Precise and efficient cleavage of recombinant fusion proteins using mammalian aspartic proteases

Blanka Kühnel1, Joanel Alcantara2, Joseph Boothe1, Gijs van Rooijen3 and Maurice Moloney1,2,4

1SemBioSys Genetics Inc., 110 2985 23rd Avenue NE, Calgary AB T1Y 7L3, 2Department of Biosciences, University of Calgary, 2500 University Drive NW, Calgary AB T2N 1N4 and 3GenomePrairie, 115 3553 31st Street NW, Calgary AB T2L 2K7, Canada

Introduction

The production of recombinant proteins as fusion proteins is widely used on a laboratory scale where it provides two major advantages. First, expression of the desired protein as a fusion with an affinity tag may greatly facilitate purification of the fusion by affinity chromatography (Ford et al., 1991). Secondly, incorporation of a highly stable protein at the N-terminus of a fusion may in some cases, increase the accumulation of the fusion protein and increase its solubility (Georgiou and Valax, 1996). In order to recover a recombinant protein that has been expressed as a fusion in its native form, cleavage of the fusion is required. Chemical or enzymatic cleavage of a specific peptide bond can be employed (Coligan et al., 2001). Chemical cleavage often requires elevated temperatures and toxic compounds, which denature the target protein and complicate downstream purification. Enzymatic cleavage is a milder procedure that in general does not denature the protein (Coligan et al., 2001).

For the production of pharmaceutical proteins, the authenticity of the polypeptide chain is essential. This limits the number of available cleavage agents to those that can accommodate a variety of residues on the C-terminal side of the scissile bond and exhibit sufficient specificity in their recognition sequence to avoid internal cleavage of the target protein. Factor Xa (Nagai et al., 1985) and enterokinase (LaVallie et al., 1993) are two of the most widely used cleavage systems on a laboratory scale, with expression vectors and enzymes commercially available from a variety of suppliers. Over the years, several improvements have been made to both systems. The proteases have been expressed as recombinant proteins and have themselves been fused to affinity tags to facilitate their removal after the cleavage reaction. An alternative system based on self-splicing intein reactions. An alternative system based on self-splicing intein sequences that does not require addition of a protease (Chong et al., 1997) is also commercially available. Regardless of the system used, successful cleavage defined by precision, specificity and efficiency depends on the overall conformation of the fusion protein, containing the three components: N-terminal fusion partner, cleavable linker and C-terminal fusion partner.

Here, we describe a novel cleavage method based on the maturation of the aspartic protease chymosin. Chymosin is the primary enzyme responsible for digestion of milk proteins in infant mammals and is widely used in the cheese industry. Like other aspartic proteases, chymosin has a bilobal structure with a substrate binding cleft to which each lobe contributes one catalytic aspartate residue (Tang, 1977). In addition to the cleavage of κ-casein used in milk clotting, chymosin catalyzes its own precursor maturation. The maturation of the inactive precursor involves an activation step, elicited by low pH, followed by a processing step in which the N-terminal pro-peptide is cleaved (Figure 1A). Processing, which results in the formation of the mature protease (Foltmann, 1988), can occur through both intramolecular and intermolecular cleavage of the pro-peptide. The 'heterologous cleavage' method described here utilizes the chymosin pro-peptide (chypro) as a cleavable linker (Figure 1B). The precision of cleavage obtained with this method has been determined for four model proteins expressed in Escherichia coli. Furthermore, it was found that the precision of cleavage could be optimized by mutagenesis of the chymosin pro-peptide. The heterologous cleavage system was further tested with fusion proteins immobilized on an oilbody particle. Cleavage of immobilized fusion proteins has the advantage of combining the cleavage reaction and the removal of the undesired fusion partner into one step. It is a more rigorous test of the robustness of a cleavage system, because of potential limitations of accessibility.
In vivo maturation

| Pre pro chymosin | ER | Calf stomach acid |

Fig. 1. Schematic representation of in vivo chymosin maturation (A) and ‘heterologous cleavage’ mediated by the chymosin pro-peptide (B). In vivo the zymogen pro-peptide is cleaved off when the molecule enters the low-pH environment of the stomach. Both intramolecular and intermolecular cleavage takes place. The heterologous cleavage is mediated by the intermolecular reaction.

Oleosins are structural proteins that are tightly associated with oilbodies, the natural oil storage organelles of plant seeds (Huang, 1996). They can be used as carriers and purification vehicles for industrial scale production of recombinant proteins (van Rooijen and Moloney, 1995; Moloney, 1997). The N- and C-termini of the oleosins are exposed on the surface of the oilbody, and the highly conserved hydrophobic central domain is embedded into the oilbody matrix (Parmenter et al., 1995; Huang, 1996). Proteins expressed in the seeds of transgenic plants as translational fusions with oleosins are correctly targeted to and tightly associated with the oilbody. Oilbodies and proteins associated with them can be easily separated from the majority of other seed cell components by flotation centrifugation (van Rooijen and Moloney, 1995; Moloney, 1997). Two of the model proteins, cystatin and thioredoxin, were expressed in Arabidopsis as oleosin fusions with chymosin pro-peptide between the fusion partners. These were used to test whether after isolation of the oilbodies, incubation with chymosin and flotation centrifugation, the mature proteins could be recovered in a purified form from the oilbody particle.

Materials and methods

Bacterial expression

In-frame fusions of the 126 bp DNA fragment coding for chymosin pro-sequence (Foltmann et al., 1997) with hirudin variant 1 (Bagdy et al., 1976; Dodt et al., 1986), rice cystatin (Chen et al., 1992), Arabidopsis thioredoxin (Rivera-Madrid et al., 1993) and carp growth hormone (cGH) (Chao et al., 1989; Koren et al., 1989) were polymerase chain reaction (PCR)-amplified by overlap extension (Horton et al., 1993) and inserted into the NotI site of pGEX-4T-2 expression vector. The final constructs were verified by sequencing. Supplementary data are available at Protein Engineering online.

Bacterial expression vectors were transformed into E.coli cells BL21 (DE3). Cells were grown and induced with isopropyl-β-D-thiogalactoside and the resultant glutathione S-transferase (GST) fusion proteins were purified according to the manufacturer’s protocol (Amersham Pharmacia Biotech).

Renaturation of GST-carp growth hormone from inclusion bodies

After cell lysis, inclusion bodies were collected by centrifugation, resuspended into 30 ml of distilled water and kept on ice, then 0.1 M sodium hydroxide solution was added slowly with constant stirring until the solution became transparent (~2 ml). A 4 ml volume of 10X phosphate-buffered saline was added and the solution was held on ice for 2–4 h. Subsequently, the solution was titrated to pH 8 by dropwise addition of 0.1 M HCl. The solution was centrifuged to remove particulate material and the supernatant was applied to a glutathione-Sepharose column (Amersham Pharmacia Biotech).

Expression of fusion proteins in plant seeds

The plant transformation vector, pSBS was derived from the Agrobacterium binary plasmid pPZP (Hajdukiewicz et al., 1994). The gentamycin resistance gene of pPZP221 was replaced with a parsley ubiquitin promoter driving the transcription of phosphinothricine acetyltransferase gene and terminated by a parsley ubiquitin terminator sequence. This conferred resistance in transgenic plants to the herbicide glufosinate ammonium (Wohleben et al., 1988). A promoter from the bean storage protein gene Phaseolin (Slightom et al., 1983) was used to drive the expression of the oleosin fusion proteins in a seed-specific manner. An origin of replication from pBR322 controls the replication of the plasmid in E.coli and Agrobacterium. A spectinomycin resistance gene confers antibiotic resistance to the bacterial hosts. In planta transformation of Arabidopsis was performed essentially as described by Clough and Bent (Clough and Bent, 1998). Oilbodies were isolated and washed by flotation centrifugation as described previously (van Rooijen and Moloney, 1995). Briefly, seed was ground with a mortar and pestle in 5 vol. of oilbody extraction buffer (0.5 M NaCl, 0.4 M sucrose, 50 mM Tris–Cl, pH 8.0). After centrifugation, the oilbodies were resuspended in high-stringency washing buffer (8 M urea, 100 mM Na2CO3) for the cystatin seed lines and in 50 mM Tris–Cl, pH 8.0, for the thioredoxin seed lines. The oilbodies were further washed by two subsequent centrifugation and resuspension steps in 50 mM Tris–Cl, pH 8.0.

Chymosin cleavage and analysis by SDS–PAGE

Cleavage reactions were performed in 100 mM sodium phosphate, pH 4.5, at 37°C for 2 h or at room temperature for 16 h. The optimal target-to-protease mass ratio was found to be 20:1 for GST-cGH fusions and 100:1 for all other fusions. Chymosin was provided by SKW Biosystems as a single-strength solution with 4 mg/ml protein and 83 milk clotting units/ml.

SDS–PAGE was run using standard protocols and stained with Coomassie Brilliant Blue R-250 (EM Science) (Sambrook and Russell, 2001).
Hirudin activity was assayed by the reduction of the rate of conversion of the chromogenic substrate $N$-p-tosyl-Gly-Pro-Arg $p$-nitroanilide (Sigma) by thrombin (Abildgaard et al., 1977; Lottenberg et al., 1981).

Cystatin activity was assayed by the inhibition of papain digestion of azocasein. The release of trichloroacetic acid-soluble, digested azocasein was measured spectrophotometrically (Zucker et al., 1985).

Preparation of samples for mass spectrometry
After completion of the cleavage reaction, the samples were incubated at 65°C (85°C for hirudin-containing fusions) for 5–10 min and centrifuged at 16,000 g to remove denatured chymosin, GST and uncleaved fusion protein. The supernatant was dialyzed against water in a Slide-a-lyzer cassette (Pierce) with a 3500 Da molecular weight cutoff membrane for 16 h at 4°C and dried down by freeze-drying.

For the recovery of cleavage products from oilbodies, the samples were first centrifuged at 16,000 g for 15 min to remove oilbodies and then the undernatant was subjected to heat denaturation, dialysis and freeze-drying as described above.

Mass spectrometry and $N$-terminal sequencing
Mass spectrometry was performed at Spectroscopy Laboratories, Plant Biotechnology Institute, NRC in Saskatoon on an AB Voyager MALDI/TOF mass spectrometer. N-terminal microsequencing was performed using Edman degradation of PVDF-immobilized proteins and was carried out by the Protein Chemistry Centre at the University of Victoria and the Biotechnology Laboratory at the University of British Columbia.

Results
Chymosin-mediated cleavage of hirudin-containing fusion proteins
Hirudin from medicinal leeches is a potent thrombin inhibitor with excellent pharmacological properties as a blood anticoagulant (Cheng-Lai, 1999). The interaction of the N-terminus of hirudin with the active cleft of thrombin is essential for inhibition. Substitutions or additions to the N-terminus of hirudin significantly reduce its inhibitory activity (Betz et al., 1992). We have used this characteristic as an indicator of cleavage accuracy and optimal cleavage conditions of a GST-chypro-hirudin fusion protein. As shown in Figure 2A, no inhibition of thrombin activity occurs when GST-chypro-hirudin or chymosin alone is added to a thrombin-catalyzed reaction. Thrombin activity is greatly reduced when GST-chypro-hirudin has been pre-incubated with chymosin, indicating the release of biologically active hirudin. This cleavage is specific for the chymosin pro-sequence, since a fusion protein with a Factor Xa cleavage site between GST and hirudin does not produce any hirudin activity (Figure 2A). The time course and pH dependence of the cleavage reaction are shown in Figure 2B and C, respectively. To purify the cleaved hirudin for further characterization, we took advantage of its heat stability. Whereas chymosin, GST and GST-chypro-hirudin become denatured and precipitate when incubated at 85°C, hirudin remains soluble. Cleaved hirudin purified in this fashion was subjected to analysis by mass spectrometry. As shown in Figure 3A (top profile), two major molecular ions are apparent in the mass profile. The 6968 Da peak corresponds to the mass of native hirudin, and the larger 7260 Da peak corresponds to hirudin with three extra amino acids on the N-terminus, which by deduction originate from the pro-sequence (Figure 4).

Mutagenesis of the chymosin pro-peptide
To determine whether cleavage at the upstream site could be eliminated, we used both a deletion and a substitution approach. Expression vectors were constructed in which the nucleotides coding for the three amino acids interleaved between the two cleavage sites were deleted, giving rise to two new pro-peptide configurations called ΔYSG and ΔSGF. In the third modification, Y39S, the tyrosine codon was substituted by a serine codon. Cleavage of the expressed mutant fusion proteins by chymosin resulted in the release of active hirudin. Mass analysis of the cleaved products determined by mass spectrometry, showed that only precise cleavage at position +1 occurs for the deletion mutants ΔYSG and ΔSGF, while alternative cleavage sites at position –9 and –15 appeared in the Y39S mutant (Figures 3A and 4).

Heterologous cleavage of other fusion partners
To demonstrate that the heterologous cleavage system can be applied to a wider variety of proteins, rice cystatin, Arabidopsis thioredoxin and cGH were expressed as GST-chypro-
GST-ΔYSG chypro-fusions in E.coli. Like hirudin, cystatin and thioredoxin fusions were expressed in soluble form and could be purified directly from cell lysates. Carp growth hormone was found predominantly in inclusion bodies and solubilization at alkaline pH followed by neutralization was required prior to affinity purification. SDS–PAGE analysis of cleavage products of GST-chypro-thioredoxin and GST-chypro-cGH is shown in Figure 5A and B, respectively. The cleaved products were further characterized by mass spectrometry for cystatin and thioredoxin and by N-terminal sequencing for cGH. The results, summarized in Table I, show that cleavage of cystatin and thioredoxin occurs at sites identical with those found for hirudin: at the junction between the pro-peptide and the mature protein and three amino acids upstream with the wild-type pro-peptide and predominantly at the pro-peptide–mature protein junction with the ΔYSG mutated sequence.

Cleavage of cGH occurred exclusively at the upstream −3 position when it was fused to wild-type chypro peptide and an alternate cleavage site further upstream (position −9) was found with the ΔYSG pro-peptide. Mature cGH protein starts with the amino acid serine. We speculated that this proximal serine prevents correct cleavage. To test this hypothesis, a mutant was generated in which a methionine codon was inserted at the junction of ΔYSG pro-peptide and cGH sequence. In the resulting fusion protein, methionine is the first amino acid downstream of the desired cleavage site. The cleavage, as determined by N-terminal sequencing, occurred at the predicted +1 position, releasing cGH with a methionine at its N-terminus (Figure 5B and Table I).

Cleavage of oilbody-associated fusion proteins

The utility of the heterologous cleavage system was further analyzed with oilbody-associated oleosin-ΔYSG chypro-cystatin. The deduced molecular weight for cystatin is 11391 Da. The minor peak with a mass of 12038 corresponds to cleavage at position −9 in the chypro sequence and is a result of incomplete cleavage.
applicability of this system to immobilized proteins. Expression vectors that drive the seed-specific expression of oleosin-chypro-cystatin, oleosin-ΔYSG chypro-cystatin and oleosin-ΔYSG chypro-thioredoxin were introduced into Arabidopsis. Seed from transgenic plants expressed and correctly targeted the fusion proteins to the oilbody (Figure 6A for cystatin; data not shown for thioredoxin). After isolation, the recombinant oilbodies were treated with chymosin under the same conditions as determined for the bacterial GST fusion proteins. The efficiency of cleavage was determined by densitometry of Coomassie Blue-stained gels, an estimated 98% of native seed protein has been removed at this point. N-terminal sequencing of cleaved cystatin and thioredoxin showed that the cleavage pattern was identical with that obtained for bacterial fusions, i.e. cleavage at the correct site and three amino acids upstream with the wild-type pro-peptide and exclusively at the correct site with ΔYSG chypro, respectively (Table I). Mass spectrometry confirmed the correct size of mature cystatin, ruling out any heterogeneity of the C-terminus (Figure 3B). The purified cystatin was biologically active as an inhibitor of the protease papain (Figure 6B).

Discussion

Cleavable linkers are generally derived from a substrate sequence that is susceptible to proteolytic cleavage (Walsh and Swaisgood, 1996). The heterologous cleavage system described here utilizes a pro-peptide of an autocatalytically maturing protease. The chymosin pro-peptide is sufficient to direct cleavage predominantly to the junction of the pro-peptide and the heterologous protein.

The time-scale of the cleavage reaction (Figure 2B) is comparable to that of enterokinase and Factor Xa cleavage systems. The pH optimum (Figure 2C), on the other hand, is more acidic than that for enterokinase and Factor Xa. It is well suited for the cleavage of acid-stable proteins such as the cytokines transforming growth factor-α (Todaro et al., 1980) and type I interferons (Martal et al., 1998). The optimum pH for autocatalytic maturation of chymosin is 2. This coincides with the physiological environment where chymosin is active - the calf stomach. The optimum for heterologous cleavage, defined as the conditions resulting in maximum recovery of authentic product, determined here, is pH 4.5 (Figure 2C). The thrombin inhibition assay employed in determining the optimum conditions relies both on precise cleavage and the presence of full-length hirudin molecules. Cleavage carried out at lower pH resulted in very efficient, but non-specific, digestion of the fusion protein as visualized by the disappearance of the fusion protein that is not accompanied by the appearance of full-length hirudin molecules. Cleavage carried out at low pH resulted in very efficient, but non-specific, digestion of the fusion protein as visualized by the disappearance of the fusion protein that is not accompanied by the appearance of full-length hirudin molecules. Cleavage carried out at low pH resulted in very efficient, but non-specific, digestion of the fusion protein as visualized by the disappearance of the fusion protein that is not accompanied by the appearance of full-length hirudin molecules.

Precise cleavage of fusion proteins
age. By increasing the pH, chymosin proteolytic activity becomes restricted to the most sensitive site, which is the junction between the pro-sequence and hirudin (Hubbard, 1998). Chymosin remains active up to pH 6.

Authenticity of primary protein sequence is an important factor in the production of pharmaceutical proteins. Using the wild-type pro-peptide, two cleavage sites are apparent: the desired site at the pro-peptide±recombinant protein junction and an alternative site within the pro-peptide. Based on the similarity of prochymosin to the crystal structure of propepsin (Bateman et al., 1998), the comparison of pro-peptides of several related aspartic proteases (Foltmann, 1988; Koelsch et al., 1994) and on the studies of the crystal structure of free and bound chymosin binding pocket (Groves et al., 1998), mutagenesis of the prosequence was performed. This resulted in the elimination of the alternative cleavage site for the ΔYSG and ΔSGF mutants. There is some restriction as to the N-terminal amino acid sequence of the recombinant protein that will allow precise cleavage. Insertion of a methionine as the N-terminal residue restored precise cleavage, suggesting that it is the first residue that is important for precise cleavage and that some amino acids negatively affect the precision of the cleavage reaction. Modified specificity due to downstream sequences has also been shown for Factor Xa (Nagai and Thogersen, 1987; He et al., 1993) and enterokinase (Collins-Racie et al., 1995). In order to restore the authentic protein sequence for cGH using the heterologous cleavage system, the N-terminal methionine could, if needed, be removed post-cleavage by digestion with methionyl aminopeptidase (Hirel et al., 1989; Dalboge et al., 1990).

The heterologous cleavage system was employed for the cleavage of oilbody-associated fusion proteins produced in Arabidopsis. The efficiency and precision of cleavage were found to be identical with those for the cleavage of soluble fusion proteins expressed in bacteria, suggesting that the proximity of the oilbody does not pose any steric constraints on the interaction of the protease and its substrate. On the other hand, based on the activity of cleaved hirudin, cleavage of oleosin-hirudin separated by a Factor Xa cleavage site from transgenic oilbodies (Parmenter et al., 1995) was 50 times less

Table I. Summary of chymosin cleavage sites of wild-type- and pro-sequence mutants in combination with various fusion partners as determined by mass spectrometry and N-terminal sequencing

<table>
<thead>
<tr>
<th>N-terminal fusion partner</th>
<th>C-terminal fusion partner</th>
<th>N-terminal amino acid of C-terminal fusion partner</th>
<th>Cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminal chypro</td>
<td>C-terminal chypro</td>
<td>Valine</td>
<td>ΔYSG chypro</td>
</tr>
<tr>
<td>Hirudin</td>
<td></td>
<td>Methionine</td>
<td>ΔYSG chypro</td>
</tr>
<tr>
<td>Cystatin</td>
<td></td>
<td>Methionine</td>
<td>ΔYSG chypro</td>
</tr>
<tr>
<td>Thioereoxdin</td>
<td></td>
<td>Serine</td>
<td>ΔYSG chypro</td>
</tr>
<tr>
<td>Carp growth hormone</td>
<td></td>
<td>-3 and +1</td>
<td>ΔYSG chypro</td>
</tr>
<tr>
<td>Oleosin</td>
<td></td>
<td>-3 and +1</td>
<td>ΔYSG chypro</td>
</tr>
</tbody>
</table>

GST, not determined.
efficient compared with the cleavage of GST-Factor Xa.

Choylin pro-peptide is 39 amino acids in length, compared with four amino acids that constitute the Factor Xa cleavage site. It may be that the length of the chymosin pro-peptide positions the cleavage site further away from the oilbody surface, thus providing better access for the cleavage agent.

This work demonstrates that chymosin can be used as an efficient cleavage enzyme for a variety of heterologous fusion proteins including those immobilized on particles such as oilbodies. The efficiency and precision of this system appear to be superior to those of many of the routine cleavage methods used on a laboratory scale. Although the chymosin used in these studies was of bovine origin, we have successfully repeated the experiments with plant-produced bovine chymosin (van Rooijen et al., 2001). This approach will permit the use of fusion proteins for larger scale manufacturing, both by reducing the cost of the cleavage agent and by addressing safety issues arising from the use of cleavage agents of mammalian origin.

References