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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
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 19 August 2011

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 .............................................................................................................................. 283
Endoproteases .......................................................................................................................... 284
Enteropeptidase ....................................................................................................................... 285
Thrombin ..................................................................................................................................... 285
Factor Xa ..................................................................................................................................... 285
TEV protease ............................................................................................................................ 286
Rhinovirus 3C protease ............................................................................................................ 287
An advantage of affinity-tagged proteases .................................................................................. 288
Troubleshooting endoproteolytic cleavage of affinity tags .......................................................... 288
Exoproteases ................................................................................................................................ 289
Metallocarboxypeptidases ......................................................................................................... 289
Aminopeptidases ....................................................................................................................... 290
Concluding remarks .................................................................................................................. 290
Acknowledgment ...................................................................................................................... 291
References ..................................................................................................................................... 291

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
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 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., LVPFRGS and ENLYFQG in the case of thrombin and TEV protease, respectively). The Michaelis constants \( \left( k_{\text{M}} \right) \) for the two classes of enzymes are also similar, falling in the low to mid micromolar range, but the catalytic rate constants \( \left( k_{\text{cat}} \right) \) 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 \( k_{\text{cat}} \) 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.

### 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 \( \alpha \)-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 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.

### 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 | His60 | 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 | Very promiscuous |
| Factor Xa | Plasma HEK 293 cells | 42 + 17 | None | Reducing agents | LVPR;GS | Very promiscuous |
| TEV Protease | E. coli | 27 | His60 MBP GST Stepp II | Thiolykallating agents | ENLYFQ;G | P1' can vary [68] |
| Rhinovirus 3C Protease | E. coli | 27 | His60 GST His60 GST | Thiolykallating agents | LEVLFQ;GP | Same as PreScission™ protease |
| **Exoproteases** | | | | | | |
| Carboxypeptidase A | Pancreas E. coli, S. cerevisiae | 33 | His60 | Reducing 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 | C-terminal Lys and Arg | Will cleave hydrophobic residues under certain conditions [119,120] |
| DAPase | Kidney S. frugiperda (baculovirus) | 23 + 16 + 6 | His60 | Reducing agents | N-terminal dipeptides | P2 = Pro, Lys, Arg |

1 Abbreviations used: factor Xa, blood coagulation factor X; TEV, tobacco etch virus; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; AERSF, 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 mosaic 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, dipetidyl-amino- peptidase I; Qcyclease, glutamine cyclotransferase; pGAPase, pyroglytulaminopeptidase.

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].
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 (LVPRGS) has an interesting history. This sequence closely resembles that of a natural thrombin cleavage site (LVPRGS) 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, LVPRGS 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 lysine [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
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 la” or “Nla” [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 ENLYFQG/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 of 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 [63,68]. However, autolysis of TEV protease can be avoided by introducing amino acid substitutions in the vicinity of the internal cleavage site [62,68]. 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 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 Nla protein rather than the 27 kDa catalytic domain) exceeded that of the substrates, which were far below the Km 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 TNELFYQSGT 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 potviral 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
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 PMSE, 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 TMV 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. TMV protease expression vectors can be obtained from the Addgene plasmid repository (http://www.addgene.org).

**Rhinovirus 3C protease**

Because human rhinoviruses may be associated 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 enzymes 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 His6-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 His6 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 ETLFQGP 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 LEVLFGQ/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 (DSLETLFGQPVYKDL) 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 (kcat/Km) of processing by rhinovirus 3C protease [96]. Hence, it seems that the octapeptide LEVLFGQ, 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.
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 (POI) 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.

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
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]. Exopeptolytic 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 aszymogens 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 15% w/v), while octyl-(polydisperse)oligoxyethylene 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).
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 alkylating 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

---

**Fig. 4.** Relative processing efficiency ($\frac{k_{\text{cat}}}{K_M}$) 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].
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 endoproteolytic 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) 307–324.
[3] J. Arnaud, C. Laurant, 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-Lalam, 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. Fialkowski, 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. Har, H. Bergland, His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors, J. Struct. Funct. Genom. 5 (2004) 217–225.
[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) 127–132.
[8] M. Amor-Mahjoub, J.P. Suppini, N. Gomez-Virelyunk, 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, R. 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. Gabrec-Porekar, V. Menart, S. Jevsevar, A. Videnasek, 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.

D.S. Waugh / Protein Expression and Purification 80 (2011) 283–293
for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability, Protein Eng. 8 (1995) 249–259.

B.F. Le Bonnec, T. Myles, T. Johnson, C.R. Tapparelli, R. Stone, Characterization of P2’ and P3’ specificities of thrombin using fluorescence-quenched substrates and mapping of the substrates by mutagenesis, Biochemistry 35 (1996) 7114–7122.

J.M. Petraski, J.A. Williams, J. Li, C. T. C. Wong, J. T. Nakai, B. Maslak, B.J. Backes, L. Harris, A strategy to profile prime and non-prime proteolytic substrate specificity, Bioorg. Med. Chem. Lett. 15 (2005) 3162–3166.

O. Shilling, C.M. Overall, Proteome-derived database-searchable peptide libraries for identifying protease cleavage sites, Nat. Biotechnol. 26 (2008) 685–694.

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.

C.M. Jackson, T.F. Johnson, D.J. Hanahan, Studies on bovine Factor X. I. Large-scale purification of the bovine plasma protein possessing factor X activity, J. Biol. Chem. 271 (1996) 28601–28606.

A.E. Rudolph, M.P. Mullane, R. Porche-Sorbet, S. Tsuda, J.P. Miletich, Cloning and recombinant expression of mouse coagulation factor X, Thrombosis Res. 92 (1998) 33–41.

J.C. Carrington, W.G. Dougherty, A viral cleavage site cassette: Identification of an internal cleavage site and delimitation of VPg and proteinase domains, Virology 183 (1991) 449–456.

B.F. Le Bonniec, T. Myles, T. Johnson, C.G. Knight, C. Tapparelli, S.R. Stone, S. van den Berg, P.A. Löfdahl, T. Härd, H. Berglund, Improved solubility of TEV protease by directed evolution, J. Biotechnol. 100 (2003) 69–79.

J.P. Ludeman, R.N. Pike, K.M. Bromfield, P.J. Duggan, J. Cianci, B. Le Bonniec, A.E. Rudolph, M.P. Mullane, R. Porche-Sorbet, S. Tsuda, J.P. Miletich, Factor Xa active site substrate specificity with substrate selectivity for efficient cleavage of affinity tags from expressed proteins, FOCUS 16 (1994) 2–5.

S. van den Berg, P.A. Austin, J. Tözsér, S. Cherry, J.E. Tropea, T.D. Copeland, D.S. Waugh, Efficient site-specific purification of fusion proteins by tobacco vein mottling virus protease in vitro and in vitro, Protein Expr. Purif. 38 (2004) 108–115.

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 (1999) 296–300.

C. Manthoudy, 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.

R. Lottenberg, J.A. Hall, E. Faulter, A. Zupan, U. Christensen, C.M. Jackson, The active site of factor Xa and the role of the p-nitroanilide substrates: substrate selectivity and examination of hydrolysis with different reaction conditions, Biochem. Biophys. Acta 874 (1986) 326–336.

J.B. Barchiesi, J. Hebert, H. Berglund, S. Eschaghi, Exploring the activity of tobacco etch virus protease in detergent conditions, Anal. Biochem. 382 (2008) 69–71.

B.C. Smith, E.C. Seo, J.S. Kim, H.K. Park, M. Hong, S. Lee, J.P. Miletich, Structural basis for the substrate specificity of tobacco etch virus protease, J. Biol. Chem. 277 (2002) 50564–50572.

K.T. Backes, J.L. Harris, A strategy to profile prime and non-prime proteolytic substrate specificity, Bioorg. Med. Chem. Lett. 15 (2005) 3162–3166.

J.F. Bazan, R.J. Fletterick, Detection of a trypsin-like serine protease domain in poliovirus proteinase 3C: Studies on proteolytic function of turnip mosaic virus protease in vitro, J. Biol. Chem. 264 (1989) 9738–9741.

H. Kang, Y.J. Lee, J.H. Goo, W.H. Park, Determination of the substrate specificity of turnip mosaic virus Nla protease using a genetic method, J. Gen. Virol. 82 (2001) 3115–3117.

J.M. Nicklin, K.S. Harris, P.V. Pallai, E. Wimmer, Poliovirus 3C protease as a potential target for new antiviral drugs, Antimicrob. Agents Chemother. 43 (1999) 1501–1504.
G.D. Davis, Q.M. Wang, G.A. Cox, E.C. Villarreal, R.J. Lefèvre, J. Vendrell, F.X. Avilés, Overexpression of human pancreatic procarboxypeptidase A2 in <i>Escherichia coli</i> and the encoded enzyme, Biochem. J. 333 (1998) 527–537.

L. Joshi, R.J. St. Leger, Cloning, expression, and substrate specificity of MeCPA, a rat carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen <i>Metarhizium anisopliae</i>, J. Biol. Chem. 274 (1999) 19925–19933.

C. Rutter, A novel rat carboxypeptidase, CPA2: characterization, molecular cloning and evolutionary implications on substrate specificity in the carboxypeptidases A and B superfamilies, Biomed. Pharmacother. 53 (1999) 106–113.

H. Ouyang, H.-W. Park, C. Qi, W. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Zhu, In situ proteolysis to generate crystals for structure determination: and update, PLoS ONE 4 (2009) e5094.

A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: and update, PLoS ONE 4 (2009) e5094.

D. Reverter, S. Ventura, V. Villegas, J. Vendrell, C.L. Hershberger, F.X. Avilés, Overexpression of human pancreatic carboxypeptidase A2 in Pichia pastoris and detailed characterization of its activation pathway, J. Biol. Chem. 273 (1998) 3535–3541.

S. Darmis, J. Juge, C. Marino, F.X. Avilés, A. Puigserver, J.C. Chais, X.J. Guo, Cloning, sequencing and functional expression of a cDNA encoding porcine pancreatic procarboxypeptidase A1, Eur. J. Biochem. 259 (1999) 719–725.

J. Joshi, R.J. St. Leger, Cloning, expression, and substrate specificity of MeCPA, a rat carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen <i>Metarhizium anisopliae</i>, J. Biol. Chem. 274 (1999) 19925–19933.

C. Rutter, A novel rat carboxypeptidase, CPA2: characterization, molecular cloning and evolutionary implications on substrate specificity in the carboxypeptidases A and B superfamilies, Biomed. Pharmacother. 53 (1999) 106–113.

A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: and update, PLoS ONE 4 (2009) e5094.

D. Reverter, S. Ventura, V. Villegas, J. Vendrell, C.L. Hershberger, F.X. Avilés, Overexpression of human pancreatic carboxypeptidase A2 in Pichia pastoris and detailed characterization of its activation pathway, J. Biol. Chem. 273 (1998) 3535–3541.

S. Darmis, J. Juge, C. Marino, F.X. Avilés, A. Puigserver, J.C. Chais, X.J. Guo, Cloning, sequencing and functional expression of a cDNA encoding porcine pancreatic procarboxypeptidase A1, Eur. J. Biochem. 259 (1999) 719–725.

J. Joshi, R.J. St. Leger, Cloning, expression, and substrate specificity of MeCPA, a rat carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen <i>Metarhizium anisopliae</i>, J. Biol. Chem. 274 (1999) 19925–19933.

C. Rutter, A novel rat carboxypeptidase, CPA2: characterization, molecular cloning and evolutionary implications on substrate specificity in the carboxypeptidases A and B superfamilies, Biomed. Pharmacother. 53 (1999) 106–113.

A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: and update, PLoS ONE 4 (2009) e5094.