<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Going native: Complete removal of protein purification affinity tags by simple modification of existing tags and proteases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Goh, Hui Chin; Sobota, Radoslaw M.; Ghadessy, Farid J.; Nirantar, Saurabha</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/42481">http://hdl.handle.net/10220/42481</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>).</td>
</tr>
</tbody>
</table>
Going native: Complete removal of protein purification affinity tags by simple modification of existing tags and proteases

Hui Chin Goha, Radoslaw M. Sobotab, Farid J. Ghadessya, Saurabh Nirantar a, **

a p53Lab, Agency for Science Technology and Research (A*STAR), 8A Biomedical Grove, 138648, Singapore
b Systems Structural Biology Group, Institute of Molecular and Cellular Biology (A*STAR), 61 Biopolis Dr, 138673, Singapore

A R T I C L E   I N F O

Article history:
Received 6 May 2016
Received in revised form 18 August 2016
Accepted 5 September 2016
Available online 8 September 2016

Keywords:
Protein purification
Affinity tags
Protease
Enforced co-localisation
Native protein

A B S T R A C T

Protein purification typically involves expressing a recombinant gene comprising a target protein fused to a suitable affinity tag. After purification, it is often desirable to remove the affinity tag to prevent interference with downstream functions of the target protein. This is mainly accomplished by placing a protease site between the tag and the target protein. Typically, a small oligopeptide ‘stub’ C-terminal to the cleavage site remains attached to the target protein due to the requirements of sequence-specific proteases. Furthermore, steric hindrance can also limit protease efficiency. Here, we show that respectively fusing the interacting ePDZ-b/ARVCF protein-peptide pair to the target protein and a protease enables efficient processing of a minimised sequence comprising only residues N-terminal to the cleavage site. Interaction of the protein-peptide pair enforces proximity of the protease and its minimised cleavage sequence, enhancing both catalysis of a sub-optimal site and overcoming steric hindrance. This facilitates the high yield purification of fully native target proteins without recourse to specialised purification columns.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Proteins are the workhorses of living systems, catalysing reactions essential for metabolism, detecting and responding to environmental stimuli, providing structural integrity to the cell structure and detecting pathogens among numerous other functions. Many of these functions are of academic or industrial interest, necessitating the purification of the relevant proteins. Typically, this is accomplished by expressing a recombinant version of the protein of interest fused to an affinity tag that enables column purification using an appropriate affinity matrix. Whilst this process enables routine purification of large amounts of target protein, the presence of the affinity tag fused to the protein can interfere with downstream applications or function [1,2]. The tag may be removed by cleavage of a protease site placed between the target protein and the affinity tag, followed by a step to separate the protein and affinity tag. Nonetheless, most sequence-specific proteases have amino acid requirements at both the N- and C-terminal sides of the cleavage site [3], leaving a small “stub” still attached to the protein of interest. This may be unfavourable for certain downstream applications and in the case of therapeutic proteins, it may raise the possibility of an immune response.

Serial chromatography purifications of cellular lysates can be used to isolate proteins of interest based on their size, charge and pl. However this method is often laborious, non-generic and does not guarantee fully pure protein. Researchers have previously developed methods to fully remove any trace of the affinity tag. A system comprising a column-immobilized mutant subtilisin which is inactive in the absence of fluoride ions, and a target protein fused to subtilisin prodomain at its N terminus has been developed [4]. The prodomain-target protein fusion is loaded onto the mutant subtilisin column, non-specific proteins washed off and cleavage at the junction of the prodomain-protein of interest triggered by the addition of fluoride ions at high concentration. This method requires a specialised column and results in the presence of fluoride ions in the eluate which may have to be removed prior to downstream applications.

Another system developed involves the fusion of the protein of interest to a mutant intein and chitin binding domain at the C terminus [5]. The expressed protein is bound to a chitin matrix and the cellular proteins and debris are washed out. Addition of a strong
reducing agent triggers the intein, thereby severing the link between the intein and the native target protein, which is eluted. This process also requires a specialised column and results in the presence of a strong reducing agent in the elution buffer.

The TAGzyme method from Qiagen uses aminopeptidases to cleave off N terminal histidine tags, while preventing the cleavage of target protein residues by using “stop points” such as glutamine. These “stop point” containing proteins are processed by Qyclease which converts the N terminal glutamine into pyroglutamate. This can then be cleaved by an aminopeptidase (pGAPase) to yield native protein. However, the necessity for inserting genetically coded “stop points” for most proteins, and a multi-step process necessary for obtaining native protein leaves substantial room for improvement.

We have therefore developed a purification system to generate native (i.e. completely untagged) proteins of interest (Fig. 1). This requires expression of a fusion protein comprising the ePDZ-b affinity clamp domain adjacent to a sub-optimal sequence-specific protease cleavage site (preferred residues at P1’ and P2’ are removed) followed immediately by the protein of interest. The cognate protease is expressed with a C-terminal tag comprising the ARVCF peptide sequence that binds ePDZ-b very tightly (Kd < 5 nM) [6]. Typically, the presence of methionine at the P1’ position in addition to steric hindrance posed by the folded target protein would render cleavage at this site sub-optimal due to compromised protease affinity. However, fusion to the interacting protein-peptide pair enforces proximity, allowing a much faster reaction rate by increasing the effective concentration of protease in the vicinity of a sub-optimal cleavage site. The concept of enforced proximity has previously been described to generate proteins with enhanced activity at sub-optimal cleavage sites. This provides an active SplB protease, which in turn cleaves more of the fusion protein, resulting in a rapid chain reaction yielding mature SplB-His tag-ARVCF peptide.

The Protein A open reading frame (ORF) was amplified using oligonucleotides oSN684 and oSN1096. oSN1096 appends a sequence coded as “GGSGKGGWEL” at the 3’ of protein A. SplB was PCR-amplified with oSN1099 and oSN875. oSN875 appends a partial His tag to the 3’ of SplB. oSN1099 appends a sequence complementary to that of oSN1096, enabling SOE PCR of these two ampiclons to yield protein A-WELQ-SplB-truncated His tag cassette.

Simultaneously, the pET22b (kan) vector was amplified by inverse PCR using oSN1183 and oSN1184, which have complementary 5’ ends for infusion cloning, followed by intra-plasmid infusion cloning (Clontech), creating a vector which has the “PQPVDSWV” coding ARVCF peptide just downstream of the XhoI site. This vector was then linearized with Ndel/Xhol, and the fusion ORF protein A-WELQ-SplB-truncated His tag was inserted by infusion cloning yielding protein A-WELQ-SplB-His tag-ARVCF peptide in pET22b (Kan) vector.

The TEV-AP4 protein expression construct encoded the ORF for a maltose binding protein (MBP)-optimal TEV cleavage site (ENLYFQSGGSHHHHHHGGDSWV) fusion protein. The ORF was synthesized (Genscript) and inserted into the Ndel/Xhol sites in pET22b vector. The TEV expressed incorporated the S219V mutation to reduce autolysis [8]. The TEV protease auto-cleaves itself from MBP after expression and is purified using the 6-histidine tag preceding the truncated ARVCF peptide (DSWV, indicated in bold above). The vector pRK793 [8] was used to express histidine-tagged TEV comprising the S219V mutation.

2.2. Protease gene constructs

SplB proteases were expressed with a protein A domain fused at the N terminus, an SplB cleavage site “WELQ” immediately preceding the first amino acid of the mature SplB protease, and a 6X histidine tag followed by the ARVCF peptide at the C terminus (protein A-GGWELQ-mature SplB-His tag-ARVCF peptide). The protein A is removed by cleavage at the “WELQ” site for a small minority of fusion proteins during protein expression, forming active SplB protease, which in turn cleaves more of the fusion protein, resulting in a rapid chain reaction yielding mature SplB-His tag-ARVCF peptide.

2.3. SplB-His tag-truncated ARVCF peptide constructs

These were constructed by doing inverse PCR with the above-mentioned protein A-WELQ-SplB-His tag-ARVCF construct using the oligonucleotides oSN1225, 1226, 1227, 1228 and 1230 with the common reverse oligonucleotide oSN1229 to yield constructs expressing ARVCF peptides PQPVDSWV, PQVPSWV, PVPSWV, VPDWSV and DSWV respectively.

2.4. SplB-truncated His tag-DSWV constructs

The protein A-WELQ-SplB-truncated His tag-DSWV was created

---

**Fig. 1.** Optimising cleavage at suboptimal protease sites by enforced co-localisation. Top schematic depicts domain structure of fusion protein comprising target of interest and ePDZ-b connected by protease cleavage site with methionine at P1’ position. Protease is expressed with C-terminal ARVCF-pep tag. Interaction of ARVCF-pep (light magenta) with ePDZ-b (green) promotes co-localisation of protease (cyan) at sub-optimal cleavage site (grey), resulting in cleavage and generation of target protein with no exogenous amino acids at the N-terminus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
by inverse PCR of the protein A-WELQ-SplB-6x His tag-DSWV construct described above with oSN1396 and oSN1397 followed by intra-plasmid infusion to truncate the number of histidines from 6 to 5. For the construct with 4 histidines, the original protein A-WELQ-SplB-6x His tag-DSWV ORF was amplified with oSN684 and oSN1407, followed by infusion into a vector obtained by inverse PCR using the oligonucleotides oSN1229 and oSN359.

2.5. ePDZ-b -WELQ-protein constructs

The ePDZ-b ORF was PCR-amplified using tandem oligonucleotides oSN1344 and 1348 and reverse oligonucleotide oSN1386 to attach a 6x histidine tag to the N terminus of ePDZ-b and a linker coding for amino acids “WELQM” to the 3’ terminus. The ORF for LSSmOrange (OFP) was PCR-amplified using oSN1318 and oSN1312, followed by SOE PCR to create histidine tag-ePDZ-b-WELQ-OPF expression cassette. This was then inserted by infusion cloning into pET22b (kan) vector obtained by inverse PCR using oSN359-1384.

2.6. ePDZ-b -TEV site-protein constructs

The previously constructed ePDZ-b-WELQ-OPF plasmid was inverse PCR-amplified with oSN1410 and oSN1315. The OFP ORF (PCR-amplified using oSN1414-1412) was inserted into this plasmid by infusion cloning to yield His tag-ePDZ-b-ENLYFQ-OPF. This plasmid was subsequently PCR-amplified using primers oSN1415 and oSN1416 and the fibronectin 10Fn3 domain ORF (obtained by amplifying plasmid 10FN3-HA_pET22b with primers oSN1417 and 1418) inserted via infusion cloning. The resulting His tag-ePDZ-b-ENLYFQ-10Fn3 expression plasmid was subsequently amplified by inverse PCR using primer pairs oSN1419/1420 and oSN1419/1421 and products intramolecularily ligated post-phosphorylation to yield constructs expressing His-tag-ePDZ-b-ENLYFQ-MGGS-10Fn3 and His-tag-ePDZ-b-ENLYFQ-MGGS-10Fn3 respectively.

2.7. Protein expression

All protein expression was carried out in E. coli BL21 (DE3) using IPTG induction of the T7 promoter. Colonies were inoculated overnight (ON) in 10 mL TY medium supplemented with the appropriate antibiotic at 37 °C with shaking at 180 rpm. The next day, the cultures were diluted into 800 mL of fresh medium and allowed to grow at 37 °C to an OD600 of 0.5, upon which IPTG (1 mM final) was added. Induction was allowed to proceed overnight at room temperature. The next day, cells were pelleted down and stored at −20 °C until further processing.

2.8. Protein purification

Cell pellets obtained as described above were resuspended in 20 mL of PBS-Tween or other buffers as mentioned in the text followed by lysis by sonication. The lysate was centrifuged at 14800 rpm for 10 min in 50 mL Falcon tubes and the supernatant was collected. In parallel, a 1 mL FF His-trap column (GE Healthcare) was washed with 10 mL of the same buffer in which the pellet was resuspended. After collecting the cell lysate supernatant, it was run through the column, followed by washing with 10 mL of washing buffer (the same buffer used for cell pellet resuspension). Thereafter elution buffer (50 mM Sodium Phosphate + 300 mM NaCl + 500 mM Imidazole) was run through the column and 1 mL fractions were collected. All other fractions were also collected and the normalized volumes of the fractions were run on an SDS-PAGE gel followed by staining by Instant Blue for visualization.

The fractions containing high concentrations of pure protein were mixed with glycerol to a final volume of 50% glycerol and stored at −20 °C.

2.9. Go Native protein purification

Cell pellets were resuspended and lysed as described above. The cell lysate was loaded on an equilibrated His-trap column followed by washing with 10 mL of buffer. Next, 5 mL of the relevant protease (0.5 mg/mL) was loaded on the column. The column was capped and incubated at room temperature for the relevant duration. Thereafter, washing buffer was run through the column and fractions collected. Finally, imidazole elution buffer was run through the column. All fractions were collected and normalized volumes electrophoresed on an SDS-PAGE gel followed by staining with Instant Blue.

2.10. Mass spectrometry/N-terminal protein sequencing

To elucidate cleavage sequence specificity mass spectrometry analysis was performed. Following SDS-PAGE separation and Coomassie visualisation, gel band with protease cleaved product was excised and subjected to “in-gel” digestion [9]. Following desalting, sample was injected and separated using Orbitrap Fusion mass spectrometer coupled to nanoUHPLC system Easy LC1000 (Thermo) in a 45 min gradient. To generate peak list Proteome Discoverer 1.4 software was used. Searches were done using Sequest HT algorithm against Human Uniprot database appended with specific target sequence (87458 entries) with following parameters: precursor mass tolerance (MS) 10 ppm, (MS/MS) IT-MS/MS 0.6 Da; no-enzyme; 2 miss cleavages; Static modifications: Carboxamidomethyl (C). Variable modifications: Oxidation (M), Deamidated (NQ). N-terminal sequencing of gel-purified protein [10] was carried out by Biosynthesis (USA).

3. Results

We first tested the principle using ePDZ-b fused to the target protein (orange fluorescent protein, OFP) and ARVCF peptide fused to SplB protease (SplB-AP). SplB protease cleaves after the sequence WELQ with methionine at the P1’ position poorly tolerated [11,12]. When combined with potential steric exclusion by the protein of interest being purified, methionine at P1’ will pose barriers to optimal SplB protease cleavage. The WELQ peptide sequence was introduced between ePDZ-b and OFP. Incubation of the fusion substrate (ePDZ-b-WELQ-OPF) with a stoichiometric excess of SplB protease showed clear improvements in activity between ePDZ-b and ARFCP peptide may result in prolonged tethering of protease to ePDZ-b after cleavage of target protein. This would reduce “turn-over” of the protease, necessitating use of higher stoichiometric amounts. We tested this hypothesis by reducing the affinity of the ePDZ-b-ARFCP peptide interaction by serially truncating the ARFCP peptide fused to SplB from 8 to 4 amino acids (QPVDSWV to DSWV). At high protease concentration, no variation was observed in cleavage efficiency (Fig. 2, lanes 2–5). At sub-stoichiometric amounts, 3 and 4 amino acid truncations of the ARVCF peptide showed clear improvements in activity.
compared to full-length peptide (Fig. 2, compare lane 16 with 18–19). Furthermore, the overall activity compared to the SplB-CON and SplB-COM was significantly enhanced (compare lanes 18–19 and 20–21).

We next applied the principle to TEV protease, one of the most ubiquitous enzymes used to remove affinity tags that optimally cleaves the consensus sequence ENLYFQ [13,14]. We constructed a fusion substrate wherein this sequence was truncated to ENLYFQ, and placed between the ePDZ-b and OFP components. TEV is able to cleave when serine at P1’ position is replaced by methionine with marginal loss of activity [15]. However, steric constraints due to the tightly folded N-terminal region of OFP could impact negatively on cleavage. The results show clearly improved cleavage when TEV is fused to the optimised 4-amino acid truncated ARVC peptide (TEV-AP4) (Fig. 3A). Notable cleavage was observed after only 30 min incubation with majority of substrate being cleaved after 2.5 h. In comparison, TEV protease only showed significant cleavage after 24 h incubation. Note that both the TEV-AP4 and TEV proteins used comprise the S219V autolysis resistance mutation [8]. A control experiment using a fusion substrate comprising the full TEV consensus sequence (ENLYFQS) led to equivalent cleavage by both wild-type TEV and TEV-AP4 (Fig. 3B). The endogenous TEV protease (minus the S219V mutation) showed even poorer cleavage when compared to TEV-AP4 for the same substrate (Fig. S1). We next assayed TEV-AP4 on a fusion substrate where OFP was replaced with the fibronectin 10Fn3 domain (ePDZ-b-ENLYFQ1-10Fn3). As before, significantly enhanced cleavage was obtained in very short times when using TEV-AP4 compared to TEV (Fig. 4A). Given the highly compact 10Fn3 fold [16], it is likely that steric constraints inhibit cleavage by TEV. Introduction of GGS or GGSGGS linkers (to yield ePDZ-b-ENLYFQ1-MGGS-10Fn3 and ePDZ-b-ENLYFQ1-MGGSGGS-10Fn3 respectively) resulted in improved cleavage by TEV, yet this was still considerably less than for TEV-AP4 (Fig. 4B,C). As before, endogenous TEV also performed poorly using this substrate set (Fig. S2).

We next explored whether the enforced-proximity concept was applicable to conventional on-column cleavage and purification protocols using histidine tagged proteins. Complete immobilisation of protease via its histidine tag could reduce turnover during on-column cleavage of a co-immobilised substrate, necessitating use of increased amounts for efficient cleavage. Addition of 30 mM imidazole alleviated this constraint, resulting in improved cleavage efficiencies using histidine tagged ePDZ-b-WELQ-OFP and SplB-AP proteins (Fig. S3). These conditions were used for the on-column
Fig. 4. Improved cleavage of fibronectin 10Fn3 domain fusion proteins comprising TEV cleavage site with methionine at P1’ position. A) The ePDZ-b-ENLYFQ-10Fn3 fusion protein was incubated with either TEV-AP4 (lanes 2–9) or TEV protease (lanes 14–21) for indicated times. Native 10Fn3 is arrowed red. Lanes 11 and 23 show untreated fusion substrate. Lanes 12 and 24 show purified TEV-AP4 and TEV (dotted arrows) respectively. B) Same as in A, except using the fusion protein substrate ePDZ-b-ENLYFQ-MGGS-10Fn3. C) Same as in A, using substrate ePDZ-b-ENLYFQ-MGGSGGS-10Fn3. Note TEV used here comprises the S219V autolysis resistance mutation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
catalysis [20]. This principle has previously been described to bonds between two or more molecules need to have the ability to Cdc2 kinase which is directed to its targets by cyclins[19]. Similarly, MGGS-10Fn3 and TEV (S219V). Asterisk shows MGGS-10Fn3 protein.

Proteases. (B) Same as in A, except using fusion protein substrate HIS-ePDZ-b-ENLYFQ-OFP. Asterisk) in

Fig. 5. Improved on-column cleavage of a recalcitrant fusion protein substrate by TEV-AP4. (A) The HIS-ePDZ-b-ENLYFQ-OPF fusion substrate protein and either HIS-TEV-AP4 (left gel) or HIS-TEV (right gel) were co-immobilised and on-column cleavage carried out for 2.5 h. Lanes 2 + 11: Bacterial cell-lysate. Lanes 3–5/12–14: non-specific proteins eluted after three washes post loading. Lanes 6 + 15: native OFP (highlighted by asterisk) in flow-through post protease incubation. Lanes 7 + 16: Proteins eluted from column post-digestion by imidazole. Lanes 9 + 18: HIS-TEV-AP4 and HIS-TEV proteases. (B) Same as in A, except using fusion protein substrate HIS-ePDZ-b-ENLYFQ-MGGS-10Fn3 and TEV (S219V). Asterisk shows MGGS-10Fn3 protein.

cleavage of histidine tagged ePDZ-b-ENLYFQ-OPF protein by histidine tagged TEV-AP4. Upon elution with PBS, the yield of native OFP was significantly increased when histidine tagged TEV-AP4 was used compared to histidine tagged endogenous TEV (Fig. 5A, compare lanes 6 and 15). Both mass spectrophotometry and N-terminal sequencing analysis confirmed correct cleavage by TEV-AP4 to yield OFP with an N-terminal methionine (Figs. S4 and S5). We also carried out on-column cleavage of the histidine tagged ePDZ-b-ENLYFQ-MGGS-10Fn3 substrate, this time comparing with histidine-tagged TEV(S219V). As before, TEV-AP4 showed improved cleavage on-column (Fig. 5B).

4. Discussion

Examples are abound in nature of proteins that have evolved to serve as platforms to bring reactants into close proximity [17]. This mechanism improves the affinity (i.e. Km) component of catalysis, which can often be rate limiting. An example is the Ste5 platform in budding yeast, which serves as a platform to co-localize various reactants, enabling faster signalling [18]. Another example is the Cdc2 kinase which is directed to its targets by cyclins [19]. Similarly, enzymes such as polymerases involved in the formation of covalent bonds between two or more molecules need to have the ability to bind the reactants and bring them into close proximity before catalysis [20]. This principle has previously been described to enhance the efficiency of N-degron activation by TEV and to select for TEV variants with increased P1’ substrate tolerance in vivo [21,22]. Here, the interacting SF3b153-381–424 protein domain and human spliceosome subunit p14 were respectively tethered to the N-degron and protease.

We have extended this concept using ePDZ-b and a minimised ARVCF peptide pair to enhance protease cleavage for in vitro removal of affinity tags. ePDZ-b comprises a synthetic bi-domain protein “affinity clamp” evolved for high affinity binding to the ARVCF peptide [6]. Results using SplB protease indicated that enforced proximity can overcome issues of steric hindrance arising from the presence of a bulky hydrophobic amino acid at P1’ and/or where the structured N-terminus of a protein can limit access of protease to an adjacent cleavage site. Whilst TEV has been shown to tolerate a range of amino acids at the P1’ position, we have shown greatly improved activity when methionine is present at this position by promoting co-localisation. This result also indicates that the ePDZ-b component of the substrate fusion protein is not contributing to steric hindrance, as reported for other fusion partners [23]. TEV activity is ~10-fold reduced at 4 °C (commonly used for prolonged on-column digestion) compared to its activity at 30 °C [24]. Our data shows that cleavage to yield native proteins is rapidly expedited by enforced co-localisation at room temperature (~22 °C). This opens up the option of incubation at temperatures closer to the TEV optimum for shorter periods that preserve integrity of target proteins.

In sum, we have described facile use of the ePDZ-b/ARVCF peptide pair to optimize sub-optimal cleavage by site-specific proteases and generate completely tag-free target proteins. We are currently evaluating the “Go Native” concept in the context of other commonly used proteases and affinity tags to further determine its broad applicability.

Funding

This research was funded by: Agency for Science, Technology and Research Singapore (12302FG015), core funding from IMCB and Young Investigator Grant YIG (A*STAR), NMRC MS-CETSA platform grant MOHIAFCAT2/004/2015 to RMS.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.pep.2016.09.001.

References


