A semisynthetic peptide-metalloporphyrin responsive matrix for artificial photosynthesis

Sun, Z.

Chapter 5

Purification of recombinant variants of tCsmA

5.0 Abstract

The tCsmA was derived from the natural system by removing the flexible segments at the N and C-terminal, leaving the two functional central α-helix segments. Without a large fusion tag, tCsmA rapidly degrades. Two tags were studied to prevent the degradation of tCsmA, TRX and MBP. With TRX the stability can be increased, but the sample precipitates making further treatment difficult. With the MBP tag a strongly associated soluble tCsmA-MBP construct is obtained. While SEC can separate the tCsmA and MBP with the help of 6 M GuHCl, and the tCsmA denatures, a small fraction of tCsmA was recovered during the dialysis. This paves the way for further improvement towards obtaining recombinant variants of tCsmA.

5.1 Introduction

The harvesting of sunlight using artificial photosynthesis has been widely studied for decades. In nature, materials for light harvesting, charge separation, and catalysis are integrated into high-density matrix arrays with proteins.[1] High-density arrays of pigments inspired by nature have been explored for the design of artificial photosynthetic units.[2-5] In particular, the chemically robust ZnPP, a compound found in red blood cells, is often used for studies into artificial light harvesting.[6-8] To build a semi-artificial matrix around the ZnPP chromophore, we employ a truncated version of the natural 59-residue CsmA as the peptide motif to bind ZnPP and form a semisynthetic 2D scaffold. The semisynthetic tCsmA construct, containing 43 amino acids, was derived from the natural system by removing the flexible segments from the N and C-terminal which attach to the membrane of the chlorosome and the FMO complex, leaving the functional central α-helix.[9]

NMR, X-ray crystallography and EM are major techniques to study the structure of proteins.[10] As is shown in chapter 4, it is possible to obtain at moderate solution for the tCsmA-ZnPP complex with cryo-EM, however, it is not possible to obtain dynamic information with this method.[11-12] On the other hand, NMR can be used to obtain abundant microscopic information, *e.g.* angles, distances, coupling constants, chemical shifts, relaxation rates, to study the activation thermodynamics.[13] Generally, the isotope labelled of proteins is necessary for NMR. SPPS is too costly to obtain labelling protein.[14-15] However, via protein biosynthesis large or isotope labelled proteins are more easily obtained. Protein expression in *E.coli* has been one of the most popular means of producing recombinant proteins for over the past two decades.[16] *E.coli* is a well-established host that offers short culturing time, easy genetic manipulation and low cost media. In this chapter, our specific aim is using *E.coli* to express the tCsmA peptide.

5.2 Experimental section

5.2.1 Materials

Bacterial strains and plasmids

All cloning experiments were performed in the Leiden Cell observatory facility by its support staff. For the expression experiments two bacterial strains BL21(DE3) and BL21(DE3)/pLys (CamR) were used. Plasmid vectors pETMBP1a, pETMBP1b, pETTRX1a, pETTRX1b (EMBL, Hamburg) and pET16SK-SUMO2 are used for expression to get variants of tCsmA to ensure that at least one of them works as well as tCsmA from SPPS. All these plasmids carry the kanamycine resistance gene.[17] For production of tCsmA with a His tag plasmid pET3His was used (AmpR).[18] The tCsmA without a tag was also tested. The sequences are listed in Table 5.1 and include leftovers of the various tags that were used. Purification was performed with the buffers, proteins and protease inhibitors shown in Table 5.2.

Table 5.1 The sequence of recombinant variants of tCsmA

Name	Sequence
GAMG-tCsmA	GAMG-FTDILAAAGRIFEVMVEGHWETVGMLFDSLGKGTMRINRNAYG
M-tCsmA	M-FTDILAAAGRIFEVMVEGHWETVGMLFDSLGKGTMRINRNAYG
tCsmA	FTDILAAAGRIFEVMVEGHWETVGMLFDSLGKGTMRINRNAYG
G-tCsmA	G-FTDILAAAGRIFEVMVEGHWETVGMLFDSLGKGTMRINRNAYG
MG-tCsmA	MG-FTDILAAAGRIFEVMVEGHWETVGMLFDSLGKGTMRINRNAYG

Table 5.2 Buffers, proteins and protease inhibitors

Buffer	Components
NPI-10	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0
NPI-250	50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, pH 8.0
Sol Buffer	50 mM Tris-HCl, 6 M 1-propanol, 2 M Urea, 0.5 mM EDTA, pH 8.5[19]
MBP-Lysis Buffer	20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, pH 7.5
MBP-Column Buffer	20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, pH 7.5
MBP-Column Elution Buffer	MBP-Column Buffer with 10 mM Maltose
Phosphate Buffer (Kpi)	50 mM KPi buffer pH 7.9 or pH 7.5
Propanol-Buffer	6M 1-propanol, 50 mM KPi pH 7.9. Phosphate Buffer: 50 mM KPi pH 7.9 or pH 7.5
Phosphate Buffer	50 mM KPi pH7.9 or pH7.5
TEV Buffer	50 mM Tris-HCl, 0.5 mM EDTA, 14 mM 2-mercaptoethanol, pH 8.0
2x Osmosis buffer	40 mM Tris-HCl, 5 mM EDTA, pH 8.0
Sucrose buffer	40 % m/v sucrose in MQ water and 2x Osmosis buffer

For bacterial protein extraction, B-PER reagent in the KPi buffer and Inclusion Body Solubilisation Reagent (IBS) were purchased from Thermo Scientific.

Pefabloc SC (Sigma-Aldrich): a stock of 200 mM in MQ was stored at 4°C and used at 1 mM final concentration, which was added to prevent proteolytic degradation of peptides. Complete (Roche) 25× stock was made by dissolving 1 tablet in 2 mL MQ and stored at -20 °C. One mini Complete Ultra EDTA-free tablet (Roche) was dissolved in 10 mL Lysis Buffer and used immediately.

AcTEV protease 10 u/µL was purchased from In Vitrogen. For cleavage 50 units/mg MBP-CsmA fusion protein was used overnight at 22 °C in the TEV Buffer. Lysozyme stock was freshly prepared 50 mg/mL in the MBP-Lysis buffer.

ZnPP was bought from Frontier Scientific. Commercial tCsmA was synthesized by GL Biochem Ltd. (Shanghai).

Equipment

Cell-free *in vitro* protein synthesis kit (PurExpress) and Gibson Assembly Master Mix (Bioké); Slide-A-Lyzer Dialysis cassettes G2 (3 mL and 0.5 mL MWCO 3.5 kDa) and Pierce 96-well Micro-dialysis Plate (3.5K MWCO) (Thermo Scientific); Amicon Ultra-0.5ml and Ultra-2 mL centrifugal filters Ultracel-3K (Sigma-Aldrich); 1 mL Ni-NTA columns (Qiagen); Superdex 75 Increase 10/300 GL and 5 mL MBPTrap HP columns (GE Healthcare); DeNovix DS-11 nanodrop spectrophotometer (DeNovix); Size Exclusion Chromatography (SEC) was done using an Äkta Explorer 10S and all other columns were used on an Äkta Start purification system (GE Healthcare).

5.2.2 Methods

The cells were grow in LB medium with 50 µg/mL ampicillin at 37 or 18 °C in an orbital shaker at 150 rpm. The frozen glycerol stocks of transformed cells were diluted 1 : 100 to grow with different culture time. When OD600 reached about 1.0, expression of the opsin was induced by the addition of IPTG to a final concentration of 1 mM. The detail of culturing and further processes of different tCsmA fusion proteins are shown as below.

Untagged-tCsmA and His-tagged tCsmA
His-tagged tCsmA producing plasmid pET28a-tCsmA was transformed into *E.coli* strains BL21 (DE3) and BL21 (DE3)/pLys. Both strains were induced and cultured at 37 °C for 2h and 18 °C overnight.

SUMO-tagged tCsmA
pET16SK-SUMO2-tCsmA was derived from MBP1b-tCsmA by PCR using primers sumoFW (CAGGGTCTCGCATGGGCTTTACCGATATTCTGGCAGCAGC) and sumoREV (GAGGGATCCTTTCAGCAAAAAACCCCTCAAGACCC). The resulting sequence is MG-CsmA.

GAMG-tCsmA, tCsmA and G-tCsmA from MBP tags
pETMBP1a-tCsmA (Fusion 6xHis-MBP-TEV-tCsmA) and pETMBP1b-tCsmA (fusion MBP-6xHis-TEV-tCsmA) produce the GAMG-tCsmA protein after TEV cleavage. pETMBP1b-ΔGAM-tCsmA and pETMBP1b-ΔGAMG-tCsmA were derived from pETMBP1b-tCsmA using Gibson assembly following the manufacturers protocol resulting in G-tCsmA and tCsmA respectively.

The pellet from 1 L cells was lysed in 22 mL MBP-Lysis Buffer containing 1 mg/mL lysozyme, 1 mM Pefabloc and 2 mM DTT. After 10 minutes incubation and 4 minutes sonication the lysate was centrifuged for 30 min at 100,000 × *g* using a 70 Ti rotor (Beckman Coulter). The supernatant was loaded and washed on a 5 mL MBP-Trap column equilibrated with the MBP-Column Buffer at 4 °C. The protein was collected by isocratic elution using the MBP-Column Buffer containing 10 mM Maltose. The fractions collected were dialyzed against 1 L TEV Buffer containing 14 mM 2-mercaptoethanol at 4 °C for 2 hours and one more time for the overnight. The purified proteins were flash frozen with nitrogen and stored at -80 °C.

After TEV cleavage GuHCl was added to get a final concentration of 6 M. Then the sample was put on a rotary shaker for 90 minutes. The 480 µL samples were injected using a 0.5 mL loop on a Superdex 75 SEC column equilibrated in the TEV buffer containing 6 M GuHCl and eluted at 0.6 mL/min to collect 0.2 mL fractions. The collection was concentrated to a final volume of 0.5 mL by Amicon-2ml centrifugal filter to dialyze and obtain the precipitated tCsmA. The sample was dialyzed first using 0.5 mL Slide-A-Lyzer in 2 L 10 mM KPi buffer pH 5.0 at room temperature for 2 hours and one more time overnight. The final dialysis proceeded for 6 hours into the next day. The sample was centrifuged 30 minutes at 30,000 × *g*,

washed with 500 µL pH 5.0 KPi buffer one time, and 500 µL pH 7.9 KPi buffer for two more times. The final pellet of protein was freeze-dried overnight.

GAMG-tCsmA from TRX tag

pETTRX1a-tCsmA (fusion 6xHis-TRX-TEV-tCsmA) and pETTRX1b-tCsmA (fusion TRX-6xHis-TEV-tCsmA) were derived from pETMBP1a-CsmA using NcoI and KpnI. The resulting protein is GAMG-tCsmA. The inclusion bodies were purified from 500 mL IPTG induced cell culture and divided into 40 mL batches. The pellets were stored at -80 °C. The Osmotic lysis and B-PER lysis were both used for the purification from inclusion bodies.

For Osmotic lysis, the cell pellet was washed gently by cold 10 mL 10 mM pH 7.0 Tris buffer containing 30 mM NaCl. The supernatant was removed after 10 minutes centrifuge at 4000 rpm. The pellet was resuspended with 2mL sucrose buffer, and the supernatant was removed by 4 minutes centrifugation at 15,000 × *g* after 10 minutes incubation on ice. The pellet was resuspended by 2 mL Osmosis buffer. After 10 minutes incubation on ice, 80 µL complete stock was added to the solution. In the final step, after 4 minutes centrifugation at 15,000 × *g* the supernatant and pellet were separated.

For B-PER lysis, the pellet was dissolved in cold 5 mL B-PER containing Complete and was set on a rotary shaker for 10 minutes incubation on ice. The pellet was resuspended in 5 mL B-PER with 0.2 mg/mL lysozyme solution after 15 minutes centrifugation at 15,000 × *g* at 4 °C. The suspension was further incubated 5 min at room temperature after shaking vigourously for 1 min using a vortexer. 15 mL 1:10 diluted B-PER in MQ water was added and the suspension was centrifuged for 15 min at 15,000 × *g* at 4 °C after 1 minute. The pellet was resuspended in 20 mL 1:10 B-PER and centrifuged three times. In the final step, the inclusion body pellet (26 mg) was resuspended for 30 min with a rotary shaker in either 210 µL IBS with EDTA-free Complete or Sol-Buffer.

SDS-PAGE

Two types of SDS-PAGE gel were used in this chapter. Tris-glycine gel is used for separating protiens >30 KDa, whereas Tris-tricine gel is used for proteins <30 KDa. Aliquots of samples were diluted with SDS sample buffer and incubated for 30 min at 37 °C, then run on polyacrylamide gel at 20 mA for 2 hours. The gel was stained using Coomassie brilliant-blue G for 2 hours and

destained overnight in a solution which contains methanol, glacial acetic acid and water (volume ratio = 1 : 1 : 8).

5.3 Results and discussion

5.3.1 Protein expression

In Figure 5.1 the gels for expression of His-tagged tCsmA, and the untagged construct are shown. His tag is the smallest tag and although it may not be the most suitable for expression of tCsmA, it was tried to verify that expression cannot be realized by standard methods and more complicated approaches are warranted. The gels in Figure 5.1A and C are essentially identical, and in particular there is no evidence for the presence of the His-tCsmA. Likewise, SUMO-tag expression turned out to be unsuccessful (data not shown). In contrast, when the same plasmid for untagged-tCsmA was used in the *in vitro* protein synthesis system, protein expression could be observed (Figure 5.1B, Lane 2). This indicates that *in vivo* the small tCsmA peptide is probably digested by proteases. To overcome this problem, larger protective tags such as TRX and MBP tag were considered to obtain tCsmA *in vivo*.

An example of a tag that has been used successfully in the past for protein expression is TRX.[20] The most important advantage of this tag is enhanced solubility. In adddition this 12 kDa tag is much larger than His (1 kDa), which implies that it can protect the tCsmA better against digestion. As Figure 5.2A shows, pETTRX1b-tCsmA (tCsmA gene is behind TRX gene) and pETTRX1a-tCsmA (tCsmA gene is in front of TRX gene) both are able to produce the tCsmA. However, the protein obtained from the pETTRX1b-tCsmA plasmid was partially degraded, there are two protein protection bands which present TRX-tCsmA and TRX protein respectively. The degraded TRX protein fraction band is labelled with a star in Figure 5.2A Lane 2. The pETTRX1a-tCsmA on the other hand remains intact and was used for further steps.

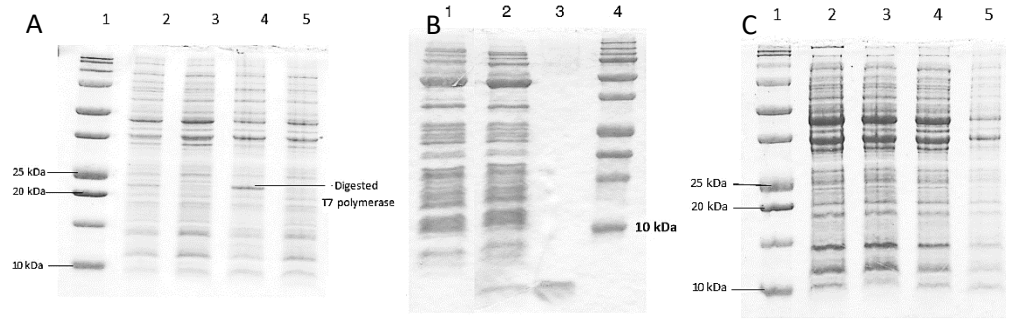


Figure 5.1 A : Protein gel showing the expression of His-tCsmA from BL21 His-tCsmA at 37 °C. Lane 1: Tricine gel containing MW marker; Lane 2: cell pellets from BL21 His-tCsmA 37 °C induced 0 h; Lane 3: cell pellets from BL21 His-tCsmA pLys 37 °C induced 0 h; Lane 4: cell pellets from BL21 His-tCsmA 37 °C induced 2 h; Lane 5: cell pellets from BL21 His-tCsmA pLys 37 °C induced 2 h.
B : Protein gel showing the *in vitro* translation tCsmA proteins. Lane 1: without plasmid; Lane 2: pET3-tCsmA; Lane 3: synthetic tCsmA peptide; Lane 4: MW marker.
C : Protein gel showing the expression of His-tCsmA from BL21 His-tCsmA at 18 °C. Lane 1: Tricine gel containing MW marker; Lane 2: cell pellets from BL21 His-tCsmA 18 °C induced 0 h; Lane 3: cell pellets from BL21 His-tCsmA pLys 18 °C induced 0 h; Lane 4: cell pellets from BL21 His-tCsmA 18 °C induced 2 h; Lane 5: cell pellets from BL21 His-tCsmA pLys 18 °C induced 2 h.

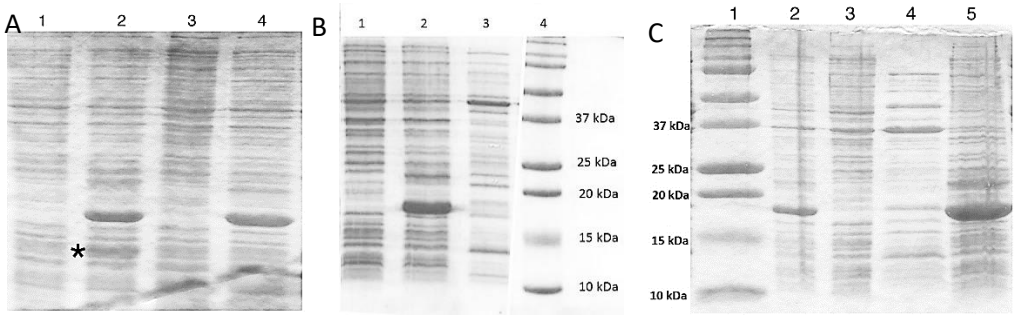


Figure 5.2 A : Protein gel showing the expression of TRX-tCsmA fusion protein. Lane 1: pETTRX1b-tCsmA -IPTG; Lane 2: pETTRX1b-tCsmA +IPTG; Lane 3: pETTRX1a-tCsmA -IPTG; Lane 4: pETTRX1a-tCsmA +IPTG.
B : Protein gel showing the effect of osmotic lysis. Lane 1: BL21(DE3)/pETTRX1a-tCsmA -IPTG; Lane 2: BL21(DE3)/pETTRX1a-tCsmA +IPTG; Lane 3: supernatant; Lane 4: MW marker.
C : Protein gel showing the effect of B-PER lysis. Lane 1: MW marker; Lane 2: BL21(DE3)/pETTRX1a-tCsmA +IPTG; Lane 3: BL21(DE3)/pETTRX1a-tCsmA -IPTG; Lane 4: supernatant; Lane 5: pellet.

The Figure 5.2B and C show the results of lysis. The fusion protein from the pETTRX1a-tCsmA plasmid only appears in the pellets obtained by centrifugation after cell lysis using B-PER and not in the supernatant (Figure 5.2C Lanes 4 and 5). Osmotic lysis gives similar results (Figure 5.2B, pellet was not shown). Thus, the TRX tag protects the tCsmA from degradation, however, it is not able to solubilize the peptide in the buffer.

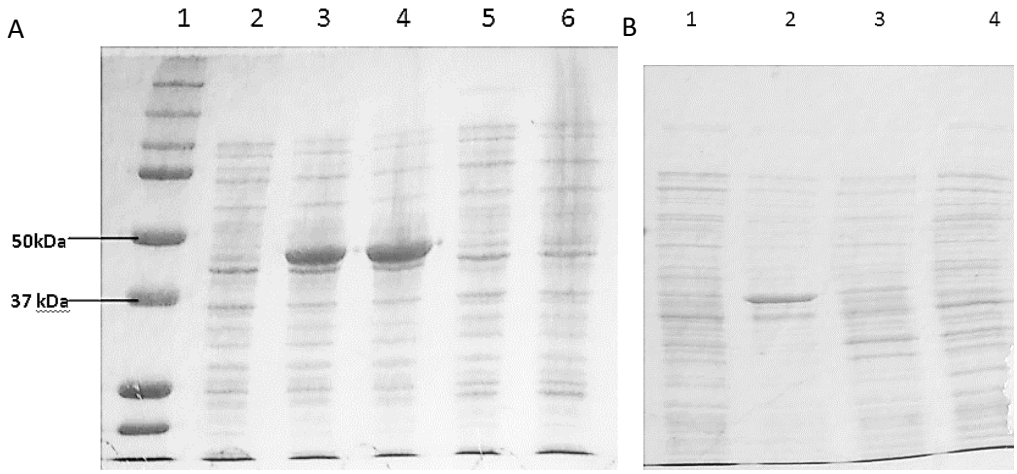


Figure 5.3 A : Protein gel showing the expression of MBP-tCsmA fusion protein in different strains. Lane 1 : MW marker; Lane 2, 3 and 4 : cell pellets from BL21 pET-MBP1a-tCsmA induced for 0 h, 1 h and 2 h; Lanes 5 and 6 : BL21 pET-MBP1a-tCsmA pLys induced for 0 h and 1 h respectively.
B : Protein gel showing the expression of MBP-tCsmA fusion protein in BL21 pET-MBP1a-tCsmA at different temperature. Lane 1 and 2 : cell pellets from BL21 pET-MBP1a-tCsmA induced at 37 °C for 0 h and 2h; Lane 3 and 4 : cell pellets from BL21 pET-MBP1-CsmA induced at 18 °C 0 h and the overnight respectively.

In Figure 5.3A, an overproduction band that is clearly present in Lane 3 and 4, corresponds with the molecular weight of MBP-tCsmA of 45 kDa. This indicates a successful expression of induced BL21(DE3) pET-MBP1a-tCsmA (tCsmA gene is behind MBP gene) in *E.coli* BL21 (DE3) strain at 37 °C. In contrast, there is no evidence for overproduction in the culture of BL21 (DE3)/pLys strain at 37 °C (Figure 5.3A Lane 5 and 6).

Two overproduction tests were performed for the BL21 (DE3) pET-MBP1a-CsmA strain at both 37 °C for 2h (Figure 5.3B Lane 1 and 2) and at 18 °C overnight (Figure 5.3B Lane 3 and 4). There appears to be no

overproduction band at 18 °C, and we take 37 °C as a suitable temperature for the culture of BL21 (DE3) pET-MBP1a-tCsmA.

5.3.2 Protein purification

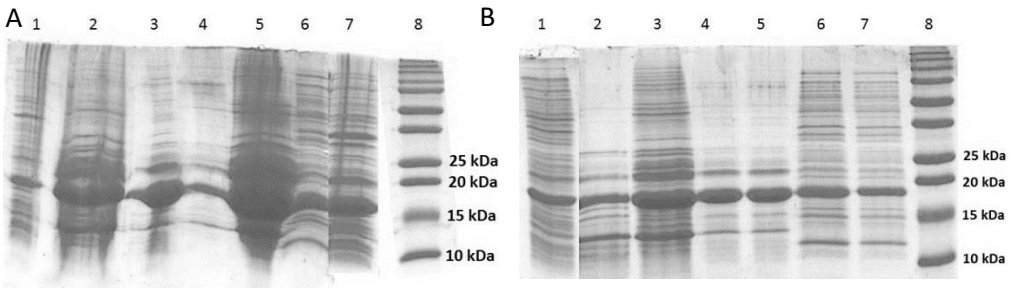


Figure 5.4 A : Protein gel showing the extraction of TRX-tCsmA fusion protein by B-PER lysis and osmotic lysis. Lane 1: BL21(DE3)/pETTRX1a-tCsmA cells +IPTG; Lane 2: supernatant after treatment of the inclusion body pellet with IBS buffer from B-PER lysis; Lane 3: supernatant after treatment the inclusion body pellet with Sol-Buffer from B-PER lysis; Lane 4: pellet after treatment of the inclusion body pellet with IBS buffer from B-PER lysis; Lane 5: pellet after treatment of the inclusion body pellet with Sol-buffer from B-PER lysis; Lane 6: supernatant after treatment of the inclusion body pellet with Sol-buffer from Osmotic lysis; Lane 7: pellet after treatment of the inclusion body pellet with Sol-buffer from Osmotic lysis; 8 : MW marker.
B : Protein gel showing the solublization of TRX-tCsmA fusion protein after extraction. Lane 1: BL21(DE3)/pETTRX1a-tCsmA cells +IPTG; Lane 2: supernatant of TRX-tCsmA after dialysis in the propanol buffer (solubilized by the IBS buffer from B-PER lysis); Lane 3: supernatant of TRX-tCsmA after dialysis in 2 M Urea, 50 mM KPi pH 7.5 (solubilized by IBS buffer from B-PER lysis); Lane 4: supernatant of TRX-tCsmA after dialysis in 6 M Urea (solubilized by the Sol buffer from the B-PER lysis); Lane 5: TRX-tCsmA from Lane 6 after dialysis in 2 M Urea, 50 mM KPi pH 7.5; Lane 6: supernatant of TRX-tCsmA after dialysis in the propanol buffer (solubilized by Sol buffer from Osmotic lysis); Lane 7: supernatant of TRX-tCsmA after dialysis in 2 M Urea, 50 mM KPi pH 7.5 (solubilized by the Sol buffer from Osmotic lysis); Lane 8 : MW marker.

To purify the tCsmA from TRX fusion protein, the first step is to extract the fusion protein in a soluble state to continue the cleavage. In Figure 5.4, most fusion protein is present in the supernatant after the treatment with IBS buffer (Figure 5.4A Lane 2) and Sol-Buffer (Figure 5.4A Lane 3). According to the width of protein bands, the IBS buffer dissolved more fusion protein

than the Sol-buffer, but also contained more contaminants (Figure 5.4A Lane 2 and 3).

Propanol buffer and urea were also used to solubilize the fusion proteins. According to the Figure 5.4B, the TRX fusion protein from the IBS buffer treatment remained mostly in the supernatant after dialysis in 2 M Urea, 50 mM KPi pH 7.5 (Figure 5.4B Lane 6). The fusion proteins from Sol buffer contain much fewer contaminants, but more protein precipitated during the dialysis (Figure 5.4A Lane 3 and Lane 6). After B-PER lysis more TRX fusion protein was recovered than for Osmotic lysis (Figure 5.4A Lane 2 and Figure 5.4B Lane 2). Although the dialysis extraction from the IBS and Sol buffer with urea both yield encouraging results, for TEV cleavage another step, involving dialysis against the KPi buffer, is needed. Unfortunately, here all fusion protein precipitated.

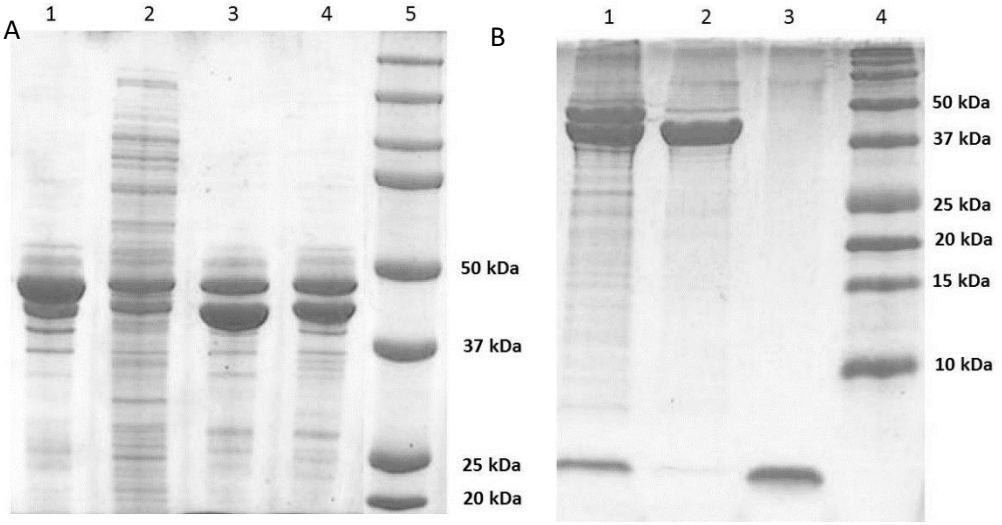


Figure 5.5 A : Protein gel showing the cleavage effect of MBP-tCsmA fusion protein. Lane 1: MBP1a-tCsmA before cleavage; Lane 2: MBP1b-tCsmA before cleavage; Lane 3: MBP1a-tCsmA after cleavage; Lane 4: MBP1b-tCsmA after cleavage; Lane 5: MW marker.
B : Protein gel showing the cleaved MBP1b-tCsmA fusion protein in Tricine gel to show the amount of tCsmA. Lane 1: pellet from MBP1b-tCsmA after cleavage; Lane 2: supernatant from MBP1b-tCsmA after cleavage; Lane 3: 1 µg commercial tCsmA; Lane 4: MW marker.

In contrast with the TRX tag, the MBP tag is able to solubilize tCsmA in the lysis buffer. MBP1a-tCsmA and MBP1b-tCsmA both show broad cleaved

protein bands which indicates a good cleavage efficiency (Figure 5.5A). Although a minor fraction of the tCsmA precipitated during cleavage, the cleaved tCsmA hardly shows bands below 10 KDa, probably most of the cleaved tCsmA still remains associated with the MBP protein in solution (Figure 5.5B Lane 1 and Lane 2).

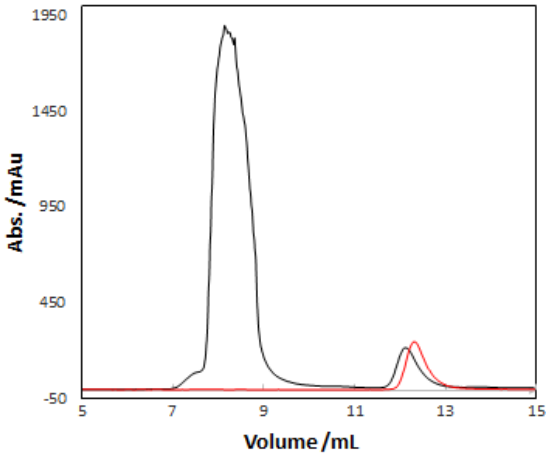


Figure 5.6 SEC elution profile (Superdex 75 10/300) of recombinant MBP-GAMG-tCsmA after TEV cleavage and commercial tCsmA peptide in the TEV Buffer containing 6 M GuHCl. red: 480 µg commercial tCsmA peptide (4.8 kDa), Ve=12.3 mL; Black: 4.8 mg MBP-GAMG-tCsmA treated with TEV protease. At Ve=8.5 mL the MBP tag (48 kDa) and non-cleaved MBP-GAMG-tCsmA (53.1 kDa) are eluting. At Ve=12.1 mL GAMG-tCsmA (5.1 kDa).

Using a Superdex-75 column with Size Exclusion Chromatography (SEC) the two proteins could be separated in the presence of GuHCl. In Figure 5.6 the SEC trace for the cleavage product is compared with benchmark data collected from a sample of synthetic tCsmA. This provides convincing evidence that the tCsmA can be separated from the MBP by SEC. However, since the running buffer of SEC contains 6 M GuHCl, another dialysis step is needed to collect the tCsmA fraction from the SEC buffer. This last step is very inefficient, with very low yield. When 300 µg of synthetic tCsmA is dissolved in the GuHCl buffer and dialyzed against the KPi buffer, ~ 160 µg is recovered after freeze drying the precipitated peptide and redissolving in the propanol buffer, while only 20 µg is recovered when a SEC step is performed in between. Since most protein was lost during the dialysis after SEC purification, most probably the SEC protocol, with purification in GuHCl

after SEC, leads to denaturation and unfolded protein, which passes through the membrane of the dialysis cassette.

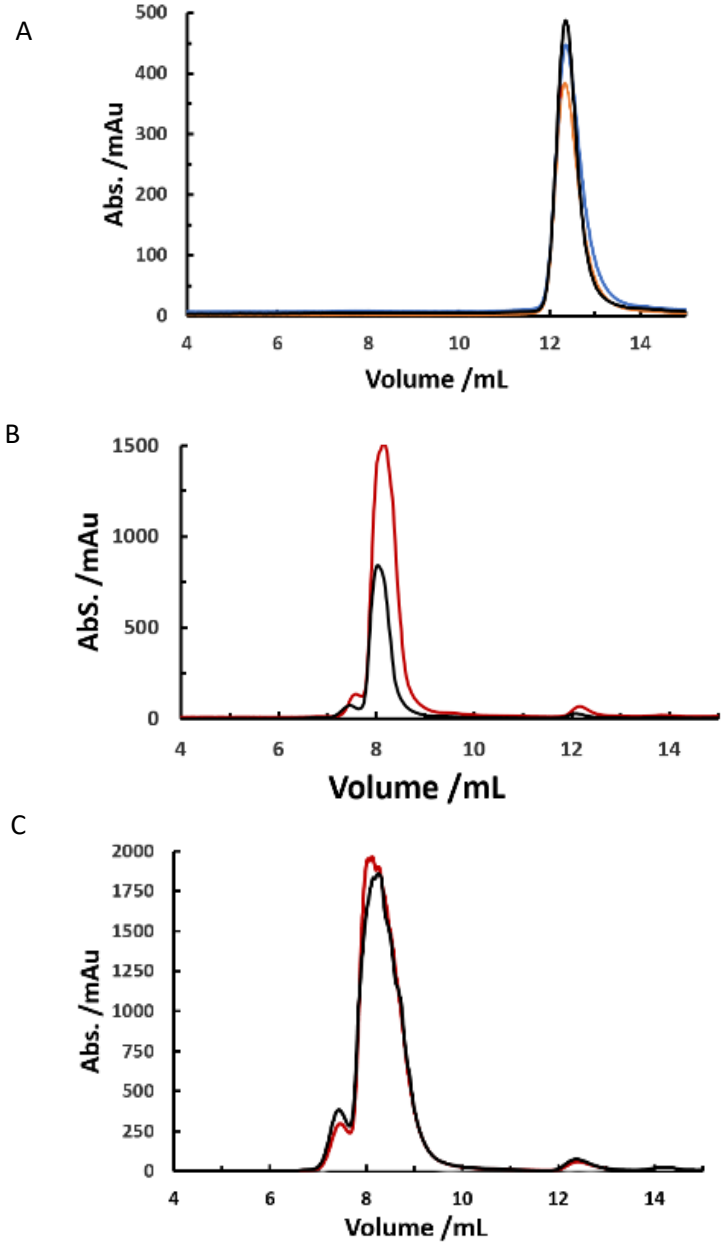
5.4 Conclusions

The tCsmA degrades during expression without the assistance of big fusion tags. TRX and MBP tags both prevent degradation of tCsmA. However, for the TRX tag, the inclusion bodies containing the fusion protein precipitate during expression. Although it is possible to extract the fusion protein in the SOL buffer, this is not a good environment for cleavage. Although the MBP tag is able to solubilize the tCsmA and the SEC can separate the tCsmA and MBP with the help of 6 M GuHCl, during the purification, the tCsmA was denatured too much, and the stretched tCsmA was lost in dialysis. According to UV analysis, a small fraction of purified tCsmA peptides can be recovered, around 10% of the total amount of protein after cleavage.

5.5 References

1. Ganapathy, S.; Oostergetel, G. T.; Wawrzyniak, P. K.; Reus, M.; Gomez Maqueo Chew, A.; Buda, F.; Boekema, E. J.; Bryant, D. A.; Holzwarth, A. R.; de Groot, H. J., *Proc Natl Acad Sci U S A* **2009,** *106* (21), 8525-30.
2. Yalamanchili, S.; Emmer, H. S.; Fountaine, K. T.; Chen, C. T.; Lewis, N. S.; Atwater, H. A., *Acs Photonics* **2016,** *3* (10), 1854-1861.
3. Hu, S.; Chi, C. Y.; Fountaine, K. T.; Yao, M. Q.; Atwater, H. A.; Dapkus, P. D.; Lewis, N. S.; Zhou, C. W., *Energ Environ Sci* **2013,** *6* (6), 1879-1890.
4. Gao, L.; Cui, Y. C.; Wang, J.; Cavalli, A.; Standing, A.; Vu, T. T. T.; Verheijen, M. A.; Haverkort, J. E. M.; Bakkers, E. P. A. M.; Notten, P. H. L., *Nano Lett* **2014,** *14* (7), 3715-3719.
5. Katterle, M.; Prokhorenko, V. I.; Holzwarth, A. R.; Jesorka, A., *Chemical Physics Letters* **2007,** *447* (4-6), 284-288.
6. Fry, H. C.; Garcia, J. M.; Medina, M. J.; Ricoy, U. M.; Gosztola, D. J.; Nikiforov, M. P.; Palmer, L. C.; Stupp, S. I., *J Am Chem Soc* **2012,** *134* (36), 14646-14649.
7. Luo, L.; Chang, C. H.; Chen, Y. C.; Wu, T. K.; Diau, E. W., *J Phys Chem B* **2007,** *111* (26), 7656-64.
8. Chromiński, M.; ó Proinsias, K.; Martin, E.; Gryko, D., *European journal of organic chemistry* **2013,** *2013* (8), 1530-1537.
9. Nielsen, J. T.; Kulminskaya, N. V.; Bjerring, M.; Linnanto, J. M.; Rätsep, M.; Pedersen, M. Ø.; Lambrev, P. H.; Dorogi, M.; Garab, G.; Thomsen, K., *Nature communications* **2016,** *7*, 12454.
10. Warren, B. E., *X-ray Diffraction*. Courier Corporation: 1990.
11. Milne, J. L.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S., *The FEBS journal* **2013,** *280* (1), 28-45.
12. Cheng, Y.; Grigorieff, N.; Penczek, P. A.; Walz, T., *Cell* **2015,** *161* (3), 438-449.
13. Coates, G. R.; Xiao, L.; Prammer, M. G., *NMR logging: principles and applications*. Haliburton Energy Services Houston: 1999; Vol. 344.
14. Synthesis—A, F. S. P. P., Practical Approach. Chan, WC: 2000.
15. Fields, G. B., *Peptide characterization and application protocols*. Humana Press: 2007; Vol. 386.
16. Miroux, B.; Walker, J. E., *Journal of molecular biology* **1996,** *260* (3), 289-298.
17. Studier, F. W., *Methods Enzymol.* **1990,** *185*, 60-89.
18. Chen, B. P.; Hai, T., *Gene* **1994,** *139* (1), 73-75.
19. Singh, S. M.; Sharma, A.; Upadhyay, A. K.; Singh, A.; Garg, L. C.; Panda, A. K., *Protein expression and purification* **2012,** *81* (1), 75-82.
20. Hirota, K., Nakamura, H., Masutani, H., & Yodoi, J. *Annals of the New York Academy of Sciences.* **2002,** 957(1), 189-99.

5.6 Appendix



SFigure 5.1 Size exclusion chromatography elution profile (Superdex 75 10/300) of recombinant MBP-GAMG-tCsmA after TEV cleavage and synthetic CsmA peptide in TEV Buffer containing 6 M GuHCl. A : commercial tCsmA with different amounts (350, 400 and 450 µL respectively); B : MBP1a-tCsmA after TEV cleavage (1 mg and 2 mg respectively); C : 3 mg MBP1b-ΔGAM-tCsmA after TEV cleavage (two times

repeat).

The accuracy of the SEC was confirmed by different experiments (SFigure 5.1). As SFigure 5.1A shows, when we injected different volumes of commercial tCsmA solution, the intensity is proportional to the volume injected. For the MBP-tCsmA sample, it also shows same result (SFigure 5.1B). The results are repeatable, the duplo sample showed the overlap in SEC spectrum (SFigure 5.1C). It indicates that the SEC is a precise tool to compare the amounts of proteins.