



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A dual protease approach for expression and affinity purification of recombinant proteins



Sreejith Raran-Kurussi, David S. Waugh*

Protein Engineering Section, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

ARTICLE INFO

Article history:

Received 8 February 2016

Received in revised form

1 April 2016

Accepted 7 April 2016

Available online 19 April 2016

Keywords:

Maltose-binding protein

Tobacco etch virus protease

Rhinovirus 3C protease

His-tag

Affinity chromatography

ABSTRACT

We describe a new method for affinity purification of recombinant proteins using a dual protease protocol. *Escherichia coli* maltose binding protein (MBP) is employed as an N-terminal tag to increase the yield and solubility of its fusion partners. The MBP moiety is then removed by rhinovirus 3C protease, prior to purification, to yield an N-terminally His₆-tagged protein. Proteins that are only temporarily rendered soluble by fusing them to MBP are readily identified at this stage because they will precipitate after the MBP tag is removed by 3C protease. The remaining soluble His₆-tagged protein, if any, is subsequently purified by immobilized metal affinity chromatography (IMAC). Finally, the N-terminal His₆ tag is removed by His₆-tagged tobacco etch virus (TEV) protease to yield the native recombinant protein, and the His₆-tagged contaminants are removed by adsorption during a second round of IMAC, leaving only the untagged recombinant protein in the column effluent. The generic strategy described here saves time and effort by removing insoluble aggregates at an early stage in the process while also reducing the tendency of MBP to “stick” to its fusion partners during affinity purification.

Published by Elsevier Inc.

Rapid advances in genomics and proteomics during the past three decades have revolutionized the fields of biotechnology and human medicine, particularly when recombinant DNA technology joined hands with structural biology. Currently, samples of proteins for structural and functional studies are routinely obtained by bioengineering [1]. Even so, protein purification remains the principal bottleneck. Conventional methods of protein purification have been almost completely supplanted by affinity-based methods that employ protein or peptide affinity tags [2–4]. The popularity of these affinity-based methods can be attributed to their generic nature in comparison with traditional approaches, which are rather protein specific. Other purification platforms such as ion exchange, hydrophobic interaction, and size exclusion chromatography are

used as auxiliary steps to further enhance the purity of the sample if necessary.

In our laboratory and many others, a dual His₆–MBP (maltose-binding protein) tag is used in an initial immobilized metal ion affinity chromatography (IMAC) step [5–7]. The reason for employing the dual tag has been discussed in detail elsewhere [8]. Briefly, MBP enhances the solubility and improves the yield of its fusion partners during overproduction but is not a particularly effective affinity tag for protein purification. Hence, the His₆ tag is included to allow affinity chromatography by IMAC. The N-terminal His₆–MBP tags are subsequently removed by tobacco etch virus (TEV) protease to generate a tag-free protein. Although in general this approach has been very successful, it is not without its problems. For instance, a significant fraction of aggregation-prone proteins that are rendered soluble by fusing them to MBP subsequently precipitate when the fusion proteins are cleaved by TEV protease [9,10]. These proteins presumably either are not properly folded or exist as soluble aggregates in partially folded forms. Usually it cannot be ascertained whether or not this will be a problem until after affinity purification. A second potential pitfall is the tendency of some proteins to stick to MBP after TEV digestion, making it difficult to separate them from each other. The interaction between MBP and its fusion partners may be related to the mechanism of solubility enhancement [11]. In these situations, it is usually necessary to employ an MBPTrap (or

Abbreviations used: MBP, maltose-binding protein; IMAC, immobilized metal ion affinity chromatography; TEV, tobacco etch virus; PCR, polymerase chain reaction; ORF, open reading frame; ChikV, Chikungunya virus; DHFR, dihydrofolate reductase; DUSP14, dual-specificity phosphatase 14; MERS-CoV 3CL^{pro}C148A, Middle East respiratory syndrome coronavirus 3C-like protease; GFP, green fluorescent protein; RBS, ribosome-binding site; IPTG, isopropyl β-D-1-thiogalactopyranoside; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TMV, tobacco vein mottling virus.

* Corresponding author.

E-mail address: waughd@mail.nih.gov (D.S. Waugh).

amylose) column, or another IMAC step to remove the MBP, thereby making the purification process more labor intensive. In the current study, we attempted to circumvent both of these problems by using a dual protease approach to achieve sequential removal of the MBP and His₆ tags.

Materials and methods

Materials

pBAD24–sfGFPx1 was a gift from Sankar Adhya and Francisco Malagon (Addgene plasmid no. 51558) [12]. The 3C protease expression vector pET/3C was a gift from Ari Geerloff (EMBL, Heidelberg, Germany). The pBLN200–GFPmut2–Car9 plasmid was a gift from François Baneyx [13].

All chemicals of the highest available purity were purchased from Sigma–Aldrich (St. Louis, MO, USA), American Bioanalytical (Natick, MA, USA), Thermo Fisher Scientific (Rockford, IL, USA), Roche Diagnostics (Indianapolis, IN, USA), or EMD Millipore (Billerica, MA, USA) unless otherwise stated. Restriction endonucleases were obtained from New England Biolabs (Ipswich, MA, USA). Fast-Link T4 DNA Ligase was purchased from Epicentre (Madison, WI, USA). All polymerase chain reactions (PCRs) were carried out using either the PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies, Santa Clara, CA, USA) or the Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific).

Construction of pSRK2721

The Gateway destination vector pSRK2721 (see Fig. S1 in online supplementary material) was constructed as follows. First, a *Sna*BI restriction site was inserted at the end of the open reading frame (ORF) encoding MBP in pKM596 [14] using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) with primers PE-2852 and PE-2853 (Table 1) to generate the plasmid pSRK2704. In the next step, the C terminus of MBP in pSRK2704 was further extended to include an in-frame 3C protease recognition site followed by a His₆ tag by PCR restriction cloning using PE-42, PE-2886, and PE-2887 primers (Table 1). The PE-42 primer was designed to anneal just proximal to the *Bgl*III site in the *malE* gene, whereas PE-2886 and PE-2887 were partially overlapping reverse primers that were employed to engineer the desired insertions (3C site and His₆ tag) followed by a *Sna*BI restriction site. The PCR amplicon, consisting of a C-terminal portion of the MBP ORF from pKM596 joined in-frame to a sequence encoding GGGLEVLFQ/GPHHHHHHYV, was digested with *Bgl*III and *Sna*BI and then cloned between the same sites in pSRK2704 to yield the destination vector pSRK2721 (Table 2). The “YV” residues shown above are cloning

artifacts derived from the *Sna*BI site. The rhinovirus 3C protease recognition site is underlined, and the cleavage site is marked by a forward slash (see above).

Construction of entry clones

The sfGFP ORF was amplified from pBAD24–sfGFPx1 [12] by a single PCR using primers PE-277 (Forward-1), PE-2829 (Forward-2), and PE-2830 (Reverse) (Table 1). The PCR product contained the N- and C-terminal attB recombination sites and a TEV protease recognition site appended to the N terminus of the sfGFP ORF. The PCR amplicon was used in the Gateway BP reaction with pDONR221 (Thermo Fisher Scientific) to generate the entry clone pSRK2661 (Table 2). Similarly, the Chikungunya virus (ChikV) protease ORF was amplified from a ChikV cDNA clone using the primers PE-277 (Forward-1), PE-2645 (Forward-2) and PE-2646 (Reverse) (Table 1). The PCR amplicon was used in the Gateway BP reaction with pDONR221 to generate the entry clone pKK2483 (Table 2). Additional entry clones encoding *Francisella tularensis* IgLc, human dihydrofolate reductase (DHFR), human dual specificity phosphatase 14 (DUSP14), the human SUMO-conjugating enzyme Ubc9, catalytically inactive Middle East respiratory syndrome coronavirus 3C-like protease (MERS-CoV 3CL^{pro}C148A), and wild-type green fluorescent protein (GFP) were constructed as described previously [8,15–17].

Construction of fusion protein expression vectors

The destination vector (pSRK2721) was recombined in a Gateway LR reaction (Thermo Fisher Scientific) with various entry clones to generate the dual protease expression vectors for further study. All of the expression vectors are listed in Table 2. A schematic representation of the MBP fusion proteins they produce is shown in Fig. 1A. The second protease recognition site corresponds to that of TEV protease, which was incorporated into the entry clones.

Construction of an IPTG-inducible, untagged 3C protease expression vector

Gateway recombinational cloning (Thermo Fisher Scientific) was used to make an untagged rhinovirus 3C protease expression vector for co-lysis experiments. The 3C protease ORF was amplified by PCR from pET/3C using primers PE-2856 and PE-2857 (Table 1). These primers incorporated the attB recombination sites along with an appropriately positioned ribosome-binding site (RBS) in the PCR product. Subsequently, the PCR product was used in a BP reaction with pDONR221 (Thermo Fisher Scientific) to generate the entry clone pSRK2703. The entry clone was recombined in an LR reaction with pDEST42 (Thermo Fisher Scientific) to generate the

Table 1
Primer sequences.

Primer	Sequence (5'–3')
PE-2852	GAA AGA CGC GCA GAC TAA TTC GTA CGT AAT CAC AAG TTT GTA CAA AAA AGC
PE-2853	GCT TTT TTG TAC AAA CTT GTG ATT ACG TAC GAA TTA GTC TGC GCG TCT TTC
PE-42	GGC ACA CGA CCG CTT TGG TGG CTA C
PE-2886	ATG CAT TAC GTA GTG ATG GTG ATG GTG ATG CGG ACC CTG GAA CAG AAC TTC
PE-2887	ACC CTG GAA CAG AAC TTC CAG ACC ACC ACC CGA ATT AGT CTG CGC GTC TTT C
PE-277	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA GAA CCT GTA CTT CCA G
PE-2829	GAG AAC CTG TAC TTC CAG ATG CGT AAA GGC GAA GAG
PE-2830	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATT TGT ACA GTT CAT CCA TAC
PE-2645	GGC TCG GAG AAC CTG TAC TTC CAG AGT AAC GCA TTC CAA AAC AAA GCC AAC GTT TGT
PE-2646	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATC CTA CAA AGG CTG CAT TCA GTT GAT
PE-2856	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT TAA GAA GGA GAT ATA CAT ATG GGA CCA AAC ACA GAA TTT G
PE-2857	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATT TTT TCT CTA CAA AAT A
pSAMRFFwd	CAA AGG ATC TTC TTG AGA TCC TGG CTT CTG TTT CTA TCA GCT GTC C
pSAMRFRRev	CGT GAG CAT CCT CTC TCG TTT CGC GGG GCA TGA CTA ACA TGA GAA

Table 2
Plasmids used in this study.

Plasmid name	Vector type/gene	Reference
pSRK2721	MBP–3C site–His ₆ destination vector	This work
pSRK2661	sfGFP entry clone	This work
pKK2483	ChikV protease entry clone	This work
pDN2349	Ubc9 entry clone	[28]
pSN1751	IgIC entry clone	[8]
pSRK2210	DHFR entry clone	[11]
pDUSP14 ^a	DUSP14 entry clone	[15]
pDN2482	MERS 3CL ^{Pro} entry clone	[16]
pKM1122	wt-GFP entry clone	[17]
pSRK2703	3C protease entry clone	This work
pSRK2730	IgIC expression	This work
pSRK2746	sfGFP expression	This work
pSRK2729	DHFR expression	This work
pSRK2728	DUSP14 expression	This work
pSRK2788	Ubc9 expression	This work
pSRK2789	MERS-CoV 3CL ^{Pro} expression	This work
pSRK2792	wt-GFP expression	This work
pSRK2795	ChikV protease expression	This work
pSRK2738	pET-24a(+)-derived vector with p15A copy control and L-arabinose inducible	This work
pSRK2751	3C protease expression (L-arabinose inducible; p15A copy control)	This work
pSRK2706	3C protease expression (IPTG inducible)	This work

^a Encodes residues 2 to 191 of DUSP14 inserted into pDONR221.

IPTG (isopropyl β -D-1-thiogalactopyranoside)-inducible 3C protease expression vector pSRK2706 (Table 2).

Construction of an L-arabinose-inducible, untagged 3C protease expression vector

A kanamycin-resistant plasmid with a p15A origin of replication and an L-arabinose-inducible promoter was created as described

previously with slight variations [18]. In the initial step, the desired insertion sequence (p15A ori) was PCR-amplified using a high-fidelity DNA polymerase. The forward and reverse primers for creating a “p15A mega primer” were pSAMRFFwd and pSAMRFRev, respectively (Table 1). The reaction setup, cycling method, and template DNA were the same as described previously [18]. This mega primer amplicon was used as complementary primers in a subsequent QuikChange Lightning reaction (Agilent Technologies)

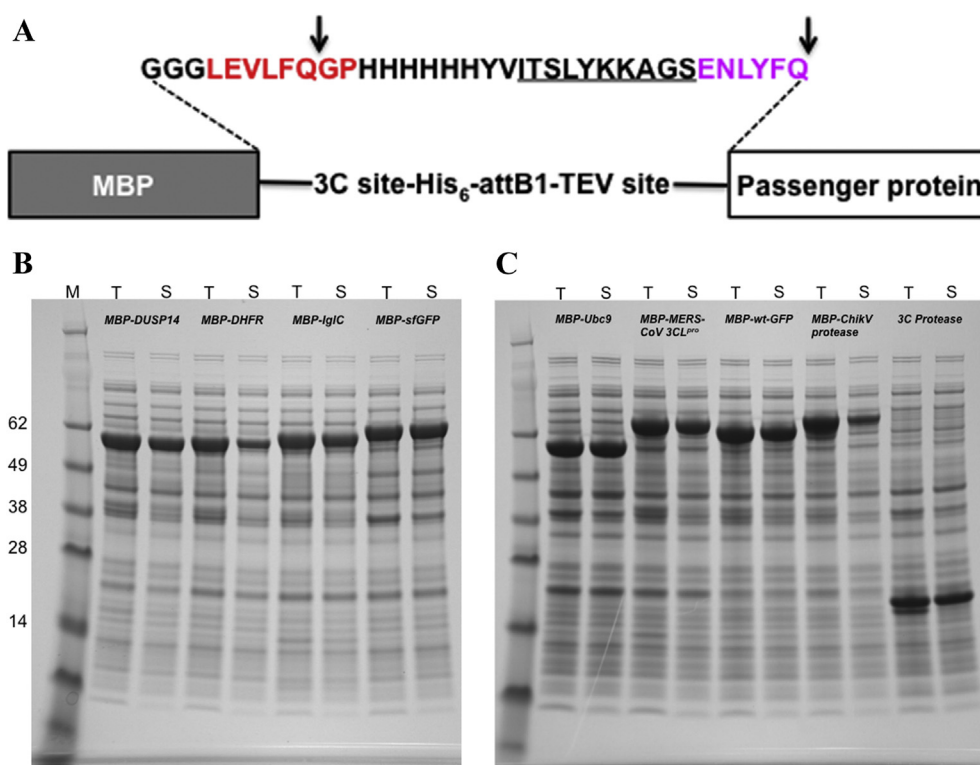


Fig. 1. Design of fusion proteins in dual protease format. (A) A schematic representation of the MBP fusion proteins (not to scale). The red and magenta letters show the recognition sequences for 3C protease and TEV protease, respectively. The downward arrows show the sites of protease cleavage. The residues encoded by the attB1 recombination site are underlined. (B,C) Solubility of MBP fusion proteins. SDS-PAGE analysis of total (T) and soluble (S) proteins from the cells expressing MBP fusion proteins. The last two lanes in panel C illustrate 3C protease solubility. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

to bring about the desired change in the target plasmid. The QuikChange Lightning reaction mix consisted of the target DNA template (50 ng) and the mega primer amplicon (350 ng) along with other PCR components. The cycling method was as follows: 95 °C for 30 s (segment 1); 95 °C for 30 s, 52 °C for 1 min, 68 °C for 7 min (segment 2 for 5 cycles); 95 °C for 30 s, 55 °C for 1 min, 68 °C for 7 min (segment 3 for 13 cycles). The target plasmid in the QuikChange reaction was pBLN200–GFPmut2–Car9 [13] in our reaction mix instead of pET-28a(+) in the earlier method [18]. The pBLN200–GFPmut2–Car9 is a derivative of pET-24a(+) and carries the arabinose-inducible P_{BAD} promoter and the *araC* gene instead of the T7 promoter. Thus, effectively the ColE1 replicon in pBLN200–GFPmut2–Car9 was replaced with the p15A replicon. The modified vector, pSRK2738 (Table 2), was confirmed by PCR and DNA sequencing. Induction of the vector with l- arabinose was confirmed by monitoring the fluorescence of the GFPmut2 protein in *E. coli* BL21(DE3) cells (Agilent Technologies). Next, the 3C protease ORF, bracketed by *NdeI* and *XhoI* sites, was amplified by PCR, cleaved with these restriction enzymes, and inserted between the same sites in pSRK2738 to generate the l- arabinose-inducible, untagged 3C protease expression vector pSRK2751 (Table 2).

Protein expression and purification

E. coli BL21–CodonPlus(DE3)–RIL cells (Agilent Technologies) were used for all protein expression experiments except coexpression experiments involving two different plasmids, in which case BL21(DE3) cells were used instead.

In protein expression experiments, cells were grown with shaking at 37 °C in Luria broth containing the appropriate antibiotics and 0.2% glucose to mid-log phase, at which point overexpression of the fusion proteins was induced at 30 °C for 4 h by the addition of 1 mM IPTG. In cells coexpressing a fusion protein with 3C protease from pSRK2751, 0.01–0.2% l- arabinose was added after a delay of 3 h following IPTG induction of the MBP fusion proteins at 30 °C. The culture was incubated for an additional 1 h after the addition of l- arabinose (total 4-h induction). The induced cells were pelleted by centrifugation and stored at –80 °C.

To assess the solubility of recombinant proteins, *E. coli* cells overexpressing MBP fusions or 3C protease (produced from pSRK2706) were cultured separately and resuspended in approximately 0.2–0.4 culture volumes of ice-cold sonication buffer (50 mM Tris–HCl [pH 8.0] and 150 mM NaCl). The cells were lysed by sonication, and the insoluble material was pelleted by centrifugation. Samples of the total and soluble fractions were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). The cells were normalized to an OD₆₀₀ of 10.00 (arbitrary setting) in sonication buffer before lysis. To check the 3C protease activity, equal volumes of the soluble extracts from cells expressing an MBP fusion protein and 3C protease were combined and incubated for 15 min at 4 °C. Following the incubation, the supernatants were analyzed by SDS–PAGE before and after centrifugation. In this manner, the propensity of the passenger protein to precipitate after protease digestion (i.e., after separation from MBP) could be monitored prior to affinity chromatography. The titration experiment to determine the optimal amount of 3C cells to be used in the co-lysis protocol was determined by varying the amount of soluble extract containing 3C protease in the above experiment.

Large-scale protein purifications were performed at 4–8 °C. For the combined cell lysis (co-lysis) method of protein purification, *E. coli* cell pastes of MBP fusion protein (~5.0 g) and 3C protease (~0.5 g) were suspended together in 150 ml of buffer A (50 mM Tris [pH 7.6], 200 mM NaCl, and 5% [v/v] glycerol) containing 25 mM imidazole and Complete EDTA (ethylenediaminetetraacetic acid)-

free protease inhibitor cocktail tablets (Roche Diagnostics). The mixture of cell suspensions was lysed with an APV-1000 homogenizer (Invensys APV Products, Albertslund, Denmark) at 69 MPa and centrifuged at 30,000 g for 30 min. The supernatant was filtered through a 0.2- μm polyethersulfone membrane and applied onto a 5-ml HisTrap FF column (GE Healthcare Life Sciences). The column was washed to baseline with buffer A containing 25 mM imidazole and eluted with a linear gradient of imidazole in buffer A to 250 mM. The column effluent contained the MBP and 3C protease. The eluted fractions containing the His₆-tagged proteins were pooled and concentrated using a 10-kDa MWCO (molecular weight cutoff) membrane (Millipore). The concentrated protein was diluted with buffer A to reduce the imidazole concentration to approximately 25 mM and was digested overnight with His₆-tagged TEV protease [19] at 4 °C. The digest was applied to a 5-ml HisTrap FF column to capture the cleaved His₆ tag and the His₆ TEV. The column effluent contained the pure protein of interest. The effluent was concentrated and incubated overnight with 10 mM dithiothreitol at 4 °C. The reduced sample was loaded onto a HiPrep 26/60 Sephacryl S-200 HR column (GE Healthcare Life Sciences) that was equilibrated with 25 mM Tris (pH 7.6), 150 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine (TCEP), and 5% (v/v) glycerol.

Results

Rationale and vector design

The fusion proteins used in this study have the following modular architecture: MBP tag–3C protease site–His₆ tag–attB1 site–TEV protease site–passenger protein (Fig. 1A). The purpose of the MBP tag is to enhance the solubility and promote the proper folding of the passenger protein [4,11,20]. The 3C protease site allows the MBP moiety to be separated from the His₆-tagged passenger protein prior to IMAC. This has two benefits. First, it enables one to determine whether or not the passenger protein will precipitate when it is released from MBP before any chromatographic steps are performed. Second, it helps to alleviate the propensity of MBP to stick to and copurify with some passenger proteins, probably because such an interaction is less likely to persist once the two fusion partners are no longer covalently attached to one another. After IMAC purification of the His₆-tagged passenger protein, the His₆ tag is removed by His₆-tagged TEV protease. In the final step, the cleaved His₆ tag and the TEV protease are adsorbed by another IMAC step, leaving just the pure untagged passenger protein in the column effluent.

The removal of MBP by 3C protease can be accomplished in either of two ways. In one approach, *E. coli* cell paste containing an overproduced fusion protein is mixed with cell paste containing 3C protease and lysed together. The 3C protease, which is highly active at low temperature [21], will rapidly cleave the fusion protein during processing of the sample. Alternatively, 3C protease can be coexpressed with the fusion protein substrate. For this purpose, an arabinose-inducible 3C protease vector is used, enabling the production of 3C protease to be controlled independently of the IPTG-inducible fusion protein expression vector. The ability to delay the induction of 3C protease expression is an important advantage because prolonged association of MBP with its passenger proteins stimulates their proper folding [22].

Solubility analysis of MBP fusions with dual protease recognition sites

Studies have shown that the length and sequence of the linker that connects MBP to its passenger proteins can have a profound

impact on the solubility of MBP fusion proteins [10,23]. Hence, it was important to assess the solubility of the MBP fusion proteins with a linker composed of a 3C protease cleavage site followed by a His₆ tag, an attB1 recombination site (an artifact of Gateway cloning), and a TEV protease site, respectively. Except for sfGFP, Ubc9, and MERS-CoV 3CL^{Pro}, all of the other passenger proteins used in this study are poorly soluble when expressed in an unfused form or as GST fusion proteins in *E. coli* [8,14,20,24]. As shown in Fig. 1B and C, all of the fusion proteins with the modified linker sequence were expressed at a very high level, and most of them were also highly soluble. The partial solubility of the DHFR and ChikV protease fusion proteins is comparable to that of their counterparts with conventional linkers [11,25]. The untagged 3C protease solubility is shown in Fig. 1C as well.

Cleavage of fusion proteins by 3C protease

To assess the ability of 3C protease to cleave fusion protein substrates in crude cell extracts, equal volumes of soluble extracts from cells expressing an MBP fusion protein and 3C protease were combined and incubated for 15 min at 4 °C. The results confirmed that the fusion proteins were cleaved nearly to completion during the short incubation period (Fig. 2A and B). On prolonged incubation, the 3C protease also cleaved the TEV recognition site in the MBP fusion proteins (see [supplementary material](#)). This observation was unexpected considering the reportedly strict requirement for Gly and Pro residues in the P1' and P2' positions, respectively, of a 3C protease recognition site [26]. In a subsequent titration experiment, it was determined that an *E. coli* cell weight ratio of 10:1 (fusion protein cells/3C protease cells) can remove the MBP tag efficiently without causing any nonspecific cleavage at the TEV site (Fig. 3). Thus, generally speaking, 1 L of cells expressing 3C protease is required to cleave the MBP fusion protein in cells obtained from 10 L of medium.

We also experimented with the processing of MBP fusion proteins inside intact cells by coexpression of 3C protease. For this purpose, we used a 3C protease expression vector that can be induced by L-arabinose independently of the IPTG-inducible MBP fusion protein expression vector. As shown in Fig. 4, even in the absence of L-arabinose, enough 3C protease was produced to cleave the majority of the MBP–sfGFP fusion protein. Nearly complete digestion was obtained in the presence of 0.01% arabinose. Higher concentrations of L-arabinose led to unintended cleavage of the TEV

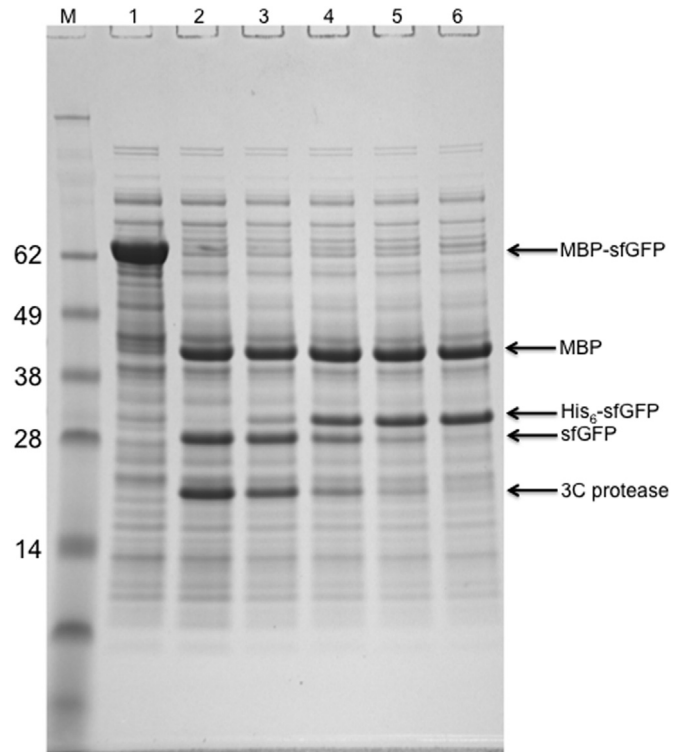


Fig. 3. 3C protease titration. A titration analysis using the MBP–sfGFP protein and 3C protease was performed. Lane 1: uncut MBP–sfGFP; lanes 2 to 6: samples with increasing substrate-to-enzyme (S/E) ratios, that is, S/E = 1 (lane 2), 2 (lane 3), 5 (lane 4), 10 (lane 5), and 20 (lane 6).

protease recognition site by 3C protease. Cleavage of the TEV protease site by 3C protease was confirmed by mass spectrometry (see [Supplementary Table 1](#)).

A dual protease strategy for protein purification

The dual protease protocol for protein purification consists of four main steps. First, cells expressing 3C protease are combined with cells expressing an appropriately designed MBP fusion protein

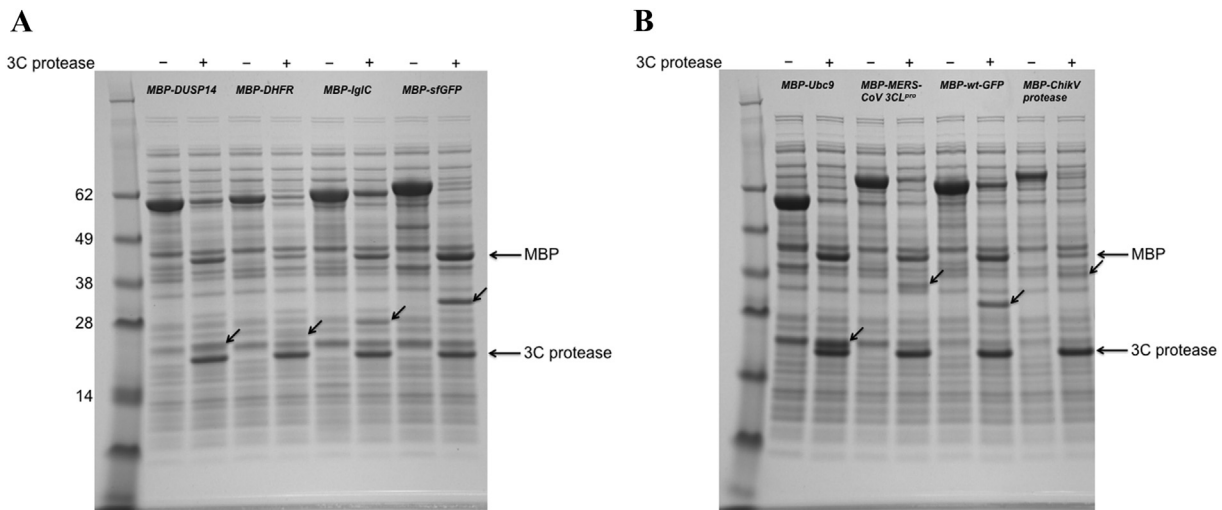


Fig. 2. SDS–PAGE analysis of 3C protease cleavage. (A,B) Lysates of cells expressing MBP fusion proteins and 3C protease (separately) were mixed together and incubated (see Materials and Methods). The soluble intracellular proteins, in either the presence (+) or absence (–) of 3C protease, are shown on the gel. The slanting arrows indicate the positions of the cleaved passenger proteins.

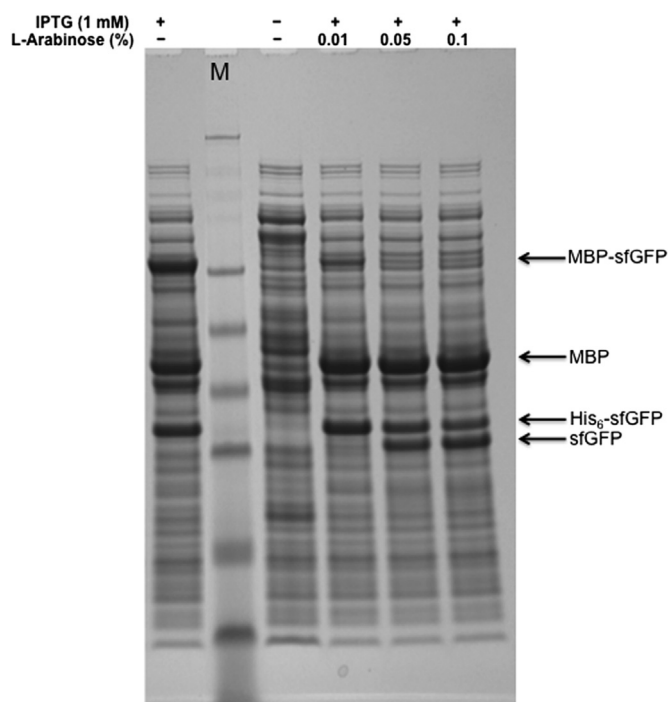


Fig. 4. In vivo (intracellular) processing by L-arabinose induction of 3C protease. The MBP–sfGFP fusion protein was digested in vivo by coexpressing 3C protease to varying degrees as indicated. The presence (+) or absence (–) of IPTG or L-arabinose is marked. The leaky expression of 3C protease causes partial digestion of the MBP–sfGFP (first lane from left). M, molecular weight markers.

(Fig. 1A) and lysed together to cleave off the MBP tag. Second, the soluble His₆-tagged passenger protein is purified by IMAC on a HisTrap column. Third, the affinity-purified protein is digested with His₆-tagged TEV protease to remove the His₆ tag and the translated attB1 region, thereby generating the native protein devoid of any tags and cloning artifacts with the possible exception of an N-terminal glycine residue. Finally, the pure protein is collected in the effluent from a second HisTrap column, which serves to remove the His tag, the TEV protease, and any residual undigested His-tagged passenger protein. This second IMAC step also aids in the removal of any nonspecific proteins that appeared in the first IMAC elution. A gel filtration column is recommended as a final polishing step and a means of removing any soluble aggregates that may be present.

To evaluate the efficacy of this generic protocol, we purified GFP, sfGFP, IgLc, Ubc9, ChikV protease, MERS-CoV 3CL^{Pro}, and DUSP14, all of which were produced as MBP fusion proteins of the type depicted in Fig. 1A. At every step, the purity of the proteins was monitored by SDS–PAGE (Fig. 5). The cleavage products generated by 3C protease and TEV protease were also analyzed by electrospray ionization mass spectroscopy along the purification route. The phosphatase activity or the visible green fluorescence confirmed the proper folding of the purified DUSP14 or GFP, respectively (data not shown). Following a final buffer exchange with a gel filtration column to enhance the homogeneity of the purified proteins, they were judged to be more than 95% pure by SDS–PAGE (Fig. 5). Their molecular weights were confirmed by electrospray ionization mass spectroscopy. The final yield of the purified passenger proteins ranged from 0.5 to 5.0 mg per gram of wet *E. coli* cell weight.

Discussion

The extraordinary ability of MBP to enhance the yield and promote the solubility of its fusion partners is well documented

[14,20,23], but it is not a particularly good affinity tag for protein purification. Therefore, we and others have employed a generic strategy for protein expression and purification that combines the solubility-enhancing benefit conferred by MBP with the powerful advantage of immobilized metal affinity chromatography by using a polyhistidine tag in a tandem configuration with MBP (His₆–MBP) [5–7]. Although frequently successful, this approach has two potential pitfalls. First, many aggregation-prone proteins that can be rendered soluble by fusing them to MBP subsequently precipitate after they are separated from the solubility enhancer. Second, MBP has a tendency to form non-covalent complexes with some passenger proteins, which complicates their downstream purification. The first problem occurs when passenger proteins are unable to fold properly, either spontaneously or with the assistance of endogenous chaperones, once the competing kinetic pathway of aggregation is blocked by fusing them to MBP [11]. Currently, there is no generally effective solution to this problem. However, the method reported here, in which MBP is cleaved from the passenger protein in the crude cell extract, enables many of these “dead end” cases to be spotted very early, prior to purification of the fusion protein by affinity chromatography. Still, occasionally a passenger protein may remain soluble but be biologically inactive after it is released from MBP, possibly because fusion to MBP enables certain proteins to evolve into kinetically trapped folding intermediates that are no longer susceptible to aggregation. Therefore, it is advisable to employ a biological assay, if possible, at an early stage to confirm that the passenger protein is properly folded. The second advantage of cleaving the MBP fusion protein prior to affinity purification is that this reduces the tendency of MBP to “stick” to the cleaved passenger protein during its purification. This characteristic of MBP may be related to the mechanism by which it promotes solubility and inhibits aggregation, which is thought to involve the formation of transient complexes between MBP and partially folded passenger proteins [11].

Although the amino acids surrounding the 3C protease cleavage site are the same in all of the fusion proteins, they were not all cleaved with equal efficiency (Fig. 2A and B). The most likely explanation for this is that some of the fusions exist in the form of soluble aggregates, thereby masking the cleavage site. We have also observed cases in which a His₆–MBP fusion protein binds readily to IMAC (e.g., Ni-NTA) resin, whereas the corresponding His₆ fusion protein generated after 3C cleavage does not. This could be due to obstruction of the His₆ tag when it is present on the N terminus of the passenger protein as opposed to when it is attached to the N terminus of MBP. This is exemplified by DUSP14 (see Fig. S3A in supplementary material). Inefficient binding of a His₆-tagged passenger during IMAC is not necessarily indicative of an unfolded protein, however, because at least some of the unbound His₆ DUSP14 was enzymatically active (data not shown).

Donnelly and coworkers have described a similar dual protease approach using tobacco vein mottling virus (TVMV) and TEV proteases [27]. However, the method presented here has two noteworthy advantages. First, 3C protease is used instead of TVMV protease to separate MBP from the His-tagged passenger protein. 3C protease is superior to TVMV protease because it is highly active at low temperature [21]. Consequently, the reaction proceeds efficiently on ice. Moreover, in Donnelly and coworkers' method the fusion protein is cleaved by TVMV protease inside intact cells. Yet expression of the TVMV protease is constitutive, resulting in the rapid cleavage of MBP fusion proteins following their translation. There is significant evidence to suggest that the longer a protein remains fused to MBP, the more likely it is to fold correctly [22]. Although delayed induction of the 3C protease gene on the plasmid

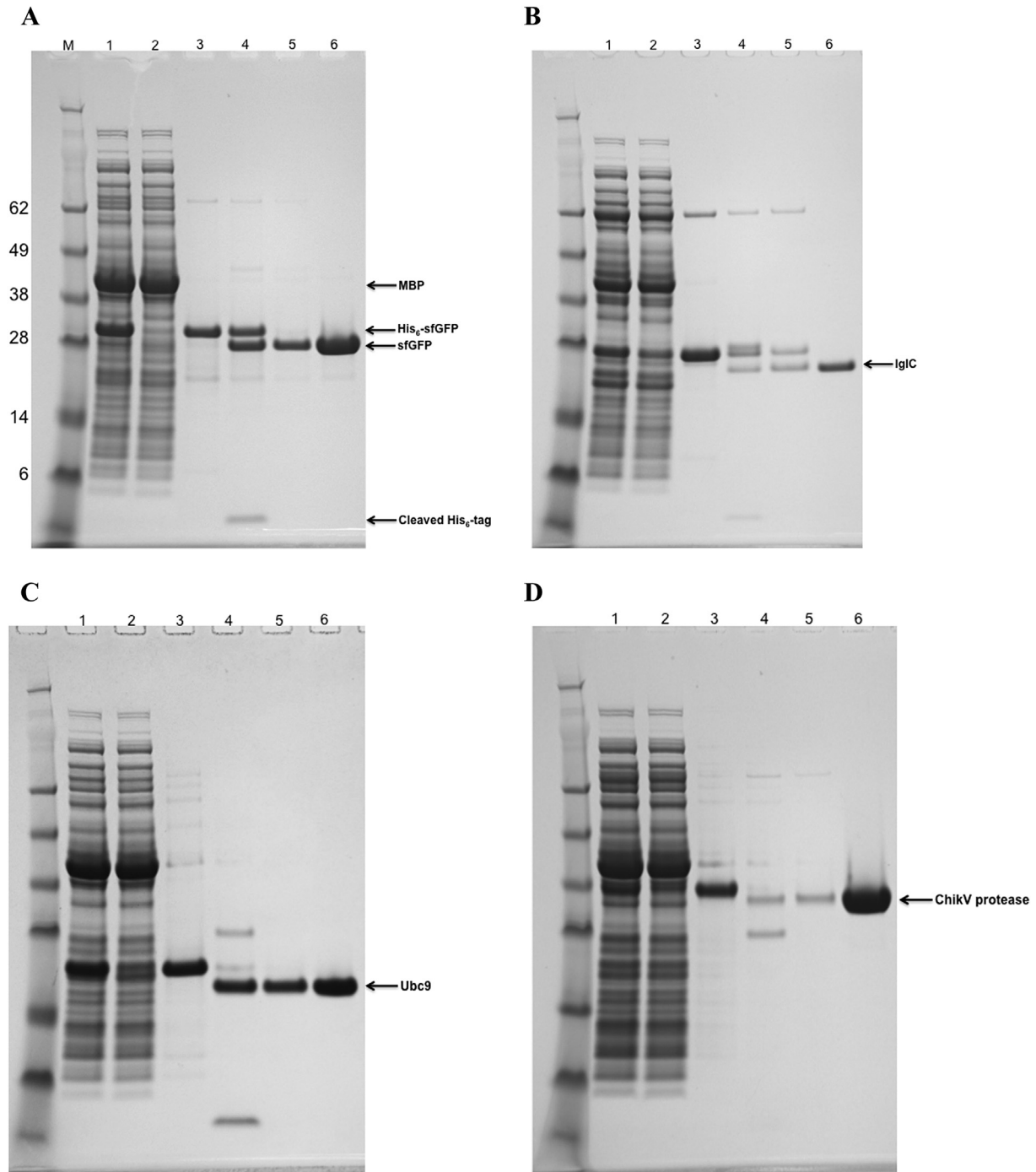


Fig. 5. Protein purification using the dual protease method. Purification was monitored by SDS–PAGE at different stages. In all four panels (A–D), the following gel-loading pattern applies. Lane 1: soluble lysate (crude); lane 2: flow-through from first IMAC column (unbound); lane 3: eluate from first IMAC column; lane 4: products of TEV protease digest; lane 5: flow-through from second IMAC column; lane 6: protein after final gel filtration column. The passenger proteins are marked in all panels. The pattern of digestion products is marked in panel A.

pSRK2751 with arabinose was intended to solve this problem, unfortunately enough protease was produced in the absence of inducer to cleave a substantial portion of a well-expressed MBP–sfGFP fusion protein *in vivo* (Fig. 4). Hence, we recommend employing the novel co-lysis method of 3C protease cleavage instead.

The dual protease strategy for protein expression and purification described here is simple and universally applicable. For example, MBP fusion proteins expressed in insect cells or other

hosts could be co-lysed with *E. coli* cells containing 3C protease if desired; it is not necessary to develop a means of coexpressing 3C protease in each different host. Moreover, although the attrition rate may seemingly appear higher than that associated with the use of a dual His₆–MBP tag and a single protease, proteins that cannot be purified using the dual protease approach very likely would also prove to be problematic when the more conventional approach is employed, only at a later stage and after a considerably greater expenditure of time and effort.

Acknowledgments

We thank Karina Keefe and Danielle Needle for constructing the ChikV protease and the Ubc9 entry clones, respectively. This project was funded by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. We thank the Biophysics Resource in the Structural Biophysics Laboratory, Frederick National Laboratory, for assistance with electrospray ionization mass spectroscopy. The authors declare that they have no conflicts of interest.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ab.2016.04.006>.

References

- [1] Z.S. Derewenda, The use of recombinant methods and molecular engineering in protein crystallization, *Methods* 34 (2004) 354–363.
- [2] J.J. Lichty, J.L. Malecki, H.D. Agnew, D.J. Michelson-Horowitz, S. Tan, Comparison of affinity tags for protein purification, *Protein Expr. Purif.* 41 (2005) 98–105.
- [3] K. Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, *Appl. Microbiol. Biotechnol.* 60 (2003) 523–533.
- [4] D.S. Waugh, Making the most of affinity tags, *Trends Biotechnol.* 23 (2005) 316–320.
- [5] S.N. Hewitt, R. Choi, A. Kelley, G.J. Crowther, A.J. Napuli, W.C. Van Voorhis, Expression of proteins in *Escherichia coli* as fusions with maltose-binding protein to rescue non-expressed targets in a high-throughput protein-expression and purification pipeline, *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 67 (2011) 1006–1009.
- [6] X. Jing, J. Jaw, H.H. Robinson, F.D. Schubot, Crystal structure and oligomeric state of the RetS signaling kinase sensory domain, *Proteins* 78 (2010) 1631–1640.
- [7] S. Nallamsetty, D.S. Waugh, A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His₆–maltose binding protein fusion tag, *Nat. Protoc.* 2 (2007) 383–391.
- [8] S. Nallamsetty, B.P. Austin, K.J. Penrose, D.S. Waugh, Gateway vectors for the production of combinatorially-tagged His₆–MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*, *Protein Sci.* 14 (2005) 2964–2971.
- [9] Y. Nomine, T. Ristriani, C. Laurent, J.F. Lefevre, E. Weiss, G. Trave, A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein, *Protein Eng.* 14 (2001) 297–305.
- [10] D.R. Smyth, M.K. Mrozkiewicz, W.J. McGrath, P. Listwan, B. Kobe, Crystal structures of fusion proteins with large-affinity tags, *Protein Sci.* 12 (2003) 1313–1322.
- [11] S. Raran-Kurussi, D.S. Waugh, The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated, *PLoS One* 7 (11) (2012) e49589.
- [12] F. Malagon, RNase III is required for localization to the nucleoid of the 5' pre-rRNA leader and for optimal induction of rRNA synthesis in *E. coli*, *RNA* 19 (2013) 1200–1207.
- [13] B.L. Coyle, F. Baneyx, A cleavable silica-binding affinity tag for rapid and inexpensive protein purification, *Biotechnol. Bioeng.* 111 (2014) 2019–2026.
- [14] J.D. Fox, K.M. Routzahn, M.H. Bucher, D.S. Waugh, Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers, *FEBS Lett.* 537 (2003) 53–57.
- [15] G.T. Lountos, J.E. Tropea, S. Cherry, D.S. Waugh, Overproduction, purification, and structure determination of human dual-specificity phosphatase 14, *Acta Crystallogr. D. Biol. Crystallogr.* 65 (2009) 1013–1020.
- [16] D. Needle, G.T. Lountos, D.S. Waugh, Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity, *Acta Crystallogr. D. Biol. Crystallogr.* 71 (2015) 1102–1111.
- [17] K.M. Routzahn, D.S. Waugh, Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins, *J. Struct. Funct. Genomics* 2 (2002) 83–92.
- [18] S. Sathiamoorthy, J.A. Shin, Boundaries of the origin of replication: creation of a pET-28a-derived vector with p15A copy control allowing compatible coexistence with pET vectors, *PLoS One* 7 (10) (2012) e47259.
- [19] R.B. Kapust, J. Tozser, J.D. Fox, D.E. Anderson, S. Cherry, T.D. Copeland, D.S. Waugh, Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency, *Protein Eng.* 14 (2001) 993–1000.
- [20] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, *Protein Sci.* 8 (1999) 1668–1674.
- [21] S. Raran-Kurussi, J. Tozser, S. Cherry, J.E. Tropea, D.S. Waugh, Differential temperature dependence of tobacco etch virus and rhinovirus 3C proteases, *Anal. Biochem.* 436 (2013) 142–144.
- [22] R.B. Kapust, D.S. Waugh, Controlled intracellular processing of fusion proteins by TEV protease, *Protein Expr. Purif.* 19 (2000) 312–318.
- [23] S. Raran-Kurussi, K. Keefe, D.S. Waugh, Positional effects of fusion partners on the yield and solubility of MBP fusion proteins, *Protein Expr. Purif.* 110 (2015) 159–164.
- [24] P. Sun, J.E. Tropea, D.S. Waugh, Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner, *Methods Mol. Biol.* 705 (2011) 259–274.
- [25] S. Raran-Kurussi, D.S. Waugh, Expression and purification of recombinant proteins in *Escherichia coli* with a His₆ or dual His₆–MBP tag, *Methods Mol. Biol.* (2016) (in press).
- [26] D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, *Protein Expr. Purif.* 80 (2011) 283–293.
- [27] M.I. Donnelly, M. Zhou, C.S. Millard, S. Clancy, L. Stols, W.H. Eschenfeldt, F.R. Collart, A. Joachimiak, An expression vector tailored for large-scale, high-throughput purification of recombinant proteins, *Protein Expr. Purif.* 47 (2006) 446–454.
- [28] W.M. Hewitt, G.T. Lountos, K. Zlotkowski, S.D. Dahlhauser, L.B. Saunders, D. Needle, J.E. Tropea, C. Zhan, G. Wei, B. Ma, R. Nussinov, D.S. Waugh, J.S. Schneekloth Jr, Insights into the allosteric inhibition of the SUMO E2 enzyme Ubc9, *Agnew Chem. Int. Ed. Engl.* 55 (2016) 5703–5707.