Protein Expression and Purification 186 (2021) 105919

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Milligram scale expression, refolding, and purification of *Bombyx mori* cocoonase using a recombinant *E. coli* system

Chanrith Phoeurk [a,b], Ameeq Ul Mushtaq [a], Per Rogne [a,*], Magnus Wolf-Watz [a,*]

[a] Department of Chemistry, Umeå University, Umeå, Sweden
[b] Department of Bio-Engineering, Royal University of Phnom Penh, Phnom Penh, Cambodia

ARTICLE INFO

Keywords:
Silk moth (*Bombyx mori*)
Cocoonase
Serine protease
Refolding
Escherichia coli

ABSTRACT

Silk is one of the most versatile biomaterials with signature properties of outstanding mechanical strength and flexibility. A potential avenue for developing more environmentally friendly silk production is to make use of the silk moth (*Bombyx mori*) cocoonase, this will at the same time increase the possibility for using the byproduct, sericin, as a raw material for other applications. Cocoonase is a serine protease utilized by the silk moth to soften the cocoon to enable its escape after completed metamorphosis. Cocoonase selectively degrades the glue protein of the cocoon, sericin, without affecting the silk-fiber made of the protein fibroin. Cocoonase can be recombinantly produced in *E. coli*, however, it is exclusively found as insoluble inclusion bodies. To solve this problem and to be able to utilize the benefits associated with an *E. coli* based expression system, we have developed a protocol that enables the production of soluble and functional protease in the milligram/liter scale. The core of the protocol is refolding of the protein in a buffer with a redox potential that is optimized for formation of native and intramolecular di-sulfide bridges. The redox potential was balanced with defined concentrations of reduced and oxidized glutathione. This *E. coli* based production protocol will, in addition to structure determination, also enable modification of cocoonase both in terms of catalytic function and stability. These factors will be valuable components in the development of alternate silk production methodology.

1. Introduction

Silks are a general term for widely different protein polymers spun by a range of different arthropods with a range of functions [1]. Textile silk is extracted from the cocoon of the silkworm *Bombyx mori* (*B. mori*). With a global production of more than 150,000 tons a year, from 1 million tons of fresh cocoons, silk is one of the most valuable natural fibers in the world [2], with an annual production value of more than 3 billion dollars in 2019 (fao.org/faostat/, FAO 2020).

One crucial step in the process of turning the cocoon into fabric is the removal of the protein glue, sericin that holds the fibroin fibers together in the cocoon, this process is called degumming [3]. Degumming is heavily energy consuming and it is of interest to develop "green" degumming methods that are more environmentally friendly. In nature, the silk moth uses the protease cocoonase to degrade sericin to soften the cocoon. This is essential for the fully developed moth to be able to escape the cocoon [4]. Cocoonase has a high potential for improving the effectivity of degumming, while at the same time reducing the environmental impact of the silk industry.

The cocoon from *B. mori* contains a number of proteins. Sericin make up approximately one fourth of the weight of the cocoon. However, most abundant are the fibroins, the water-insoluble core of the silk fiber that is processed into the silk-threads, which makes up about two-thirds [5,6] of the weight of the cocoon. In addition to these two main proteins, two so-called seroins (BmSER 1 and BmSER 2) and two protease inhibitors (BmSPI 1 and BmSPI 2) have been isolated from the cocoon [7]. These last four proteins are involved in protecting the pupae against diseases [8,9]. Fibroin is made up of three different polypeptides, the heavy chain, the light chain, and a glycoprotein called P25 [10]. The heavy and the light chains are bound to each other with disulfide bonds while P25 is associated via non-covalent interactions [10].

Sericin is a group of proteins encoded by four different genes, *sericin*

Abbreviations: B. mori, *Bombyx mori*; BmCoc, *Bombyx mori* cocoonase; BPTI, Bovine pancreatic trypsin inhibitor; CD, Circular dichroism; CSE, Cocoon shell extract; HSQC, Heteronuclear single quantum coherence; SBTI, Soybean trypsin inhibitor; TEV, Tobacco etch virus; TLCK, *N*-α-p-Tosyl-L-lysine chloromethyl ketone; TROSY, Transverse relaxation optimized spectroscopy.

* Corresponding authors. Umeå University, Department of Chemistry, SE-901 87, Umeå, Sweden.
E-mail addresses: per.rogne@umu.se (P. Rogne), magnus.wolf-watz@umu.se (M. Wolf-Watz).

https://doi.org/10.1016/j.pep.2021.105919
Received 19 March 2021; Received in revised form 28 April 2021; Accepted 20 May 2021
Available online 24 May 2021

1 through *4* [11]. Sericin 1 and 3 are the main cocoon sericin, the other sericin covers silk fibers that have other functions such as attaching the larvae or cocoon to leaves or tree branches [12]. The sericin genes undergo alternative splicing to give rise to differently sized polypeptides in the range of 150–400 kDa [11]. However, other reports suggest sericin polypeptides with as low mass as 10 kDa [13] and these smaller peptides may be a result of the different methods used to extract sericin from the cocoon. Presently, sericin are mostly discarded as waste during silk fiber production [2]. However, sericin has several important applications where the medical industry has, for example, shown interest due to its mitogenic properties for wound healing and bone tissue engineering [14]. The cosmetic industry is also a potential application, in moisturizers, for example [14,15].

Degumming is one of the steps in the standard processing approach of the cocoons into silk fibers. Degumming can be accomplished by soaking the cocoons in hot water, however, this is a lengthy process that causes damage to the fibroin [2]. Most industrial silk processes use soaps in an alkaline water solution to hydrolyze and remove the sericin [3,16]. These methods do, however, require large quantities of heat and water and are therefore resulting in a significant environmental impact. In addition, the waste products can pollute the environment of the silk processing site [3].

An attractive route to make the degumming process more environmentally friendly is to utilize enzymatic degumming [17]. In enzymatic degumming, sericin is hydrolyzed and removed by using proteases. Proteases like alcalase, savinase [18], and bromelain [19] have been explored for this role. Enzymatic reactions are generally conducted at ambient temperatures and in non-toxic solutions directly reducing both the energy consumption and the pollution problem. When using proteases, it is important to consider the specificity of the enzymes since the fiber is also made of protein (fibroin) and is, generally, susceptible to degradation. Cocoonase has been shown to out-perform commercially available alcalase by providing a more complete sericin removal while at the same time causing less damage to the silk fiber [20]. Increased understanding of the properties of this enzyme will open avenues to using cocoonase in the silk industry and as a product of the knowledge the industrial potential of the enzyme might be increased by biotechnological improvements of its properties, i.e., increased stability or activity.

B. mori cocoonase (BmCoc) is a serine protease, similar to trypsin, and is produced as an inactive zymogen by the silk moth during the pupa state [21]. The zymogen is exported and stored as a dry deposit on the head of the moth [21]. At the time of break-out of the silk moth, the deposited enzyme is dissolved in a buffer produced by the moth and activated by cleavage of a 34 amino acid residues long N-terminal peptide to its active form [20]. The cleavage of the zymogen into the active BmCoc can be accomplished by different proteases like trypsin or subtilisin [21]. However, the main route of activation is cleavage by already activated BmCoc [22] in a self-propelling mechanism.

BmCoc can be extracted from its natural source, and early investigations [23–25] relied upon this method. However, the amounts of enzymes that can be obtained are very small, a protocol published by Fukumori et al. [25] showed that less than 1 μg of the pure enzyme could be obtained from one silk moth.

To efficiently obtain protein on a scale necessary for in depth biophysical investigations, including structure determination, a recombinant system is required. Using a yeast system based on *Pichia pastoris (P. pastoris)*, Unajak and co-workers [26] were able to express and purify BmCoc with a yield of 5.4 mg/l media. However, an expression system based on *Escherichia coli* (*E. coli*) has several advantages compared to a system based on yeast (eukaryotic cells). The cloning and transformation of *E. coli* cells are quick and easy in comparison and there are a multitude of developed expression vectors. The cells grow quickly on minimal media, which is important if isotopically labeled proteins are required. Isotopically labeled proteins are for example necessary when properties on the atomic scale are studied using nuclear magnetic resonance (NMR) spectroscopy. Effective NMR spectroscopy on proteins larger than 10 kDa requires proteins that are labeled with either ^{13}C, ^{15}N, ^{2}H, or a combination of these isotopes [27]. This can be obtained by growing the expression cells in a media that is enriched with the desired isotopes. *E. coli* can grow on minimal media where the sources of nitrogen and carbon can be closely controlled. In the commonly used M9 media nitrogen and carbon is supplied in the form of ammonium chloride ($^{15}NH_4Cl$) and ^{13}C enriched glucose, respectively. These chemicals are available, relatively affordable, enriched with the sought-after isotopes.

The disadvantage with using an *E. coli*-based expression system to express eukaryotic proteins is the lack of "folding system" which can result in the expressed proteins being misfolded and aggregated as so-called inclusion bodies. The problem is especially significant for proteins containing intramolecular di-sulfide bridges. BmCoc contains six cysteine residues that might form up to three such bridges. If the cysteines are solvent-exposed, either due to misfolding or during the folding process, inter-molecular di-sulfide bridges can form, permanently trapping the expressed proteins in a misfolded state. In order to refold active protease from inclusion bodies, they must be solubilized by unfolding and refolded into their native fold. A significant challenge to a refolding protocol is to develop a method that favors the formation of native di-sulfide bridges while dis-favoring non-native di-sulfide bridges and higher-order oligomers.

Here we present a robust and reproducible method to recombinantly produce large (mg/liter) amounts of functional BmCoc in *E. coli* by unfolding and refolding of protein expressed as inclusion bodies. The core of the protocol is that refolding is performed in an oxygen-free environment and with a redox potential that is optimized for the formation of native di-sulfide bridges. This favorable redox potential is accomplished with a balance of reduced and oxidized glutathione.

Two different versions of BmCoc were produced, one including a short N-terminal peptide (GAM-BmCoc) and one with a native N-terminal (En-BmCoc). The GAM peptide is used to reduce the autoproteolytic activity of cocoonase, giving a more stable enzyme for structural investigations. Quantitative analysis of the purified protein demonstrated that the enzyme is active and monomeric. Additionally, the GAM variant is stable over extended periods of time. En-BmCoc was shown to effectively degrade sericin on solid cocoon particles but displayed self-cleavage when incubated over longer time periods. In the manuscript we refer to the two variants (GAM-BmCoc and En-BmCoc) when appropriate, or alternatively BmCoc when joint properties of the two variants are discussed. The results represent a significant step forward and will enable both fundamental biochemical and structural investigations, as well as the initial steps toward green production of silk fibers.

2. Materials and methods

2.1. Cloning of BmCoc into the pET-His vector

Using the amino acid sequence for BmCoc (NCBI accession ID: BAJ46146) the corresponding cDNA, codon optimized for *E. coli* expression [28], was synthesized (Eurofins Genomics, Germany). The BmCoc cDNA was cloned into either (1) a pETHis-1a vector, providing a tobacco etch virus (TEV) protease cleavable N-terminal poly histidine tag (His-tag) or (2) a pET-NHis vector providing a N-terminal His-tag with an enterokinase cleavable linker (introduced by using PCR). Both vectors were kindly provided by Günther Stier, EMBL, Heidelberg, Germany. The TEV protease cleaves the recognition sequence (ENLYFQG) between the glutamine (Q) and the glycine (G) resulting in a primary sequence with the non-native N-terminal residue "G". In addition, the restriction enzyme NcoI is used to cleave the vector in order to insert the gene and the recognition site for NcoI (CCATGG) introduces codons for alanine (A) and methionine (M) between the TEV cleavage site and the N-terminal of the protein. Since TEV cleavage would not

produce a native N-terminal, the inclusion of these two extra amino acid residues was not considered problematic. This version of BmCoc will be referred to as GAM-BmCoc. Since enterokinase cleaves C-terminally of its recognition site (DDDDK), a native N-terminal is obtained, this version will be called En-BmCoc. For large parts of the expression, refolding, purification, and characterization the two variants (GAM-BmCoc and En-BmCoc) are treated equally, in these cases, they are called BmCoc.

2.2. Expression and refolding of recombinant BmCoc

The recombinant plasmids were transformed into BL21 (DE3) *E. coli* cells by heat shock and grown overnight on lysogeny broth (LB) -plates supplemented with 50 μg/ml kanamycin. Transformed cells were picked and grown at 37 °C in 15 ml LB containing 50 μg/ml kanamycin overnight. One milliliter of the over-night culture was used to inoculate 20 ml of fresh LB (50 μg/ml kanamycin) that was subsequently cultured at 37 °C for 2 h. The cells were harvested by gentle centrifugation and resuspended in 1 L of M9 medium (see Supplementary Table 1 for complete recipe). The cells were grown to an optical density at 600 nm (OD_{600}) in the interval 0.6–0.7 and then induced with 0.75 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and further grown at 37 °C for an additional 20–21 h. The cells were harvested by centrifugation at 3,300 *g for 30 min and the pellet was resuspended in 100 mM Tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl) at pH 7.5. Cell lysis was accomplished by using a Branson 450 Digital Sonifier (BRANSON Ultrasonics Corporation, USA). The lysate was centrifuged at 21,000 *g for 45 min, and the overexpressed recombinant BmCoc was found exclusively in the pellet fraction.

To refold and recover the recombinant BmCoc from inclusion bodies, a previously described protocol was modified [29]. The pellet was dissolved in 12 ml of 6 M guanidine hydrochloride (GuHCl), 100 mM Tris-HCl at pH 7.5 containing 30 mM dithiothreitol (DTT) and was incubated for 2 h at 37 °C. The dissolved pellet solution was centrifuged at 14,000 *g for 10 min to remove undissolved cell debris. 300 ml of refolding buffer (0.9 M GuHCl, 100 mM Tris-HCl at pH 7.5) was prepared and deoxygenated by purging the solution with nitrogen gas with stirring for an hour in a cold room. Reduced and oxidized glutathione was added to the deoxygenized refolding buffer to a final concentration of 5 mM and 1 mM, respectively. The soluble, reduced and unfolded protein solution was rapidly diluted into the refolding buffer with a volume ratio of one to thirty. This dilution lowers the DTT concentration to below 1 mM, lower than the reduced and oxidized glutathione. The bottle with the refolding solution was rapidly closed and sealed with parafilm; and the refolding reaction was allowed to proceed for 16–18 h at 6.5 °C with gentle stirring. Finally, the refolded sample was diluted two-fold in 50 mM Tris at pH 7.5, 500 mM NaCl with 10 mM imidazole. Precipitated protein were removed by centrifugation at 33,000 *g for 20 min followed by filtering with a 0.45 μm filter.

2.3. Purification of His-tagged BmCoc

Refolded BmCoc was applied to a Ni-NTA gravity flow column (Ni Sepharose™ 6 Fast Flow, GE Healthcare, USA) equilibrated with 50 mM Tris-HCl at pH 7.5, 500 mM NaCl, 5 mM imidazole. After washing the column with the equilibration buffer, the protein was eluted with 50 mM Tris-HCl, 500 mM NaCl, 400 mM imidazole at pH 7.5. The eluted BmCoc containing fractions were analyzed with SDS-PAGE, and concentrated using a 15 ml Amicon® Ultra 10K centrifugal filter (Merck Millipore Ltd., Ireland) for the following purification step by size exclusion chromatography (SEC). SEC was performed using a HiLoad™ 16/600 Superdex™ 75 prep grade column (GE Healthcare, USA) equilibrated with 50 mM NaCl, 2 mM DTT, 30 mM sodium phosphate buffer at pH 7.0. The purity and concentration of protein fractions were determined with SDS-PAGE and the absorption at 280 nm using an extinction coefficient of 24785 M^{-1} cm^{-1} and 21805 M^{-1} cm^{-1} for His-tag GAM-BmCoc and En-BmCoc respectively.

2.4. Removal of His-tag

To remove the His-tag, the GAM-BmCoc solution (with the TEV protease cleavage site) was buffer exchanged into 0.5 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, 50 mM Tris-HCl, pH 8.0. TEV protease was added in the ratio of 1:50 (w/w) and incubated overnight at room temperature. Proteolytic removal of the His-tag was verified with SDS-PAGE. The buffer was exchanged, by repeated concentration in a 15 ml Amicon® Ultra 10K centrifugal filter (Merck Millipore Ltd., Ireland) and dilution, and loaded on the Ni-NTA gravity flow column equilibrated with 50 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole, pH 7.5 to remove the His-tagged TEV protease that is of equal size to the cocoonase, 27 kDa and 24 kDa, respectively. This step also removes the cleaved off His-tag.

For the En-BmCoc, the buffer was exchanged to 50 mM Tris-HCl, pH 7.75. The protein was activated by auto-proteolytic (i.e. an intermolecular cleavage process in the absence of enterokinase) cleavage of the linker by incubating the solution at 37 °C for 15–17 h. This can be rationalized with that the linker contains a lysine residue that together with arginines are known targets of certain serine proteases The cleaved protein was polished with a second round of SEC to separate it from His-tag residues. Finally, both GAM-BmCoc and En-BmCoc containing solutions were buffer exchanged to 20 mM NaPi, 50 mM NaCl, pH 6.0 for NMR experimental analysis. Purified BmCoc, was stored at −20 °C.

2.5. Circular dichroism (CD) spectroscopy

Far-UV CD spectra were recorded in the range of 195–250 nm at 20 °C using a Jasco J-715 spectropolarimeter (Jasco inc, Japan). The protein concentration was 20 μM in 10 mM sodium phosphate buffer at pH 7.0. The spectra were acquired using a 0.1 cm quartz cuvette, a bandwidth of 2 nm, and the data were recorded and averaged based on ten repetitions. Background correction was performed by subtracting the buffer signal. Thermal unfolding experiments were conducted using the same sample conditions by following the CD signal at 205 nm wavelength. The temperature was raised from 20 °C to 75 °C at a rate of 1 °C per minute. The melting point (T_M) and enthalpy of unfolding (ΔH_U) was quantified by non-linearly fitting equation (1) [30] to the acquired data the Origin software (OriginLab, USA)

$$CD(T) = \frac{(CD_n + m_n T) + (CD_d + m_d T) * e^{\frac{\Delta H_U *}{R}\left(\frac{1}{T_M} - \frac{1}{T}\right)}}{1 + e^{\frac{\Delta H_U *}{R}\left(\frac{1}{T_M} - \frac{1}{T}\right)}} \quad (1)$$

Where CD (T) is the signal as a function of the temperature, T, CD_n is the CD signal of the native protein at 0 K, m_n is the change of the native CD signal with temperature, CD_d and m_d are the corresponding parameters for the denatured state and R is the gas constant.

2.6. Protein NMR spectroscopy

Protein NMR spectra were acquired on a Bruker Avance III HD 600 MHz spectrometer equipped with a 5 mm BBO cryoprobe and a SampleJet sample changer, using a ^{15}N labeled GAM-BmCoc sample at a concentration of 0.8 mM and containing 5% D_2O (v/v). ^{1}H–^{15}N-TROSY-HSQC spectra were acquired in 20 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 308 K. The pulse programs were obtained from the Topspin 3.6.2 library. Two-dimensional ^{1}H–^{15}N TROSY experiments were acquired with 8 scans, time-domain sizes of 256(^{15}N) × 2048(^{1}H) complex points and sweep widths of 8417.51 Hz and 1946.30 Hz along the ^{1}H and ^{15}N dimensions, respectively. All NMR experimental data were processed and analyzed using the TopSpin 3.6.2 program (Bruker Biospin, Switzerland).

NMR data for the qualitative protein stability assessment were acquired on a Bruker Avance III HD 850 MHz spectrometer equipped with a z-gradient cryoprobe, using a 1 mM GAM-BmCoc sample containing 5% D_2O (v/v). 3 mM protease inhibitor, phenylmethyl sulfonyl fluoride (PMSF) was added to prevent the self-degradation of protein. $^{1}H–^{15}N$-TROSY-HSQC spectra were recorded in identical conditions and parameters other than the sweep widths of 10204 Hz and 2757 Hz along the ^{1}H and ^{15}N dimensions, respectively.

2.7. Activity of cocoonase from NMR and model substrates

The enzymatic activities of GAM-BmCoc and En-BmCoc were determined by measuring the initial rate of the enzyme-catalyzed hydrolysis reaction of three individual model substrates: (1) Z-L-Lys-ONp hydrochloride, (2) Nα-benzoyl-L-arginine 4-nitroanilide hydrochloride (Sigma Aldrich, USA), and (3) Sericin3 peptide (acetylated-SDEKSSQS-NHMe, synthesized by GenScript Biotech, Netherlands). For each experiment 2 mM substrate and between 50 nM and 5 μM BmCoc were used. All experiments were carried out in a 20 mM sodium phosphate buffer at pH 6.0, supplemented with 50 mM NaCl and 5% D_2O (v/v). Trypsin from bovine pancreas (Sigma Aldrich, USA) was used as a control. Quantitative NMR data of enzyme kinetic assays were acquired on a Bruker Avance III HD 600 MHz spectrometer equipped with a 5 mm BBO cryoprobe and a SampleJet sample changer. One-dimensional ^{1}H NMR experiment was first performed on the sample containing only substrate(s) for resonance identification. One dimensional (1D) ^{1}H NMR experiment with excitation sculpting (es) for water suppression (Topspin3.6 zgesgp pulse program) was used. Each 1D proton NMR spectrum was recorded in the same buffer conditions at 298 K, and collected with a spectral width of 16.02 ppm (9615.385 Hz) over 24998 points yielding an acquisition time of 1.3 s per experiment. A relaxation delay of 1 s was used between each scan and 112 of scans were averaged for the signal. This experiment was assembled to collect FIDs continuously in a time series for example, a total of 50 such experiments (50 × 4.52 min = 226 min) were performed.

2.8. In situ, cocoon assay

Bombyx mori cocoons were obtained from the Silk Center, Royal University of Phnom Penh, Cambodia. The pupas were removed from the cocoons by cutting the cocoons with a clean pair of scissors. All handling of the cocoons was performed at room temperature. Seventy milligram of silkworm cocoon was weighed and cut into approximately 0.5–1 mm pieces and extensively washed with the reaction buffer (20 mM sodium phosphate, 50 mM NaCl at pH 6.0). Thereafter the cocoon pieces were suspended in 5 ml reaction buffer and incubated with either 1 μM or 10 μM En-BmCoc at 25 °C for 24 h with shaking at 170 rpm. Released peptide fragments were observed by one-dimensional ^{1}H NMR spectra acquired at 298 K. The experiment was designed with a negative control (no enzyme added) and a positive control (by adding 1 μM trypsin). The principle of the experiment is that while the cocoon is a macroscopic object and does therefore not give rise to observable NMR signals, released peptide fragments that tumble fast in solution give rise to observable signals. This principle is demonstrated by the absence of observable signals in the negative control while there exist numerous observable resonances after incubation with both En-BmCoc and trypsin.

3. Results and discussion

To fulfill the demands of structural studies and industrial applications, recombinant BmCoc production is required. Therefore, we have developed a robust and reproducible protocol for production of milligram quantities of BmCoc in *E. coli*. We have focused on two variants of the enzyme one with a completely native and activated sequence (En-BmCoc) and one with three non-native and N-terminal amino acid residues (GAM-BmCoc). GAM-BmCoc serves as a proxy of the inactivated zymogen and was designed to enable structural studies since it is expected to have a lower likelihood of self-cleavage. *Bombyx mori* cocoonase is a serine protease and it displays strong sequence similarity to trypsin with 41% sequence identity. Trypsin contains six intramolecular di-sulfide bridges [31] (Fig. 1A). In order to determine the number of putative native di-sulfide bridges in BmCoc a structural homology model of BmCoc was created using Robetta modeling [32] with trypsin (PDB ID 2FI5) as template (Fig. 1B). From the model, we identified three putative di-sulfides that all are conserved in relation to trypsin. In BmCoc the di-sulfides are formed from the following pairs of cysteines: C63:C79, C187:C200 and C211:C235, corresponding to the conserved cysteine pairs of trypsin: C42:C58, C168:C182 and C191:C220 in trypsin. Thus in development of a protocol for production of BmCoc care must be taken such that the three expected native and intramolecular di-sulfides are properly formed. These di-sulfides poses a challenge to protein production in *E. coli* since its cytoplasm has a reducing redox potential.

3.1. Sub-cloning of BmCoc into pET plasmids

In order to have complete control of the gene, the BmCoc gene corresponding to the active enzyme (residues 35–260) were synthesized (Eurofins Genomics, Germany) with an N-terminal His-tag and the two different linkers, TEV and enterokinase cleavable respectively (Fig. 2). To improve the expression, the genes were codon optimized for *E. coli* [33]. The genes was cloned into a pET vector (pET His 1a) under control of the T7 promotor.

3.2. Production and refolding of BmCoc

The two versions showed a similar behavior during expression, refolding, and initial purification. Significant differences did not occur until the His-tag removal step.

The plasmids were transformed into *E. coli* BL21(DE3) cells using the heat-shock procedure. Transformed cell colonies were picked and the cells were cultivated in LB-media to an optical density at 600 nm (OD_{600}) of 0.7 before induction with IPTG at a final concentration of 0.75 mM. After 20 h of expression at 37 °C, SDS-PAGE analysis on the whole cells showed that both recombinant GAM-BmCoc (Fig. 3A) and En-BmCoc was highly overexpressed. In addition we demonstrated a

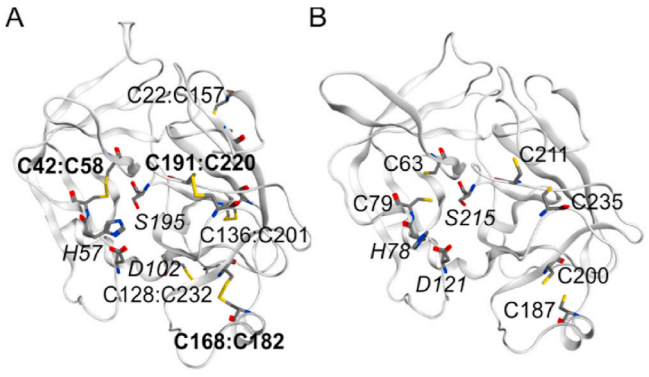

Fig. 1. Di-sulfides and catalytic triad in BmCoc from a homology model based on trypsin. (**A**) Structure of trypsin (PDB ID 2FI5) with the six di-sulfide bonds indicated (named as residueX:residueY), the bonds indicated in bold have conserved counterparts in BmCoc, together with the catalytic triad (H57, D102 and S195, italic). (**B**) Homology model of BmCoc obtained by using the Robetta server [32]. The cysteine residues involved in putative di-sulfide bridges are labeled together with the catalytic triad (H78, D121 and S215). The resulting sequence alignment from the comparative modelling is shown in Supplementary Fig. 1. The figure was created using MOE (Chemical computing group, Canada).

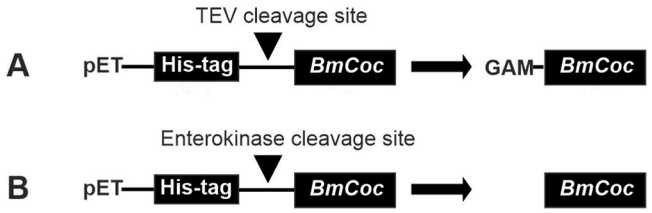

Fig. 2. *Bombyx mori* cocoonase plasmid constructs. Schematic illustration of the two plasmid constructs utilized for production in *E. coli.* BmCoc, was cloned with an N-terminal His-tag with either (A) a TEV protease- or (B) an enterokinase-cleavage site to facilitate removal the tag. Cleavage with the TEV protease leaves the non-native sequence GAM sequence N-terminal to the native isoleucine 35 of activated BmCoc. On the other hand, cleavage at the enterokinase site produces the correct and native activated enzyme with an N-terminus corresponding to ILE35.

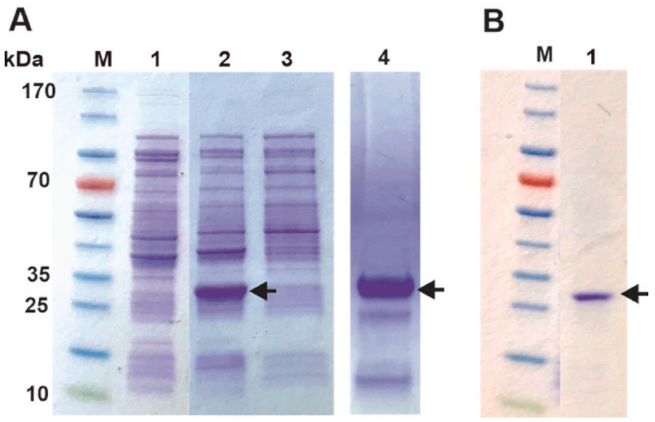

Fig. 3. Production and solubilization of GAM-BmCoc. Both production and a successful solubilization of insoluble GAM-BmCoc was analyzed with SDS-PAGE. (A) Lane 1: Approximately 400 μg of cells prior to induction, Lane 2: approximately 400 μg of cells 20 h after induction with IPTG, Lane 3: soluble fraction of approximately 400 μg of cells after lysis, Lane 4: insoluble fraction of approximately 2 mg of cells after lysis. GAM-BmCoc (25 kDa) production was induced with 0.75 mM IPTG and 20 h of overexpression. (B) Lane 1: 12 μl out of 600 ml (1:50000) of the soluble fraction after the refolding step. The arrow shows the protein band at the size of His-tagged GAM-BmCoc, indicating that GAM-BmCoc stays soluble in a native buffer after the refolding step. En-BmCoc showed similar expression and solubilization characteristics. The migration of BmCoc is indicated with an arrow. PageRulerTM prestained protein ladder was used as marker in both A and B (lanes M).

high level of overexpression also in M9 minimal media supplemented with either ^{15}N or $^{15}N/^{13}C$, this is relevant since it enables future structure determination and also quantification enzyme dynamics with solution state NMR spectroscopy [34]. After cell lysis by sonication, the expressed protein, with the expected molecular weight of 26 kDa, was absent from the soluble fraction and present exclusively as inclusion bodies in the cell debris pellet (Fig. 3A). The insoluble proteins from the inclusion bodies were resolubilized by resuspending the pellet in 6 M GuHCl at 37 °C for 2 h. Di-sulfide bridges were reduced by adding DTT to a final concentration of 30 mM. Both versions of BmCoc were successfully refolded by diluting the unfolded protein 30 times (v/v) in a buffer containing 5 mM reduced and 1 mM oxidized L-glutathione. This ratio of reduced vs. oxidized glutathione is optimized in order to keep the redox balance environment such that native and internal di-sulfide bridges are favored while intermolecular di-sulfide bridges or non-native and solvent exposed bridges would not form. The protocol is based on a previously established methods that was developed for the Human Cationic Trypsinogen [29]. The soluble material from this refolding step was analyzed on a SDS-PAGE gel showing a high concentration of almost pure cocoonase (Fig. 3B). The efficacy of the refolding step was assessed by comparing the absorbance of the eluted monomer peak to dimer/-oligomer peaks in the size exclusion chromatography (SEC) to around 90%.

3.3. Purification of refolded BmCoc

The refolded BmCoc was first purified in a batch-mode by virtue of the N-terminal His-tag by loading the supernatant from the refolding step onto a nickel affinity column. After washing the column with the binding buffer, bound proteins were eluted with 400 mM imidazole. The eluated fractions were qualitatively analyzed with SDS-PAGE gel indicating reasonably pure GAM-BmCoc (Fig. 4), from a single purification step. En-BmCoc exhibited a similar purification pattern.

The protein was further purified by size exclusion chromatography (SEC) using a HiLoad™ 16/600 Superdex™ 75, The GAM-BmCoc containing chromatography peak is symmetric with an apparent Gaussian shape, which is indicative of mono-disperse material (Fig. 5A). SDS-PAGE analysis of the fractions in the main peak show a single protein band (Fig. 5A, inset) indicating that the protein is purified to homogenetity. To determine the aggregation status of the eluted GAM-BmCoc we turned to quantitative analysis of the SEC chromatogram and using the empiric linear relationship between log(Mw) and elution volume (V_e) for spherical particles [35]. On basis of a calibration curve generated with the proteins human albumin (66 kDa), Ovalbumin (44 kDa) and Myoglobulin (17 kDa), and the V_e of GAM-BmCoc (75.1 ml) we computed an apparent molecular weight of 25 kDa for GAM-BmCoc (Fig. 5B). En-BmCoc shows a similar behavior to GAM-BmCoc with V_e of 75.5 ml and an estimated molecular weight of 25 kDa. Since these quantified molecular weights are virtually similar to the weight computed from the primary sequence (24 kDa), it is evident that both versions of refolded BmCoc are monomeric in solution. This is a significant result since it indicates that only internal di-sulfides have been formed during refolding.

For GAM-BmCoc the N-terminal His-tag was removed by adding TEV protease in ratio of 1:50 (w/w). After 15–17 h incubation at room temperature, the cleavage was analyzed with SDS-PAGE (Fig. 6A) showing that the protein was completely cleaved. The protease and the remaining His-tag were removed using a Ni-column by collecting the

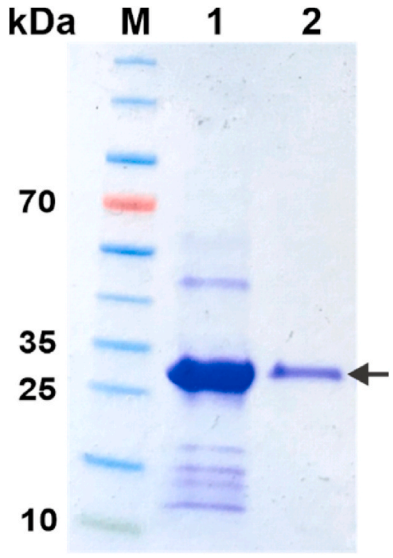

Fig. 4. Analysis of His-tag purification. The initial purification step and subsequent removal of the His-tag was analyzed with SDS-PAGE. Analysis of his-tagged GAM-BmCoc fractions eluted with 400 mM imidazole, first and second 5 ml fractions. The black arrow indicates the His-tagged GAM-BmCoc. Lane M: 3 μl PageRuler™ prestained protein ladder.

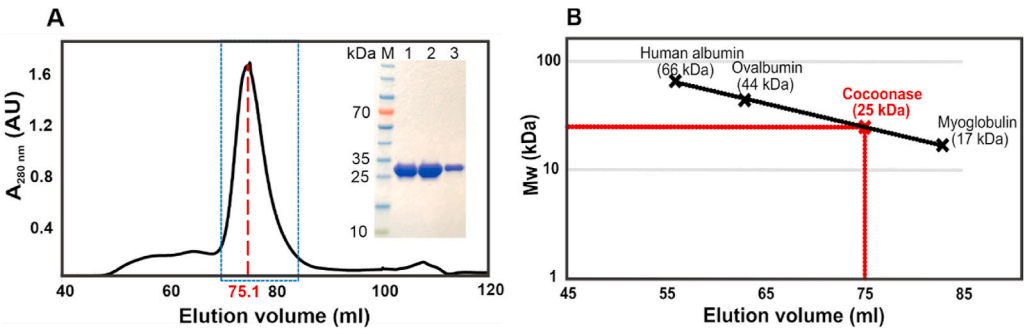

Fig. 5. GAM-BmCoc is monomeric in solution. The apparent molecular weight of GAM-BmCoc was quantified with quantitative size exclusion chromatography (SEC) (A) SEC profile of refolded GAM-BmCoc. Three fractions from the main peak, eluted at 75.1 ml (highlighted with the box), was analyzed by SDS page gel (insert). (B) The apparent molecular weight was determined by comparing the elution volume to three standard proteins. The elution volumes of the standard proteins were obtained from the manufacturer of the Superdex 75 pg 16/600 pre-packed column. A plot of log(Mw) of the standard proteins versus elution volume (dotted line) rendered a linear correlation that was used as a standard curve to determine the size of GAM-BmCoc. The size of GAM-BmCoc was determined from the elution volume (75.1 ml) to be 25 kDa. Since the molecular weight computed from the primary sequence is 24 kDa it is evident that GAM-BmCoc is monomeric in solution.

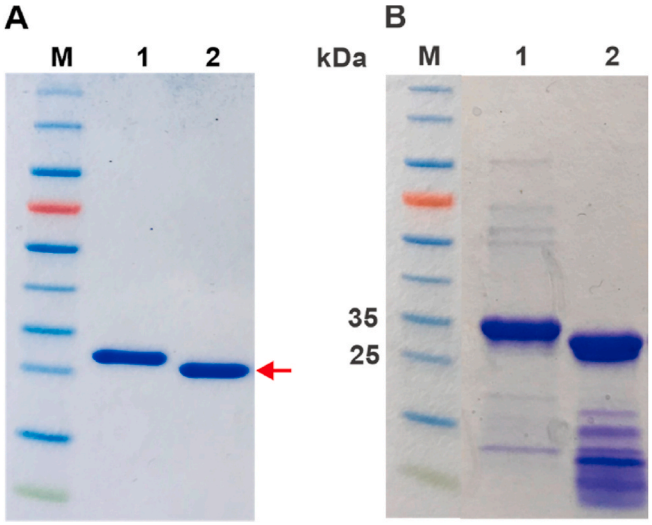

Fig. 6. His-Tag removal from BmCoc. The His-tag was removed by either TEV protease, from the GAM-BmCoc (A), or by autolysis, from En-BmCoc (B). The SDS-PAGE gel shows the protein before (lane A1 and B1) and the protein after cleavage (A2 and B2). The change in migration indicates that the molecular weight has be reduced by the removal of the His-Tag. The autolysis of En-BmCoc does not only remove the His-tag (too small to be visible on the gel) but also degrades BmCoc itself as evident by the low molecular weight bands visible on the gel.

flow-through fraction. The final yield obtained from this construct were 15 mg/l, 8 mg/l, and 7 mg/l respectively from three different culture media (LB, ^{15}N labeled M9, and double ^{15}N & ^{13}C labeled M9 media) (Table 1).

After the SEC purification step, En-BmCoc protein, was buffer exchanged to 50 mM Tris-HCl, pH 7.75 and incubated at 37 °C for 15–17 h. We observe that active BmCoc itself can cleave the His-tagged precursor protein and therefore a chain reaction is initiated after buffer exchange by autolyzed BmCoc molecules (Fig. 6B). Therefore, enterokinase is not required for removal of the tag. The cleaved protein was polished with another round of SEC to separate from His-tag residues and degraded protein. For En-BmCoc, the final yield was about, 5 mg/l ^{15}N in labeled M9 media, a bit lower than GAM-BmCoc, 8 mg/l. A problem with active En-BmCoc is that it also displays auto-cleavage resulting in degradation of the native enzyme, as evident by the smaller peptides seen on the SDS-PAGE gel of the cleavage (Fig. 6B). This process is however slow enough to allow quantification of activity as discussed below (see below). Taken together the developed protocols enables the production of large amounts of both unlabeled but also of isotopically enriched GAM/En BmCoc.

Table 1
Summary of protein yields during purification of recombinant *Bombyx mori* cocoonase (GAM-BmCoc).

Step	Total protein (mg) from 1 L cell culture		
	LB media	^{15}N-M9 media[a]	^{13}C-M9 media[b]
Ni-NTA affinity chromatography	47	11	N/A
Size exclusion chromatography	18	9.5	9.1
Ni-NTA affinity chromatography	15	8.0	7.0

N/A = not checked.
[a] ^{15}N labeling medium.
[b] Double ^{15}N and ^{13}C labeling medium.

3.4. Biophysical characterization of BmCoc

Although the SEC analysis provides an indication that the enzyme is not only monomeric but also properly folded, we wanted to test this feature further and on the molecular level with an extended biophysical characterization. First, we quantified the thermal unfolding of the enzyme by monitoring the CD signal at 205 nm in response to a temperature scan (Fig. 7A). The temperature dependency of the CD signal displays a cooperative transition over a narrow temperature interval. This behavior is a solid indication that the protein has secondary structure at ambient temperatures, while it is unfolded at temperatures well above the infliction point (i.e. the melting temperature, T_M). The main driving force in thermal unfolding of a protein is the difference in enthalpy (ΔH_U) between folded and unfolded states. A fit of the CD data to equation (1) revealed that the ΔH_U for unfolding of His-tag GAM-BmCoc was 390 ± 50 kJ mol^{-1}. The absolute magnitude of the unfolding ΔHu is comparable to other proteins of similar size, for instance adenylate kinase from *E. coli* with a size of 24 kDa has a ΔH_U of 570 ± 20 kJ mol^{-1} [36]. The quantitative analysis of the CD data also provided the T_M that was found to be 60 ± 0.5 °C. Thus the T_M is well above the temperature where cocoonase is active in its natural environments. The CD signal in the far-UV region is predominantly sensitive to secondary structure elements [37] and to take the characterization further and to the level of individual amino acid residues, we turned to protein NMR spectroscopy on ^{15}N enriched material. The ^{1}H–^{15}N TROSY-HSQC NMR spectra of ^{15}N labeled GAM-BmCoc, was well dispersed in both ^{1}H and ^{15}N dimensions, with approximately 200 backbone amide resonances highly resolved (Fig. 7B). This well-dispersed NMR spectrum is strongly indicative of a homogenous, folded globular protein. The enzyme was shown to be active (see below) and hence the fold must be correct. This is the first heteronuclear 2D spectrum of *Bombyx mori* cocoonase published, reflecting on that we have solved the difficulties to obtain

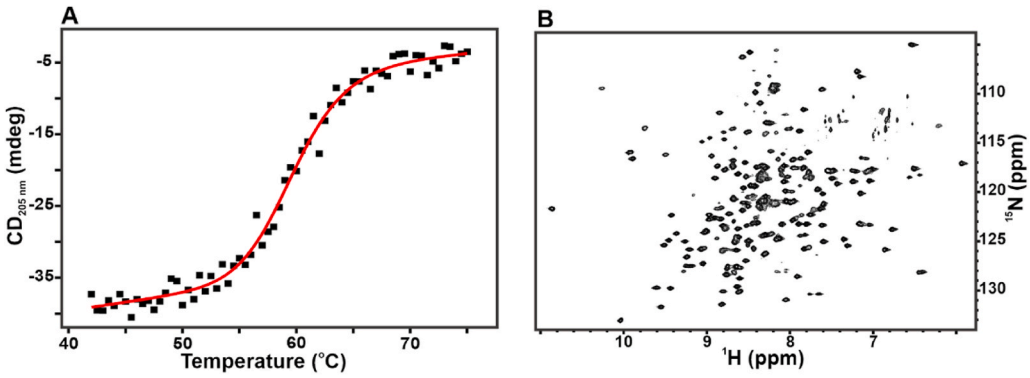

Fig. 7. The refolded BmCoc is a folded protein. (A) The thermal unfolding curve of His-tagged BmCoc shows a cooperative unfolding indicating a well folded protein. The unfolding temperature (T_m) and the enthalpy of unfolding (ΔH) were determined to 60 ± 0.5 °C and −390 ± 50 kJ mol^{-1} respectively. The unfolding was monitored by circular dichroism (CD) at 205 nm and the data was fitted to a two-state reaction (equation (1)). The sample contained 20 μM of His-tagged GAM-BmCoc in 10 mM NaPi buffer at pH 7.0. (B) The well dispersed TROSY-HSQC NMR spectrum of ^{15}N labeled GAM-BmCoc is indicative of a folded protein. The spectrum was recorded on a 0.8 mM GAM-BmCoc sample in a 20 mM NaPi buffer at pH 6.0 supplemented with 50 mM NaCl, at 308K and 600 MHz.

sufficient amount for structural studies of cocoonase.

Since En-BmCoc showed signs of self-degradation (Fig. 6B), we used NMR spectroscopy to investigate this. To this end the magnitude of the En-BmCoc signals were monitored over time and we did observe spectral changes that indicated self-degradation (Supplementary Fig. 2A). However, the degradation is sufficiently slow to allow quantitative analysis of enzymatic activity (see below). On the other hand, stability over extended periods is essential for structural studies with NMR, and we, therefore, investigated the inhibitory effect of different known protease inhibitors. In the apo state, En-BmCoc showed an extensive degradation within 22 h at 35 °C; therefore, it is necessary, in order to obtain the long-term stability needed for the acquisition of structural data to use an effective protease inhibitor. Unajak and co-workers [26] reported that PMSF, a serine protease specific inhibitor, could strongly inhibit the protease activity of recombinant BmCoc. Our results showed that PMSF provided only a limited inhibitory effect. Although cocoonase belongs to the same family of serine protease as trypsin and chymotrypsin, the inhibitory effects by a wide range of serine protease inhibitors are known to be relatively low [25]. Previous studies has showed that the inhibitory effect of a range of trypsin inhibitors, such as bovine pancreatic trypsin inhibitor (BPTI), soybean trypsin inhibitor (SBTI), cocoon shell extract (CSE), and *N*-α-p-Tosyl-L-lysine chloromethyl ketone (TLCK), was lower for BmCoc compared to trypsin [25]. Zhao and co-workers found that PMSF could inhibit the total larval mid-gut proteases, high-alkaline trypsin, low-alkaline trypsin and chymotrypsin [38]. Diisopropyl fluorophosphate was proved as a potent inhibitor of cocoonase [24]; however, it is about 200 times less sensitive with cocoonase than chymotrypsin [39,40]. Based on these previous investigations, it is not surprising that PMSF has a limited inhibitory effect on cocoonase.

We have also used the cOmplete™ protease inhibitor cocktail (Roche, USA) containing both reversible and irreversible protease inhibitors to increase the stability of BmCoc (Supplementary Fig. 2B). The inhibitor stabilizes the enzyme for at least 3 days; which will enable the time consuming NMR experiments required for structure determination.

3.5. Enzymatic activity

As mentioned above BmCoc and trypsin have a high sequence similarity (41% identity), therefore it was postulated that BmCoc would have a similar substrate selectivity as for trypsin, cleaving the peptides on the C-terminal side of lysine and arginine amino acid residues (MEROPS database [41]). In this study, two lysine peptides (Z-L-Lys-ONp and a peptide derived from Sericin3) and one arginine peptide (Nα-Benzoyl-L-arginine 4-nitroanilide hydrochloride, L-BAPNA) were designed and used as model substrates to determine the activity of En-BmCoc. Real-time enzyme kinetics were monitored by quantitative ^{1}H NMR at 25 °C. The assay is a modification of our previously established assay based on ^{31}P NMR designed for nucleoside phosphate catalysis [42]. In short, the assay provides k_{cat} values by determining the initial slopes of the changes in either substrate or product concentrations, determined by the intensities of the NMR resonances (Supplementary Fig. 3). Since NMR is a quantitative technique, the intensity of resonances are directly correlated to the concentrations, so the maximum reaction velocity (V_{max}) was determined as the initial slopes of the peak intensities as a function of time after enzyme addition. The k_{cat}, values were determined from the maximum reaction rate divided by the concentration of the enzyme, i.e $k_{cat} = V_{max}/[E]_{tot}$ (Table 2). The design of experiment included trypsin as a positive control for benchmarking of quantified k_{cat} values. With Z-L-Lys-ONp was used as a substrate, the k_{cat} value of the active En-BmCoc (20 s^{-1}) is equal to trypsin within the experimental uncertainty. On the other hand, the k_{cat} of En-BmCoc is 10 fold lower than trypsin for cleavage of L-BAPNA substrate indicating that cocoonase have a higher specificity for a lysine peptide compared to trypsin. This results indicates that BmCoc has a preference for cleavage after lysine over arginine. Although the activity of En-BmCoc is comparable to trypsin for the model compounds, the k_{cat} for cleavage of the peptide derived from the natural sericin substrate is significantly slower (k_{cat} is 0.25 s^{-1}). With this somewhat unexpected observation we decided to develop an in-situ assay for Sericin degradation of cocoons (see below).

The enzymatic activities are so far discussed for native, En-BmCoc. It is known that the activation mechanism of trypsin is dependent on the insertion of the N-terminal residue (Isoleucine 16 in BmCoc and isoleucine 35 in trypsin) to enable indirect contacts with the active site [43]. This feature is not possible for the zymogen where the N-terminal peptide blocks insertion of Ile16/Ile35. Even though we have not studied the zymogen per se we reasoned that the GAM-BmCoc could acts as proxy for the zymogen since the N-terminal is non-native and likely hinders Ile35 to insert into the body of the enzyme. The k_{cat} value of GAM-BmCoc was low with all substrates; for instance, the hydrolysis of

Table 2
Enzyme kinetics constant (k_{cat}, turnover number) of *BmCoc* and trypsin using three different synthesized peptides.

Enzyme	*Substrate*, k_{cat} (s^{-1})		
	Z-L-Lys-ONp	Nα-Benzoyl-L-arginine 4-nitroanilide	Sericin3 peptide
GAM-BmCoc	0.11 ± 0.01	0.01 ± 0.001	0.03 ± 0.002
En-BmCoc	20.0 ± 0.51	0.04 ± 0.01	0.25 ± 0.02
Trypsin	20.0 ± 0.23	0.36 ± 0.05	0.56 ± 0.03

The k_{cat} values are mean values of two independent experiments with four technical repetitions.

Z-L-Lys-ONp catalyzed by this GAM-BmCoc was approximately 180 times slower than En-BmCoc (Table 2). This results indicates that leaving GAM residues at the N-terminal significantly interferes the activity of BmCoc and that the determined k_{cat} values is an approximation of the activity of the zymogen. The quantification of activity with model substrates demonstrated that BmCoc is an active enzyme which is the ultimate proof that our refolding protocol is functional.

3.6. In-situ assay for softening of cocoons

Although the experiments with model substrates are useful, the main goal of our research is to study the efficacy of En-BmCoc towards its natural substrate, i.e. the sericin proteins embedded in cocoons. GAM-BmCoc was not tested because it showed low activity against the peptides. In its natural environment, BmCoc is activated by flushing of excreted solvent over the solid BmCoc deposit on the head of the moth. This wetting mechanism enables the direct contact between the enzyme and the surface of the cocoon. In order to mimic these circumstances in a laboratory setting, we developed a NMR detected experiment where recombinant En-BmCoc was added to cocoon fragments that were pre-wetted with buffer. Pieces of cocoon were suspended and pre-incubated in a buffer consisting of 20 mM NaPi, 50 mM NaCl at pH 6.0 for 24 h at 25 °C. In order to define the time equal to zero point a ^{1}H 1D NMR spectrum was recorded before adding the enzyme. To initiate hydrolysis of sericin either 1 μM or 10 μM of En-BmCoc were added to 70 mg of cocoon. As a control, 1 μM trypsin was added in a separate experiment. The principle of detection in the experiment is that while the intact cocoon is too large to yield detectable NMR signals [27], released peptide fragments tumble fast enough to be readily detected. The hydrolyzed products released into the solution were analyzed by serial ^{1}H NMR spectra up to 24 h after the enzyme was added. Resonances in the amide proton region, between 6.7 and 8.7 ppm, originating from the peptides released upon sericin break-down of the cocoon were monitored (Fig. 8). The spectra showed that En-BmCoc had a sizeable activity in hydrolyzing the sericin, which is notable since the action on a sericin model peptide was very slow (Table 2). The amount of the released sericin peptides scaled with enzyme concentration indicating a Michaelis-Menten like behavior. It should be noted that although we perform a quantitative detection of released peptides it is not possible to provide a meaningful V_{max} or k_{cat} since the substrate is a solid particle. In addition, trypsin displayed detectable cleavage hydrolysis of the cocoon but in comparison to En-BmCoc the activity of trypsin is considerably lower. Thus, we can conclude that although the k_{cat} for cleavage of model peptides is similar, BmCoc has evolved with an increased efficacy towards sericin embedded in the cocoon substrate in comparison to trypsin. These differences may be encoded in the surface properties of BmCoc and therefore it is important to solve its three dimensional structure to enable identification of the physiochemical mechanism of the increased efficacy of BmCoc, The fact that our recombinantly produced and refolded BmCoc is active on the natural substrate opens the door for studies not only limited to the enzyme itself but also to the properties of the de-gummed silk fiber with for instance electron microscopy [26].

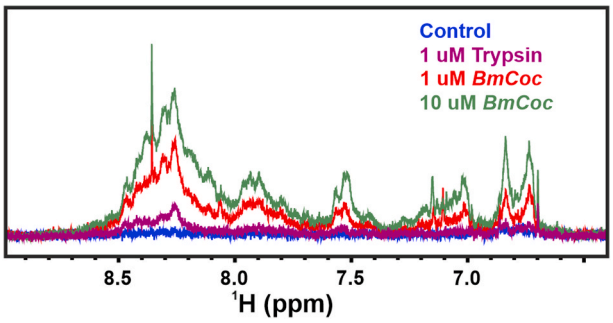

Fig. 8. In situ assay of cocoon softening by BmCoc. Proteolytic activity of BmCoc on *B. mori* cocoons was detected with ^{1}H NMR spectroscopy. The 600 MHz ^{1}H NMR spectra of peptide fragments released from the cocoon by the proteolytic activity of fully active En-BmCoc and trypsin (control) are shown. The proteolytic reaction was occurring for 24 h and thereafter NMR spectra were acquired for the soluble fraction. Proteolytic cleavage was performed on cocoons in 20 mM NaPi, 50 mM NaCl buffer, pH 6.0 and measured at 25 °C. The color coding of the spectra is as follows: red; 1 μM BmCoc, green; 10 μM BmCoc, purple; 1 μM trypsin, and blue; the negative control without enzyme.

5. Conclusions

Here we describe a stable and reproducible method for the production of large amounts of recombinant *Bombyx mori* cocoonase (BmCoc). By combining refolding under carefully controlled redox conditions with efficient *E. coli* expression, we were able to express and purify fully active BmCoc. Using this protocol, a yield of up to 15 mg of functional BmCoc per liter media was obtained. The function of the enzyme was demonstrated for model substrates, but also and importantly for *Bombyx mori* cocoons. A biophysical characterization revealed that the enzyme is stable, monomeric and well folded. However, a still unsolved problem with our method is the auto-proteolysis of the enzyme resulting in reduced yield of active enzyme. In the light of the results with the non-active, non-auto-proteolytic GAM variant, there might be a possibility to develop a variant with an N-terminal that is active but not auto-proteolytic.

An *E. coli* based system has a number of benefits in the characterization and potential industrial use of cocoonase. It is fast and gives a generally high yield, in addition, *E. coli* based systems are well studied and there are tools available for efficient genetic manipulation. For biophysical characterization using NMR, ^{15}N and/or ^{13}C labeled proteins are necessary, and an *E. coli* based system makes this feasible. This is because *E. coli* can be effectively grown in minimal media where carbon and nitrogen are provided in the form of single chemicals. In M9 medium, commonly used for *E. coli* cultivation, carbon and nitrogen are provided as glucose and ammonium chloride, respectively. Uniformly labeled variants of these chemicals are commercially available, at a relatively affordable price. Using labeled minimal media, the yield was not as good as for expression in rich, LB, media, but sufficient for NMR analysis.

BmCoc has large potential to improve the environmental profile and increase the economic benefit of the silk industry. A more gentle approach, using enzymes, to silk degumming might at the same time provide the medical and cosmetic industries with a potential raw material in the leftover sericins. Other enzymes than BmCoc have been used for enzymatic degumming [18–20], however they have not shown the same selectivity and efficiency [20]. Our investigation showed that trypsin was as efficient in cleaving sericin peptides as BmCoc, however, it was less effective when applied to the natural substrate, the cocoon. With the help of our new expression and purification system, the molecular properties of the BmCoc can be determined and also optimized, and using the well-developed tools for *E. coli* genetics the possibility of exploring these properties.

Author statement

Chanrith Phoeurk: Overall experimental work and writing of manuscript.

Ameeq Ul Mushtaq: Recording and analysis of NMR experiments.

Per Rogne: Assistance in experimental work and writing of manuscript.

Magnus Wolf-Watz: Conceptualizing and writing of the manuscript.

Acknowledgements

We thank The Protein Expertise Platform at Umeå University for

providing reagents and expertise for protein overexpression and purification. NMR experiments were performed at "NMR for life", Swedish NMR centre at Umeå University. This study was financially supported by grants from The Swedish Research Council (2017–04203), the Kempe Foundation (JCK-1417), and the Sweden-RUPP Bilateral Program (11599). The Silk Center, Royal University of Phnom Penh, Cambodia is acknowledged for *Bombyx mori* cocoons.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pep.2021.105919.

References

[1] G.H. Altman, F. Diaz, C. Jakuba, T. Calabro, R.L. Horan, J. Chen, H. Lu, J. Richmond, D.L. Kaplan, Silk-based biomaterials, Biomaterials 24 (2003) 401–416.
[2] P. Aramwit, T. Siritientong, T. Srichana, Potential applications of silk sericin, a natural protein from textile industry by-products, Waste Manag. Res. 30 (2012) 217–224.
[3] G. Freddi, R. Mossotti, R. Innocenti, Degumming of silk fabric with several proteases, J. Biotechnol. 106 (2003) 101–112.
[4] T. Gai, X. Tong, M. Han, C. Li, C. Fang, Y. Zou, H. Hu, H. Xiang, Z. Xiang, C. Lu, F. Dai, Cocoonase is indispensable for Lepidoptera insects breaking the sealed cocoon, PLoS Genet. 16 (2020), e1009004.
[5] N.M. Mahmoodi, F. Moghimi, M. Arami, F. Mazaheri, Silk degumming using microwave irradiation as an environmentally friendly surface modification method, Fibers Polym. 11 (2010) 234–240.
[6] S.C. Kundu, B. Kundu, S. Talukdar, S. Bano, S. Nayak, J. Kundu, B.B. Mandal, N. Bhardwaj, M. Botlagunta, B.C. Dash, C. Acharya, A.K. Ghosh, Nonmulberry silk biopolymers, Biopolymers 97 (2012) 455–467.
[7] X. Nirmala, K. Mita, V. Vanisree, M. Žurovec, F. Sehnal, Identification of four small molecular mass proteins in the silk of *Bombyx mori*, Insect Mol. Biol. 10 (2001) 437–445.
[8] Z. Dong, P. Zhao, C. Wang, Y. Zhang, J. Chen, X. Wang, Y. Lin, Q. Xia, Comparative proteomics reveal diverse functions and dynamic changes of *Bombyx mori* silk proteins spun from different development stages, J. Proteome Res. 12 (2013) 5213–5222.
[9] C.P. Singh, R.L. Vaishna, A. Kakkar, K.P. Arunkumar, J. Nagaraju, Characterization of antiviral and antibacterial activity of *Bombyx mori* seroin proteins, Cell Microbiol. 16 (2014) 1354–1365.
[10] S. Inoue, K. Tanaka, F. Arisaka, S. Kimura, K. Ohtomo, S. Mizuno, Silk fibroin of Bombyx mori is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6 : 6 : 1 molar ratio, J. Biol. Chem. 275 (2000) 40517–40528.
[11] Z. Dong, K. Guo, X. Zhang, T. Zhang, Y. Zhang, S. Ma, H. Chang, M. Tang, L. An, Q. Xia, P. Zhao, Identification of *Bombyx mori* sericin 4 protein as a new biological adhesive, Int. J. Biol. Macromol. 132 (2019) 1121–1130.
[12] Y. Takasu, T. Hata, K. Uchino, Q. Zhang, Identification of Ser2 proteins as major sericin components in the non-cocoon silk of *Bombyx mori*, Insect Biochem. Mol. Biol. 40 (2010) 339–344.
[13] Y.-Y. Jo, H. Kweon, J.-H. Oh, Sericin for tissue engineering, Appl. Sci. 10 (2020).
[14] L. Lamboni, M. Gauthier, G. Yang, Q. Wang, Silk sericin: a versatile material for tissue engineering and drug delivery, Biotechnol. Adv. 33 (2015) 1855–1867.
[15] S.C. Kundu, B.C. Dash, R. Dash, D.L. Kaplan, Natural protective glue protein, sericin bioengineered by silkworms: potential for biomedical and biotechnological applications, Prog. Polym. Sci. 33 (2008) 998–1012.
[16] M.N. Padamwar, A.P. Pawar, Silk sericin and its applications: a review, J. Sci. Ind. Res. 63 (2004) 323–329.
[17] S. Suwannaphan, E. Fufeungsombut, A. Promboon, P. Chim-anage, A serine protease from newly isolated *Bacillus sp*. for efficient silk degumming, sericin degrading and colour bleaching activities, Int. Biodeterior. Biodegrad. 117 (2017) 141–149.
[18] J. Kim, M. Kwon, S. Kim, Biological Degumming of silk fabrics with proteolytic enzymes, J. Nat. Fibers 13 (2016) 629–639.
[19] U. Ninpetch, M. Tsukada, A. Promboon, Mechanical properties of silk fabric degummed with bromelain, J. Eng. Fibers Fabr. 10 (2015) 69–78.
[20] P. Rodbumrer, D. Arthan, U. Uyen, J. Yuvaniyama, J. Svasti, P. Y. Wongsaengchantra, Functional expression of a *Bombyx mori* cocoonase: potential application for silk degumming, Acta Biochim. Biophys. Sin. 44 (2012) 974–983.
[21] E. Berger, F.C. Kafatos, R.L. Felsted, J.H. Law, Cocoonase 3. Purification, preliminary characterization, and activation of zymogen of an insect protease, J. Biol. Chem. 246 (1971) 4131–4137.
[22] R.L. Felsted, J.H. Law, A.K. Sinha, M.S. Jolly, Properties of the Antheraea mylitta cocoonase, Comp. Biochem. Physiol. B Comp. Biochem. 44 (1973) 595–609.
[23] F.C. Kafatos, C.M. Williams, Enzymatic mechanism for escape of certain moths from their cocoons, Science 146 (1964) 538–540.
[24] F.C. Kafatos, A.M. Tartakoff, J.H. Law, Cocoonase 1. Preliminary characterization of a proteolytic enzyme from silk moths, J. Biol. Chem. 242 (1967) 1477–1487.
[25] H. Fukumori, S. Teshiba, Y. Shigeoka, K. Yamamoto, Y. Banno, Y. Aso, Purification and characterization of cocoonase from the silkworm *Bombyx mori*, Biosci. Biotechnol. Biochem. 78 (2014) 202–211.
[26] S. Unajak, S. Aroonluke, A. Promboon, An active recombinant cocoonase from the silkworm Bombyx mori: bleaching, degumming and sericin degrading activities, J. Sci. Food Agric. 95 (2015) 1179–1189.
[27] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, Protein NMR Spectroscopy, Academic Press inc., San Diego, USA, 1996.
[28] J. Bogomolovas, B. Simon, M. Sattler, G. Stier, Screening of fusion partners for high yield expression and purification of bioactive viscotoxins, Protein Expr. Purif. 64 (2009) 16–23.
[29] L. Szilagyi, E. Kenesi, G. Katona, G. Kaslik, G. Juhasz, L. Graf, Comparative in vitro studies on native and recombinant human cationic trypsins - Cathepsin B is a possible pathological activator of trypsinogen in pancreatitis, J. Biol. Chem. 276 (2001) 24574–24580.
[30] L. Swint, A.D. Robertson, Thermodynamics of unfolding for Turkey ovomucoid 3rd domain- thermal and chemical denaturation, Protein Sci. 2 (1993) 2037–2049.
[31] E. Zakharova, M.P. Horvath, D.P. Goldenberg, Functional and structural roles of the cys14–cys38 Disulfide of bovine pancreatic trypsin inhibitor, J. Mol. Biol. 382 (2008) 998–1013.
[32] Y. Song, F. DiMaio, R.Y.-R. Wang, D. Kim, C. Miles, T.J. Brunette, J. Thompson, D. Baker, High-resolution comparative modeling with RosettaCM, Structure 21 (2013) 1735–1742.
[33] N.A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, O. Gileadi, Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study, Protein Expr. Purif. 59 (2008) 94–102.
[34] M. Kovermann, P. Rogne, M. Wolf-Watz, Protein dynamics and function from solution state NMR spectroscopy, Q. Rev. Biophys. 49 (2016).
[35] D.G. Rhodes, R.E. Bossio, T.M. Laue, Determination of size, molecular weight, and presence of subunits, in: second ed., in: R.R. Burgess, M.P. Deutscher (Eds.), Guide to Protein Purification, vol. 463, Elsevier Academic Press Inc, San Diego, 2009, pp. 691–723.
[36] U. Olsson, M. Wolf-Watz, Overlap between folding and functional energy landscapes for adenylate kinase conformational change, Nat. Commun. 1 (2010).
[37] K.E. van Holde, W.C. Johnson, P.S. Ho, Principles of Physical Biochemistry, Prentice Hall, Upper Saddle River, New Jersey, USA, 1998.
[38] A. Zhao, Y. Li, C. Leng, P. Wang, Y. Li, Inhibitory effect of protease inhibitors on larval midgut protease activities and the performance of *Plutella xylostella* (Lepidoptera: Plutellidae), Front. Physiol. 9 (2019).
[39] E.F. Jansen, A.K. Balls, The inhibition of b- and y-chymotrypsin and trypsin by diisopropyl fluorophosphate, J. Biol. Chem. 194 (1952) 721–727.
[40] E.F. Jansen, M.D.F. Nutting, R. Jang, A.K. Balls, Inhibition of the proteinase and esterase activities of trypsin and chymotrypsin by diisopropyl fluorophosphate: crystallization of inhibited chymotrypsin, J. Biol. Chem. 179 (1949) 189–199.
[41] N.D. Rawlings, A.J. Barrett, P.D. Thomas, X. Huang, A. Bateman, R.D. Finn, The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database, Nucleic Acids Res. 46 (2017) D624–D632.
[42] P. Rogne, T. Sparrman, I. Anugwom, J.-P. Mikkola, M. Wolf-Watz, Realtime ^{31}P NMR investigation on the catalytic behavior of the enzyme adenylate kinase in the matrix of a switchable ionic liquid, ChemSusChem 8 (2015) 3764–3768.
[43] J.P. Abita, M. Delaage, M. Lazdunski, Mechanism of activation of trypsinogen- role of 4 N-terminal apartyl residues, Eur. J. Biochem. 8 (1969) 314–324.