

Production and activation of recombinant papain-like cysteine proteases

Dieter Brömme,^{a,*} Ferez S. Nallaseth,^a and Boris Turk^b

^a Mount Sinai School of Medicine, Department of Human Genetics, New York, NY 10029-6574, USA

^b Department of Biochemistry and Molecular Biology, J. Stefan Institute, SI-1000 Ljubljana, Slovenia

Accepted 23 June 2003

Abstract

Papain-like cysteine proteases are the most numerous family of the cysteine protease class. They are expressed throughout the animal and plant kingdoms as well as in viruses and bacteria. More recently, this protease family has drawn attention as a potential pharmaceutical drug target in diseases characterized by excessive extracellular matrix degradation such as in osteoporosis, arthritis, vascular diseases, and cancer. Moreover, papain-like cysteine proteases have been identified as critical components of the life cycle and invasive potential of various human and live stock pathogens as well as major allergens. Therefore, this protease class is rigorously studied and requires sufficient amounts of protease protein to analyze structure–activity relationships, their 3-D structures as well as to screen for and optimize potent and selective inhibitors. This review summarizes approaches to generate active papain-like cysteine proteases by heterologous expression in a variety of expression systems.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Cysteine proteases; Papain-like; Cathepsins; Expression; Activation; Purification; *Pichia*; *E. coli*; Baculovirus

1. Introduction

Papain-like cysteine proteases have been identified as key proteolytic activities in degenerative, invasive, and immune system related disorders [1,2]. For example, cathepsin K is the major bone-degrading activity in osteoclasts and its selective inhibition is thought to be beneficial for the treatment of osteoporosis and certain forms of arthritis [3–5]. Cathepsin S has been demonstrated to be critical in MHC class II dependent antigen presentation. Inhibition of cathepsin S significantly decreases the response to antigens and thus cathepsin S is evaluated as a novel drug target for asthma and certain auto-immune diseases [6,7]. Cathepsins also have been implicated in tumor invasion and metastasis [8,9]. Increasing interest is focused on cysteine proteases expressed in a wide range of pathogenic parasites causing malaria, Chagas' disease, and schistosomiasis [1,10]. Papain-like cysteine proteases in parasitic organisms have indispensable roles in growth, cell differentiation,

signaling, and host invasion, and frequently act as virulence factors by proteolytically degrading the host immune system [11].

Functional and structural evaluations of cysteine proteases as well as high throughput screening for potent protease inhibitors require sufficient amounts of protein. Whereas some of the cathepsins (e.g., cathepsins B, L, and H) that are ubiquitously expressed at high levels in relatively easily accessible organs have been isolated from human and other mammalian species, the scarcity of the tissue and cell-type specifically expressed cathepsins requires the production of recombinant protein. Furthermore, it is also advisable to obtain cysteine proteases from hazardous and pathogenic organisms such as from various human and animal parasites by heterologous expression. Thus, the availability of efficient protein expression systems is critical for the further understanding of this biologically and pharmaceutically important protease family. In this review, we will summarize methods of heterologous expression of papain-like proteases, their activation into catalytically active proteins, their purification, and storage.

* Corresponding author. Fax: 1-212-849-2508.

E-mail address: Dieter.Bromme@mssm.edu (D. Brömme).

2. Description of methods

2.1. Expression of recombinant papain-like cysteine proteases in *Escherichia coli*

Escherichia coli is generally the first choice for the heterologous expression of proteins due to the ease with which it is manipulated and its rapid growth in inexpensive media. The major concerns for producing active protease in *E. coli* are ensuring high transcription and translation rates, controlling toxicity of the heterologous gene product, maintaining solubility of the protease, and preventing proteolytic degradation of the expressed protein. Most of these concerns can be addressed by the selection of expression vectors with strongly inducible promoters and strong ribosomal binding sites, targeting sequences which direct the gene product to the periplasmic compartment, protease-deficient *E. coli* hosts, and the introduction of peptide “tags” that allow easy purification. A major breakthrough in the development of bacterial expression systems was the introduction of the bacteriophage T7 promoter to control the expression of recombinant proteins in *E. coli* [12]. This system is widely used for the expression of papain-like proteases. The first protein expressed in such a system was cathepsin L [13], however, the protein produced was a deletion mutant with very low yields. The first two proteins successfully expressed and produced in high quantity were the plant proteases, papain [14], and caricain (papaya protease omega [15]). Since then, the system has been successfully exploited for the expression of mammalian cathepsins, such as cathepsin B [16], cathepsin L [17,18], cathepsin S [19,20], cathepsin K [21,22], cathepsin O [23], and cathepsin X [24] (see Table 1).

Initial attempts to express active papain-like cysteine proteases in *E. coli* failed. However, these experiments revealed that (i) the removal of the toxic signal sequence is essential for protein expression [25] and (ii) the presence of the prodomain is required for correct folding to obtain fully active enzyme [23,24]. A direct comparison with a factorial screen for cathepsin S has demonstrated that although it is possible to express the enzyme without its prodomain, yields are extremely low when compared with the expression of the proenzyme [20]. Similar results were obtained for cathepsin L [26]. An exception seems to be falcipain-2, which only needs the N-terminal extension, unique for falcipain, for proper folding, but not the prodomain [27]. Therefore, successful protocols use protease cDNAs coding for the proenzymes, which are cloned into an expression vector (pET or pGEX vectors are most commonly used). BL21(DE3) pLysS *E. coli* strains are most commonly used as hosts [12] and are grown in LB medium in the presence of ampicillin (50–100 µg/ml) and chloramphenicol (25 µg/ml) at 37 °C until an A_{600} of 0.6–1.0 is reached. Protein expression is

then induced by the addition of IPTG (0.4–1 mM final concentration). Cells are harvested 3–4 h post-induction.

The T7 promoter system gives protein yields as high as 50–220 mg/L culture for papain-like cysteine proteases. However, most proteins are produced in insoluble inclusion bodies. In order to solubilize the proteins, cells are initially resuspended in a small volume of a slightly alkaline buffer (50 mM Tris, pH 8.0), containing 5 mM EDTA and 5% sucrose (v/v) or 150 mM NaCl instead of sucrose can be used [16,20]. Following cell lysis by freeze–thaw cycles and sonication, inclusion body pellets are washed in similar buffers as described above (Triton X-100 and 1.0 M urea can be added; see [16,19,20]), and solubilized. Solubilization requires the denaturation of the protein with 6–7 M guanidine hydrochloride (or 8 M urea) in a buffer solution (imidazole, Tris, etc.). At this step, the –SH groups in the proteins may be protected by sulfonation with disodium 2-nitro-5-thiosulphobenzoate [28], which has been shown to be beneficial [16,18,19], although not essential.

A purification step preceding the refolding was found to give higher yields. This could be done either by affinity chromatography using Ni-chelating chromatography, if the construct was made with a His-tag [22,27], or by gel filtration [17,19,21].

Refolding is the most critical step in the process. Protein is rapidly diluted (50–100-fold) into the refolding buffer to a final concentration of 10–50 µg/ml. At higher protein concentrations, yields were dramatically decreased [16,19]. An alkaline pH seems to be essential for the proper disulfide formation [16,20]. A typical refolding buffer is a 50–100 mM Tris buffer, pH 8.0–8.6, containing 5 mM EDTA. Addition of glycerol (10–20% v/v) as well as the addition of Arg (0.3–0.7 M) to increase polarity [19–22] is highly recommended to increase the refolding yield. Furthermore, the correct redox pair ratio of the buffers used is critical. Two combinations have been described: (i) cysteine/cystine redox pair (3–5 mM/0–1 mM) [16,17], and (ii) reduced (10 mM) and oxidized (1 mM) glutathione [20–22,27]. The addition of detergents is optional (Brij, Chaps) and may improve final yields [20,22]. The optimal temperature for refolding is +4 °C [16,20] and the time of incubation can vary between 15 and 24 h. After the refolding step, the samples are concentrated and dialyzed against the refolding buffer in the absence of the redox pair components. To avoid autoactivation, proenzymes can be stored in this buffer (neutral to alkaline pH is essential) at –70 °C (at least –20 °C) for several months. In order to increase the purity of the final product, proenzymes can be purified by gel filtration, ion exchange chromatography or affinity chromatography on Ni–NTA–Sepharose or glutathione–Sepharose. The latter method is only applicable if the appropriate fusion proteins containing the His-tag or glutathione *S*-transferase have been expressed.

Table 1
Heterologous expression of papain-like cysteine proteases

Host species	Protease	Host strain/expression vector or construct	Purification	Yield (mg/L)	References
<i>E. coli</i>	Cathepsin B (human)	BL21(DE3)pLyS/pET3A	Sephacrose S	3.0	[16]
	Cathepsin B (<i>L. mexicana</i>)	m15pREP4/pGL180(CPB2.88CTE)	Ni ²⁺ -NTA-Agarose or Mono Q, Sephadex G50	>3.5	[43]
	Cathepsin K (human)	LW29(DE3)/pET16b	Superdex 75	50	[44]
		JM109/pQE-30	Ni ²⁺ -NTA-Sephacrose	54	[22]
	Cathepsin L (human)	BL21(DE3)/pCtsL	Sephacdex G-75	1.3	[17]
		BL21(DE3)pLysS/pET811	Sephacryl S-200		[45]
	Cruzain (<i>T. cruzi</i>)	DH5 α /pCheYTc	DEAE-Sephacrose	20–30	[46]
	Cathepsin S (human)	BL21(DE3)pLysS/pKGG2-1	Sephacryl S-200	2	[19]
		K12(Δ DE3)/pT73.3	Thiopropyl-Sephacrose 6B	14	[20]
	Cathepsin C (bovine)	BL21/p8PM	ND	0 (deg)	[29]
	Caricain (<i>C. papaya</i>)	BL21(DE3)/pET3aPpQI	ND	ND	[15]
	Falcpain-2 (<i>P. falciparum</i>)	AD(DE3)pLysS/pRSET M159pREP4/pQE	Ni ²⁺ -NTA, Q-Sephacrose	ND	[47]
Yeast	Cathepsin C (bovine)	<i>C. boidinii</i> SK741/pFCTC-S1	ND	12	[29]
	Cathepsin B (<i>L. major</i>)	<i>P. pastoris</i> GS115/pPIC9	ND	33	[48]
	Cathepsin B2 (<i>S. mansoni</i>)	<i>P. pastoris</i> X33/pPICZ α	PD10, Mono S	40–60	[49]
	Cathepsin F (human)	<i>P. pastoris</i> GS115/pPIC9	Source 15S, Superdex-75	0.17	[50]
			HiTrap SP	1.0	[51]
	Cathepsin K (human)	<i>P. pastoris</i> GS115/pPIK9K	<i>n</i> -Butyl-Sephacrose	38	[52]
	Cathepsin L (human)	<i>P. pastoris</i> GS115/pPIC9	HPLC on vydac C4	10	[53,54]
	Cathepsin L (<i>C. sinensis</i>)	<i>P. pastoris</i> X33/pPICZ α	Ni ²⁺ -NTA agarose, Sephacryl S-200	19.7	[55]
	Cathepsin X (human)	<i>P. pastoris</i> GS115/pPIC9	CM-Sephacdex	5	[56]
	Cathepsin L (human)	<i>S. cerevisiae</i>	S-Sephacrose	0.0025	[53]
		BJ3501/YpDC222			[57]
	Cruzain-2 (<i>T. cruzi</i>)	<i>S. cerevisiae</i> BJ3501/YpDC280	Thiopropyl-Sephacrose 6B, Con A-Sephacrose	0.029	[58]
	Cathepsin S (human)	<i>S. cerevisiae</i> BJ3501/YpDC222	Thiopropyl-Sephacrose 6B	0.17	[40]
	Papain (<i>C. papaya</i>)	<i>S. cerevisiae</i> BJ3501/YpDC222	Thiopropyl-Sephacrose 6B	ND	[59]
Baculo-virus	Cathepsin B (human)	<i>Sj9/pFastBac1</i> (Bacmid)	SP-Sephacrose	>25	[60]
	Cathepsin B1 (<i>S. mansoni</i>)	<i>Sj9/ND</i>	ND	ND	[61]
	Cathepsin C (human)	High Five/pVL1393	<i>n</i> -Butyl-Sephacrose FF, Q-Sephacrose FF	10–15; 20–30	[39]
	Cathepsin C (rat)	High Five/pVL1393	Ni ²⁺ -NTA, butyl-Sephacrose FF, Q-Sephacrose FF	50	[62]
	Cathepsin C (<i>S. japonicum</i>)	High Five/pMHis + BacPAK6 (Bacmid)	Ni ²⁺ -NTA		[63]
	Cathepsin K (<i>M. fascicularis</i> ; monkey)	<i>Sj21/vBacMneatK</i>	S-Sephacrose	7.5	[64]
	Cruzain (<i>T. cruzi</i>)	<i>Sj21/pFastBac HTb</i>	Ni ²⁺ -NTA or Con A-Sephacrose, Mono Q, Mono P	2	[65]
	Cathepsin S (mouse)	<i>Sj9/pFastBac1</i> (Bacmid)	Ni ²⁺ -NTA or HiPrepQXL	>1	[66]
	Cathepsin S (rat)	<i>Sj9/pFastBac1</i> (Bacmid)	Ni ²⁺ -NTA or HiPrepQXL	>1	[66]
	Cathepsin S (humanized rod enz)	<i>Sj9/pFastBac1</i> (Bacmid)	Ni ²⁺ -NTA or HiPrepQXL	>1	[66]
Mammalian cells	Cathepsin K (<i>Rhesus</i> monkey)	CHO/pcDNA3.1	Unpurified cell extract	ND	[36]
	Cathepsin B (human)	HeLa/vTF7EMCV	Cathepsin B mab Affi-gel10	0.2/10 ⁸ cells	[35]
	Cathepsin L (human)	A431 epidermal carcinoma cell line/pHa-MDR1 + pHaMEPh	Immune precipitation, protein A-Sephacrose or Con A-Sephacrose	15 $\times$ basal AU/ 10 ⁶ cells	[45]
	Cathepsin L (mouse)	NIH3T3 or KNIH/pHaMDR1 + pcosMEP5A	Protein A-Sephacrose or Con A-Sephacrose	3 $\times$ basal AU/ 10 ⁶ cells	[45]

ND, not described; degr., degraded.

2.2. Expression of recombinant papain-like cysteine proteases in yeast

Three yeast based expression systems have been used for the production of recombinant papain-like cysteine proteases: (i) *Saccharomyces cerevisiae*, (ii) *Pichia pastoris*, and (iii) *Candida boidinii*. All three systems allow post-translational modifications such as disulfide bond formation, proteolytic processing, and glycosylation similar to mammalian cells, although high mannose glycosylation is sometimes a problem using *S. cerevisiae*. Cysteine protease cDNAs are ligated to the appropriate leader sequence in transfer vectors (prepro- α -factor for *Saccharomyces* and *Pichia* expression and signal sequence of proteinase A for the *Candida* system) with the first codon coding for the first amino acid of the pro-region. Whereas, the expression vectors for the *Saccharomyces* system exploit *ADH* or α -*MF* inducible promoters, the *Pichia* system uses the *AOX1* inducible promoter or in case of the pGAP *Pichia* plasmids, the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter. The latter expression system does not require methanol for induction. In the *Candida* system, the *FDH1* promoter is used. Our experience with the *Pichia* system is focused on the use of the pPIC9K vector and GS115 cells. Particular attention should be given to the preparation of the linearized vector/insert construct. The restriction fragment retaining sequences that (i) regulate the expression of the protein encoded by the insert and (ii) facilitate homologous recombination with the genomic *AOX1* locus, is gel purified twice. This fragment (0.3–3 μ g DNA) is then electroporated into GS115 electro-competent host cells. Terminal *AOX1* sequences allow for preferred integration into the *AOX1* locus. These plasmid integrants facilitate optimal expression of the *AOX1* promoter driven cathepsin coding sequences. *AOX1* locus specific integrants have a Mut^S (slow growing) phenotype on methanol since they are dependent on the weaker *AOX2* regulatory locus, while those integrated at ectopic loci (e.g. *AOX2* or insertion integrants at *AOX1* loci) have a Mut⁺ (rapidly growing) phenotype on methanol due to an intact and more active *AOX1* locus. The two types of integrants are scored by replica streaking primary colonies from minimal dextrose agar (MDA) plates onto minimal methanol agar (MMA) and MDA plates. Since the productivity of Mut⁺ and Mut^S transformants is not predictable, at least six to eight colonies of each phenotype should be screened for levels of heterologous protein/activity expression. For example, 7/8 Mut^S and 2/8 Mut⁺ clones displayed high expression of cathepsin K (unpublished FN, DB). Screening can be performed in 50 ml culture tubes. Briefly, a 6 ml culture is grown at 30 °C, 250 rpm for 3 days in MDA, then centrifuged, and resuspended in 2 ml BMGY. Half the aliquot (1 ml) is frozen away in 15% glycerol-BMGY. The remaining 1 ml is brought up

to 40 ml BMGY and incubated at 250 rpm, 30 °C for 2–3 days. The OD₆₀₀ should be ~20. Cells are centrifuged at 1500 rpm for 10 min and the pellets resuspended in 4 ml BMMY. The pH values can drop significantly (down to pH 4) within the first 36 h and must be readjusted to a pH range of pH 6–6.5. Similar drops in pH have also been described for the *Candida* expression system [29].

The key variables for maximizing yields of cysteine protease expression are: choice of culture medium, cell density, maintaining optimal pH range (which is affected by both the previous parameters), and timing of the activity assays. Low pH values lead to autocatalytic activation of the zymogen form of the expressed cysteine protease and subsequent autodegradation and denaturation of the mature protease. Autoactivation of lysosomal cathepsins can be prevented at pH values higher than 6. We adjust the pH to 6.5 with 5% ammonium hydroxide. After reaching maximal cell density the pH remains stable in BMMY media but not in BMM. Methanol induction should be performed every 12 h when a continuous feeding mode is not available. The addition of antifoam is recommended to minimize denaturation of the secreted protease. Activity assays should be performed twice a day after methanol induction. The supernatant will be harvested after reaching a maximum of activity. This time point must be determined empirically, since it depends on the expression and secretion rate of the appropriate protease and its stability in the medium.

A major advantage of yeast for the expression of cysteine proteases is that it does not express and secrete endogenous cysteine endoproteases. The only endogenous cysteine protease produced in yeast is the cytosolic aminopeptidase, bleomycin hydrolase [30], which can be easily distinguished from heterologously expressed papain-like cysteine proteases. A further specific advantage of the *Pichia* system is its very low level of secretion of endogenous proteins which leave the heterologously expressed protein, the only major protein in the production media, and thus dramatically simplify the purification procedure [31]. Yeast systems have been extensively used to express papain-like cysteine proteases (see Table 1).

2.3. Expression of recombinant papain-like cysteine proteases in insect cells

Baculovirus expression vectors have been used for the production of several mammalian, including human thiol-dependent cathepsins and various parasite derived papain-like cysteine proteases (Table 1). Similar to yeast expression systems, they allow post-translational modifications similar to mammalian cells. To express a functional protease, the complete coding region for the protease, including the ATG start codon, has to be cloned into a shuttle vector (e.g., pVL1392.93, pBlue-

Bac, and pFastBac) and then inserted into the genomes of polyhedros virus, AcNPV or a Bacmid DNA by homologous recombination. Expression transfer vectors are available with His-tag for simple purification or epitope tags for easy detection. Host cells for *AcNPV* infections are cell lines (*Sf9* and *Sf21*) derived from the fall armyworm (*Spodoptera frugiperda*) and from egg cells from *Trichoplusia ni* (High Five). A distinct disadvantage of insect cells for the expression of cysteine proteases is that they express several endogenous cathepsin-like activities which might interfere in the analysis of the recombinant protease. In addition, a cathepsin B-like cysteine protease is encoded in the baculovirus [32]. Care must be taken to remove all endogenous cathepsin-like activities during the purification procedure. The insect cell expression system allows high level production in the milligram range as described for cathepsin S, although other related cathepsins revealed significantly lower yields (see Table 1).

2.4. Expression of recombinant papain-like cysteine proteases in mammalian cells

Heterologous expression of cysteine proteases in mammalian cells is mostly used to study intracellular transport mechanisms, the secretion of zymogens, or regulatory aspects of protease biology [33,34] and less for large scale production of recombinant protein. This is due to the high costs of mammalian cell cultures and the availability of alternative expression systems as discussed above. Similar to the insect cells, mammalian cells express numerous endogenous cathepsins which have to be eliminated in the purification procedure. Sloane and co-workers [35] have expressed functional recombinant human procathepsin B in monkey kidney and HeLa cells using the *Vaccinia* expression system. Rhesus monkey cathepsin K has been expressed in Chinese hamster ovary (CHO) cells [36].

2.5. Activation of precursors

Papain-like cysteine proteases contain a proregion which has a tri-fold function: (i) acting as a chaperone domain for the transport to the endosomal/lysosomal compartment, (ii) ensuring the correct folding of the catalytically mature protein, and (iii) inhibition of the activity of the proenzyme. The length of the proregion varies significantly and comprises peptide length between 38 amino acids for cathepsin X and 251 amino acids for cathepsin F. Most heterologous expression systems for cysteine proteases produce proteolytically inactive proenzymes which are stable at slightly acidic to slightly alkaline pH values. In order to generate proteolytically active proteases the propeptide region must be removed. The proregion can be removed autocatalytically by incubating the precursor proteins

in acetate buffer at pH 4–4.5. Autocatalytic activation can be facilitated in the presence of negatively charged glycosaminoglycans with dextran sulfate being most commonly used (5–20 µg/ml; [18,37]). The only enzymes where no autoactivation was observed were the exopeptidases, cathepsins X [38] and C [39]. They both required an additional protease for full activation. Cathepsin C could be activated during the purification procedure, but this was likely due to trace amounts of a cathepsin L-like protease from the host insect cells [39]. Autocatalytically processed cysteine proteases normally have *N*-terminal extensions of 2–6 amino acid residues [16,19,40]. However, these extensions do not interfere with catalytic properties of the enzymes.

The processing of procathepsins can be dramatically enhanced by the addition of porcine pepsin (0.6 mg/ml). Here, the crude supernatants of cathepsin producing expression cultures are adjusted to a pH of 4.0 prior to the pepsin addition and the mixture is incubated at 37 °C. Aliquots of the reaction mixture are analyzed for the rate of hydrolysis of a suitable substrate (e.g., Z-FR-MCA). Aliquots should be tested at 5 min intervals to determine the maximum level of activity. Autocatalytic as well as pepsin activation lead to a partial degradation of the recombinant enzyme and thus the optimal time point for a balance between maximal activation and minimal degradation must be determined. Pepsin activation or autocatalytic processing can be terminated by raising the pH to 5.5. Samples of the time course of activation should be saved and analyzed by Western blot analysis if a suitable antibody for the appropriate recombinant protease is available. This will verify whether the processing of the proenzyme has been completed. Pepsin-activated cysteine proteases normally have the correct processing site (YPro). An additional advantage of the use of pepsin is the proteolytic clearance of most yeast culture proteins present in the supernatant of the culture.

2.6. Purification of recombinant papain-like cysteine proteases

Autoactivation or activation of secreted protease zymogens by pepsin hydrolyzes most proteins present in the culture medium. This proteolytic clearance simplifies subsequent purification and frequently allows to obtain electrophoretically homogenous protein after a single chromatography step. Depending on the physico-chemical parameters of the recombinant protease any combination of chromatographic resins available can be selected. This includes cationic and anionic resins, various resins based on hydrophobic interactions, as well as affinity chromatography. Since many cysteine proteases generated in yeast, insect or mammalian expression systems are glycosylated, Con

A-Sepharose columns can be used. Thiopropyl-Sepharose is uniquely suited to bind to activated cysteine residues as present in the ion pair of active sites in these proteases. At acidic pH, only cysteine proteases bind to this resin and can be easily eluted from the column at neutral pH with an excess of a reducing agent such as cysteine or dithiothreitol. However, it should be mentioned that not all cathepsins bind equally efficiently to thiopropyl-Sepharose. In our experience, thiopropyl-Sepharose was less effective for cathepsins K, V, and F from supernatants of *Pichia* cultures. Chromatography steps used for the purification of various papain-like cysteine proteases are compiled in Table 1.

2.7. Activity assays for papain-like cysteine proteases

Mammalian cysteine proteases from the papain family are expressed in lysosomes and have pH optima at pH 5.5 ± 0.5 . With the exception of cathepsin S they are unstable at neutral or alkaline pH. The standard assay is performed in an acetate or phosphate buffer, pH between 5 and 6 containing 1–2.5 mM dithiothreitol (or 5–10 mM cysteine) and 1–2.5 mM EDTA. The concentrations of buffer, dithiothreitol and EDTA, vary between protocols but do not seem to have a significant effect on the enzyme activities. The protease concentration is between 0.1 and 10 nM. Z-FR-MCA can be used for most mammalian cathepsins with an endopeptidase activity (cathepsins B, L, S, K, V, F, and O) and the hydrolysis of the substrates can be continuously recorded with a fluorimeter using 380 nm for excitation and 450 nm for emission. Alternatively, the chromogenic substrate Z-FR-pNA can be used and recorded at 405 nm. For some of the endopeptidolytically active cathepsins more selective and potent substrates are available. Z-RR-MCA is selective for cathepsin B, Z-GPR-MCA is only cleaved by cathepsins K and B, Z-VVR-MCA is approximately 8–10 times more efficient than Z-FR-MCA for cathepsin S, and Z-LR-MCA is twice as potent for cathepsin K than Z-FR-MCA. It should be noted that these substrates are not selective for cathepsins S and K but provide an alternative if the availability of the protease amount is an issue.

Cathepsins H and C exhibit an amino and dipeptidyl aminopeptidase activity and thus require N-terminally unprotected peptide substrates. Arg-MCA (or Arg-pNA) can be used for cathepsin H and GF-MCA (or a pNA analogue) for cathepsin C. Cathepsin X is a carboxypeptidase and internally quenched fluorescence substrates such as Abz-FRF-(4NO₂) are recommended.

It should be noted that the measured activities of papain-like cysteine proteases derived from heterologous expression should always be confirmed by the in-

hibition with E64, a selective but broad-specificity irreversible inhibitor for all mammalian cathepsins with the exception of cathepsin X. This is in particular importance when expression systems are used which contain tryptic and aminopeptidase activities (e.g., the insect cell baculovirus system, any mammalian expression systems, and yeast expression systems). Z-XR-MCA substrates are hydrolyzed by trypsin-like activities and R-MCA (pNA) is degraded by various non-serine and metallo-dependent amino peptidases. Those overlapping activities are particularly a problem when the heterologously expressed cysteine protease has to be isolated from the host cell extract. Attention must also be given to endogenous cathepsin activities in mammalian and insect cell expression systems. Endogenously expressed cysteine protease activities are frequently very similar to the recombinant enzyme activities. For example, baculovirus expresses a cathepsin B-like activity [32] as well as several endopeptidase activities with one being very similar to cathepsin K (unpublished observation, DB).

Various plant and parasite-derived cysteine protease activities have a pH optimum of pH 7.5 ± 0.5 and can be analyzed in appropriate Tris/HCl, Hepes or related neutral to slightly alkaline buffer systems.

2.8. Storage of recombinant cathepsins

In general, cysteine proteases are relatively unstable enzymes which are prone to autocatalytic degradation, oxidation, and protein denaturation. Exposure of most recombinant cathepsins to room temperature and in sub-micromolar enzyme concentration leads to a rapid loss of enzyme activity. According to our observations, some cathepsins such as cathepsin K are unstable at -20°C and must be stored at -80°C . Micromolar stocks of cathepsins L, K, V, S, and F in acetate buffer, pH 5.5, containing 1 mM dithiothreitol and 1 mM EDTA retained their full activity at -80°C over several years. It is strongly suggested that the protease stock solutions are frozen in small aliquots to avoid frequent melting and freezing which results in significant losses of enzyme activity. It has been reported that cathepsins B, L, and H can be stored over long periods of time in the form of a reversible mercury complexes [41]. In this form the protein is also stable at room temperature. Prior to use the protease has to be reactivated in the presence of 10 mM EDTA. However, it should be noted that not all cathepsins can be stored by this method. Unpublished observations suggest an irreversible denaturation of cathepsins K and S in the presence of mercury ions. Alternatively, papain-like cysteine proteases can also be reversibly inactivated by the addition of methyl methane thiosulfonate (MMTS) in a molar ratio of 1.25:1 (inhibitor:enzyme) [42].

3. Concluding remarks

The purpose of this review was to provide an overview of the various protein expression systems for the generation of papain-like cysteine proteases, their purification, and activity assays. Each of the expression systems discussed have specific advantages as well as disadvantages and the researcher has to decide which systems suit his/her application most. Based on our experience, the *Pichia* system appears to be the most efficient when high yields of active protease are required at minimal production costs. A clear advantage of *Pichia* and yeast expression systems in general is the lack of endogenous cysteine proteases, the high yields, and the low content of proteins in the production media all of which facilitate a simple purification protocol. If glycosylation is a problem for crystallographic studies, potential N-glycosylation sites can be easily inactivated by site-directed mutagenesis. *E. coli* based expression is clearly advantageous if large amounts of inactive proteins are needed with a minimal investment in time and costs. Inclusion of body-based protease protein is suitable for antigen production to generate antibodies. The bacterial expression system is also recommended when a workable refolding protocol is on hand which permits the generation of active proteases. For most human cathepsins, detailed methods of expression and purification have been described. However, the increasing number of parasite-derived cysteine proteases with pharmaceutical relevance may require the development of modified and novel expression and purification protocols.

Acknowledgments

This work was supported by the NIH Grants AR46182 and 48669, by a Biomedical Science grant from the Arthritis Foundation (DB), and by the Ministry of Schools, Science and Sports of the Republic of Slovenia (BT). We thank Dr. Marko Dolinar for fruitful discussions.

References

- [1] F. Lecaille, J. Kaleta, D. Brömme, *Chem. Rev.* 102 (2002) 4459–4488.
- [2] D. Brömme, J. Kaleta, *Curr. Pharm. Des.* 8 (2002) 1639–1658.
- [3] M. Gowen, J.G. Mery, S. Kumar, *Emerging Drugs* 5 (2000) 1–43.
- [4] D.S. Yamashita, R.A. Dodds, *Curr. Pharm. Des.* 6 (2000) 1–24.
- [5] W.-S. Hou, W. Li, G. Keyszer, E. Weber, R. Levy, M.J. Klein, E.M. Gravalles, S.R. Goldring, D. Brömme, *Arthritis Rheum.* 46 (2002) 663–674.
- [6] N. Cimerman, P.M. Brguljan, M. Krasovec, S. Suskovic, J. Kos, *Pflugers Arch.* 442 (2001) R204–R206.
- [7] R.J. Riese, R.N. Mitchell, J.A. Villadangos, G.P. Shi, J.T. Palmer, E.R. Karp, G.T. De Sanctis, H.L. Ploegh, H.A. Chapman, *J. Clin. Invest.* 101 (1998) 2351–2363.
- [8] I.M. Berquin, B.F. Sloane, *Perspect. Drug Discovery Design* 2 (1995) 371–388.
- [9] N. Levicar, T. Strojnik, J. Kos, R.A. Dewey, G.J. Pilkington, T.T. Lah, *J. Neurooncol.* 58 (2002) 21–32.
- [10] J.H. McKerrow, *Int. J. Parasitol.* 29 (1999) 833–937.
- [11] M. Sajid, J.H. McKerrow, *Mol. Biochem. Parasitol.* 120 (2002) 1–21.
- [12] F.W. Studier, *J. Mol. Biol.* 219 (1991) 37–44.
- [13] S.M. Smith, M.M. Gottesman, *J. Biol. Chem.* 264 (1989) 20487–20495.
- [14] M.A. Taylor, K.A. Pratt, D.F. Revell, K.C. Baker, I.G. Sumner, P.W. Goodenough, *Protein Eng.* 5 (1992) 455–459.
- [15] D.F. Revell, N.J. Cummings, K.C. Baker, M.E. Collins, M.A. Taylor, I.G. Sumner, R.W. Pickersgill, I.F. Connerton, P.W. Goodenough, *Gene* 127 (1993) 221–225.
- [16] R. Kuhelj, M. Dolinar, J. Pungercar, V. Turk, *Eur. J. Biochem.* 229 (1995) 533–539.
- [17] M. Dolinar, D.B. Maganja, V. Turk, *Biol. Chem. Hoppe Seyler* 376 (1995) 385–388.
- [18] D. Barlic-Maganja, M. Dolinar, V. Turk, *Biol. Chem.* 379 (1998) 1449–1452.
- [19] G. Kopitar, M. Dolinar, B. Strukelj, J. Pungercar, V. Turk, *Eur. J. Biochem.* 236 (1996) 558–562.
- [20] D.A. Tobbell, B.J. Middleton, S. Raines, M.R. Needham, I.W. Taylor, J.Y. Beveridge, W.M. Abbott, *Protein Expr. Purif.* 24 (2002) 242–254.
- [21] K.J. D'Alessio, M.S. McQueney, K.A. Brun, M.J. Orsini, C.M. Debouck, *Protein Expr. Purif.* 15 (1999) 213–220.
- [22] H.S. Hwang, H.S. Chung, *Protein Expr. Purif.* 25 (2002) 541–546.
- [23] G. Velasco, A.A. Ferrando, X.S. Puente, L.M. Sanchez, C. Lopez-Otin, *J. Biol. Chem.* 269 (1994) 27136–27142.
- [24] I. Santamaria, G. Velasco, A.M. Pendas, A. Fueyo, C. Lopez-Otin, *J. Biol. Chem.* 273 (1998) 16816–16823.
- [25] T. Vernet, D.C. Tessier, F. Laliberte, D. Dignard, D.Y. Thomas, *Gene* 77 (1989) 229–236.
- [26] T. Ogino, T. Kaji, M. Kawabata, K. Satoh, K. Tomoo, T. Ishida, H. Yamazaki, K. Ishidoh, E. Kominami, *J. Biochem. (Tokyo)* 126 (1999) 78–83.
- [27] P.S. Sijwali, B.R. Shenai, P.J. Rosenthal, *J. Biol. Chem.* 277 (2002) 4.
- [28] T.W. Thannhauser, Y. Konishi, H.A. Scheraga, *Anal. Biochem.* 138 (1984) 181–188.
- [29] T. Komeda, K. Tazumi, H. Shimada, K. Kano, T. Hayashi, H. Saito, H. Tsumura, N. Kato, Y. Sakai, K. Kondo, *Appl. Microbiol. Biotechnol.* 59 (2002) 252–258.
- [30] H.E. Xu, S.A. Johnston, *J. Biol. Chem.* 269 (1994) 21177–21183.
- [31] J.M. Cregg, T.S. Vedvick, W.C. Raschke, *Biotechnology (N. Y.)* 11 (1993) 905–910.
- [32] D. Brömme, K. Okamoto, *Biol. Chem. Hoppe Seyler* 376 (1995) 611–615.
- [33] B. Wang, G.P. Shi, P.M. Yao, Z. Li, H.A. Chapman, D. Brömme, *J. Biol. Chem.* 273 (1998) 32000–32008.
- [34] T. Wex, F. Buhling, H. Wex, D. Gunther, P. Malfertheiner, E. Weber, D. Brömme, *J. Immunol.* 167 (2001) 2172–2178.
- [35] W.P. Ren, R. Fridman, J.R. Zabrecky, L.D. Morris, N.A. Day, B.F. Sloane, *Biochem. J.* 319 (1996) 793–800.
- [36] J. Guay, D. Riendeau, J.A. Mancini, *Bone* 25 (1999) 205–209.
- [37] J. Rozman, J. Stojan, R. Kuhelj, V. Turk, B. Turk, *FEBS Lett.* 459 (1999) 358–362.
- [38] D.K. Nagler, R. Zhang, W. Tam, T. Sulea, E.O. Purisima, R. Menard, *Biochemistry* 38 (1999) 12648–12654.
- [39] S.W. Dahl, T. Halkier, C. Lauritzen, I. Dolenc, J. Pedersen, V. Turk, B. Turk, *Biochemistry* 40 (2001) 1671–1678.

- [40] D. Brömme, P.R. Bonneau, P. Lachance, B. Wiederanders, H. Kirschke, C. Peters, D.Y. Thomas, A.C. Storer, T. Vernet, *J. Biol. Chem.* 268 (1993) 4832–4838.
- [41] A.J. Barrett, H. Kirschke, *Methods Enzymol.* 80 (1981) 535–561.
- [42] D.D. Roberts, S.D. Lewis, D.P. Ballou, S.T. Olson, J.A. Shafer, *Biochemistry* 25 (1986) 5595–5601.
- [43] S.J. Sanderson, K.G. Jpollock, J.D. Hilley, M. Meldal, P.St. Hilaire, M.A. Julianos, L. Julianos, J.C. Motttram, G.H. Coombs, *Biochem. J.* 347 (2000) 383–388.
- [44] K.J. D'Alessio, M.S. McQueney, K.A. Brun, M.J. Orsini, C.M. Debouck, *Protein Expr. Purif.* 15 (1999) 213–220.
- [45] L. Smith, C. Gilligan, in: W.A. Peck (Ed.), *Bone and Mineral Research*, vol. 6, Elsevier, New York, 1989, pp. 139–173.
- [46] A.E. Eakin, A.A. Mills, G. Harth, J.H. McKerrow, C.S. Craik, *J. Biol. Chem.* 267 (1992) 7411–7420.
- [47] P.J. Rosenthal, P.S. Sijwali, A. Singh, B.R. Shenai, *Curr. Pharm. Des.* 8 (2002) 1659–1672.
- [48] V.J. Chan, P.M. Selzer, J.H. McKerrow, J.A. Sakanari, *Biochem. J.* 340 (1999) 113–117.
- [49] C.R. Caffrey, J.P. Salter, K.D. Lucas, D. Khiem, I. Hsieh, K.C. Lim, A. Ruppel, J.H. McKerrow, M. Sajid, *Mol. Biochem. Parasitol.* 121 (2002) 49–61.
- [50] J.R. Somoza, J.T. Palmer, J.D. Ho, *J. Mol. Biol.* 322 (2002) 559–568.
- [51] E.M. Gravallesse, Y. Harada, J.T. Wang, A.H. Gorn, T.S. Thornhill, S.R. Goldring, *Am. J. Pathol.* 152 (1998) 943–951.
- [52] C. Linnevers, S.P. Smekens, D. Brömme, *FEBS Lett.* 405 (1997) 253–259.
- [53] E. Carmona, E. Dufour, C. Plouffe, S. Takebe, P. Mason, J.S. Mort, R. Menard, *Biochemistry* 35 (1996) 8149–8157.
- [54] H.E. Khouri, T. Vernet, R. Menard, F. Parlati, P. Laflamme, D.C. Tessier, B. Gour-Salin, D.Y. Thomas, A.C. Storer, *Biochemistry* 30 (1991) 8929–8936.
- [55] S.Y. Park, K.H. Lee, Y.B. Hwang, K.Y. Kim, S.K. Park, H.A. Hwang, J.A. Sakanari, K.M. Hong, S.I. Kim, H. Park, *J. Parasitol.* 87 (2001) 1454–1458.
- [56] D.K. Nagler, R. Zhang, W. Tam, T. Sulea, E.O. Purisima, R. Menard, *Biochemistry* 38 (1999) 12648–12654.
- [57] R. Menard, E. Carmona, S. Takebe, E. Dufour, C. Plouffe, P. Mason, J.S. Mort, *J. Biol. Chem.* 273 (1998) 4478–4484.
- [58] A.P.C.A. Lima, F.C.G. dos Reis, C. Serveau, G. Lalmanach, L. Juliano, R. Menard, T. Vernet, D.Y. Thomas, A.C. Storer, J. Scharfstein, *Mol. Biochem. Parasitol.* 114 (2001) 41–52.
- [59] T. Vernet, P.J. Berti, C. de Montigny, R. Musil, D.C. Tessier, R. Menard, M-C. Magny, A.C. Storer, D.Y. Thomas, *J. Biol. Chem.* 270 (1995) 10838–10846.
- [60] P.M. Steed, D. Lasala, J. Liebman, A. Wigg, K. Clark, A.K. Knap, *Protein Sci.* 7 (1998) 2033–2037.
- [61] B. Gotz, M.Q. Klinkert, *Biochem. J.* 290 (Part 3) (1993) 801–806.
- [62] C. Lauritzen, J. Pedersen, M.T. Madsen, J. Justesen, P.M. Martensen, S.W. Dahl, *Protein Expr. Purif.* 14 (1998) 434–442.
- [63] L. Hola-Jamriska, L.T. King, J.P. Dalton, V.H. Mann, J.G. Aaskov, P.J. Brindley, *Protein Expr. Purif.* 19 (2000) 384–392.
- [64] M.S. McQueney, J. Field, C.R. Hanning, K. Brun, K. Ramachandran, J. Connor, F. Drake, C.S. Jones, B.Y. Amegadzie, *Protein Expr. Purif.* 14 (1998) 387–394.
- [65] V. Alvarez, F. Parussini, L. Aslund, J.J. Cazzulo, *Protein Expr. Purif.* 26 (2002) 467–475.
- [66] C.S. Mason, M.B. Lamers, I.M. Henderson, T. Monk, D.H. Williams, *Protein Expr. Purif.* 23 (2001) 45–54.