Protein Expression and Purification xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Review

An overview of enzymatic reagents for the removal of affinity tags

David S. Waugh *

Protein Engineering Section, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA

ARTICLE INFO

Article history:
Received 2 August 2011
and in revised form 4 August 2011
Available online xxxx

Keywords:
Carboxypeptidase
Aminopeptidase
PreScission protease
Affinity tag
Polyhistidine tag
His-tag
TEV protease
Enterokinase
Enteropeptidase
Thrombin
3C protease
Factor Xa

ABSTRACT

Although they are often exploited to facilitate the expression and purification of recombinant proteins, every affinity tag, whether large or small, has the potential to interfere with the structure and function of its fusion partner. For this reason, reliable methods for removing affinity tags are needed. Only enzymes have the requisite specificity to be generally useful reagents for this purpose. In this review, the advantages and disadvantages of some commonly used endo- and exoproteases are discussed in light of the latest information.

Published by Elsevier Inc.

Contents

Introduction

Affinity tags have become essential tools for the production of recombinant proteins in a wide variety of settings, from basic research to high-throughput structural biology. Not only do they facilitate the detection and purification of their fusion partners, as originally intended, but they may also have a beneficial impact on the yield of recombinant proteins and, in some cases, increase their solubility and even promote their proper folding [2,3].

Despite these important advantages, the Achilles heel of the affinity tagging strategy always has been and remains the removal

* Fax: +1 301 846 7148.
E-mail address: waughd@mail.nih.gov

1046-5928/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.pep.2011.08.005

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

Table 1
Endo- and exoproteases for the removal of affinity tags.

Enzyme	Source(s)	Molecular weight (kDa)	Tagged forms	Inhibitors	Recognition Site	Notes
Endoproteases						
Enteropeptidase	Duodenum *E. coli* *S. cerevisiae*	110 + 35	His_6	Reducing agents	DDDDK↓	P1′ ≠ Pro, Trp 35 KDa light chain is active by itself
Thrombin	Plasma CHO cells	32 + 4.5	None	Reducing agents	LVPR↓GS	
Factor Xa	Plasma HEK 293 cells	42 + 17	None	Reducing agents Chelating agents Phosphate ions	LVPR↓GS	Very promiscuous
TEV Protease	*E. coli*	27	His_6 MBP GST Strep II	Thiol alkylating agents	ENLYFQ↓G	P1′ can vary [68] P2′ ≠ Pro Ac-TEV™ = S219 V mutant
Rhinovirus 3C Protease	*E. coli*	27	His_6 GST His_6-GST	Thiol alkylating agents	LEVLFQ↓GP	Same as PreScission™ protease
Exoproteases						
Carboxypeptidase A	Pancreas *E. coli* *S. cerevisiae* *S. frugiperda* (baculovirus)	33	His_6	Reducing agents Chelating agents	C-terminal amino acids except Pro, Lys and Arg	Asp, Glu, Gly cleaved slowly
Carboxypeptidase B	Pancreas *E. coli* *P. pastoris*	35	none	Reducing agents Chelating agents	C-terminal Lys and Arg	Will cleave hydrophobic resides under certain conditions [119,120]
DAPase	Kidney *S. frugiperda* (baculovirus)	23 + 16 + 6	His_6	Reducing agents Thiol alkylating agents	N-terminal dipeptides	P2 ≠ Pro, Lys, Arg P1 ≠ Pro

of tags. Whereas many tagged proteins retain their structural integrity and biological activity, others clearly do not, e.g., [4–11]. Therefore, whenever possible, it is prudent to remove tags from recombinant proteins. Although both chemical and enzymatic methods have been used to cleave fusion proteins at designed sites, only the natural proteolytic enzymes have the requisite specificity to be broadly useful reagents for this purpose. Because they are not as versatile and therefore generally less useful than *trans*-acting reagents, neither the self-cleaving inteins [12] and self-cleaving variants of subtilisin [13] will be discussed here, nor will the Ulp1 protease since it only cleaves SUMO tags [14]. Rather, this review will focus on the most generally applicable and commonly used enzymatic reagents for the removal of affinity tags (Table 1). Since the last comprehensive review of this topic [3] much research on these reagents has been conducted. As a result, a wealth of new information has accumulated on the advantages, disadvantages, and biochemical characteristics of various reagents.

Endoproteases

For many years, serine proteases such as activated blood coagulation factor X (factor Xa),[1] enterokinase (hereafter referred to by its more appropriate moniker enteropeptidase), and α-thrombin were the reagents of choice for removing affinity tags, yet the literature is replete with reports of fusion proteins that were cleaved by these proteases at locations other than the designed site. Over the last decade or so, it has become increasingly evident that certain viral proteases have far more stringent sequence specificity, which has led to an upsurge in their popularity. These enzymes have a chymotrypsin-like fold with an atypical catalytic triad in which cysteine replaces serine, and they exhibit an absolute requirement for glutamine in the P1[2] position of their substrates. The nuclear inclusion protease from tobacco etch virus (TEV) is probably the best-characterized enzyme of this type. The other is the human rhinovirus 3C protease. In stark contrast to factor Xa, enteropeptidase and thrombin, there have been very few if any reports of cleavage at noncanonical sites in designed fusion proteins by these viral proteases.

The stringent specificity of the viral proteases probably can be attributed to their low turnover rates. The number of substrate residues that are recognized by the serine proteases and the viral proteases is similar (e.g., LVPRGS and ENLYFQS in the case of thrombin and TEV protease, respectively). The Michaelis constants (K_M) for the two classes of enzymes are also similar, falling in the low to mid micromolar range, but the catalytic rate constants (k_{cat}) of the viral proteases are on the order of 100-fold lower than those of the serine proteases, resulting in much slower turnover rates [15–19]. Each class of protease undoubtedly associates transiently with suboptimal recognition sites, but on average, a catalytic event is far more likely to occur when a serine protease does so because its k_{cat} is so much greater than that of the typical viral protease. The practical ramification of this observation is that one must use considerably more viral protease than serine protease to digest a fixed amount of fusion protein at a similar rate. However, this is not a significant handicap because, unlike the serine proteases, large quantities of recombinant viral proteases can easily be produced in *Escherichia coli*. This advantage, coupled with their more stringent sequence specificity, has made viral proteases the reagents of choice for endoproteolytic removal of affinity tags.

[1] *Abbreviations used:* factor Xa, blood coagulation factor X; TEV, tobacco etch virus; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; AEBSF, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride; PEG, polyethylene glycol; GST, glutathione S-transferase; EGTA, ethylene glycol tetraacetic acid; EDTA, ethylenediaminetetraacetic acid; MBP, maltose binding protein; ATCC, American Type Culture Collection; TVMV, tobacco vein mottling virus; SARS, severe acute respiratory syndrome; BAP, biotin acceptor peptide; POI, protein of interest; BoCPA, bovine carboxypeptidase A; BoCPB, bovine carboxypeptidase B; SDS, sodium dodecyl sulfate; DAPase, dipeptidyl-aminopeptidase I; Qcyclase, glutamine cyclotransferase; pGAPase, pyroglutamylaminopeptidase.

[2] The nomenclature used here to describe individual amino acids in protease recognition sites and corresponding amino acid-binding sites in proteases was introduced by Schechter and Berger [1].

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

Enteropeptidase

Native enteropeptidase is a disulfide-linked heterodimer composed of "heavy" and "light" chains (with apparent molecular weights of 110 and 35 kDa, respectively), which are extensively glycosylated [20]. Although originally purified from natural sources [20,21], recent advances have facilitated the production of a recombinant enteropeptidase light chain in the periplasm of *E. coli* [22,23], making it more economical to manufacture and yielding a product free of contaminating proteases. Moreover, the 26 kDa light chain (which, when expressed in *E. coli*, is devoid of glycosylation) retains the specificity of the native enzyme and is even more active on fusion protein substrates than the heterodimer [22]. The availability of recombinant enteropeptidase also presents an opportunity for the attachment of affinity tags to the enzyme to facilitate its purification and separation from the products of an enteropeptidase digest [24]. One drawback of enteropeptidase, however, is that the light chain contains 4 disulfide bonds that are essential for catalytic activity, thus this enzyme is incompatible with buffers containing reducing agents like dithiothreitol (DTT). On the other hand, a recent study showed enteropeptidase to be relatively insensitive to a wide variety of detergents [25], which may enhance its value as a tool for the production of membrane proteins.

A significant advantage of enteropeptidase is that no critical specificity determinants are located on the C-terminal side of the scissile bond in its substrates (Table 1). Consequently, when an affinity tag is joined to the N-terminus of the protein of interest, in most cases enteropeptidase is able to generate a digestion product with a native N-terminus. A comprehensive study examining the relative processing rates of otherwise identical fusion proteins with all twenty possible amino acids in the P1′ position confirmed this property of enteropeptidase and revealed the rank order of processing efficiency [26]. Only proline and tryptophan were not well tolerated in the P1′ position. Another study examining the importance of the P1–P5 positions concluded that the P1 lysine was the most important specificity determinant, followed by the aspartate residues in the P2, P3, P5 and P4 positions, respectively, with the latter position contributing very little to specificity [27]. Interestingly these investigators found that the sequence DDDDR was cleaved more efficiently than the canonical DDDDK.

The principal drawback of enteropeptidase is its promiscuity [28,29], which is particularly troublesome when a cryptic cleavage site is located within the protein of interest. Although efforts have been made to mitigate this problem by varying the recognition site [30] and tinkering with the enzyme [31], a satisfactory solution has yet to present itself.

Thrombin

Thrombin is typically purified from bovine plasma [32]. Although recombinant human thrombin has been produced in Chinese hamster ovary cells for clinical applications [33], as of now no convenient procedure for the expression and purification of recombinant thrombin at the bench-level scale has been described. The lack of a ready source of genetically engineered thrombin is a noteworthy disadvantage because, unlike enteropeptidase and the viral proteases (below), no affinity tags can be added to facilitate its removal following the digestion of a fusion protein.

Like many serine proteases, thrombin can be inactivated by phenylmethylsulfonyl fluoride (PMSF) or 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF). Alternatively, a biotinylated form of thrombin that can be adsorbed on avidin or streptavidin resin is commercially available (Novagen, Madison, WI). Like enteropeptidase, thrombin is a disulfide-linked heterodimer. It also has three intramolecular disulfide bonds in one of its two chains [34], rendering it sensitive to reducing agents. The optimum temperature for thrombin activity is 45 °C and the enzyme maintains 20% of its maximal activity at 15 °C [35]. Thrombin is active over a pH range of 5–10, with a pH optimum of about 9.5 in the absence of NaCl and 8.3 in the presence of 1 M NaCl [35]. At pH 8.3, its activity increases with rising NaCl concentration up to at least 1 M. Glycerol and polyethylene glycol (PEG 600) enhance the stability of thrombin. Like enteropeptidase, thrombin is relatively resistant to a wide variety of detergents [25].

The thrombin cleavage site most frequently used in fusion protein substrates (LVPR↓GS) has an interesting history. This sequence closely resembles that of a natural thrombin cleavage site (LVPR↓GF) in human factor VIII. Yet the factor VIII sequence was not cleaved as efficiently as others in a comparative study [36], raising the question of why it was viewed as a particularly good thrombin site in the first place. Later, in their seminal paper describing the first application of glutathione S-transferase (GST) as an affinity tag, Smith and Johnson replaced the P2′ phenylalanine in the factor VIII sequence with a serine residue purely for the purpose of creating a BamHI restriction endonuclease cleavage site that would facilitate the construction of GST fusion proteins [37]. After that, LVPR↓GS was universally espoused as the canonical thrombin cleavage site, irrespective of whether or not BamHI was utilized for cloning, e.g., [38]. Therefore, it is possible this is not the linear epitope that is most efficiently cleaved by thrombin.

Specificity studies with thrombin have shown a range of sequence selectivity of 656-fold and 33-fold at the P2′ and P3′ positions, respectively [39]. Proline and the negatively charged amino acids greatly diminished processing efficiency when present in the P2′ position of peptide substrates (the "canonical" serine was not tested in this study). The identity of the residue in the P3′ position was not nearly as influential, although the negatively charged residues were most inhibitory. In another study, little bias was observed in the P2 and P3 sites, except for the exclusion of acidic residues. Interestingly, a strong preference was found for serine in the P1′ position, in contrast to the "canonical" glycine, which was a distant fourth, after alanine and threonine [40]. When a large library of different peptide sequences was screened for thrombin cleavage, the results revealed that the P1 arginine is the most conserved residue, followed by the P1′ glycine [41]. Yet a remarkable level of promiscuity was observed, consistent with reports of cleavage at cryptic sites in fusion proteins [42]. The *nominal* requirement for a Gly-Ser dipeptide in the P1′ and P2′ positions, which would result in the retention of two nonnative residues on the N-terminus of the protein of interest following thrombin digestion, is a marked disadvantage of thrombin relative to enteropeptidase.

Factor Xa

A blood clotting enzyme like thrombin, the γ-carboxylated glycoprotein factor Xa is either isolated from blood plasma or expressed recombinantly and secreted from mammalian cells [43–47]. Despite the availability of recombinant factor Xa, no affinity-tagged forms of the enzyme have been engineered to date. Factor Xa is composed of two disulfide-linked polypeptide chains with apparent molecular weights of 17 and 42 kDa, each of which contains a number of internal disulfide bonds, rendering the enzyme sensitive to reducing agents. Factor Xa also binds calcium ions and therefore should not be used in the presence of chelating agents such as ethylene glycol tetraacetic acid (EGTA) and ethylenediaminetetraacetic acid (EDTA). The sensitivity of Factor Xa to various detergents has also been studied [25].

The specificity determinants of factor Xa have been examined in some detail [48–51]. Although it is a commonly held belief that this enzyme is insensitive to the identity of the residues on the

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

C-terminal side of the scissile bond (its recognition site is usually denoted as IEGR↓), this is not the case. In the P1′ position, processing efficiency varies over a 50-fold range, with hydrophobic residues being most favorable and negatively charged residues and proline being least favorable. An even greater range of processing efficiency (160-fold) was observed in the P2′ position, with threonine and proline being particularly poorly tolerated [51]. As with enteropeptidase and thrombin, cleavage by factor Xa at undesired cryptic sites remains a common problem with no apparent solution [42].

TEV protease

Virologist William Dougherty and his colleagues were the first to characterize TEV protease in the late 1980s. Initially identified as a 49 kDa processing product of the viral polyprotein and termed "nuclear inclusion protein Ia" or "NIa" [52], later work demonstrated that the protease activity resides in its C-terminal 27 kDa domain [53]. Aligning the sequences of experimentally determined processing sites suggested that the recognition site for TEV protease is a linear epitope consisting of seven amino acid residues. This conjecture was confirmed by demonstrating that the sequence ENLYFQ↓S could function as a TEV protease cleavage site when placed in an artificial context [54] and led to the realization that TEV protease might be a useful reagent for the removal of affinity tags by site-specific endoproteolysis [55].

Early efforts to overproduce a polyhistidine-tagged form the 27 kDa catalytic domain of TEV protease in *E. coli* met with only limited success, yielding approximately 4 mg of pure protein per liter of bacterial culture [55]. In later work, the yield was improved dramatically by utilizing a solubility-enhancing fusion partner [56] in conjunction with a tRNA-encoding accessory plasmid or synonymous codon substitutions in the TEV protease open reading frame [57]. Further improvements in the production of soluble His-tagged TEV protease were subsequently realized, yielding up to 400 mg/L or 15 mg/g of cell paste [58,59]. TEV protease has also been fused to a number of other affinity tags, including glutathione S-transferase [59], maltose binding protein (MBP) [56], and the Streptag II [60]. TEV protease expression vectors can be obtained from a variety of sources, including the American Type Culture Collection (ATCC), the Addgene plasmid repository (http://www.addgene.org), and the Protein Structure Initiative Biological Materials Repository (http://psimr.asu.edu/). Straightforward procedures for the expression and purification of TEV protease have been described [61,62].

The 27 kDa catalytic domain of TEV protease readily cleaves itself near its C-terminus to generate a truncated enzyme with greatly diminished activity [63,18]. However, autolysis of TEV protease can be avoided by introducing amino acid substitutions in the vicinity of the internal cleavage site [62,18]. One of these mutants (S219V) is approximately 100-fold more resistant to autoinactivation than the wild-type protease and, fortuitously, also has moderately greater catalytic activity [18]. The S219V mutant is marketed by Invitrogen under the trade name Ac-TEV protease. Additional mutations have been introduced into TEV protease as another means of improving its solubility [64,65].

The specificity determinants of TEV protease have been thoroughly investigated. Early experiments by Dougherty and colleagues demonstrated a strong preference for glutamic acid in the P6 position of a TEV protease recognition site, little or no selectivity in the P5 position, a moderate preference for leucine in the P4 position, a strong preference for tyrosine in the P3 position, a roughly equal tolerance for phenylalanine, cysteine and isoleucine in the P2 position (with all other residues being poorly tolerated in this position), a strong bias in favor of glutamine in the P1 position, and some degree of selectivity in the P1′ position, with serine, isoleucine, and asparagine being the most favorable of the residues tested [66]. These results led to the notion that the "consensus" TEV protease cleavage site could be defined as Glu-Xaa-Xaa-Tyr-Xaa-Gln↓Ser/Gly and the assumption that those positions denoted as "Xaa" could be freely substituted. However, more recent experiments have shown that TEV protease is quite sensitive even to conservative amino acid substitutions in the P4 and P2 positions [67]. In another study, the P1′ specificity of TEV protease was systematically assessed in the context of a model fusion protein [68]. These investigators found that the enzyme was surprisingly tolerant of a wide variety of residues in this position. Small aliphatic residues such as glycine, alanine, serine, methionine, and cysteine were exceptionally well tolerated, whereas the negatively charged and β-branched hydrophobic residues were inhibitory to processing. These findings upended the dogma that a serine or glycine residue is essential in the P1′ position of a TEV protease recognition site, revealing instead that, like enteropeptidase and factor Xa, in many cases TEV protease is capable of generating digestion products with native N-termini. As noted, the discrepancies between early and later results are probably due to differences in the experimental methods employed [68]. In the initial studies carried out by Dougherty and colleagues, the concentration of enzyme (which was the full-length NIa protein rather than the 27 kDa catalytic domain) exceeded that of the substrates, which were far below the K_M of the enzyme. Later experiments were conducted under more realistic reaction conditions, using pure preparations of enzyme and substrate with a substantial molar excess of the latter. An important footnote to early studies of TEV protease specificity that is sometimes overlooked by today's protein engineers is that proline (but no other residue) is decidedly inhibitory in the P2′ position of its substrates [66]. Other studies in which large combinatorial libraries of peptides have been interrogated by TEV protease are in good general agreement with the results of systematic mutagenesis experiments [69,70].

The crystal structure of catalytically inactive (C151A) TEV protease in complex with the peptide substrate TTENLYFQSGT revealed that although the P7, P2′ and P3′ residues were clearly visible in the electron density map, their side chains do not engage in any noteworthy interactions with the enzyme (no density was observed for the P8 threonine residue) [71]. Of the side chains in the P6–P1′ positions of the peptide, only that of P5 asparagine does not make intimate contact with the enzyme, consistent with the observed lack of sequence specificity at this position [66]. The tolerance for a wide variety of side chains in the P1′ position of TEV protease substrates [68] is possible because the S1′ "pocket" of the enzyme is actually a long, shallow groove on its surface. Consequently, the side chain of the P1′ residue is partially exposed to solvent rather than completely buried within the complex (Fig. 1A). This is not a general property of potyviral proteases, however. In the co-crystal structure of the related (53% amino acid identity) tobacco vein mottling virus (TVMV) protease with its canonical peptide substrate, the S1′ pocket is small and round and the side chain of the P1′ serine residue projects directly into it (Fig. 1B) [72]. Hence, from a biotechnological standpoint, it is fortuitous that TEV protease, with its relaxed P1′ specificity, emerged as a reagent of choice for removing affinity tags.

Quite a lot is known about the performance of TEV protease under various reaction conditions and in the presence of different additives. It is active over a pH range between 6 and 9, but inactive at or below pH 5 [63]. In the same study, the enzyme was reported to be relatively insensitive to NaCl concentrations between 0.1 and 2.0 M. In a later study, TEV protease activity was observed to be greatest in the absence of monovalent salt, but decreased only moderately at NaCl concentrations up to 200 mM [73]. Although the optimum temperature for TEV protease activity is 30–34 °C, the enzyme retains significant activity at 4 °C [73,74]. Activity

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

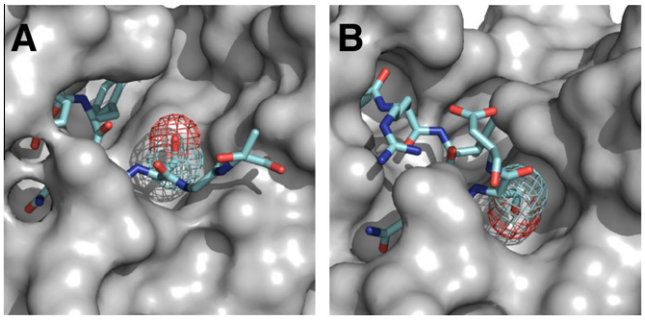

Fig. 1. Space-filling representations of the S1′ pockets of TEV protease (A) and TVMV protease (B). The canonical peptides that were co-crystallized with each enzyme are shown as stick representations. The side chains of the P1′ serine residues are encased in mesh. Note that the side chain of the P1′ serine projects along the surface of a shallow groove in TEV protease whereas the side chain of the corresponding serine points into a small, shallow pocket in TVMV protease.

drops off abruptly above 37 °C, probably due to denaturation of the enzyme [73]. TEV protease is not inhibited by 2.5% sucrose or 0.1% Triton X-100, but is completely inactivated by 0.01% sodium dodecyl sulfate [75]. The enzyme is insensitive to the protease inhibitors PMSF, AEBSF, and pepstatin A, relatively tolerant of aprotinin (up to 0.3 mg/ml) and leupeptin (up to 100 μM), and can tolerate up to 0.5 M urea [75,76]. Its catalytic activity is impeded by some detergents, however [25,77,78]. TEV protease activity is not impeded by reducing agents such as dithiothreitol (DTT) or chelators like EDTA, but, as expected, it is highly sensitive to thiol alkylating agents like iodoacetamide [75].

Although TEV protease is by far the best studied and most widely used potyviral protease, several others have also been characterized to varying degrees. These include the TVMV protease [73,79], the plum pox virus protease [80], and the turnip mosaic virus protease [81,82]. None of these enzymes offer any compelling advantages over TEV protease, however. TVMV protease expression vectors can be obtained from the Addgene plasmid repository (http://www.addgene.org).

Rhinovirus 3C protease

Because human rhinoviruses may account for the majority of common colds and have been associated with acute and chronic bronchitis along with other respiratory tract illnesses, the literature pertaining to rhinovirus 3C protease is dominated by the quest for inhibitors of this enzyme [83]. The same is true of the 3C-like proteases encoded by related viruses that cause polio, hepatitis A, and severe acute respiratory syndrome (SARS), among other diseases. Indeed, the poliovirus 3C protease was among the earliest enzyme of its type to be characterized [84–86]. It is not entirely clear why human rhinovirus 14 3C protease emerged from this group of picornaviruses as the prototypical enzyme for endoproteolytic removal of affinity tags from recombinant proteins. Similarly, there seems to be no compelling reason why TEV protease was selected for this role from among well over one hundred potyviruses that infect plants. In both cases, it may simply be a matter of happenstance. In any event, the tobacco etch virus and the human rhinovirus are members of the same viral superfamily [87,88]. Consequently, their 3C-like proteases are structurally and functionally related to one another and share many common characteristics.

Rhinovirus 14 3C protease is commercially available in the form of a glutathione S-transferase (GST) fusion protein called PreScission™ protease [89] and as a dual His_6-GST fusion protein. PreScission™ protease and rhinovirus 3C protease were identified as distinct enzymes with differing sequence specificity in the latest review article to include an in-depth survey of enzymatic methods for tag removal [3], but this is not the case. Rhinovirus 3C protease has also been fused to His_6 and MBP tags in *E. coli* [90], to DsbA [91], and very recently to the biotin acceptor peptide (BAP) [92]. The yield of BAP-tagged 3C was reported to be ~6 mg/g of wet cell paste. However, only about 60% of the BAP-tagged 3C protease was biotinylated *in vivo* (inexplicably, it was purified by conventional methods instead of by affinity chromatography on monomeric avidin resin). The yields of other tagged forms of the enzyme have not been reported. Unfortunately, rhinovirus 3C expression vectors are not presently available from open plasmid repositories, making them somewhat difficult to acquire.

Knowledge about the substrate specificity of human rhinovirus 3C protease has been gained mainly through studies of the enzyme from serotype 14. One comparative analysis of peptide substrates derived from natural 3C cleavage sites in the rhinovirus polyprotein established that the 2C/3A site was cleaved most efficiently [93] and that the specificity determinants are confined to the P5–P2′ sites [94], leading to the conclusion that the sequence ETLFQ↓GP is the optimum recognition site. At about the same time and using a very similar approach, a different group found that the 3B/3C cleavage site was the most efficiently processed of the natural sites [95] and that the specificity determinants are confined to the P5–P2′ sites [19], leading them to identify a different consensus sequence: PVVVQ↓GP. However, the canonical recognition site identified in the product literature accompanying PreScission™ protease is LEVLFQ↓GP. Walker and colleagues [90] speculated that this unnatural recognition site may have evolved in part from the observation that the peptide EVLFQPG was hydrolyzed nearly five times more efficiently than the parental ETLFQGP peptide derived from the natural 2C/3A polyprotein processing site [94]. In the same study, a longer peptide encompassing the 2C/3A processing site (DSLETLFQGPVYKDL) was cleaved about twice as efficiently as the shorter ETLFQGP peptide. Although the longer peptide includes the natural leucine residue in the P6 position, there was insufficient data to conclude that its presence (rather than some other difference between the two peptides) was responsible for the observed twofold variation in processing efficiency. Yet in a more recent study, replacing a leucine in the P6 position of a peptide substrate with an arginine led to a 10-fold reduction in the catalytic efficiency (k_{cat}/K_M) of processing by rhinovirus 3C protease [96]. Hence, it seems that the octapeptide LEVLFQPG, the so-called "PreScission protease site", approximates the optimum recognition site. Although, as discussed above, two groups arrived at somewhat different conclusions regarding the optimum sequence for a rhinovirus 3C protease cleavage site, both were in agreement on one key point: the requirement for the Gly-Pro dipeptide immediately after the scissile bond is very stringent. This is a noteworthy disadvantage of 3C compared to TEV protease; there is virtually no opportunity to produce a recombinant protein with a native N-terminus after digesting a fusion protein with 3C protease. There have been no "substrate protease" experiments performed with 3C protease (i.e., no genetic selections or screens), and consequently the identity of the "ideal" 3C cleavage site rests entirely on the systematic mutagenesis studies.

The impact of protease inhibitors, additives, and reaction conditions on the activity of rhinovirus 3C protease has been investigated [96–98], and in general, the results are very similar to those reported for TEV protease. However, rhinovirus 3C protease is rumored to have greater catalytic activity than TEV protease at low temperature (4 °C). Yet an exhaustive search of the literature failed to identify any instances in which the temperature-dependence of the two proteases has been compared directly. Instead, one finds only anecdotal statements that are unsupported by facts [99]. It would be gratifying to finally see this issue addressed experimentally.

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

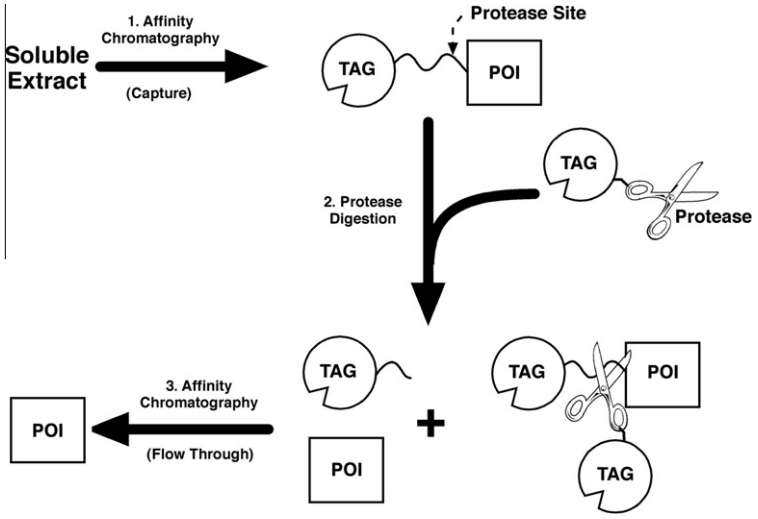

Fig. 2. A generic strategy for protein purification that utilizes an affinity-tagged endoprotease. See text for discussion.

While, as discussed above, the human rhinovirus serotype 14 3C protease has been widely used as a tool for endoproteolytic cleavage of fusion proteins, it suffers from the major disadvantage of leaving behind a Gly-Pro dipeptide on the N-terminus of a recombinant protein after digestion. The 3C proteases encoded by SARS virus [100] and poliovirus [101] also have a stringent requirement for specific residues on the C-terminal side of the scissile bond. However, the hepatitis A virus 3C protease has more relaxed specificity, as it has been shown to tolerate a variety of small aliphatic residues in the P1′ position and any residue except arginine and proline in the P2′ position [102–104]. This enzyme is straightforward to overproduce in *E. coli* and can be purified to greater than 95% homogeneity in two conventional chromatography steps, yielding more than 6 mg per liter of bacterial culture [105]. Perhaps it is worth reexamining the utility of hepatitis A 3C protease as a reagent for the endoproteolytic removal of affinity tags.

An advantage of affinity-tagged proteases

The ease of overproducing affinity-tagged forms of TEV and rhinovirus 3C proteases in *E. coli* has inspired the development of a generic protocol for protein purification [2,3] that is outlined in Fig. 2. The same affinity tag is attached to both the protein of interest and to an endoprotease. The tagged protein of interest (POI) is first purified from a soluble extract by some type of affinity chromatography, depending on the tag that is used. In the second step, the affinity tag is removed from the POI by the correspondingly tagged endoprotease. Finally, the digestion products are subjected to the same form of affinity chromatography a second time, in this case to remove undigested fusion protein substrate, the tagged protease, the cleaved tag, and any endogenous proteins that bound to the affinity resin during the first round of chromatography, leaving only the untagged POI in the unbound effluent. Variations of this strategy have been developed by a number of groups and have found widespread use in the structural genomics community, e.g., [106–109].

Troubleshooting endoproteolytic cleavage of affinity tags

As mentioned above, the removal of affinity tags has always been the Achilles heel of the fusion protein strategy. The problem of cleavage at secondary sites has been mitigated to a large degree by the use of viral proteases with very stringent sequence specificity. Still, it is not uncommon to encounter a situation where a fusion protein is cleaved very inefficiently by the chosen protease. There are several possible reasons for this. The most trivial explanation can be attributed to a failure to recognize the incompatibility of certain proteases with buffer components and additives. For example, a calcium-dependent protease such as factor Xa should never be used to cleave a fusion protein in a phosphate buffer (the most common buffer used for the purification of His-tagged proteins) because of the potential for the formation of insoluble calcium phosphate.

The inability of a protease to cleave a fusion protein may also be caused by steric hindrance, which can be of several types. For example, the cleavage site may be too close to ordered structure in the target protein. An example of such a situation is illustrated in Fig. 3. In this case, the substrate is an MBP-SycH fusion protein. SycH is an export chaperone for the protein tyrosine phosphatase YopH in *Yersinia pestis*, the plague-causing bacterium. When SycH is expressed in *E. coli* as a fusion to the C-terminus of MBP, with a TEV protease recognition site between the two domains, the fusion protein is not cleaved very efficiently (Fig. 3A). (In this experiment, the fusion protein was cleaved *in vivo*, using a separate plasmid expression vector to produce TEV protease [110].) In a second-generation construct, five glycine residues were inserted between the TEV protease recognition site and the N-terminus of SycH. As shown in Fig. 3B, this results in much more efficient processing of the fusion protein by TEV protease. However, no crystals of SycH with the additional glycine residues appended to its N-terminus were ever obtained. Instead, the form of SycH that was eventually crystallized had no extra residues added to its N-terminus. The structure revealed that the N-terminus of SycH is an integral part of the folded protein [111], consistent with the hypothesis that the close proximity of the TEV protease recognition site to the (structured) N-terminus of SycH in the original fusion protein was responsible for its failure to be cleaved efficiently and that the addition of five glycine residues alleviated this impediment.

The second level of steric hindrance applies to oligomeric proteins. It may be intuitively obvious how the presence of a tag like the 42 kDa MBP on a multimeric protein might create steric problems for affinity tag removal, yet even a tag as small as polyhistidine can present similar difficulties [112].

Finally, steric hindrance can sometimes result from the use of solubility-enhancing affinity tags like MBP [56] and NusA [113]. While these solubility enhancers can maintain their fusion partners in a soluble state in *E. coli*, and maybe in other heterologous hosts as well, their fusion partners may or may not be folded into their native conformations. At least some of these fusion proteins probably exist in the form of soluble aggregates that may not be

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

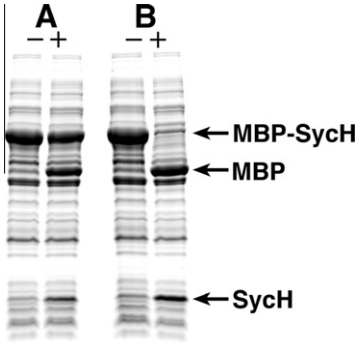

Fig. 3. Overcoming steric hindrance by inserting "spacer" residues between a protease recognition site and the N-terminus of the protein of interest. (A) Expression of MBP-SycH fusion protein with a TEV protease recognition site immediately adjacent to the N-terminal residue of SycH in the absence (−) or presence (+) of TEV protease. (B) Expression of an otherwise identical MBP-SycH fusion protein in which five consecutive glycine residues were inserted between the TEV protease recognition site and the N-terminal residue of SycH in the absence (−) and presence (+) of TEV protease.

readily accessible to an endoprotease, as has been shown to be the case for one MBP fusion protein [114].

Exoproteases

Exoproteases (aminopeptidases and carboxypeptidases) are not used nearly as frequently as endoproteases for the removal of affinity tags. This is mainly because most of the time affinity tags are joined to the N-terminus of a protein of interest, where they can best exert a positive effect on the yield and solubility of their fusion partners [2], and endoproteases like enteropeptidase and TEV protease can remove N-terminal affinity tags without leaving behind any nonnative residues. Yet sometimes there are compelling reasons to add a tag to the C-terminus of a recombinant protein. For example, a C-terminal tag is less likely to interfere with the removal of signal peptides and the secretion of proteins. However, endoproteolytic removal of C-terminal tags is complicated by the fact that the principal specificity determinants of endoproteolytic enzymes (e.g., Factor Xa, thrombin, enteropeptidase, tobacco etch virus and rhinovirus 3C proteases) are located on the N-terminal side of the scissile bond. Consequently, the removal of a C-terminal tag by any of them would leave behind a significant number of non-native residues (six in the case of TEV protease, which is equivalent in number to a hexahistidine tag). One solution to this problem is to use an exopeptidase, specifically a carboxypeptidase, to remove short C-terminal tags such as polyhistidine.

Aminopeptidases have also been used to remove short affinity tags from the N-termini of recombinant proteins [3], although not very frequently because endoproteases can accomplish this task quite well.

Metallocarboxypeptidases

Of the many varieties of carboxypeptidases, those belonging to the M14A subfamily of zinc-dependent metallocarboxypeptidases, historically known as the "digestive carboxypeptidases," have proven to be particularly useful for the processive hydrolysis of C-terminal residues from peptides and globular proteins [115,116]. Digestive carboxypeptidases can be classified on the basis of their sequence specificity. For example, type A carboxypeptidases preferentially remove C-terminal amino acid residues having aromatic or branched aliphatic side chains [117] whereas type B carboxypeptidases exhibit a strong preference for basic amino acids [118]. However, carboxypeptidase B will remove aromatic residues under certain conditions [119,120]. The A-type carboxypeptidases are further subdivided into type A1 and type A2 isoforms in rodents and humans [121]. Carboxypeptidase A1 preferentially catalyzes the removal of aliphatic residues from peptide substrates, while the A2 isoforms show higher specificity toward aromatic residues like phenylalanine and tryptophan. Carboxypeptidase A2 is not present in the bovine pancreas. Instead, the single bovine carboxypeptidase A (BoCPA) has relatively broad substrate specificity [121]. Exoproteolytic removal of C-terminal His-tags and polyarginine tags from globular proteins has been achieved using BoCPA and bovine carboxypeptidase B (BoCPB), respectively [116,122–129]. Neither enzyme will digest a C-terminal proline residue. Thus, incorporation of a proline at the appropriate location is one way to ensure that the digestion product will have a homogeneous endpoint.

A- and B-type carboxypeptidases are synthesized as zymogens with an N-terminal signal peptide that targets them for secretion from cells. Following secretion, the non-catalytic domain (termed the prodomain) is removed by proteolytic cleavage to give rise to the catalytically active carboxypeptidase. The prodomains not only maintain the enzymes in a catalytically dormant state, but they are also required for their proper folding [130]. The mature ~30–35 kDa BoCPA and BoCPB contain one and three disulfide bonds, respectively. Consequently, it is not advisable to use them in the presence of reducing agents.

As zinc-dependent enzymes, the digestive carboxypeptidases are inhibited by chelating agents like EDTA and 1,10-phenanthroline [120]. Sodium dodecyl sulfate (SDS) strongly inhibits carboxypeptidase A but hardly affects carboxypeptidase B activity (up to 1% w/v), while octyl-(polydisperse)oligooxyethylene and lauryl glutamate have only a mild inhibitory effect on both enzymes [131]. Carboxypeptidases A and B are optimally active between pH 7 and 9; they are inhibited by mercury but stimulated by cobalt [132,133].

Because of their complex folding and activation pathways, the metallocarboxypeptidases have been challenging enzymes to overproduce in heterologous systems. Even now, virtually every commercial preparation is obtained from natural sources. One group described the production of pre-pro-BoCPA in the yeast *Saccharomyces cerevisiae* as a secreted zymogen and its subsequent activation by trypsin, but the yield of recombinant protein was not reported [134]. More recently, proBoCPA was produced as a thioredoxin fusion protein in *E. coli*, but the yield of purified fusion protein (prior to activation and purification of BoCPA) was only 0.8 mg per liter of bacterial culture and most of the fusion protein accumulated in the form of insoluble aggregates [135]. Efforts have also been made to overproduce mammalian carboxypeptidase A subtypes A1 and A2 in yeast [130,136–138], but their narrower specificity makes these enzymes less appealing than BoCPA for the removal of C-terminal affinity tags from recombinant proteins. Moreover, none of the recombinant mammalian type A carboxypeptidases was engineered to include an affinity tag. Recently, a polyhistidine-tagged type A carboxypeptidase (MeCPA) from the fungal entomopathogen *Metarhizium anisopliae* has been secreted from baculovirus-infected insect cells and purified to homogeneity [117]. However, the yield of pure, active enzyme was a disappointing 0.25 mg per liter of conditioned medium. Yet this material proved to be highly active and, as predicted [139], the specificity of MeCPA was found to be even broader than that of BoCPA (Fig. 4). The same group has subsequently developed a strategy to produce MeCPA in *E. coli* that may be generally applicable to metallocarboxypeptidases. The proMeCPA zymogen is fused to the C-terminus of MBP to maintain it in a soluble form in the cytosol while the protein disulfide isomerase DsbC is co-expressed from a separate plasmid. This procedure yields about 1 mg of pure MeCPA per liter of bacterial culture (Austin et al., in preparation).

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

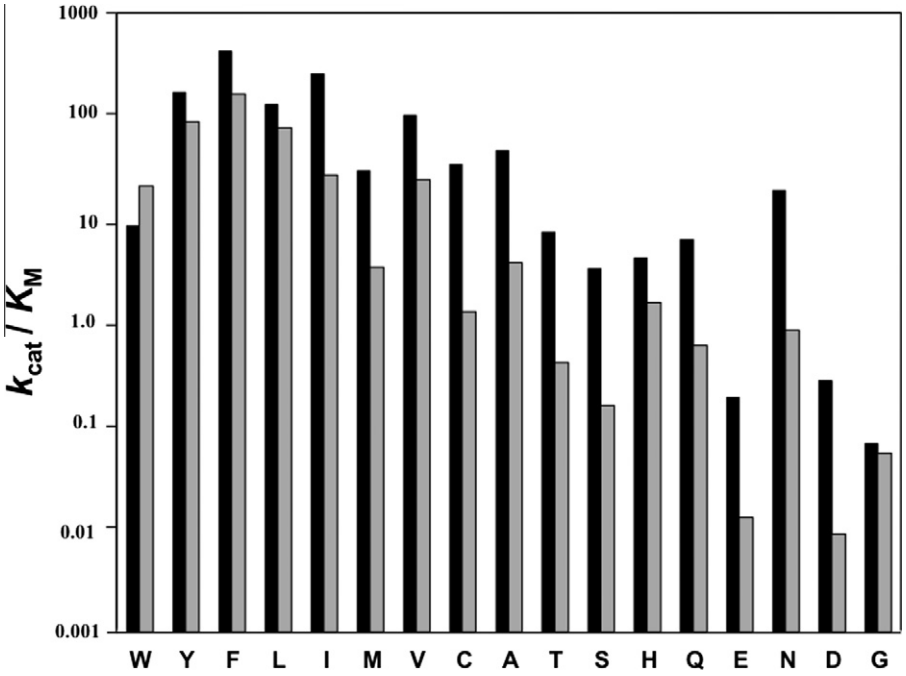

Fig. 4. Relative processing efficiency (k_{cat}/K_M) of peptide substrates by MeCPA (black bars) and BoCPA (gray bars). The peptide substrates used were VSQNPKX, wherein X was the variable amino acid. No processing of peptides terminating in Arg, Lys, or Pro was observed for either enzyme. Data were compiled from [117].

Porcine carboxypeptidase B has been expressed at very high levels (250 mg per liter) by secretion of the zymogen from the yeast *Pichia pastoris* [140]. Rat procarboxypeptidase B has been expressed in *E. coli*, refolded, activated by trypsin and purified to yield 40 mg per liter of culture medium [141]. However, neither of these enzymes included an affinity tag to facilitate their purification or their removal from the products of a carboxypeptidase digest. There is currently no known source of recombinant BoCPB.

Aminopeptidases

Interestingly, there are no known aminopeptidase counterparts to the digestive carboxypeptidases. However, endoproteases that exhibit little or no specificity for residues on the C-terminal side of the scissile bond (e.g., enteropeptidase and TEV protease) can be used to remove N-terminal affinity tags without leaving any extra residues on the N-terminus of the digestion product in most cases. Nevertheless, N-terminal exopeptidases are occasionally used for the removal of affinity tags. The TAGZyme™ system, a commercial product, is presently the only noteworthy tool of this type [142].

The TAGZyme system consists of three enzymatic reagents: dipeptidyl-aminopeptidase I (DAPase), glutamine cyclotransferase (Qcyclase), and pyroglutamylaminopeptidase (pGAPase). The principal enzyme is DAPase, which catalyzes the stepwise excision of dipeptides from the N-terminus of a polypeptide. Natural "road blocks" to DAPase digestion are a lysine or arginine in the first position of a dipeptide or a proline in either position. Consequently, DAPase cannot remove any N-terminal affinity tag. For instance, it could only remove a FLAG-tag (DYKDDDDK) if the tag were modified so that the proximal lysine residue did not occupy the first position of the second dipeptide, and it could not remove a polyarginine tag (RRRRR) or a Strep II tag (WSHPQFEK) under any circumstances. Instead, it is mainly used as a reagent for removing N-terminal polyhistidine tags, which need not contain any road-blocking residues. Incorporation of an arginine or lysine residue following the polyhistidine sequence will terminate DAPase digestion at that point, leaving behind the basic residue on the N-terminus of the digestion product. Alternatively, if a true "native" N-terminus is desired, then this can be accomplished by inserting a glutamine residue immediately after the polyhistidine sequence and utilizing, in addition to DAPase, the other two enzymes, Qcyclase and pGAPase. When glutamine is exposed at the N-terminus of the polypeptide chain, it can be cyclized to pyroglutamate by the action of Qcyclase, rendering it resistant to further digestion by DAPase. The pyroglutamyl residue is then excised by pGAPase to yield a product with a native N-terminus.

DAPase , which is also known as cathepsin C, is composed of three polypeptide chains, at least two of which are linked together by one or more disulfide bonds [143]. Consequently, its enzymatic activity is sensitive to reducing agents. Moreover, as a cysteine protease, DAPase is inhibited by thiol akylating agents. DAPase is active over a pH range of 4–8 and exhibits maximal activity at pH 5.5 [143]. A polyhistidine-tagged form of DAPase has been secreted from baculovirus-infected insect cells, yielding approximately 50 mg of active enzyme per liter [144].

The TAGZyme™ system has been further refined such that all three enzymatic reagents are modified by the addition of polyhistidine tags, thereby facilitating their removal from a digestion product [145]. Although clever and elegant, this approach is complicated by the fact that it requires, in most cases, three distinct enzymes to produce a digestion product with a native N-terminus. Additionally, care must be taken to remove the DAPase prior to the addition of pGAPase. Vectors and methods for the production of these reagents are proprietary and therefore not readily available to researchers. As a result, the cost of these reagents is a potential impediment to their use on a large scale, such as would be required for structural biology.

Concluding remarks

Considering its high specificity yet relatively good tolerance for a wide variety of residues in the P1′ position of its recognition site, coupled with its ease and economy of production and the ready availability of expression vectors from open sources (including affinity-tagged variants), TEV protease is probably the single best

endoprotease for the removal of N-terminal affinity tags. The ideal reagent would have all of these attributes along with totally relaxed P1′ specificity. Perhaps such an enzyme can be identified from among those encoded by other potyviruses or be engineered by rational design. As for the digestive carboxypeptidases, it would be useful to investigate whether a combination of A- and B-type enzymes can be used to remove a wider variety of short affinity tags from the C termini of recombinant proteins. Moreover, since the termini of proteins are often disordered, the ability to trim unstructured residues (apart from affinity tags) from their C termini might improve their propensity to crystallize. Yet it is unclear if the activity of these carboxypeptidases will be deterred upon encountering ordered structure in a globular protein. This needs to be tested. In principle, DAPase could be used in a similar fashion to remove disordered residues from the N-termini of proteins, subject to the same caveat. Proline is a roadblock for all of these exopeptidases and it would be useful to find a means of circumventing this issue. Finally, there are at least two ways in which endoproteases and exoproteases might be used in concert with each other that have yet to be explored. First, large C-terminal affinity tags (globular proteins) could be removed first by endoproteolytic cleavage at a designed site and then by treatment of the product with one or more carboxypeptidases to remove the remnants of the endoprotease recognition site from its C-terminus. The second potential application pertains to X-ray crystallography. It has been shown that the addition of trace amounts of relatively nonspecific endoproteases (e.g., chymotrypsin, subtilisin, thermolysin) to concentrated solutions of proteins prior to setting them up in crystallization trials sometimes enhances the growth of crystals [146,147]. In some cases, the endoproteolytic cleavage events occur within disordered internal loops in the protein, in which case further trimming of the "loose ends" left by the endoproteases might be affected by exopeptidases.

Acknowledgment

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

[1] I. Schechter, A. Berger, On the size of the active site in proteases, Biochem. Biophys. Res. Com. 27 (1967) 157–162.
[2] D.S. Waugh, Making the most of affinity tags, Trends Biotechnol. 23 (2005) 316–320.
[3] J. Arnau, C. Lauritzen, G.E. Petersen, J. Pedersen, Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins, Protein Expr. Purif. 48 (2006) 1–13.
[4] B.B. Sun-Lailam, J.M. Hevel, Efficient cleavage of problematic tobacco etch virus (TEV)-protein arginine methyltransferase constructs, Anal. Biochem. 387 (2009) 130–132.
[5] J. Wu, M. Filutowicz, Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale, Acta Biochim. Pol. 46 (1999) 591–599.
[6] E.A. Woestenenk, M. Hammarström, S. van den Berg, T. Härd, H. Berglund, His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors, J. Struct. Funct. Genom. 5 (2004) 217–229.
[7] A. Chant, C.M. Kraemer-Pecore, R. Watkin, G.G. Kneale, Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site, Protein Expr. Purif. 39 (2005) 152–159.
[8] M. Amor-Mahjoub, J.P. Suppini, N. Gomez-Vrielyunck, M. Ladjimi, The effect of the hexahistidine-tag in the oligomerization of HSC70 constructs, J. Chromatogr. B 844 (2006) 328–334.
[9] F. Renzi, G. Panetta, B. Vallone, M. Brunori, M. Arceci, I. Bozzoni, P. Laneve, E. Caffarelli, Large-scale purification and crystallization of the endoribonuclease XendoU: trouble-shooting with His-tagged proteins, Acta Crystallogr. F 62 (2006) 298–301.
[10] V. Gaberc-Porekar, V. Menart, S. Jevsevar, A. Vidensek, A. Stalc, Histidines in affinity tags and surface clusters for immobilized metal-ion affinity chromatography of trimeric tumor necrosis factor alpha, J. Chromatogr. A 852 (1999) 117–128.
[11] H. Horchani, S. Quertani, Y. Gargouri, A. Sayari, The N-termninal His-tag and the recombination process affect the biochemical properties of *Staphylococcus aureus* lipase produced in *Escherichia coli*, J. Mol. Catal. B Enzyme 61 (2009) 194–201.
[12] Y. Li, Self-cleaving fusion tags for recombinant protein production, Biotechnol. Lett. 33 (2011) 869–881.
[13] B. Ruan, K.E. Fisher, P.A. Alexander, V. Doroshko, P.N. Bryan, Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification, Biochemistry 43 (2004) 14539–14546.
[14] T. Panavas, C. Sanders, T.R. Butt, SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems, Methods Mol. Biol. 497 (2009) 303–317.
[15] G. Isetti, M.C. Maurer, Employing mutants to study thrombin residues responsible for factor XIII activation peptide recognition: a kinetic study, Biochemistry 46 (2007) 2444–2452.
[16] K.M. Bromfield, N.S. Quinsey, P.J. Dugan, R.N. Pike, Approaches to selective peptidic inhibitors of factor Xa, Chem. Biol. Drug Des. 68 (2006) 11–19.
[17] M.E. Gasparian, V.G. Ostapchenko, A.A. Schulga, D.A. Dolgikh, M.P. Kirpichnikov, Expression, purification, and characterization of human enteropeptidase catalytic subunit in *Escherichia coli*, Protein Expr. Purif. 31 (2003) 133–139.
[18] R.B. Kapust, J. Tözsér, 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.
[19] A.C. Long, D.C. Orr, J.M. Cameron, B.M. Dunn, J. Kay, A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase, FEBS Lett. 258 (1989) 75–78.
[20] J.J. Liepnieks, A. Light, The preparation and properties of bovine enterokinase, J. Biol. Chem. 254 (1979) 1677–1683.
[21] L.E. Anderson, K.A. Walsh, H. Neurath, Bovine enterokinase, purification, specificity, and some molecular properties, Biochemistry 16 (1977) 3354–3360.
[22] L.A. Collins-Racie, J.M. McColgan, K.L. Grant, E.A. DiBlasio-Smith, J.M. McCoy, E.R. LaVallie, Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory partner DsbA, Biotechnology 13 (1995) 982–987.
[23] L. Huang, H. Ruan, W. Gu, Z. Xu, P. Cen, L. Fan, Functional expression and purification of bovine enterokinase light chain in recombinant *Escherichia coli*, Prep. Biochem. Biotechnol. 37 (2007) 205–217.
[24] S.I. Choi, H.W. Song, J.W. Moon, B.L. Seong, Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology, Biotechnol. Bioeng. 75 (2001) 718–724.
[25] J.M. Vergis, M.C. Wiener, The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal, Protein Expr. Purif. 78 (2011) 139–142.
[26] T. Hosfield, Q. Lu, Influence of the amino acid residue downstream of (Asp)4Lys on enterokinase cleavage of a fusion protein, Anal. Biochem. 269 (1999) 10–16.
[27] Y.-T. Kim, W. Nishii, M. Matsushima, H. Inoue, H. Ito, S.J. Park, K. Takahashi, Substrate specificities of porcine and bovine enteropeptidases toward the peptide Val-(Asp)4-Lys-Ile-Val-Gly and its analogs, Biosci. Biotechnol. Biochem. 72 (2008) 905–908.
[28] K.T. Boulware, P.S. Daugherty, Protease specificity determination by using cellular libraries of peptide substrates (CLiPS), Proc. Natl Acad. Sci. USA 103 (2006) 7583–7588.
[29] O.W. Liew, J.P. Ching Chong, T.G. Yandle, S.O. Brennan, Preparation of recombinant thioredoxin fused N-terminal proCNP: analysis of enterokinase cleavage products reveals new enterokinase cleavage sites, Protein Expr. Purif. 41 (2005) 332–340.
[30] M.E. Gasparian, M.L. Bychkov, D.A. Golgikh, M.P. Kirpichnikov, Strategy for improvement of enteropeptidase efficiency in tag removal processes, Protein Expr. Purif. (2011) Epub ahead of print.
[31] H. Chun, K. Joo, J. Lee, H.C. Shin, Design and efficient production of bovine enterokinase light chain with higher specificity in *E. coli*, Biotechnol. Lett. 33 (2011) 1227–1232.
[32] W.G. Owen, C.T. Esmon, C.M. Jackson, The conversion of prothrombin to thrombin. I. Characterization of the reaction products formed during the activation of bovine prothrombin, J. Biol. Chem. 249 (1974) 594–605.
[33] P.D. Bishop, K.B. Lewis, J. Schultz, K.M. Walker, Comparison of recombinant human thrombin and plasma-derived human alpha-thrombin, Semin. Thromb. Hemostasis 32 (2006) 86–97.
[34] L.A. Bush-Pelc, F. Marino, Z. Chen, A.O. Pineda, F.S. Mathews, E. Di Cera, Important role of the cys-191 cys-220 disulfide bond in thrombin function and allostery, J. Biol. Chem. 282 (2007) 27165–27170.
[35] S. Le Borgne, M. Graber, Amidase activity and thermal stability of human thrombin, Appl. Biochem. Biotechnol. 48 (1994) 125–135.
[36] J.-Y. Chang, Thrombin specificity: requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate, Eur. J. Biochem. 151 (1985) 217–224.
[37] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase, Gene 67 (1988) 31–40.
[38] A. Houbrechts, B. Moreau, R. Abagyan, V. Mainfroid, G. Preaux, A. Lamproye, A. Poncin, E. Goormaghtigh, J.M. Ruysschaert, J.A. Martial, K. Goraj, Second - generation octarellins: two new de novo (beta/alpha)8 polypeptides designed

for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability, Protein Eng. 8 (1995) 249–259.
[39] B.F. Le Bonniec, T. Myles, T. Johnson, C.G. Knight, C. Tapparelli, S.R. Stone, Characterization of the P2' and P3' specificities of thrombin using fluorescence-quenched substrates and mapping of the subsites by mutagenesis, Biochemistry 35 (1996) 7114–7122.
[40] H.M. Petrassi, J.A. Williams, J. Li, C. Tumanut, J. Ek, T. Nakai, B. Masick, B.J. Backes, J.L. Harris, A strategy to profile prime and non-prime proteolytic substrate specificity, Bioorg. Med. Chem. Lett. 15 (2005) 3162–3166.
[41] O. Shilling, C.M. Overall, Proteone-derived database-searchable peptide libraries for identifying protease cleavage sites, Nat. Biotechnol. 26 (2008) 685–694.
[42] R.J. Jenny, K.G. Mann, R.L. Lundblad, A critical review of the methods for cleaving fusion proteins with thrombin and factor Xa, Protein Expr. Purif. 31 (2003) 1–11.
[43] C.M. Jackson, T.F. Johnson, D.J. Hanahan, Studies on bovine Factor X. I. Large-scale purification of the bovine plasma protein posessing factor X activity, Biochemistry 7 (1968) 4492–4505.
[44] H.-H. Heidtmann, R.E. Kontermann, Cloning and recombinant expression of mouse coagulation factor X, Thrombosis Res. 92 (1998) 33–41.
[45] A.E. Rudolph, M.P. Mullane, R. Porche-Sorbet, S. Tsuda, J.P. Miletich, Factor Xa St. Louis II: identification of a glycine substitution at residue 7 and characterization of the recombinant protein, J. Biol. Chem. 271 (1996) 28601–28606.
[46] A.E. Rudolph, M.P. Mullane, R. Porche-Sorbet, J.P. Miletich, Expression, purification and characterization of recombinant human factor X, Protein Expr. Purif. 10 (1977) 373–378.
[47] C. Manithody, L. Yang, A.R. Rezaie, Role of basic residues of the autolysis loop in the catalytic function of factor Xa, Biochemistry 41 (2002) 6780–6788.
[48] R. Lottenberg, J.A. Hall, E. Pautler, A. Zupan, U. Christensen, C.M. Jackson, The action of factor Xa on peptide p-nitroanilide substrates: substrate selectivity and examination of hydrolysis with different reaction conditions, Biochim. Biophys. Acta 874 (1986) 326–336.
[49] J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, Proc. Natl Acad. Sci. USA 97 (2000) 7754–7759.
[50] H.-J. Hsu, K.-C. Tsai, Y.-K. Sun, H.-J. Chang, Y.-J. Huang, H.-M. Yu, C.-H. Lin, S.-S. Mao, A.-S. Yang, Factor Xa active site substrate specificity with substrate phage display and computational molecular modeling, J. Biol. Chem. 283 (2008) 12343–12353.
[51] J.P. Ludeman, R.N. Pike, K.M. Bromfield, P.J. Duggan, J. Cianci, B. Le Bonniec, J.C. Shjisstock, S.P. Bottomley, Determination of the P1', P2' and P3' subsite-specificity of factor Xa, Int. J. Biochem. Cell Biol. 35 (2003) 221–225.
[52] J.C. Carrington, W.G. Dougherty, Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease, J. Virol. 61 (1987) 2540–2548.
[53] W.G. Dougherty, T.D. Parks, Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains, Virology 183 (1991) 449–456.
[54] J.C. Carrington, W.G. Dougherty, A viral cleavage site cassette: Identification of amino acid sequences required for tobacco etch virus polyprotein processing, Proc. Natl Acad. Sci. USA 85 (1988) 3391–3395.
[55] T.D. Parks, K.K. Leuther, E.D. Howard, S.A. Johnston, W.G. Dougherty, Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase, Anal. Biochem. 216 (1994) 413–417.
[56] 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.
[57] R.B. Kapust, K.M. Routzahn, D.S. Waugh, Processive degradation of nascent polypeptides, triggered by tandem AGA codons, limits the accumulation of recombinant tobacco etch virus protease in *Escherichia coli* BL21(DE3), Protein Expr. Purif. 24 (2002) 61–70.
[58] S. van den Berg, P.-A. Löfdahl, T. Härd, H. Berglund, Improved solubility of TEV protease by directed evolution, J. Biotechnol. 121 (2006) 291–298.
[59] P.G. Blommel, B.G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease, Protein Expr. Purif. 55 (2007) 53–68.
[60] B. Miladi, H. Bouallagui, C. Dridi, A. El Marjou, G. Boeuf, P. Di Martino, F. Dufour, A. Elm'Selmi, A new tagged-TEV protease: Construction, optimisation of production, purification and test activity, Protein Expr. Purif. 75 (2011) 75–82.
[61] J.E. Tropea, C. Cherry, D.S. Waugh, Expression and purification of soluble His(6)-tagged TEV protease, Methods Mol. Biol. 498 (2009) 297–307.
[62] L.J. Lucast, R.T. Batey, J.A. Doudna, Large-scale purification of a stable form of recombinant tobacco etch virus protease, BioTechniques 30 (2001) 544–0546.
[63] T.D. Parks, E.D. Howard, T.J. Wolpert, D.J. Arp, W.G. Dougherty, Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form, Virology 210 (1995) 194–201.
[64] L.D. Cabrita, D. Gilis, A.L. Robertson, Y. Dehouck, M. Rooman, S.P. Bottomley, Enhancing the stability and solubility of TEV protease using in silico design, Protein Sci. 16 (2007) 2360–2367.
[65] S. van den Berg, P.A. Löfdahl, T. Härd, H. Berglund, Improved solubility of TEV protease by directed evolution, J. Biotechnol. 121 (2006) 291–298.
[66] W.G. Dougherty, S.M. Cary, T.D. Parks, Molecular genetic analysis of a plant virus polyprotein cleavage site: a model, Virology 171 (1989) 356–364.
[67] J. Tözsér, J.E. Tropea, S. Cherry, P. Bagossi, T.D. Copeland, A. Wlodawer, D.S. Waugh, Comparison of the substrate specificity of two potyvirus proteases, FEBS J. 272 (2005) 514–523.
[68] R.B. Kapust, J. Tözsér, T.D. Copeland, D.S. Waugh, The P1' specificity of tobacco etch virus protease, Biochem. Biophys. Res. Commun. 294 (2002) 949–955.
[69] G. Kostallas, P. Samuelson, Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates, Appl. Environ. Microbiol. 76 (2010) 7500–7508.
[70] K.T. Boulware, A. Jabaiah, P.S. Daugherty, Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics, Biotechnol. Bioeng. 106 (2010) 339–346.
[71] J. Phan, A. Zdanov, A.G. Evdokimov, J.E. Tropea, H.K. Peters 3rd, R.B. Kapust, M. Li, A. Wlodawer, D.S. Waugh, Structural basis for the substrate specificity of tobacco etch virus protease, J. Biol. Chem. 277 (2002) 50564–50572.
[72] P. Sun, B.P. Austin, J. Tözsér, D.S. Waugh, Structural determinants of tobacco vein mottling virus substrate specificity, Protein Sci. 19 (2010) 2240–2251.
[73] S. Nallamsetty, R.B. Kapust, J. Tözsér, S. Cherry, J.E. Tropea, T.D. Copeland, D.S. Waugh, Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vito and in vitro, Protein Expr. Purif. 38 (2004) 108–115.
[74] D.A. Polayes, T.D. Parks, S.A. Joghston, W.G. Dougherty, Application of TEV protease in protein production, Methods Mol. Med. 13 (1998) 169–183.
[75] D.A. Polayes, A. Goldstein, G. Ward, A.J. Hughes Jr., TEV protease, recombinant. A site-specific protease for efficient cleavage of affinity tags from expressed proteins, FOCUS 16 (1994) 2–5.
[76] W.G. Dougherty, D. Parks, S.M. Cary, J.F. Bazan, R.J. Fletterick, Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase, Virology 172 (1989) 302–310.
[77] A.K. Mohanty, C.R. Simmons, M.C. Wiener, Inhibition of tobacco etch virus protease activity by detergents, Protein Expr. Purif. 27 (2003) 109–114.
[78] A.K. Lundbäck, S. van den Berg, J. Hebert, H. Berglund, S. Eschaghi, Exploring the activity of tobacco etch virus protease in detergent solutions, Anal. Biochem. 382 (2008) 69–71.
[79] D.C. Hwang, D.H. Kim, B.H. Kang, B.D. Song, K.Y. Choi, Molecular cloning, expression and purification of nuclear inclusion A protease from tobacco vein mottling virus, Mol. Cells 10 (2000) 148–155.
[80] N. Zheng, J. de Jesús Pérez, Z. Zhang, E. Domínguez, J.A. Garcia, Q. Xie, Specific and efficient cleavage of fusion proteins by recombinant plum pox virus NIa protease, Protein Expr. Purif. 57 (2008) 153–162.
[81] D. Kim, B.H. Kang, J.S. Han, K.Y. Choi, Temperature and salt effects on the proteolytic function of turnip mosaic potyvirus nuclear inclusion protein a exhibiting a low-temperature optimum activity, Biochim. Biophys. Acta 1480 (2000) 29–40.
[82] H. Kang, Y.J. Lee, J.H. Goo, W.H. Park, Determination of the substrate specificity of turnip mosaic virus NIa protease using a genetic method, J. Gen. Virol. 82 (2001) 3115–3117.
[83] Q.M. Wang, S.-H. Chen, Human rhinovirus 3C protease as a potential target for the development of antiviral agents, Curr. Protein Pept. Sci. 8 (2007) 19–27.
[84] M.J. Nicklin, H.G. Kräusslich, H. Toyoda, J.J. Dunn, E. Wimmer, Poliovirus polypeptide precursors: expression and in vitro processing by exogenous 3C and 2A proteinases, Proc. Natl Acad. Sci. USA 84 (1987) 4002–4006.
[85] M.J. Nicklin, K.S. Harris, P.V. Pallai, E. Wimmer, Poliovirus proteinase 3C: large-scale expression, purification, and specific cleavage activity on natural and synthetic substrates in vitro, J. Virol. 62 (1988) 4586–4593.
[86] P.V. Pallai, F. Burkhardt, M. Skoog, K. Schreiner, P. Bax, K.A. Cohen, G. Hansen, D.E.H. Palladino, K.S. Harris, M.J. Nicklin, E. Wimmer, Cleavage of synthetic peptides by purified poliovirus 3C proteinase, J. Biol. Chem. 264 (1989) 9738–9741.
[87] A.E. Gorbalenya, A.P. Donchenko, V.M. Blinov, E.V. Koonin, Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold, FEBS Lett. 243 (1989) 103–114.
[88] J.F. Bazan, R.J. Fletterick, Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses, Virology 171 (1989) 637–639.
[89] L.E. Leong, The use of recombinant fusion proteases in the affinity purification of recombinant proteins, Mol. Biotechnol. 12 (1999) 269–274.
[90] P.A. Walker, L.E. Leong, P.W. Ng, S.H. Tan, S. Waller, D. Murphy, A.G. Porter, Efficient and rapid affinity purification of proteins using recombinant fusion proteases, Biotechnology 12 (1994) 601–605.
[91] Y. Zhang, D.R. Olsen, K.B. Nguyen, P.S. Olson, E.T. Rhodes, D. Mascarenhas, Expression of eukaryotic proteins in soluble form in *Escherichia coli*, Protein Expr. Purif. 12 (1998) 159–165.
[92] J. Youell, D. Fordham, K. Firman, Production and single-step purification of EGFP and a biotinylated version of the human rhinovirus 14 3C protease, Protein Expr. Purif. (2011) Epub ahead of print.
[93] M.G. Cordingley, R.B. Register, P.L. Callahan, V.M. Garsky, R.J. Colonno, Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in *Escherichia coli*, J. Virol. 63 (1989) 5037–5045.
[94] M.G. Cordingley, P.L. Callahan, V.V. Sardana, V.M. Garsky, R.J. Colonno, Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro, J. Biol. Chem. 265 (1990) 9062–9065.

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005

[95] D.C. Orr, A.C. Long, J. Kay, B.M. Dunn, J.M. Cameron, Hydrolysis of a series of synthetic peptide substrates by the human rhinovirus 14 3C proteinase, cloned and expressed in *Escherichia coli*, J. Gen. Virol. 70 (1989) 2931–2942.
[96] G.J. Davis, Q.M. Wang, G.A. Cox, R.B. Johnson, M. Wakulchik, C.A. Dotson, E.C. Villarreal, Expression and purification of recombinant rhinovirus 14 3CD proteinase and its comparison to the 3C prorteinase, Arch. Biochem. Biophys. 346 (1997) 125–130.
[97] Q.M. Wang, R.B. Johnson, G.A. Cox, E.C. Villarreal, R.J. Loncharich, A continuous colorimetric assay for rhinovirus-14 3C protease using peptide p-nitroanilides as substrates, Anal. Biochem. 252 (1997) 238–245.
[98] Q.M. Wang, R.B. Johnson, Activation of human rhinovirus-14 3C protease, Virology 280 (2001) 80–86.
[99] V. Rubio, Y. Shen, Y. Saijo, Y. Liu, G. Gusmaroli, S.P. Dinesh-Kumar, X.W. Deng, An alternative tandem affinity purification strategy applied to *Arabidopsis* protein complex isolation, The Plant J. 41 (2005) 767–778.
[100] K. Fan, L. Ma, X. Han, H. Liang, P. Wei, Y. Liu, L. Lai, The substrate specificity of SARS coronavirus 3C-like proteinase, Biochem. Biophys. Res. Comm. 329 (2005) 934–940.
[101] J.R. Weidner, B.M. Dunn, Development of synthetic peptide substrates for the poliovirus 3C proteinase, Arch. Biochem. Biophys. 286 (1991) 402–408.
[102] J.R. Petithory, F.R. Masiarz, J.F. Kirsch, D.V. Santi, B.A. Malcom, A rapid method for determination of endoproteinase substrate specificity: specificity of the 3C proteinase from hepatitis A virus, Proc. Natl Acad. Sci. USA 88 (1991) 11510–11514.
[103] C. Huitema, L.D. Eltis, A fluorescent protein-based biological screen of proteinase activity, J. Biomol. Screen. 15 (2010) 224–229.
[104] D.A. Jewell, W. Swietnicki, B.M. Dunn, B.A. Malcom, Hepatitis A virus 3C proteinase substrate specificity, Biochemistry 31 (1992) 7862–7869.
[105] B.A. Malcom, S.M. Chin, D.A. Jewell, J.R. Stratton-Thomas, K.B. Thudium, R. Ralston, S. Rosenberg, Expression and characterization of recombinant hepatitis A virus 3C proteinase, Biochemistry 31 (1992) 3358–3363.
[106] Y. Kim, L. Bigelow, M. Borovilos, I. Dementieva, E. Duggan, W. Eschenfeldt, C. Hatzos, G. Joachimiak, H. Li, N. Maltseva, R. Mulligan, P. Quartey, A. Slather, L. Stols, L. Volkart, R. Wu, M. Zhou, A. Joachimiak, High-throughput protein purification for X-ray crystallography and NMR, Adv. Protein Chem. Struct. Biol. 75 (2008) 85–105.
[107] 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.
[108] J.E. Tropea, S. Cherry, S. Nallamsetty, C. Bignon, D.S. Waugh, A generic method for the production of recombinant proteins in *Escherichia coli* using a dual hexahistidine-maltose-binding protein affinity tag, Methods Mol. Biol. 363 (2007) 1–19.
[109] 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.
[110] R.B. Kapust, D.S. Waugh, Controlled intracellular processing of fusion proteins by TEV protease, Protein Expr. Purif. 19 (2000) 312–318.
[111] J. Phan, J.E. Tropea, D.S. Waugh, Structure of the Yersinia pestis type III secretion chaperone SycH in complex with a stable fragment of YscM2, Acta Crystallogr. D 60 (2004) 1591–1594.
[112] M. Kenig, S. Peternel, V. Gaberc-Porekar, V. Menhart, Influence of protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins, J. Chromatogr. A 110 (2006) 293–306.
[113] G.D. Davis, C. Elisee, D.M. Newham, R.G. Harrison, New fusion protein systems designed to give soluble expression in *Escherichia coli*, Biotechnol. Bioeng. 65 (1999) 382–388.
[114] Y. Nominé, T. Ristriani, C. Laurent, J.F. Lefèvre, E. Weiss, G. Travé, Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein, Protein Expr. Purif. 23 (2001) 22–32.
[115] J. Vendrell, E. Querol, F.X. Avilés, Metallocarboxypeptidases and their protein inhibitors: structure, function and biomedical properties, Biochim. Biophys. Acta 1477 (2000) 284–298.
[116] D. Fernández, I. Pallarès, J. Vendrell, F.X. Avilés, Progress in metallocarboxypeptidases and their small molecular weight inhibitors, Biochimie 92 (2010) 1484–1500.
[117] B.P. Austin, J. Tözsér, P. Bagossi, J.E. Tropea, D.S. Waugh, The substrate specificity of *Metarhizium anisopliae* and *Bos taurus* carboxypeptidases A: insights into their use as tools for the removal of affinity tags, Protein Expr. Purif. 77 (2011) 53–61.
[118] A.K. Tan, D.L. Eaton, Activation and characterization of procarboxypeptidase B from human plasma, Biochemistry 34 (1995) 5811–5816.
[119] N. Zisapel, N. Kurn-Abramowitz, M. Sokolovsky, Basic and non-basic substrates of carboxypeptidase B, Eur. J. Biochem. 35 (1973) 507–511.
[120] E. Wintersberger, D.J. Cox, H. Neurath, Bovine pancreatic procarboxypeptidase B. I. Isolation, properties, and activation, Biochemistry 1 (1962) 1069–1078.
[121] S.J. Gardell, C.S. Craik, E. Clauser, E.J. Goldsmith, C.-B. Stewart, M. Graf, W.J. Rutter, A novel rat carboxypeptidase, CPA2: characterization, molecular cloning and evolutionary implications on substrate specificity in the carboxypeptidase gene family, J. Biol. Chem. 263 (1988) 17828–17836.
[122] G. Lilus, M. Persson, L. Bulow, K. Mosbach, Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails, Eur. J. Biochem. 198 (1991) 499–504.
[123] K.C. Garcia, M.D. Tallquist, L.R. Pease, A. Brunmark, C.A. Scott, M. Degano, E.A. Stura, P.A. Peterson, I.A. Wilson, L. Teyton, αβ T cell receptor interactions with syngeneic and allogeneic ligands: affinity measurements and crystallization, Proc. Natl Acad. Sci. USA 94 (1997) 13838–13843.
[124] K. Persson, L. Åslund, B. Grahn, J. Hanke, O. Heby, *Trypanosoma cruzi* has not lost its S-adenosylmethionine decarboxylase: characterization of the gene and the encoded enzyme, Biochem. J. 333 (1998) 527–537.
[125] J. Celia, E. Wilson-Kubalek, R.A. Milligan, L. Teyton, Structure and function of a membrane-bound murine MHC class I molecule, Proc. Natl Acad. Sci. USA 96 (1999) 5634–5639.
[126] J.C. Smith, R.B. Derbyshire, E. Cook, L. Dunthorne, J. Viney, S.J. Brewer, H.M. Sassenfeld, L.D. Bell, Chemical synthesis and cloning of a poly(arginine)-coding gene fragment designed to aid polypeptide purification, Gene 32 (1984) 321–327.
[127] S.M. Fuchs, R.T. Raines, Polyarginine as a multifunctional fusion tag, Protein Sci. 14 (2005) 1538–1544.
[128] M. Rickert, X. Wang, M.J. Boulanger, N. Goriatcheva, K.C. Garcia, The structure of interleukin-2 complexed with its alpha receptor, Science 308 (2005) 1477–1480.
[129] J. Hirabayashi, K. Kasai, Arginine-tail method, an affinity tag procedure utilizing anhydrotrypsin agarose, J. Chromatogr. 597 (1992) 181–187.
[130] M.A. Phillips, W.J. Rutter, Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1, Biochemistry 35 (1996) 6771–6776.
[131] M. Kamo, A. Tsugita, Carboxypeptidase digestion in the presence of detergents, J. Biochem. 102 (1987) 243–246.
[132] D.S. Auld, B.L. Vallee, Kinetics of carboxypeptidase A. The pH dependence of tripeptide hydrolysis catalyzed by zinc, cobalt, and manganese enzymes, Biochemistry 9 (1970) 4352–4359.
[133] X.-Y. Zhang, S.-X. Li, Q.-S. Yuan, The renaturation of procarboxypeptidase B by urea gradient gel filtration and some properties of recombinant carboxypeptidase B, Protein Peptide Lett. 12 (2005) 671–676.
[134] J.H. Goo, K.H. Kim, K.Y. Choi, Cloning, sequencing and expression of the gene encoding a major allotypic preprocarboxypeptidase A from bovine pancreas, Gene 165 (1995) 333–334.
[135] R. Seddi, J.C. Chaix, A. Puigserver, X.J. Guo, Expression of a soluble and activatable form of bovine procarboxypeptidase A in *Escherichia coli*, Protein Expr. Purif. 27 (2003) 220–228.
[136] R.M. Laethem, T.A. Blumenkopf, M. Cory, L. Elwell, C.P. Moxham, P.H. Ray, L.M. Walton, G.K. Smith, Expression and characterization of human pancreatic preprocarboxypeptidase A1 and preprocarboxypeptidase A2, Arch. Biochem. Biophys. 332 (1996) 8–18.
[137] D. Reverter, S. Ventura, V. Villegas, J. Vendrell, F.X. Avilés, Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway, J. Biol. Chem. 273 (1998) 3535–3541.
[138] S. Darnis, J. Juge, C. Marino, F.X. Avilés, A. Puigserver, J.C. Chaix, X.J. Guo, Cloning, sequencing and functional expression of a cDNA encoding porcine pancreatic preprocarboxypeptidase A1, Eur. J. Biochem. 259 (1999) 719–725.
[139] L. Joshi, R.J. St. Leger, Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into ifected tissues by the fungal entomopathogen *Metarhizium anisopliae*, J. Biol. Chem. 274 (1999) 9803–9811.
[140] S. Ventura, V. Villegas, J. Sterner, J. Larson, J. Vendrell, C.L. Hershberger, F.X. Avilés, Mapping the pro-region of carboxypeptidase B by protein engineering, J. Biol. Chem. 274 (1999) 19925–19933.
[141] S.-X. Li, Y.-J. Zhang, L.-P. Tian, Q.-S. Yuan, Y. Gong, Cloning and expression of a new rat procarboxypeptidase B gene in *Escherichia coli* and purification of recombination carboxypeptidase B, Protein Peptide Lett. 10 (2003) 581–590.
[142] J. Pedersen, C. Lauritzen, M.T. Madsen, S.W. Dahl, Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases, Protein Expr. Purif. 15 (1999) 389–400.
[143] I. Dolenc, B. Turk, G. Pungercic, A. Ritonja, V. Turk, Oligomeric structure and substrate induced inhibition of human cathepsin C, J. Biol. Chem. 270 (1995) 21626–21631.
[144] C. Lauritzen, J. Pedersen, M.T. Madsen, J. Justesen, P.M. Martensen, S.W. Dahl, Active recombinant rat dipeptidyl aminopeptidase I (cathepsin C) produced using the baculovirus expression system, Protein Expr. Purif. 14 (1998) 434–442.
[145] F. Schäfer, A. Schäfer, K. Steinert, A highly specific sysem for efficient enzymatic removal of tags from recombinant proteins, J. Biomol. Tech. 13 (2002) 158–171.
[146] A. Dong, X. Xu, A.M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G.V. Avvakumov, A. Bochkarev, L.-G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M.D. Herman, B.-S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H.-W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J.R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng, H. Zhu, In situ proteolysis for protein crystallization and structure determination, Nat. Methods 4 (2007) 1019–1021.
[147] A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: and update, PLoS ONE 4 (2009) e5094.

Please cite this article in press as: D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.08.005