PURIFICATION AND CLEAVAGE OF FUSION PROTEIN CONTAINING THE PHOTOSYSTEM I SUBUNIT PSI-N USING AFFINITY CHROMATOGRAPHY AND TEV PROTEASE

Martin Bengtsson
1. Sammanfattning

PSI-N är en subenhet i proteinkomplexet PSI som är ett av fyra led i den elektrontransportkedja i tylakoidmembranen som fotosyntesen ger upphov till. Då funktionen hos PSI-N fortfarande inte är helt klarlagd finns stora potentiella vinster med att kunna isolera proteinet för att sedan kunna bestämma struktur och funktion. I den här rapporten beskrivs en metod för att kunna erhålla PSI-N från *E. coli* celler transformerade med plasmid innehållande psaN, genen som kodar för PSI-N. Odling och inducering av rekombinanta celler ger ett fusionsprotein där en region utgörs av PSI-N. Genom kromatografi på nickelpelare erhålls fusionsproteinet isolerat i vattenlösning och PSI-N kan sedan klyvas från fusionsproteinet av TEV proteas. Problem vid lysering av celler visade sig ge upphov till ett delvis degraderat PSI-N men försök med TEV proteas klyvning gav önskat resultat.
1. Sammanfattning ........................................................................................................................... 2
2. Introduction ........................................................................................................................................ 4
3. Materials and method .................................................................................................................. 6
   3.1 Expression of TEV protease ........................................................................................................ 6
   3.2 Purification of TEV protease ....................................................................................................... 6
   3.3 Expression of fusion protein containing PSI-N ................................................................. 6
   3.4 Purification of fusion protein containing PSI-N ........................................................................ 7
   3.5 Determining cleavage efficiency ........................................................................................... 7
4. Results .......................................................................................................................................... 7
   4.1 Expression and purification of TEV protease ........................................................................ 7
   4.2 Expression and purification of fusion protein containing PSI-N ........................................... 8
   4.3 Determining cleavage efficiency ............................................................................................ 9
5. Discussion ...................................................................................................................................... 10
6. Conclusions ................................................................................................................................... 12
7. Acknowledgements ..................................................................................................................... 13
8. References ................................................................................................................................... 13
9. Appendix .................................................................................................................................... 14
   9.1 Chromatogram from TEV protease purification ............................................................... 14
   9.2 Chromatogram from fusion protein purification ............................................................ 15
   9.3 Buffers .................................................................................................................................. 16
2. Introduction

Photosynthesis in green plants, algae and cyanobacteria is the biochemical process of generating carbohydrates from water and carbon dioxide with the use of energy harvested from light. It is a process that takes place in the thylakoid membranes where four protein complexes with the use of redox reactions transfers electrons from water to NADP. Water acts as the electron donor and is oxidized to O₂ and four H⁺ ions by photosystem II (PSII). The electrons are then shuttled through an electron transport chain consisting of photosystem II, cytochrome b6/f, photosystem I (PSI) and FNR. In PSI, solar energy is absorbed by a membrane-associated antenna complex. The energy is shuttled to the PSI reaction complex where the energy is used to translocate electrons from plastocyanin at the lumen side of the membrane, to ferredoxin at the stroma side of the membrane (Ben-Shem et al. 2003) (Amunts et al. 2007).

In plant PSI, 17 subunits have been discovered denoted PSI-A to PSI-L and PSI-N to PSI-P. Out of these aforementioned subunits PSI-A, -B and –C are directly involved in translocating electrons across the membrane. PSI-F, a transmembrane protein present on both sides of the membrane, is active in the binding of plastocyanin to PSI (Scheller et al. 2001). In 1990 a group led by Ikeuchi discovered two previously unknown proteins in PSI. (Ikeuchi et al. 1990) The smaller of the two, a 9kDa protein, was later identified through sequencing as a separate subunit of PSI and given the name PSI-N (Ikeuchi & Inoue, 1991).

The precise role of subunit N in PSI remains unclear and the production of an isolated form of the protein is vital in understanding the interaction between plastocyanin and PSI. To achieve this Rökaeus produced five different vector constructs with psaN, the gene coding for PSI-N, and of these one were used to express the protein as a fusion protein together with thioredoxin. Using the proteolytic enzyme factor Xa to cleave PSI-N from the fusion protein it was discovered that a cleavage site for the enzyme was present within the PSI-N sequence resulting in a degraded protein (Rökaeus, 2007). To solve this Ek & Halltorp produced two vector constructs with a cleavage site for TEV protease instead of factor Xa between PSI-N and the fusion protein. The cleavage site for TEV protease has a higher specificity in comparison to factor Xa reducing the risk for cleavage inside PSI-N (Ek & Halltorp, 2007).

The structure of TEV protease is similar to those of serine proteases, utilizing a “catalytic triad” of residues to catalyze peptide hydrolysis. Where serine proteases use the triad Ser-Asp-His TEV protease has cystein instead of serine as the nucleophilic residue. The cleavage site recognized by TEV protease is generally E-Xaa-Xaa-Y-Xaa-Q-(G/S) with cleaving occurring between Q and G or Q and S. Early results showed that the protease accepts many different residues in the positions P5, P4 and P2 but today the sequence E-N-L-Y-F-Q-S is considered to be the most efficient (Waugh, 2008). The production of TEV protease in E. coli was until recently hampered by low solubility. To solve this van den Berg and colleagues subjected the gene encoding TEV protease to directed evolution resulting in a mutant called TEVSH. The mutant increase the yield of purified and active TEV protease by five times compared to its parental gene and the addition of a His6-tag provides an easy way of purification (van den Berg S et al. 2006).

In the primary structure of PSI-N there are four cystein residues present which gives the possibility of disulfide bonds within the tertiary structure. In 2006 Motohashi
and Hisabori found that in the presence of DTT, PSI-N gradually converts from an oxidized form to a reduced in a two step process, with concentrations higher than 250 µM DTT required for complete reduction of PSI-N. This was performed in thylakoids in the presence of the system responsible for reducing disulfide bonds in the chloroplasts (Motohashi & Hisabori, 2006). The typical reaction conditions as described by Waugh uses DTT to provide the protease with reducing power (Waugh, 2008). DTT is known to reduce disulfide bonds of proteins, and because of the possibility of disulfide bonds within the tertiary structure of PSI-N both DTT and glutathione should be tested in the reaction buffer in future studies. According to Waugh a redox buffer of glutathione/reduced glutathione should maintain the disulfide bonds, while providing enough reducing power for the protease to work (Waugh, 2008). As there should be no difference between DTT and glutathione in terms of protease activity only DTT was used here. The assumption that PSI-N form disulfide bonds was not supported by the publication of the PSI supercomplex structure at 3.4 Å resolution in 2007 by Amunts, Drory and Nelson (Amunts et al. 2007). It should be noted that their work involved some second guessing in regards to the structure of PSI-N as they modeled the known amino acid sequence of PSI-N into an electron density that could only be ascribed to this subunit. Until more structures at higher resolutions are published the possibility of disulfide bonds within PSI-N should not be excluded.

Ek & Halltorp produced two different constructs containing psaN, and transformed these into two different types of E.coli strains resulting in four expression systems named A, B, Ak and Bk. A and B differs in that A has an extra aminoacid (G) between the cleavage site and PSI-N. Ak and Bk differs from A and B in that Ak and Bk has been transformed into cells (BL21 (DE3)pLysS) which has an extra plasmid that codes for T7 lysozyme, which in turn represses the transcription of T7-RNA polymerase. T7-RNA polymerase synthesis is induced by IPTG and binds to the T7-Lac promoter on the vector to initiate transcription of the gene. This way transcription of genes that are believed to be toxic to the host cell can be regulated more closely. As the extra plasmid containing the gene for T7 lysozyme also codes for a protein which gives the host resistance to chloramphenicol, this should be present in the media to minimize the risk of losing the extra plasmid (Ek & Halltorp, 2007). From the four vector constructs by Ek & Halltorp, B was used to continue the work of isolating PSI-N. This was based on the assumption that the extra Glycine in construct A would not give any advantages in the process of expression and purification of PSI-N. Construct B would produce a protein more resembling PSI-N’s native structure without the extra G. Bk was discarded because no positive effects were seen over B when expressing the fusion protein.

Purification of the fusion protein will be achieved using affinity chromatography utilizing the incorporation of a histidine-tag (his-tag) in the fusion protein. Histidine, an amino acid, has a side chain consisting of an imidazole structure which at neutral pH has an affinity to bind to positively charged ions. The standard ion used here is Ni$^{2+}$, and a Ni$^{2+}$ column will allow the fusion proteins to bind while unwanted proteins are washed out. Adding a buffer with a high concentration of imidazole will then allow the fusion protein to be released resulting in a solution containing the purified fusion protein (Hengen, 1995).
3. Materials and method

3.1 Expression of TEV protease

Recombinant *E. coli* cells containing plasmids coding for the TEV protease mutant TEV$_{SH}$ were obtained from Göteborgs Universitet. Frozen cells were transferred to an agar plate with LB media containing ampicillin (100µg/ml) and chloramphenicol (170µg/ml) and incubated for 48 hours at 37°C. A clone was selected from the plate and transferred to 25 ml liquid LB media and incubated over night at 37°C and 175 rpm. 500 ml (2x250 ml) was inoculated with 10 ml of the resulting cell-culture and growth continued until OD$_{550}$ reached 0, 5 – 0, 8. Protein expression was induced by adding 1mM IPTG and the cultures were incubated for 20 hours at 20°C and 175 rpm. The cells were harvested by centrifugation at 6000xg for 10 minutes at 20°C (Beckman Coulter Avanti J-25, rotor JA-10) and the supernatant was discarded. The pellet was resuspended in 5 ml Bug Buster Master Mix (Novagen)/ gram of wet cellpaste and incubated for 20 minutes on a rotating platform at room temperature. The soluble fraction was obtained through centrifugation at 5445xg for 20 minutes at 4°C (Beckman Allegra 21R, rotor S4180) and stored at -20°C.

3.2 Purification of TEV protease

The soluble fraction containing TEV protease was centrifuged at 5445xg for 3 minutes at 4°C (Beckman Allegra 21R, rotor S4180) to remove possible particles. The supernatant was then diluted with an equal amount of binding buffer (see Appendix 8.3) before being loaded on a Ni$^{2+}$-Sepharose column (GE Healthcare Histrap 5 ml FF) and washed with 100ml of binding buffer. TEV protease was then eluted with 25 ml of elution buffer (see Appendix 8.3) and 2mM DTT and 10mM EDTA was directly added. The 10 ml of the eluate was dialyzed against 1L 50 mM Tris-Hcl pH 8, 2 mM EDTA and 2mM DTT before A$_{280}$ was measured to 0,921 with a spectrophotometer. 10% of glycerol was added before the sample was divided into 1ml aliquots and frozen at -20°C.

3.3 Expression of fusion protein containing PSI-N

Cells containing plasmids coding for the fusion protein was transferred to agar plates with LB media containing ampicillin (100µg/ml) and incubated over night at 37°C. A clone was selected from the agar plate and transferred to 25 ml liquid LB media and incubated over night at 37°C and 175 rpm. 20 ml of the resulting cell-culture was used to inoculate 1L (4x250 ml) of liquid LB media and growth was continued until OD$_{550}$ reached 0, 5-0, 8. Protein expression was induced by adding 1mM IPTG. Ek & Halltorp reported that 2 hours of incubation at 37°C and 175 rpm would be sufficient (Ek & Halltorp, 2007) but after testing the incubation time was altered to 4 hours (see chapter 5). The cells were harvested at 6000xg for 10 minutes at 20°C (Beckman Coulter Avanti J-25, rotor JA-10) and the supernatant was discarded. The pellet was resuspended in 5 ml Bug Buster Master Mix (Novagen)/gram of wet cellpaste and incubated for 20 minutes on a rotating platform at room temperature. The soluble fraction was obtained through centrifugation at 5445xg for 20 minutes at 4°C (Beckman Allegra 21R, rotor S4180) and stored at -20°C.
3.4 Purification of fusion protein containing PSI-N

The soluble fraction containing the fusion protein was centrifuged at 5445xg for 3 minutes at 4°C (Beckman Allegra 21R, rotor S4180) to remove possible particles. The supernatant was diluted with an equal amount of binding buffer (see Appendix 8.3) before being loaded on a Ni²⁺-Sepharose column (GE Healthcare Histrap 5 ml FF) and washed with 100 ml of binding buffer. The fusion protein was eluted with 25 ml of elution buffer (see Appendix 8.3) and 8 ml of the eluate was dialyzed against 1L 50 mM Tris-HCl buffer for 4 hours before Aₒ₈₀ was measured to 2.253 with a spectrophotometer.

3.5 Determining cleavage efficiency

In order to find efficient cleavage conditions a series of tests was set up. In all tests the cleavage was performed in 50mM Tris-HCl pH 8.0, 0.5 mM EDTA and 1mM DTT at 4°C, and fusion protein was added to 80 µM. TEV protease was added in two different concentrations 0.25 and 2.5 µM giving a OD₂₈₀ ratio of 100:1 and 10:1 between fusion protein and TEV protease. Samples were taken after 2, 6 and 18 hours and evaluated using electrophoresis. This led to a final test with two different configurations at two different concentrations, named R1, K1, R2 and K2. The OD₂₈₀ ratio was changed to 3:1 (2) and 5:1 (1), cleavage performed for 10 (2) and 20 (1) hours respectively. Both configurations were tested at room temperature (R) and at 4°C (K).

Samples were diluted 1:1 in Tris-Tricine sample buffer (see appendix 8.3). A ladder with known molecular standards (Polypeptide SDS-PAGE Molecular Weight Standard, Bio-Rad) was used to identify proteins. One sample of the purified fraction and one sample of purified TEV protease were also prepared to ease evaluation of the gels. The samples and standard was loaded onto a 10-20% Tris-Tricine/Peptide Ready Gel (Bio-Rad) and the gel was run at 30 V for 30 minutes during loading and then at 100 V for 105 minutes to achieve desired separation. The proteins were fixed in 10% (v/v) acetic acid and 40% (v/v) methanol at room temperature for 30 minutes on a rotating platform. The gel was stained in Bio-Safe Coomassie G-250 Stain (Bio-Rad) for one hour and destained for two hours in distilled water. The gel was then stored in transparent plastic and digitized by a computer scanner.

4. Results

4.1 Expression and purification of TEV protease

IPTG induction of TEV protease proved successful and electrophoresis showed a clear band not present in samples taken before induction (data not shown). The chromatogram showed a peak representing TEV protease during elution. TEV protease appeared as a single band at approx. 30 kDa after purification. No other bands could be detected leading to the assumption of a relatively high purity.
4.2 Expression and purification of fusion protein containing PSI-N

The change in incubation time from two to four hours (see chapter 3.3 and 5) proved successful resulting in a band showing up after induction that is not present in samples taken before adding IPTG (Figure 1). The chromatogram showed a significant peak representing the fusion protein during elution (Appendix 8.2). When evaluating the eluate with SDS-PAGE, multiple bands indicated that it was not pure with a band showing a protein with nearly the same molecular weight as the fusion protein and at high concentration giving the impression of a double band at approx. 30 kDa. Reviewing the gels from earlier results showed that the double band was present before running the chromatograph (Figure 2). In order to determine if the bands were present already after induction of recombinant cells a new electrophoresis was performed. In the first electrophoresis the sample was too concentrated giving blurry bands in the 30 kDa range. This was solved by diluting the original sample 1:3, 1:9 and 1:27. A sample of the eluate showing the double band was also loaded for reference (Figure 3).

Figure 1. Gel showing from left to right; molecular weight standard, IPTG induced cell culture and cell culture before induction. A band of approx. 30 kDa is visible in the induced sample but is lacking in the sample taken before induction, indicating a positive result from induction.

Figure 2. Gel showing from left to right; molecular weight standard, purified fraction and Bug Buster Master Mix fraction. A double band is clearly visible in both samples indicating degradation.
Figure 3. Gel comparing the purified fraction (left) with IPTG induced cell culture at different concentrations. The double band appearing in the purified fraction is not visible in the diluted sample, suggesting that degradation occurs at some point after lysis.

4.3 Determining cleavage efficiency

The TEV protease activity showed no major differences in the four different tests. Cleavage with $\text{OD}_{280}$ ratio 5:1 and 3:1 did not show any advantage when performed at 20°C as opposed to 4°C. No advantage was seen in changing the concentrations of TEV protease or the time for which the protease was allowed to work. Examining the gel containing all four test configurations, purified PSI-N and purified TEV show that the multiple bands discovered after the purification of fusion protein containing PSI-N (see chapter 4.2) all disappear after cleavage. A new band at approx. 20 kDa appears in all test configurations as well as three smaller bands in the range 10-3 kDa (Figure 4).
5. Discussion

The expression and purification of TEV protease proved to be fairly straightforward. The protocol published by van den Berg (van den Berg et al. 2006) was reliable and results showed a high level of reproducibility. During purification some minor problems were experienced due to the fact that although the TEV$_{SH}$ mutant produce a five-fold increase in the yield of purified protease compared to wild-type, relatively large volumes of cell extract were required to avoid the eluate being diluted and as such difficult to handle. The previous assumption that the purified fraction of TEV protease had a high level of purity was proven to be premature, as bands showing contamination was visible during the evaluation of cleavage efficiency.

Expression of fusion protein containing PSI-N proved more of a challenge than the expression of TEV protease. Early tests in smaller scale (5-25 ml) showed high reproducibility compared to Ek & Halltorp (Ek & Halltorp, 2007). As the cell-cultures were scaled up to volumes of 250 ml, IPTG induced expression proved difficult. Evaluation using SDS-PAGE showed no dominant band in the 30 kDa region and testing the efficiency of IPTG induction hinted at the induction time being a factor. A test of different induction times and IPTG concentrations gave credibility to this assumption as induction for four hours gave a dominant band corresponding to PSI-N but induction for two hours did not. Inducing with higher IPTG concentrations did not give any noticeable

Figure 4. The gel showing results from cleavage of fusion protein. From left to right; Molecular standard, the four cleavage configurations, a purified fraction of fusion protein, and a purified fraction of TEV protease. The cleavage configurations are from left to right: 5:1 ratio for 20 hours at room temperature, 5:1 ratio for 20 hours at 4°C, 3:1 ratio for 10 hours at room temperature and 3:1 ratio for 10 hours at 4°C.
increase in fusion protein in comparison to total protein content. Because of this IPTG expression was from here on performed for four hours instead of two with 1mM IPTG.

Purification of the fusion protein containing PSI-N resulted in multiple bands showing up during SDS-PAGE. This led to the assumption that at some point after IPTG induction the PSI-N region of the fusion protein is degraded. Indeed SDS-PAGE gels with fractions from the chromatograph and cell content after IPTG induction but before lysis using Bug Buster Master Mix show that the bands appear after lysis. When comparing this result with the results presented by Ek & Halltorp it seems they did not experience this. Our methods for lysis differ only in one aspect, the temperature at which the soluble fraction containing the fusion protein is harvested by centrifugation. Where they performed centrifugation at 20°C I choose to perform it at 4°C. Although it can not be excluded, this does not seem to be the factor behind the degradation of PSI-N. If the degradation was caused by protease from *E. coli* the lower temperature used in the method presented here should inhibit protease activity rather than promoting it. The relative short time for possible proteases to cleave the fusion protein should in my opinion not give rise to such a high concentration of degraded fusion protein compared to intact. The change in induction time does not seem to be a factor as the double band can not be detected before lysis. It should be pointed out that more tests are required to give clear answers regarding large scale expression and lysis. The results shown here indicate that a different approach to achieving a soluble fraction containing the fusion protein might be necessary. Adding protease inhibitors when using Bug Buster Master Mix could provide answers regarding possible activity of *E. coli* proteases during lysis. A change of method from chemical to mechanical lysis using a French press is another possibility. If these two approaches should not solve the problem of degradation expression at smaller scales has so far been reliable.

Cleavage tests gave more credibility to the assumption of the double band present after lysis being the result of a degraded fusion protein. Comparing the samples of TEV protease processed fusion protein to the purified fraction containing the fusion protein show that four bands at approx. 30 kDa are cleaved by the protease (Figure 4). This was somewhat confusing as this pointed at even more degradation as the double band had suddenly become four bands. One possible explanation to this is that the two gels have different resolutions giving the four different bands the appearance of being only two during the first electrophoresis (Figure 2). Another explanation could be that the samples in the later electrophoresis have become more degraded as a result from freezing and thawing. The problems with degraded PSI-N experienced by both Rökeus and me could also be ascribed to PSI-N being a protein that is easily degraded possibly due to a weak peptide bond somewhere in the primary structure. Clues to support this should be evident in the gels produced by me and Rökeus to evaluate cleavage efficiency. In my gel two bands that cannot be ascribed to either PSI-N or the remaining fusion protein appear after cleavage, one at approx. 6 kDa and another at approx 3.5 kDa. Rökeus report that she found one band with a mass between 6.5 and 3.5 kDa. Although differences in size can be explained by small differences in method or equipment, the different number of bands representing possible degradation products point towards the problems experienced by me and Rökeus has different causes.
After processing by TEV protease at least four new bands appear, one at approx. 20 kDa and three in the range 10 -3 kDa. A purified form of PSI-N should after cleavage have a molecular weight of 9 kDa and the appearance of a band around is 10 kDa is encouraging. The two smaller bands showing up at approx. 6 and 3,5 kDa could be ascribed to degraded PSI-N, either as two parts which together would make up intact PSI-N, or as part of more degraded forms of PSI-N. The fourth band at approx. 20 kDa would in this interpretation be the remains of the fusion protein. Novagen reports that the fusion protein is 162 amino acids in size. (Xa/Lic Cloning kits manual) A rough calculation of the molecular size using the estimation of 110 Da per amino acid gives you a total size of approx. 19 kDa including the TEV protease recognition site. This corresponds well with the band at 20 kDa. The difference in molecular weight between the heavier bands in the purified fraction with the 20 kDa band also points to a size of approx. 10 kDa, the band that could be interpreted as PSI-N. The disappearance of the multiple bands present in the purified fraction and, the appearance of four bands at molecular weights that could be ascribed to intact PSI-N, degraded PSI-N and cleaved fusion protein are in my opinion evidence that the use of TEV protease to obtain an isolated form of PSI-N is possible.

6. Conclusions

The goal set up at the start of this project was to obtain PSI-N in an isolated form to allow for future in-depth studies of its structure and function. Although this was not achieved steps leading closer to the realization of isolated PSI-N were taken. In continuing work towards a complete method for expression, purification and isolation of PSI-N there are a number of parameters in my work that would benefit from being optimized. First and foremost, the problem of degradation must be solved. Both I and Ek & Halltorp before me have seen that this is possible and I am confident that whoever takes over from here will solve this. The work of purifying the fusion protein has been proven to be possible as well, although it cannot be regarded as complete. There is still a level of contamination in purified fractions and removal of contaminants will probably be possible once column wash and elution is optimized. In regard to processing of the fusion protein by TEV protease there is room for optimization as well. The different cleavage tests carried out here was setup to find out if cleavage was possible and because of this high concentrations of TEV protease were used. My recommendation is to start evaluation of adding TEV protease at OD_{280} ratios of 1:10 to 1:5 compared to fusion protein.
7. Acknowledgements

I would like to thank my supervisor Professor Lars-Gunnar Franzén, Örjan Hansson and Daniel Farkas at Göteborgs Universitet for providing the TEV<sub>Si</sub> clones and useful information, Helene Berglund at Karolinska Institutet for providing information concerning the TEV mutant, Kim Thelander and Jörgen Svensson at GE Healthcare for their help with technical difficulties, Jean Paul René Schnoor for proof reading, Louise Ek for providing strawberries and to my family for providing comfort.

8. References

Xa/Lic Cloning kits manual. Novagen user protocol TB 184 Rev. D 0904
9. Appendix

9.1 Chromatogram from TEV protease purification
9.2 Chromatogram from fusion protein purification
9.3 Buffers

LB media:
1% (w/v) peptone
0.5% (w/v) yeast extract
1% (w/v) NaCl

SDS loading buffer:
50 mM Tris-HCl, pH 6.8
2% (w/v) SDS
10% (v/v) glycerol
0.1% bromphenol blue
100 mM dithiothreitol

SDS running buffer:
25 mM Tris
250 mM glycine
0.1% (w/v) SDS

Binding/Wash buffer:
20 mM sodium phosphate, pH 7.4
0.5 M NaCl
40 mM imidazole

Elution buffer:
20 mM sodium phosphate, pH 7.4
0.5 M NaCl
0.5 M imidazole