

- 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) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, 192, 277–281.
- Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. -Q., Benner, J. (1998) Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucl. Acids Res.* 26, 5109–5115.
- Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M.-Q. (1998) Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J. Biol. Chem.* 273, 10567–77
- Chong, S., Shao Y., Paulus, H, Benner, J., Perler, F. B. and Xu, M.-Q. (1996) Protein splicing involving the *Saccharomyces cerevisiae* VMA intein: the steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an *in vitro* splicing system. *J. Biol. Chem.* 271, 22159–22168.
- Xu, M.-Q. and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J.* 15, 5146–5153.
- Xu, M.-Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993) *In vitro* protein splicing of purified precursor and the identification of a branched intermediate. *Cell* 75, 1371–1377.
- 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) Protein splicing elements: inteins and exteins — a definition of terms and recommended nomenclature. *Nucleic Acids Res.* 22, 1125–1127.
- Evans, T. C., Benner, J., and Xu, M.-Q. (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* 7, 2256–2264.
- Evans, T. C., Benner, J., and Xu, M.-Q. (1999) The *in vitro* Ligation of Bacterially Expressed protein Using an Intein from Methanobacterium thermoautotrophicum. *J. Biol. Chem.* 274, 3923–3926.

- Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H\*-adenosine triphosphatase. *Science* 250, 651–657.
- Dubendorff, J. W. and Studier, F. W. (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219, 45–59.
- Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., and Tanaka, H. (1994) The role of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation *J. Bacteriol.* 176, 4465–4472.

# **IMPACT User Publications:**

- Dorsch, S., Kaufmann, B., Schaible, U., Prohaska, E., Wolf, H., and Modrow, S. (2001) The VP1-unique region of parvovirus B19: amino acid variability and antigenic stability. *J. Gen. Virol.* 82 (pt 1), 191–199.
- Park, C. M., Kim, J. I., Yang, S. S., Kang, J. G., Kang, J. H., Shim, J. Y., Chung, Y. H., Park, Y. M., and Song, P. S. (2000) A Second Photochromic Bacteriophytochrome from Synechocystis sp. PCC 6803: Spectral Analysis and Down-Regulation by Light. *Biochemistry* 39, 10840–10847.
- Lykke-Andersen, J., and Christainsen, J., (1998) The C-terminal carboxy group of T7 RNA polymerase ensures efficient magnesium ion-dependentcatalysis. *Nucleic Acids Res.* 26, 5630–5635.
- 4. Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. USA* 95, 6705–6710.
- Schleper, C., Swanson, R. V., Mathur, E. J., and DeLong, E. E. (1997) Characterization of a DNA Polymerase from the Uncultivated Psychrophilic Archaeon Cenarchaeum symbiosum. *J. of Bacteriology* 179, 7803–7811.
- Wahl, M. C., Huber, R., Prade, L., Marinkovic, S., Messerschmidt, A., and Clausen, T. (1997) Cloning, purification, crystallization, and preliminary X-ray diffraction analysis of cystathionine gamma-synthase from *E.coli. FEBS Lett.* 414, 492–496.

# Appendix I: Comparison Between C-terminal and N-terminal Fusions

The IMPACT-CN System allows the fusion of a self-cleavable intein tag to either the C-terminus or N-terminus of a target protein. Different target proteins, due to certain structural constraints, may prefer one or the other fusion to allow proper folding of the precursor protein and a higher level of protein expression. The following table lists differences in usage between the C-terminal and N-terminal fusion vectors.

**Table 5:** Comparison between the C-terminal (pTYB1 and pTYB2) and N-terminal (pTYB11 and pTYB12) fusions

C-terminal pTYB1 or pTYB2	N-terminal pTYB11 or pTYB12
The C-terminus of the target protein is fused to the internation internation internation internation tag	The N-terminus of the target protein is fused to the intein tag
Met	Not restricted to Met; Pro blocks cleavage; Cys, Ser and Thr may result in protein splicing
Asp and Arg cause <i>in vivo</i> cleavage; Asn, Cys and Pro block cleavage	No restriction; less <i>in vivo</i> cleavage.
Influenced by the target protein	Less affected by the target protein
	pTVB1 or pTYB2   The C-terminus of the target protein is fused to the intein tag   Met   Asp and Arg cause in vivo cleavage; Asn, Cys and Pro block cleavage   Influenced by the target

	C-terminal pTYB1 or pTYB2	N-terminal pTYB11 or pTYB12
Purification		
Induction of on-column cleavage	4–23°C for 16 hours; efficient	At 4°C, less efficient; may require 16°C or higher for longer time
Sensitivity to pH	pH 6.0–9.0; recommended pH 8.0–9.0	pH 6.0 inhibits cleavage; recommended pH 7.5–9.0
Use of Cys for cleavage	Cys is covalently attached to the target protein	Cys is not attached to the target protein
Elution	Normally elute after an overnight incubation	May require a longer incubation time before elution; N-extein peptide (15 aa) also present
Applications for protein labeling and peptide ligation	Yes	No

# Appendix II: The Chemical Mechanism of Protein Splicing

Protein splicing is a post-translational processing event involving the excision of an internal protein segment, the intein, from a precursor protein and the concomitant ligation of the flanking N- and C-terminal regions (the exteins) (7). Sequence alignment reveals that there are highly conserved residues at the two splice junctions: a Cys or Ser residue at the N-terminus of the intein, His-Asn at the C-terminus of the intein, and Cys, Ser or Thr as the first residue of the C-terminal extein. These conserved splice junction residues are directly involved in the catalysis of peptide bond cleavage and ligation of the protein splicing reactions. The chemical mechanism of protein splicing (Figure 5) is based on the studies on inteins from the *Saccharomyces cerevisiae VMA1* gene and *Pyrococcus sp.* GB-D DNA polymerase gene (3–6). Protein splicing involving the *Sce* VMA1 intein presumably proceeds by the following four chemical steps. Step 1: Formation of a linear thioester intermediate by an N-S acyl rearrangement at Cys1 at the N-terminus of the intein: Step 2: Formation of a branched intermediate by transesterification involving attack by the side chain of Cys455 at the downstream splice junction on the thioester formed in Step 1: Step 3: Excision of the intein by peptide bond cleavage coupled to succinimide formation involving Asn454 at the C-terminus of the intein; Step 4: Spontaneous S-N acyl rearrangement of the transitory ligation product from a thioester to a stable amide bond.



*Figure 5:* The proposed mechanism of protein splicing and cleavage reactions involving the 43 intein from Saccharomyces cerevisiae VMA1 gene.

# Appendix III: The Chemical Mechanism of Thiol-inducible Cleavage Reactions

Our understanding of the mechanism of protein splicing has enabled us to modulate the protein splicing reactions and convert protein splicing into efficient, controllable peptide bond cleavage at both termini of the intein (1–5). The IMPACT-CN System utilizes an intein from the *Saccharomyces cerevisiae* VMA1 gene (the VMA intein).

In the C-terminal fusion vectors (pTYB1 and pTYB2), the C-terminus of a target protein is fused to the N-terminus (Cys1) of the intein which in turn links to the chitin-binding domain (CBD). The C-terminal residue (Asn454) of the intein has been mutated to an alanine. This blocks the splicing reaction but still allows an N-S acyl rearrangement to occur at the intein N-terminus (Cys1) resulting in the formation of a thioester linkage between the target protein and the intein. Cleavage of the thioester bond can be induced by thiol reagents, such as 1,4-dithiothreitol (DTT),  $\beta$ -mercaptoethanol or cysteine. Use of DTT or  $\beta$ -mercaptoethanol results in the formation of a thioester bond and the C-terminal residue of the target protein. This thioester is not stable and hydrolyzes to yield a free C-terminus. However, use of cysteine results in an attached C-terminal cysteine. Cysteine initially forms a thioester with the target protein but a spontaneous S-N shift leads to the formation of a peptide bond (1) (Figure 6).

In the N-terminal fusion vectors (pTYB11 and pTYB12), the N-terminus of a target protein is fused to the C-terminus (Asn454) of the intein. The CBD is inserted in a loop region of the intein without affecting its splicing and cleavage activities. A sequence containing the first 10 residues of the maltose-binding protein is used as the N-extein sequence to provide a favorable translational start for the fusion protein. The intein contains a single substitution which changes its penultimate histidine residue (His453) to an glutamine. This substitution attenuates the succinimide formation by the adjacent intein C-terminal residue (Asn454) which, in conjunction with a substitution at the first C-extein residue, allows inducible cleavage at both termini of the intein (3). In the presence of thiol reagents, such as DTT,  $\beta$ -mercaptoethanol or cysteine, the target protein along with the short N-extein sequence are cleaved from the intein tag and eluted from the chitin column. This N-extein sequence encodes a peptide of 15 residues (MKIEEGKLVIGSLEG) which has a molecular weight of ~1.6 kDa and can be separated from the co-eluted target protein by dialysis (2) (Figure 6).



**Figure 6:** The proposed chemical mechanism of the thiol-inducible cleavage reactions catalyzed by the modified Sce VMA intein. Cleavage occurs at the N-terminus of the Sce VMA intein containing the Asn454Ala substitution (which blocks cleavage at the C-terminus in pTYB1 and pTYB2). Cleavage occurs at both termini of the Sce VMA intein containing the His453Gln substitution in pTYB11 and pTYB12.

# Appendix IV: Protein Labeling and Intein-mediated Protein Ligation

#### Labeling of Target Protein

The C-terminus of the target protein can be covalently labeled using L-[<sup>35</sup>S]-cysteine or a biotinylated synthetic peptide with an N-terminal cysteine immediately following thiol-induced on-column cleavage and elution (1,8). The thiol utilized for cleavage is typically DTT, however another thiol, 2-mercaptoethanesulfonic acid (MESNA) can be used to increase the percentage of labeled product (see below). The following protocol illustrates a typical labeling experiment. Mix 4 µ d f L-[<sup>35</sup>S]-cysteine (11.0 mCi/ml, 0.0102 µmol/ml, available from NEN) or 4 µ d f biotinylated peptide (10 mM, NH<sub>2</sub>-CDPEK\*DS-COOH, the biotin was incorporated as biotinylated lysine during synthesis with a 36 µ aliquot of the freshly purified protein sample and incubate the reaction mixture (pH 8.0) at 4°C overnight. To examine the labeled protein, add 20 µ d f 3X SDS Sample Buffer (with DTT) to the protein sample can be analyzed by SDS-PAGE and autoradiography. The biotinylated protein sample can be analyzed by SDS-PAGE and autoradiography. The biotinylated protein sample can be analyzed by a Western blot with anti-biotin antibody (Cell Signaling Technology #7051-2 or #7071-2).

The use of MESNA as the thiol to induce on-column cleavage has been found to increase the percentage of total labeled protein (8). However, MESNA is not as efficient at causing on-column cleavage as DTT and so the final yield (labeled and unlabeled) of protein with MESNA may be lower. If high specific activity or high biotin incorporation is required, the use of MESNA is recommended.

#### Intein-meditated Protein Ligation (IPL)

The IPL reaction allows the ligation of a synthetic peptide or a protein with an N-terminal cysteine residue to the C-terminus of a bacterially expressed protein through a native peptide bond (8, 9). The IPL protocol employs the IMPACT vectors (pTYB1-4, pTXB1, 3 or pTWIN1,2) to express and purify a target protein and to generate a thioester at its C-terminus. The reaction applies the chemistry described for "native chemical ligation" which fuses two synthetic peptides when the N-terminal cysteine of one peptide attacks a C-terminal thioester of another peptide [Dawson et al., (1994) Science 266, 776–779; Tam et al. (1995) Proc. Natl. Acad. Sci. USA 92, 21485– 12489]. This technique has also been described as "expressed protein ligation" [Muir et al. (1998) Proc. Natl. Acad. Sci. USA 95, 6705-6710; Severinov and Muir (1998) J. Biol. Chem. 273, 16205-16209] using pTYB vectors. Our studies have indicated that pTXB vectors which contain an intein (198 residues) from the gvrA gene of Mycobacterium xenopi are more suitable for IPL because they cleave more proficiently with thiol reagents that are best for ligation, such as 2-mercaptoethanesulfonic acid (MESNA) and thiophenol (8). Cleavage of fusion proteins expressed from a pTYB vector may require a higher incubation temperature with these thiol reagents. However, it may be advantageous to express a target protein using different vectors which contain the same restriction sites in the multiple cloning regions and examine both constructs for expression and purification. The IMPACT-TWIN Kit (NEB #E6950S) is available from NEB for protein purification, IPL and cyclization. Visit <www.neb.com> for more IPL information.

# Appendix V: Research Use Assurance Statement

The buyer and user have a non-exclusive sublicense to use this system or any component thereof for RESEARCH PURPOSES ONLY, based upon agreement to the following assurances.

Transfer of the host cells that contain the cloned copy of the T7 gene 1 to third parties is explicitly prohibited. This limitation applies to strain ER2566 which is provided in combination with the IMPACT-CN System or in combination with appropriate vectors for said system.

A license to use this system or any components thereof for commercial purposes may be obtained from New England Biolabs, Inc.

Commercial Laboratory Buyer and User: Use of the host cells that contain the cloned copy of T7 gene 1, the gene for T7 RNA Polymerase, for any purpose other than in combination with the IMPACT-CN System is explicitly prohibited.

Use of the host cells that contain the cloned copy of the T7 gene 1, the gene for T7 RNA Polymerase, <u>with any other vector(s) containing a T7 promoter</u> to direct the production of RNA or protein requires a license from Brookhaven National Laboratory. Information about research-use license agreements may be obtained from the Office of Technology Transfer, Brookhaven National Laboratory, Building 475D, P.O. Box 5000, Upton, New York, 11973-5000; telephone: 516-344-7134, fax: 516-344-3729.

You may refuse this non-exclusive research license agreement by returning the enclosed materials unused. By keeping or using the enclosed materials, you agree to be bound by the terms of this sublicense.

#### Kit Components Sold Separately:

Α 10 μg		
Α 10 μg		
NA 10 μg		
NA 10 μg		
lasmid DNA 10 μg		
<i>E. coli</i> Strain ER2566* #E4130S 0.2 ml * Only available to purchasers of IMPACT Systems or replacement vectors.		
20 ml 100 ml		
Anti-Chitin Binding Domain Serum #S6654S 50 µl		
er 0.5 A <sub>260</sub> units		
mer 0.5 A <sub>260</sub> units		
imer 0.5 A <sub>260</sub> units		
verse Primer 0.5 A <sub>260</sub> units		

### **Companion Products:**

pTWIN1 Vector DNA #N6951S 10 µg pTWIN2 Vector DNA #N6952S 10 µg pTWIN-MBP1 Vector DNA #N6953S 10 µg pTYB3 Vector DNA #N6703S 10 µg pTYB4 Vector DNA #N6704S 10 µg pTXB1 Vector DNA #N6707S 10 µg pTXB3 Vector DNA #N6708S 10 µg pKYB1 Vector DNA #N6706S 10 µg Mxe Intein Reverse Primer 0.5 A<sub>260</sub> units #S1268S Ssp DnaB Intein Forward Primer #S1269S 0.5 A260 units Mth BIB1 Intein Reverse Primer #S1270S 0.5 A<sub>260</sub> units Blue Loading Buffer Pack #B7703S

#### New England Biolabs, Inc.

32 Tozer Road Beverly, MA 01915-5599 USA Telephone (978) 927-5054 Toll Free (USA orders) 1-800-632-5227 Toll Free (USA Tech) 1-800-632-7799 Fax (978) 921-1350 e-mail: info@neb.com www.neb.com

#### New England Biolabs, Ltd.

3397 American Drive, Unit 12 Mississauga, Ontario, L4V 1T8 Canada Telephone (905) 672-3370 Toll Free 1-800-387-1095 Fax (905) 672-3414 Fax Toll Free 1-800-563-3789 e-mail: info@ca.neb.com

#### New England Biolabs GmbH

Brüningstrasse 50, Geb.G 810 65926 Frankfurt am Main Germany Telephone (0)69/305 23140 Free Call 0800/246 5227 (0800 BIOLABS) Fax (0)69/305 23149 Free Fax 0800/246 5229 (0800 BIOLABX) e-mail: info@de.neb.com

#### New England Biolabs (UK) Ltd.

73 Knowl Piece Wilbury Way Hitchin, Hertfordshire SG4 0TY England, UK Telephone (01462) 420616 Call Free 0800 318486 Fax (01462) 421057 Fax Free 0800 435682 e-mail: info@uk.neb.com

