6 Applying the New Cloning Strategy: Production of a Specifically Designed PAA Diblock Containing a Zwitterionic Tail Block*

(* contains information submitted for publication, see Ref.1)

6.1 Overview

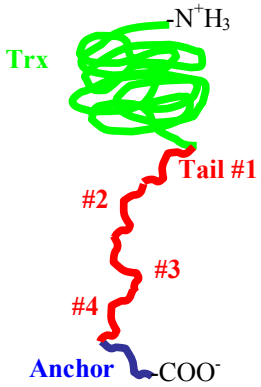


Figure 6.1 Schematic of target brush-forming PAA design.

In Chapter 5 a new universal cloning strategy for the production of specifically designed surface-active PAA's was introduced. Here, we apply this strategy to the production of a unique surface-active PAA designed to form brush layers on aluminum oxide surfaces. This chapter discusses the procedure taken in assembling the gene encoding for a PAA diblock consisting of a long zwitterionic tail block and a short acidic anchor block. Once the gene was constructed, the PAA was expressed with a HP-Thioredoxin (Trx) fusion tag in recombinant *E. coli* (see Figure 6.1). After purification, the Trx was removed and PAA product was isolated and its composition was confirmed. At the conclusion of this work, we will have demonstrated the application of the new cloning strategy towards the production of a truly unique PAA with a specified monomer sequence and composition. A manuscript submitted to *Biomacromolecules* based on the work presented in this chapter is located in Appendix A.

6.2 Specific Aim #1: *Design a PAA sequence that will self-assemble on aluminum oxide surfaces to form brush layers, and produce this PAA in genetically engineered E. coli*

Our second generation PAA was designed with several improvements in the anchor and tail blocks. Our goal was to design a PAA that forms brush layers on aluminum oxide surfaces. Therefore, the anchor block must adsorb strongly to the aluminum oxide surface and the tail block must be sufficiently soluble in the solvent (*e.g.* water) and must not adsorb onto the surface. The following sections discuss the design of each block.

[1] Henderson DB, Davis RM, Ducker WA and Van Cott KE. "A new cloning strategy for producing high molecular weight brush-forming poly(amino acids)". *Biomacromolecules* (submitted 12/2004).

6.2.1 Improved Anchor Block

The isoelectric point of the aluminum oxide surfaces used in this work was determined to occur at pH ~8.5.[1] Thus, the surface has a net positive charge close to physiological pH (6-8). Previous work has shown that poly-glutamate and poly-aspartate homopolymers will adsorb strongly onto aluminum oxide surfaces in this pH range and can even displace non-ionic PAA's.[1] As with the Pro_mGlu_n diblock, the anchor was designed to consist of a sequence of 10 glutamate residues. The tail block was fused to the amino-terminus of the anchor block to keep the carboxy-terminus of the PAA located within the adsorbing block. A *'linker'* region was introduced between the tail and anchor blocks to provide additional flexibility. Glycine (-H side chain) is the smallest amino acid and is often found in flexible loop structures in naturally occurring proteins. Thus, poly-glycine linkers are often used in order to maximize the conformational freedom of the polypeptide backbone.[2] Pinaud *et al.* used a glycine-rich linker in their peptide design, separating an 'adhesive domain' and a 'bioactive domain'.[3] In our design, four glycine residues were added at the amino-terminus of the anchor block to act as a hinge that should allow for maximum tail extension during brush formation. Alanine was also included as a result of DNA sequences that were necessary during the cloning steps. The improved anchor block sequence is given in Figure 6.2.

Anchor PAA Block:

A G G G G E E E E E E E E E E A

Figure 6.2 Amino acid sequence of the improved anchor block ($AlaGly_4Glu_{10}Ala$) which includes a Gly_4 linker. (G = Gly; E = Glu; A = Ala).

6.2.2 Possible Universal Tail Block – Zwitterionic PAA Sequence

In order for the tail block to form an effective brush, it must be hydrophilic, have a net zero charge and have no affinity for the surface. We hypothesized that the ideal tail block sequence for a surface-active PAA that forms brush layers on metal oxide surfaces would be *zwitterionic*. In order for the tail block to be zwitterionic, it must consist of an equal number of cationic and anionic residues balanced at the smallest length scales over a wide pH range. This local balance of charges leads to a vanishing net Coulomb interaction between different parts of the polymer chain, but it leaves unbalanced dipole-

dipole interactions. The adsorption of zwitterionic polymers to colloidal particles has been considered in theoretical papers by two groups: Jonsson *et al.*[4] and Rubinstein *et al.*[5,6] The mechanism for adsorption is the separation of charge within the polymer to create a net dipole within the polymer. In the limit of alternating charges, adsorption is very low due to the inability of the charges to separate.

We have designed a zwitterionic tail block in which the charges are *approximately* alternating. Thus, the charges cannot separate and therefore, the tail block should not adsorb onto aluminum oxide surfaces. Charge separation should be further hindered in our design by the small electric field emanating from the surface after the adsorption of the acidic anchor block. Thus, we predict that a zwitterionic PAA could serve as a universal tail block for all metal oxide surfaces.

Zwitterionic PAA sequences were designed consisting of triplets containing one cation (Lys or Arg), one anion (Glu or Asp) and one small neutral residue (Gly or Ser). The small, uncharged glycine and serine residues should increase flexibility in the polymer and variability in the DNA sequence. It has been shown that repetitive DNA sequences are not translated efficiently and are often deleted.[7,8] Variability was also introduced in our design by using *four* distinct tail 'modules' having the same charge structure yet with different amino acid and DNA sequences. By inserting the genes encoding for these modules sequentially, the amount of repetitiveness in the DNA sequence was minimized. Each module also begins and ends with glycine residues for two purposes. First, the glycine residues increase the flexibility between PAA blocks, which promotes tail extension. Second, and most importantly, the DNA sequences encoding for these glycines are necessary for the application of the cloning strategy used to construct the PAA gene. The amino acid sequences of the four 26-mer tail modules are given in Figure 6.3. The net charge on each module is calculated to be zero within the pH range of 5-9.

Secondary structure prediction algorithms on ExPASy predicted the anchor block and each tail block to have a random coil structure. Theory and experiment suggest that, for a diblock copolymer, the molecular masses of the anchor and tail block should be in the ratio 1:20 for maximum tail extension.[9,10] As mentioned previously during the Pro_mGlu_n work, it was not possible to accomplish this using conventional cloning

methods. Therefore, we determined that in order to produce a high molecular weight PAA that approached this optimum ratio, the gene encoding for the tail block must be constructed by linking together smaller DNA modules sequentially. This process can be repeated until a gene encoding for a tail block of sufficient length and composition was achieved. We accomplished this using the new cloning strategy introduced in Chapter 5. The following sections discuss how the cloning strategy was employed to construct the gene encoding for a unique PAA that was subsequently expressed and purified.

Zwitterionic PAA Block #1:

G	K	S	E	G	E	K	R	S	D	K	E	G	K	E	G	G	D	R	K	E	S	D	G	R	G
0	+	0	-	0	-	+	+	0	-	+	-	0	+	-	0	0	-	+	+	-	0	-	0	+	0

Zwitterionic PAA Block #2:

G	K	E	G	E	K	G	E	G	R	G	R	E	R	G	D	G	D	R	G	K	E	K	D	G	G
0	+	-	0	-	+	0	-	0	+	0	+	-	+	0	-	0	-	+	0	+	-	+	-	0	0

Zwitterionic PAA Block #3:

G	D	R	G	K	D	S	S	K	D	R	D	S	S	D	R	G	E	K	S	D	K	D	K	G	G
0	-	+	0	+	-	0	0	+	-	+	-	0	0	-	+	0	-	+	0	-	+	-	+	0	0

Zwitterionic PAA Block #4:

G	E	S	R	D	K	G	G	E	K	R	D	S	D	R	S	E	R	S	D	R	S	S	E	R	G
0	-	0	+	-	+	0	0	-	+	+	-	0	-	+	0	-	+	0	-	+	0	0	-	+	0

Figure 6.3 Amino acid sequence and charge structure of each of the four zwitterionic tail blocks. G = Gly; S = Ser; E = Glu; D = Asp; K = Lys; R = Arg.

6.3 Application of Cloning Strategy #1

As discussed in the previous section, a PAA diblock consisting of a long zwitterionic tail block and a short acidic anchor block should self-assemble on aluminum oxide surfaces to form brush layers. For this work, we wanted a PAA consisting of 5-10

mol% anchor, which would require a ~100-mer tail block. This would be a significant improvement over the $Pro_{39}Glu_{10}$ diblock designed in Chapter 3 (~20 mol% anchor). Due to the limitations of the conventional cloning methods used in Chapter 3, we were unable to build a poly-proline tail block long enough to achieve the diblock composition necessary for optimum brush formation. However, through the application of **Cloning Strategy #1** (Section 5.4.2), we were able to assemble a gene encoding for a much longer tail block.

Choosing Expression Vector

As stated in Chapter 5, we found that the commercially available expression vector pBAD/Thio-TOPO® (Invitrogen) was compatible with our cloning strategy. Using this vector, recombinant proteins can be expressed by *E. coli* with an N-terminal HP-Thioredoxin (Trx) fusion tag. The vector contains the *ara*BAD promoter, which results in a low basal level expression but very high levels of expression when the inducer (L-(+) arabinose) is present. [11,12] The HP-Thioredoxin tag simplifies purification and offers an epitope for immunodetection using a commercially available antibody (Anti-Thio™). Most importantly, the vector does not contain any internal *Sfo*I recognition sites, which is essential for the application of our cloning strategy. Finally, TOPO® TA cloning facilitates the insertion of the initial DNA module.

Insertion of the Initial DNA Module: Introduction of a Unique Sfo*I *Cloning Site

The sequence of the initial DNA module inserted into the expression vector is given in Figure 6.4 and encodes for the PAA sequence given in Figure 6.5. This PAA sequence was designed based on "Zwitterionic PAA Block #1" given in Figure 6.3. A unique *Sfo*I recognition site (5' *ggc*|*gcc* 3') was incorporated into the module's DNA sequence near the 3' end. This allowed for subsequent insertions of DNA modules *downstream* of this initial module. The base pairs encoding for a hydroxylamine chemical cleavage site (Asn-Gly) were included at the 5' end of the module to allow for the removal of the N-terminal fusion tag after PAA expression. Finally, a STOP codon was included at the 3' end of the module, which signals to the protein translation machinery of the host organism to stop making the polypeptide strand at that point.

Tail DNA Module #1 was constructed by annealing two overlapping single-stranded synthetic oligonucleotides. After hybridization, Sequenase™ was added to begin the 'fill-in' reaction. 3' A-overhangs were added to the double-stranded product with *Taq* DNA polymerase (for TOPO® TA cloning). A schematic of the module construction and its insertion into the expression vector is located in Figure 6.6. Orientation of the module was confirmed by restriction double digest (using *Nco*I and *Sfo*I) and DNA sequencing (see Section 6.11.1). The resulting vector, denoted as pTrxT$_1$, now has the unique *Sfo*I site required for the application of the new cloning strategy. The only requirement is that the *Sfo*I site must be regenerated after each insertion.

Tail DNA Module #1:

5'- aac ggc aag agc gaa ggc gag aag cgc agc gat aaa gag ggt aaa gag ggc ggt gac cgc aaa gag agc gat ggt cgc ggc | gcc tga -3'

Figure 6.4 Tail DNA Module #1 sequence. Underlined base pairs represent the *Sfo*I recognition site and "|" represents the cut site.

PAA Sequence (encoded by Tail DNA Module #1):

N G K S E G E K R S D K E G K E G G D R K E S D G R G A *

Figure 6.5 Amino acid sequence encoded by Tail DNA Module #1 . "NG" represents Asn-Gly hydroxylamine cleavage site; 'blue' represents acidic residues, E=Glu, D=Asp; 'red' represents basic residues, R=Arg, K=Lys; "G A" represents residues encoded by *Sfo*I cloning site; and "*" represents STOP codon.

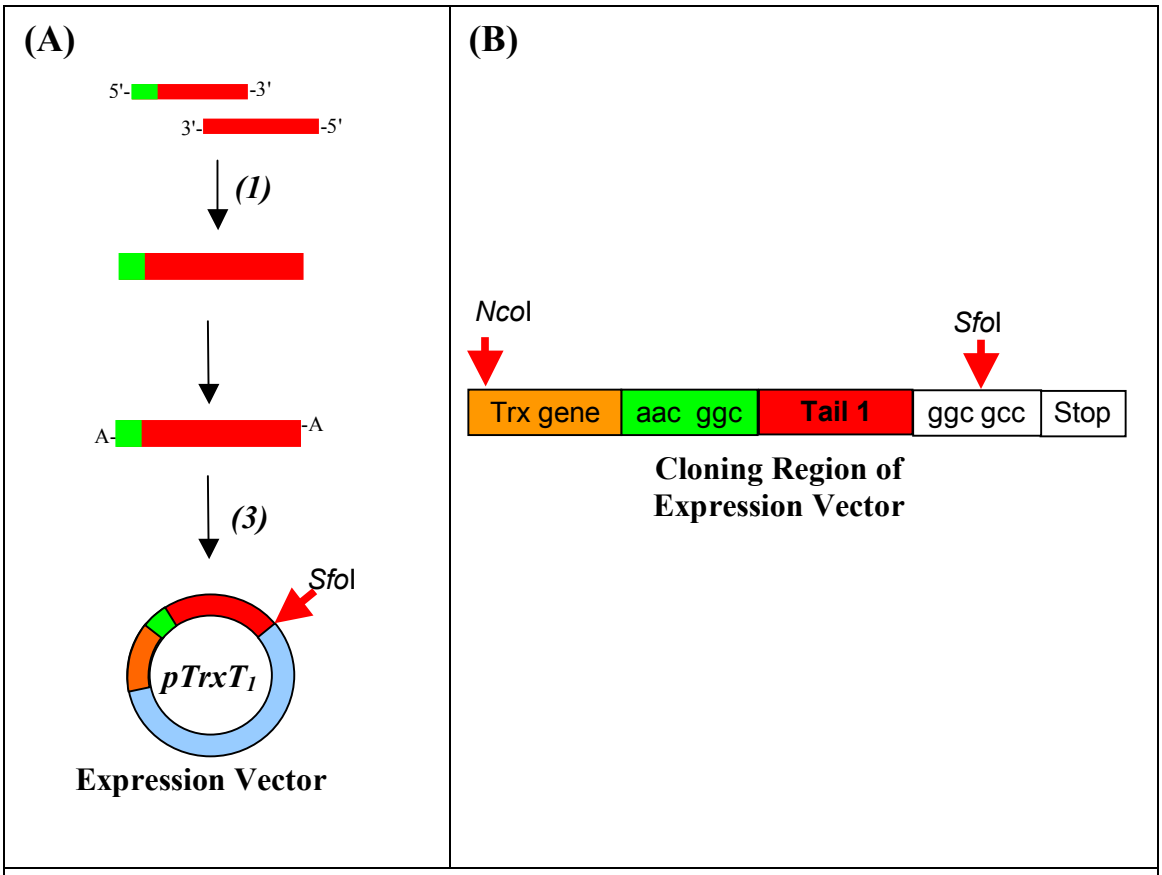


Figure 6.6 Insertion of initial DNA module into the expression vector. (A) Construction of Tail DNA Module #1 and its insertion into the expression vector. (1) Hybridization and fill-in of single-stranded synthetic oligonucleotides; (2) Addition of 3' A-overhangs with *Taq* DNA Polymerase; (3) TOPO® TA cloning into pBAD/Thio-TOPO®. (B) Schematic of cloning region showing: Trx gene (orange), Asn-Gly dipeptide sequence for hydroxylamine cleavage (green), Tail DNA Module #1 (red) and RE recognition sites (red arrows).

Insertion of Tail DNA Modules into the Expression Vector

In order to build the long DNA sequences required for the production of a high molecular weight tail block, we sequentially inserted DNA modules encoding for zwitterionic PAA sequences directly into the expression vector. The cloning procedure is summarized in Figure 6.7. All cloning steps were conducted using OneShot® MAX Efficiency® DH5αT_1^R chemically competent *E. coli* to minimize possible recombination events that may result in the loss of the PAA gene. As shown in Figure 6.7, each recombinant vector could also be transformed into the TOP10 *E. coli* strain for PAA expression.

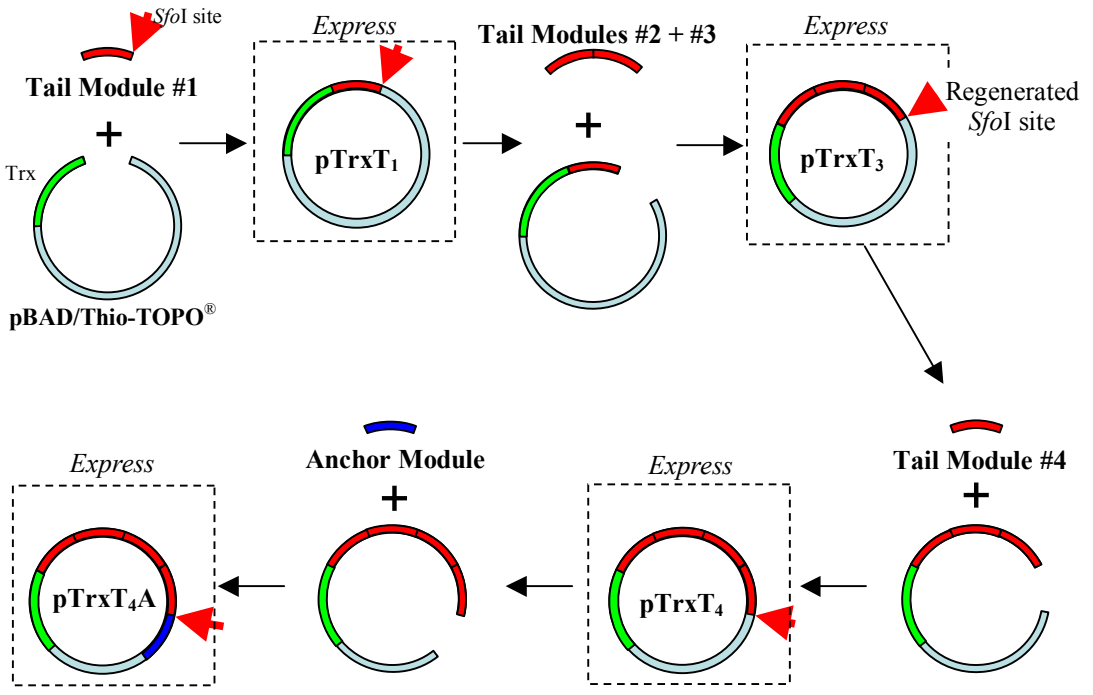


Figure 6.7 Schematic showing the application of Cloning Strategy #1.

The Tail DNA Modules inserted were based on Zwitterionic PAA Blocks #2-4 given in Figure 6.3. The DNA sequences encoding for these modules are given in Figure 6.8-Figure 6.10. From Figure 6.7, it can be seen that the second insertion into the expression vector was a '*dimer*', consisting of Tail DNA Modules #2 and #3. This module was obtained by PCR amplification of Tail DNA Modules A and B from the vector pHis$_6$-T$_2$A constructed in Chapter 5 (see Section 5.4.3) with the appropriate synthetic primers. Here, Tail DNA Modules #A and #B will be denoted as Tail DNA Modules #2 and #3, respectively. This process is summarized in Figure 6.11. The successful cloning of the dimer module demonstrated the ability to clone larger DNA modules (160 bp) using this method. Thus, the vector generated, denoted as pTrxT$_3$, contains a DNA sequence encoding for a ~75-mer tail block.

Tail DNA Module #2:

5’-ggc aaa gag ggc gag aag ggg gag ggc cgt ggc cgc gag cgc ggc gat ggc gat cgt ggt aag gaa aaa gac ggt ggc -3’

Figure 6.8 Tail DNA Module #2 sequence encoding for zwitterionic PAA block #2 (given in Figure 6.3).

Tail DNA Module #3:

5’-ggc gac cgt ggc aaa gac agc agc aag gat cgt gac agc tct gac cgt ggt gaa aaa agc gac aag gac aag ggc ggc-3’

Figure 6.9 Tail DNA Module #3 sequence encoding for zwitterionic PAA block #3 (given in Figure 6.3).

Tail DNA Module #4:

5’- ggc gaa tct cgc gac aag ggc ggt gaa aaa cgt gat tct gat cgc tct gaa cgt agc gat cgc agc agc gaa cgc ggc -3’

Figure 6.10 Tail DNA Module #4 sequence encoding for zwitterionic PAA block #4 (given in Figure 6.3).

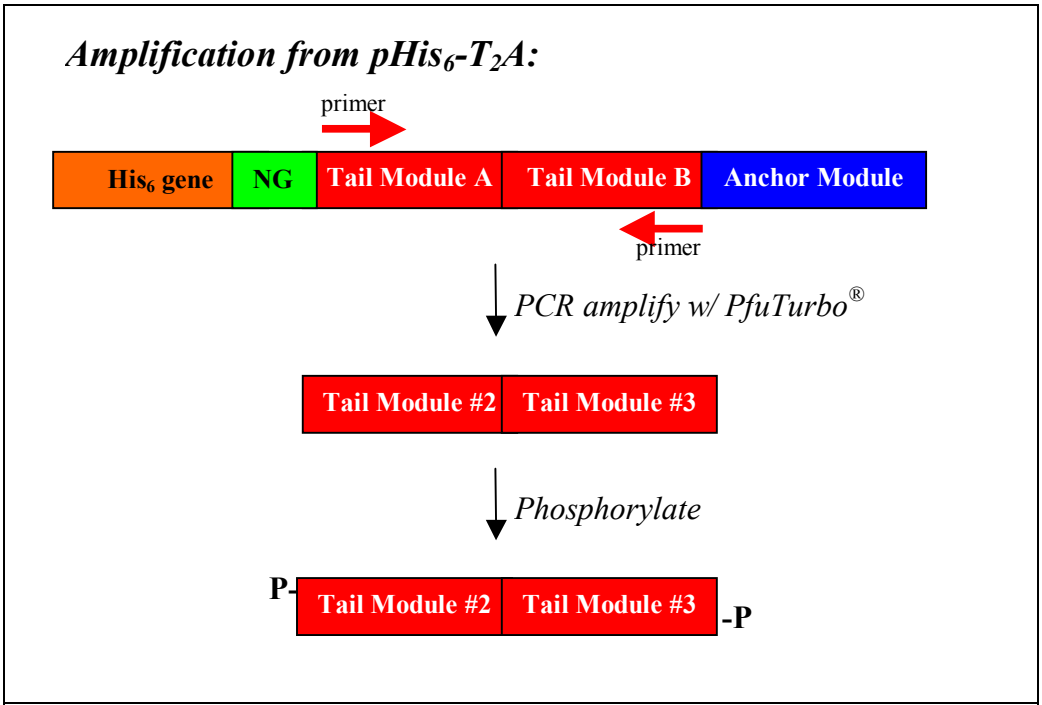


Figure 6.11 Schematic illustrating the construction of the dimer module consisting of Tail DNA Modules #2 and #3. The tail dimer module was PCR amplified from an existing recombinant vector, pHis$_6$-T$_2$A (see Chapter 5). The blunt-ended PCR product was then phosphorylated prior to insertion into the linearized pTrxT$_1$ vector. Note: Tail Modules A and B are denoted as Tail Modules #2 and #3, respectively, in this chapter.

Next, Tail DNA Module #4 was constructed from the hybridization of two single-stranded synthetic oligonucleotides and PCR amplified. The DNA sequence of this module is given in Figure 6.10. The blunt-ended module was inserted into pTrxT$_3$ generating a vector, denoted as pTrxT$_4$, containing a DNA sequence encoding for a ~100-mer tail block (Figure 6.7). Hence, we have successfully constructed a gene encoding for a PAA tail block of desired length.

Insertion of the Anchor DNA Module into pTrxT$_4$

After completing the construction of the gene encoding for the tail block, a module encoding for the anchor block was inserted (see Figure 6.7). The Anchor DNA Module (Figure 6.12) was designed based on the PAA sequence in Figure 6.2 and was constructed from the hybridization of two single-stranded synthetic oligonucleotides and PCR amplified. The blunt-ended module was inserted into pTrxT$_4$, generating the vector denoted as pTrxT$_4$A.

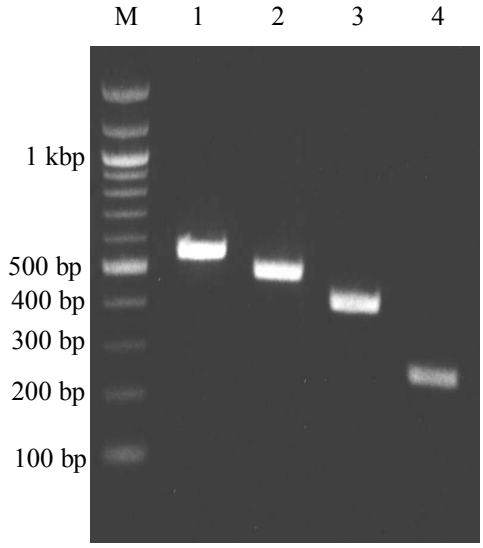
Anchor DNA Module:

5'-gcc ggc gga ggc ggt gaa gag gag gag gaa gag gaa gaa gag gaa gcc -3'

Figure 6.12 Anchor DNA Module sequence encoding for the PAA sequence given in Figure 6.2.

Analysis of the Insert Size and Orientation Using Colony PCR

It is possible for multiple insertions to take place during the blunt-end ligation step. Even though we minimized this event by optimizing the insert:vector ratio used during the ligation reaction, it is still necessary monitor the size of the insert. This can be easily accomplished by colony PCR after transformation, using primers (TrxF and pBADRev2, see Table 6.11) designed to hybridize upstream and downstream of the insert. Therefore, after each module is inserted, the PCR product will increase by the size of that module. This is shown in the agarose electrophoresis gel in Figure 6.13. For example, after Tail DNA Module #4 was inserted into the vector $pTrxT_3$, the PCR product increased from 429 to 507 bp. This increase corresponds to the insertion of the 78 bp DNA module.



Lane	Vector	Predicted PCR Product Size (bp)
1	$pTrxT_4A$	564
2	$pTrxT_4$	507
3	$pTrxT_3$	429
4	$pTrxT_1$	273

Figure 6.13 Agarose electrophoresis gel showing PCR screening of recombinant vectors constructed using Cloning Strategy #1. (1% agarose, visualized with ethidium bromide): Lane M, 100 bp DNA ladder (NEB).

It is also important to note that blunt-end ligations are non-directional. Therefore, there is a 50% chance that the modules will be inserted in a reversed orientation. These incorrect orientations will not be observed during the colony PCR screening described above. However, by using a different set of synthetic primers (given in Table 6.12), colony PCR can still be used (data not shown). In this case, the forward primer hybridizes to a DNA sequence upstream of the PAA gene. The reverse primer hybridizes to a DNA sequence *within* the most recently inserted module. Therefore, if inserted correctly, a PCR product will form. If the module is inserted incorrectly, no PCR product will be formed. This analysis is summarized in Figure 6.14.

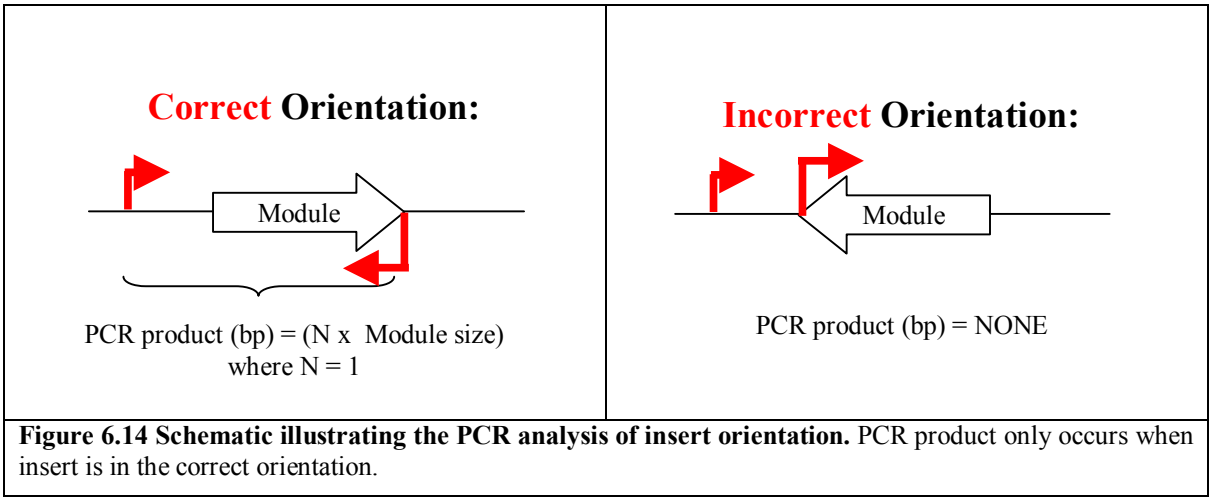


Figure 6.14 Schematic illustrating the PCR analysis of insert orientation. PCR product only occurs when insert is in the correct orientation.

The screening methods described above were used after each insertion to identify recombinant vectors containing inserts of the correct size and orientation. In addition, as with all cloning methods, the final recombinant vectors were isolated and sequenced prior to expression experiments. Samples of the recombinant vectors generated ($pTrxT_1$, $pTrxT_3$, $pTrxT_4$ and $pTrxT_4A$) were submitted to Cleveland Genomics Inc. for sequencing and all sequences were confirmed to be correct (see Section 6.11). Stocks of DH5α T_1^R *E. coli* containing each plasmid were stored at -80°C in 15% glycerol.

6.4 Expression Experiments

Once the cloning work was completed, we moved on to the product expression experiments. The vectors generated in Section 6.3 were transformed into One Shot® TOP10 Chemically Competent *E. coli*, which contain the genetic information to express

the PAA genes under the control of the *ara*BAD promoter. The PAA's expressed by each recombinant plasmid are shown in Figure 6.15

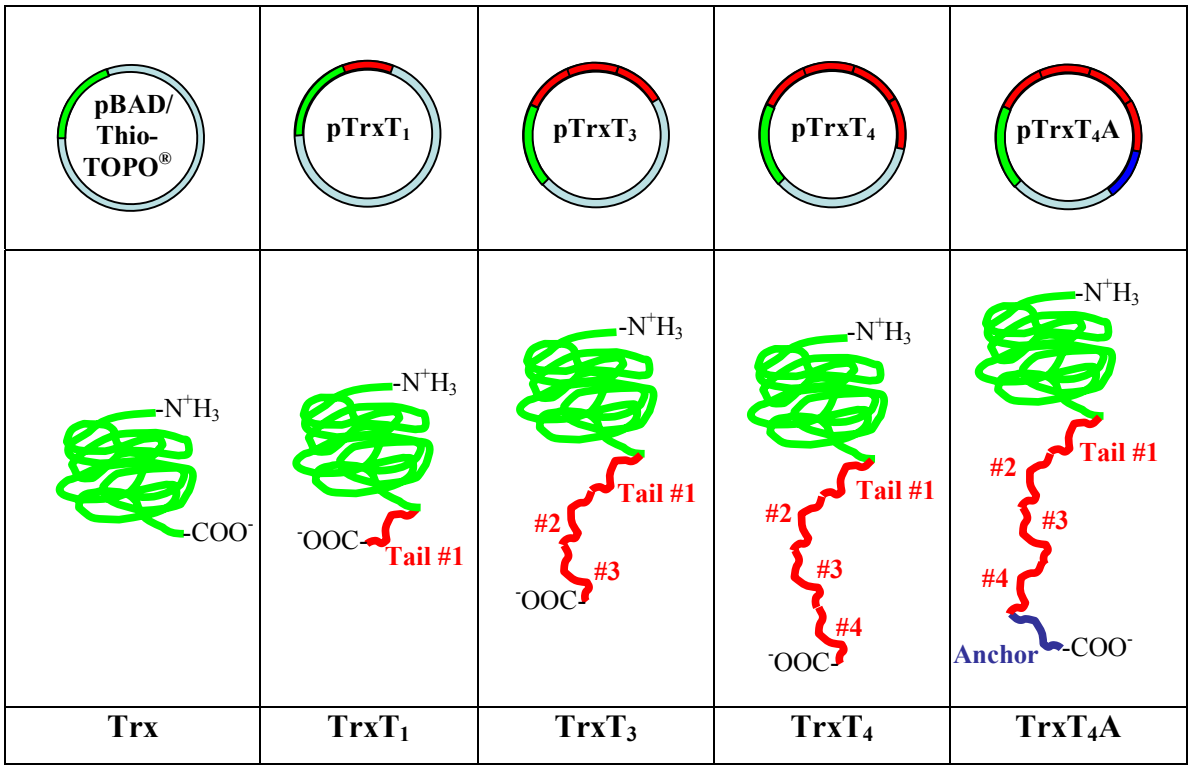
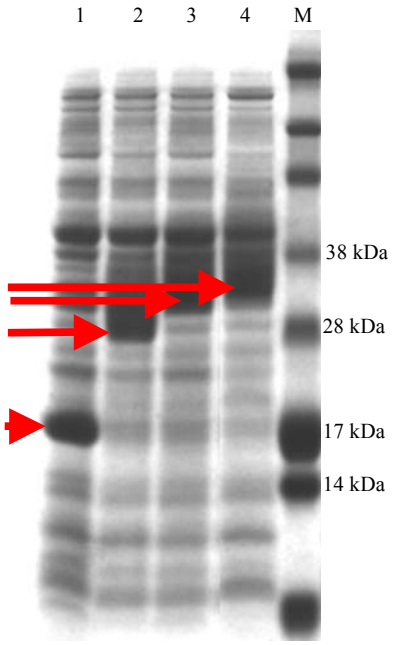


Figure 6.15 Schematic showing the PAA product expressed by each recombinant vector. The amino (N-) and carboxy (C-) termini of each PAA is indicated.

Initial expression experiments were conducted in 50 mL conical tubes. Positive transformants were grown to mid-log phase (at 37°C and with shaking) in 10 mL of LB media containing 100 µg/mL ampicillin. At mid-log (OD_{600nm} ~ 0.5), expression was induced from the *ara*BAD promoter by adding L-(+)-arabinose to a final concentration of 0.02% and the temperature was lowered to 30°C. The lower temperature was used to increase the stability of the expression plasmid. Expression cultures were incubated for 5 hrs before the cells were harvested by centrifugation (3000 x g). Cell pellets were lysed and extract samples were analyzed by SDS-PAGE (Figure 6.16).



Lane	PAA Product	Apparent MW (kDa)	Predicted MW (kDa)
1	$TrxT_1$	~17	16
2	$TrxT_3$	~27	21
3	$TrxT_4$	~30	24
4	$TrxT_4A$	~33	26

Figure 6.16 SDS-PAGE analysis of PAA expression. (NuPAGE® Novex 10% Bis-Tris polyacrylamide gels visualized by staining with Colloidal Blue Staining Kit). Lane M, SeeBlue® Plus2 Pre-Stained Protein Standard. Arrows indicate bands corresponding to each PAA product.

The arrows in Figure 6.16 show the protein bands corresponding to the PAA products: $TrxT_1$, $TrxT_3$, $TrxT_4$ and $TrxT_4A$. The band for each PAA product occurs at a MW slightly higher than the expected MW (see Section 6.11). This indicates that these PAA products most likely do not bind SDS as effectively as 'normal' proteins, causing them to migrate through the gel slower, resulting in a higher apparent MW. These expression experiments demonstrate the successful biosynthesis of the desired PAA products as Trx fusions. The following sections will focus on a larger scale production of the $TrxT_4A$ product and its subsequent purification and analysis.

6.5 Production of the $TrxT_4A$ Product

The PAA product containing a ~100-mer zwitterionic tail block and a 10-mer acidic anchor block should self-assemble on aluminum oxide surfaces to form optimized brush layers. Therefore, we conducted a larger scale production of the $TrxT_4A$ product to

determine the feasibility of producing and isolating the T_4A product. This section discusses the expression, purification and characterization of the target PAA product.

6.5.1 Expression and Cell Lysis

Expressions were conducted at a larger scale to produce sufficient quantities of the $TrxT_4A$ product. These expressions were carried out in four 2 L shaker flasks, each containing 250 mL of LB media (total of 1 L) with 100 μg/mL ampicillin. The cultures were inoculated with 10 mL of an overnight culture of TOP10 *E. coli* containing the $pTrxT_4A$ plasmid. The cultures were incubated at 37°C with shaking until they reached mid-log phase (OD_{600nm} ~ 0.5). The temperature was lowered to 30°C and expression was induced with the addition of L-(+)-arabinose to a final concentration of 0.02%. Expression was continued for 5-6 hrs under these conditions before the cells were harvested by centrifugation (3000 x g). The cell pellets were stored at -80°C.

The cell extract was obtained by reconstituting ~1 g of wet cell pellet in 10 mL 1X IMAC Binding Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.8). The cell suspension was lysed using sonication and several rapid freeze/thaw cycles. The cell debris was pelleted by centrifugation (13,000 x g) . The cell debris was discarded and the soluble cell extract containing the $TrxT_4A$ product was kept and stored at -80°C.

6.5.2 IMAC Purification of the $TrxT_4A$ Product

Immobilized metal affinity chromatography (IMAC) was used to purify the $TrxT_4A$ product from the cell extract. During these experiments, Ni-NTA spin columns were used for rapid purification. These spin columns contain a solid silica-based support functionalized with NTA (nitrilotriacetic acid) ligands, which chelate Ni^{+2} ions. Each column, pre-charged with Ni^{+2}, was loaded with 600 μL of cell extract. The flow-through was collected by centrifugation (700 x g for 2 min). Weakly bound proteins were eluted with two 600 μL washes of IMAC Wash Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 6). Finally, the purified $TrxT_4A$ product was eluted with two 200 μL washes with IMAC Elution Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 4). The eluate was quickly adjusted to pH ~ 7 with 1 M TrisHCl pH 8.0. The purification process was monitored by SDS-PAGE. The gel in Figure 6.17 shows the successful first-step

purification of the $TrxT_4A$ product. The $TrxT_4A$ product yield was estimated to be ~15 mg/L culture.

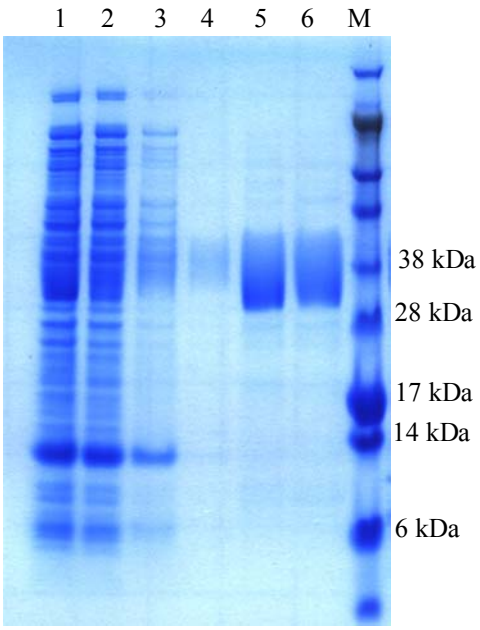


Figure 6.17 SDS-PAGE analysis of $TrxT_4A$ purification using Qiagen Ni-NTA spin columns. (NuPAGE® Novex 10% Bis-Tris polyacrylamide gels visualized by staining with Colloidal Blue Staining Kit) Lane 1, cell extract [2x dilute]; Lane 2, flow-through [2x dilute]; Lane 3, pH6 Wash 1; Lane 4, pH 6 Wash 2; Lanes 5, pH4 Elution 1; Lane 6, pH4 Elution 2; Lane M, SeeBlue® Plus2 Pre-Stained Protein Standard

6.5.3 Product Polishing Using HP-IEC

To effectively analyze the $TrxT_4A$ product, an additional purification step was necessary. To accomplish this, high performance ion exchange chromatography (HP-IEC) was employed. The pH 4 elution fraction from the Ni-NTA spin column purification step was dialyzed against deionized water and then loaded onto a Vydac DEAE 301 VHP column. Separation was obtained by using an increasing salt gradient and fractions were collected. A chromatogram showing the elution profile at 280 nm is shown in Figure 6.18.

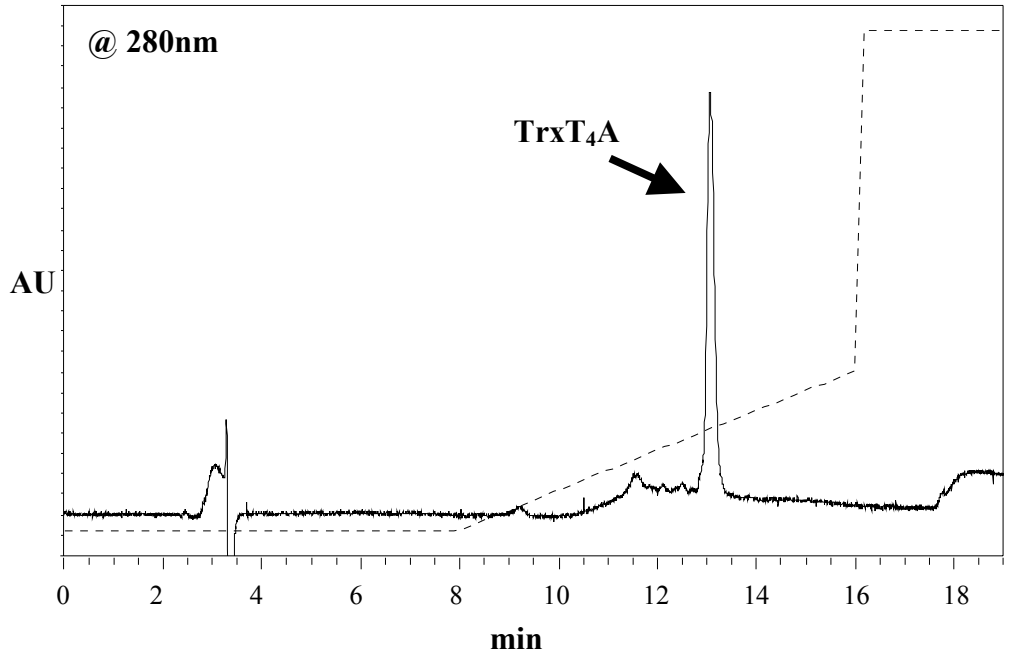


Figure 6.18 Chromatogram showing the HP-IEC purification of the TrxT$_4$A product eluted from the spin columns. Amount loaded: 100 µL. Column: Vydac DEAE 301VHP with Vydac CPF10 prefilter. 5 µm, 900 Å, 7.5 mm ID x 50 mm L. Conditions: 24°C, load at 0.5 mL/min., UV absorbance at 280 nm. Mobile phase: A = 7.5 mM Tris-HCl in 25% acetonitrile, pH 8.0. B = 500 mM NaCl in Buffer A. Gradient: Linear 0 to 32%B over 8 min at 1 mL/min. Then held at 100%B for 4 min. Collection: 0-25 min, 1 min/tube.

The strong peak observed at t = ~13 min represents the purified TrxT$_4$A product. The fusion will absorb UV at 280 nm due to residues within the Trx domain (Trp, Tyr and Cys). The confirmation of the identity of this peak was obtained by MALDI-TOF MS and is discussed in the next section.

6.5.4 MALDI-TOF MS Analysis of the TrxT$_4$A Product

Mass spectrometry was used to confirm that the molecular weight of the purified TrxT$_4$A product agreed with the expected value calculated from its amino acid sequence. A sample of the purified product was analyzed by MALDI-TOF Mass Spectrometry using an Applied Biosystems Voyager DE-Pro in linear positive ion detection mode. The results are shown below in Figure 6.19.

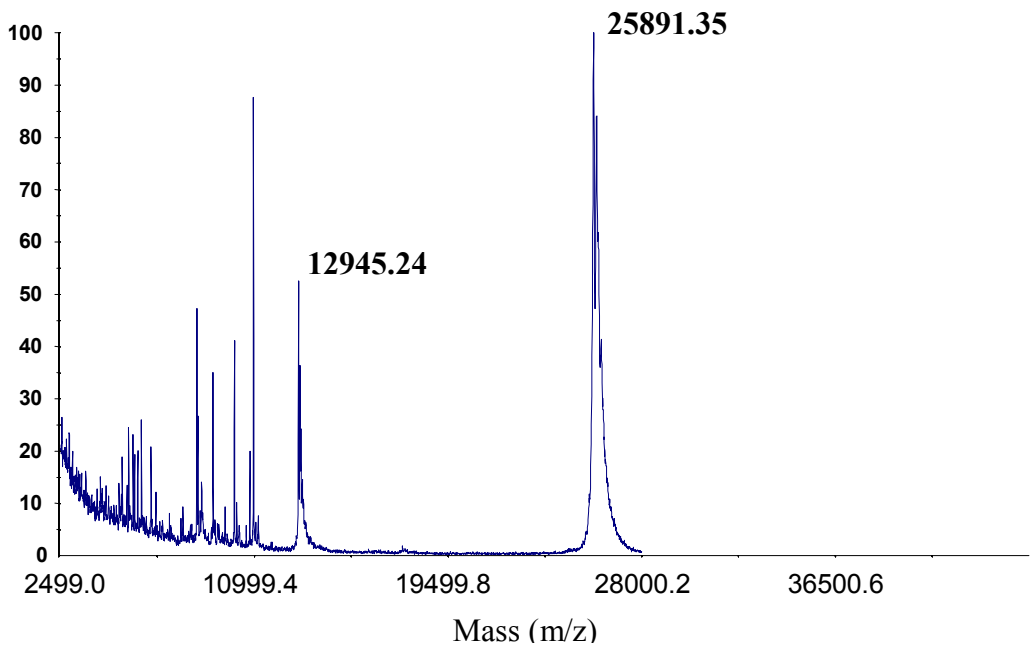


Figure 6.19 Results of MALDI-TOF MS analysis of the purified TrxT$_4$A product. Sample was mixed 1:1 with Sinapic Acid (Fluka) matrix solution (10 mg/mL Sinapic Acid in 50% acetonitrile/0.05% TFA). The plates were spotted with 1 uL of sample/matrix, and allowed to air dry. A minimum of 400 shots per sample were acquired to obtain spectrum.

A strong signal was obtained at 25,891.35 m/z, along with the corresponding doubly charged ion at 12,945.24 m/z (Figure 6.19). This was in good agreement with the calculated $[M+H]^+$ m/z of 25,889.91 m/z (MW_{avg}).

Identity	MALDI Results	Predicted[a] [M+H]+, MW_{avg}	% Diff[b]
Trx-T$_4$A	**25,891.35 m/z**	**25,889.91 m/z**	**<0.01%**

Table 6.1 Results from MALDI-TOF MS analysis of the purified Trx-T$_4$A product.

[a] Molecular weight calculated using PeptideMass software
[b] Difference relative to predicted value

Here we have described the successful production and isolation of the TrxT$_4$A product. The next step is to remove the Trx tag. Two separate methods were attempted, one enzymatic and one chemical.

6.6 Enzymatic Digestion

For the removal of the Trx fusion partner, an enzymatic recognition site was included between the tag and the T_4A sequence. Enterokinase is a highly specific enzyme that cleaves at the C-terminus of the peptide sequence DDDDK.

$TrxT_4A$ was digested with 1U EKMax™ and the reaction was monitored by SDS-PAGE (Figure 6.20A). In addition to some uncleaved fusion protein, one main peptide band appears in the digestion sample. This band appears at ~12 kDa, which is the expected size of the Trx fragment. This band was confirmed to be Trx by a Western blot analysis using an antibody against HP-Thioredoxin (Anti-Thio™) (see Figure 6.20B).

A band representing the T_4A product (~13 kDa) cannot be seen on the gel in Figure 6.20A. This may be for several reasons. First, since the Trx fragment is so close in size, the two products may not be resolved on this SDS-PAGE gel. Second, it is unclear whether a polypeptide of the composition of T_4A will stain with conventional protein staining methods. Lastly, it is possible that the T_4A is just not present after the enzyme digestion. To determine if the T_4A product was present after digestion, we turned to HPLC and mass spectrometry analysis.

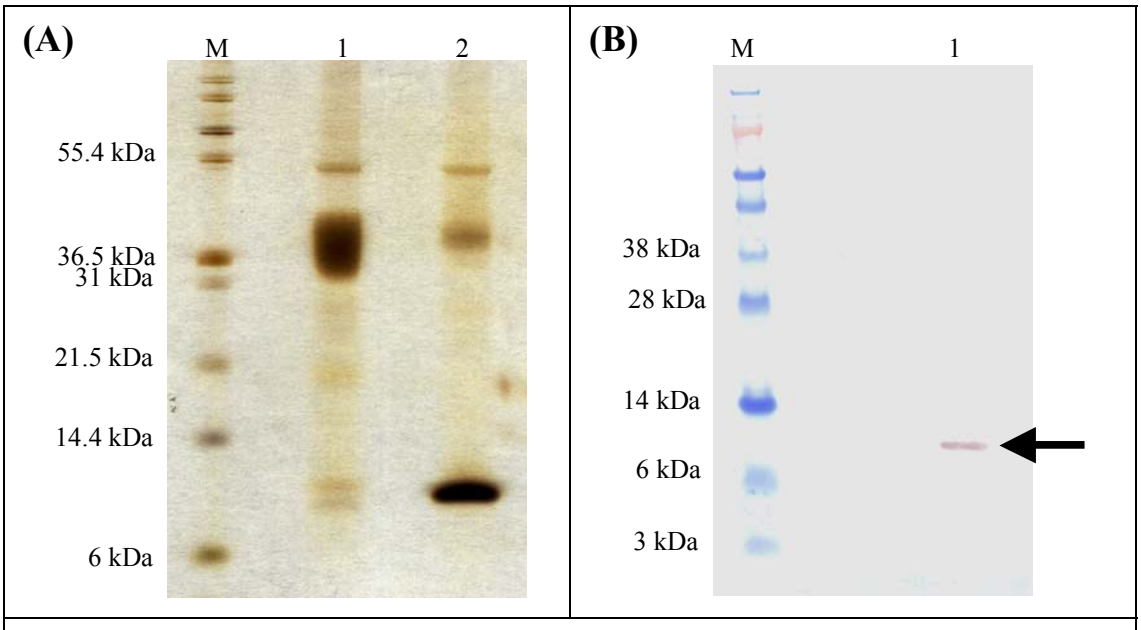


Figure 6.20 (A) SDS-PAGE analysis of the enzymatic removal of Trx using EKMax™. (NuPAGE® Novex 10% Bis-Tris polyacrylamide gels visualized by staining with SilverQuest™ Silver Staining Kit). Lane M, Mark12™ Unstained Standard; Lane 1, $TrxT_4A$ (t=0 hrs); Lane 2, $TrxT_4A$ (t= 8 hrs). **(B) Western blot analysis.** (PVDF, using Anti-Thio™). Lane M, SeeBlue® Plus2 Pre-Stained Protein Standard; Lane 1, $TrxT_4A$ (t=8 hrs).

6.6.1 MALDI-TOF MS Analysis of the Digestion Reaction

Since the T_4A product could not be visualized by SDS-PAGE or Western Blot, the enzymatic digestion reaction was analyzed using mass spectrometry. A sample of the digestion reaction was analyzed by MALDI-TOF Mass Spectrometry using a Bruker OmniFlex in linear positive ion detection mode. The results are shown in Figure 6.21.

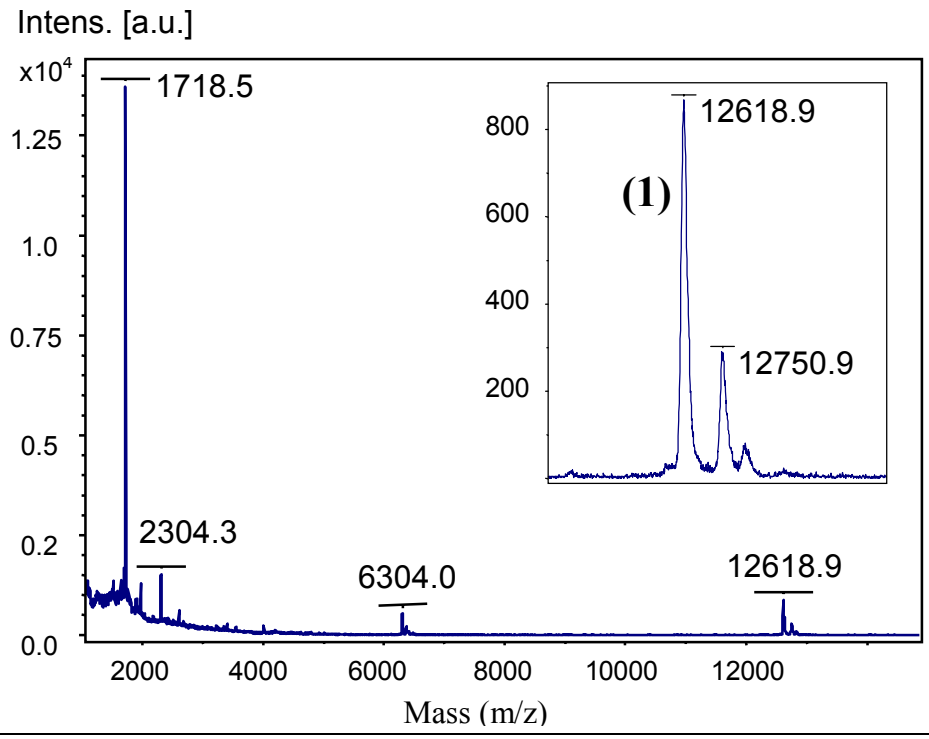


Figure 6.21 Results of MALDI-TOF MS analysis of the enzymatic digestion of $TrxT_4A$. Samples of the digest reaction were diluted 1:10 in 0.1% TFA and mixed 1:1 with sinapinic acid matrix (10 mg/mL) in 50:50 water:acetonitrile. The plates were spotted with 1 uL of sample/matrix, and allowed to air dry. A minimum of 500 shots were taken to obtain each spectrum.

A signal was obtained at 12,618.9 m/z, along with the corresponding doubly charged ion at 6,304 m/z (Figure 6.21). This was in good agreement with the calculated $[M+H]^+$ m/z of 12,671.43 m/z (MW_{avg}). These results are summarized in Table 6.2. There are several additional peaks of less intensity that appear in the mass spectra nearby peak **(1)**. These peaks may be attributed to various adducts that may form between the Trx fragment and salts or the matrix.

Peak	Identity	MALDI Results	Predicted[a] $[M+H]^+$, MW_{avg}	% Diff[b]
(1)	Trx	12,618.9 m/z	12,671.43 Da	0.41%

Table 6.2 Results from MALDI-TOF MS analysis of the enymatic digestion reaction.

[a] Molecular weight calculated using PeptideMass software
[b] Difference relative to predicted value

The calculated $[M+H]^+$ m/z of the expected T_4A fragment is 13,237.51 m/z (MW_{avg}). A peak corresponding to this fragment does not appear. This may be due to the charge structure of the T_4A construct. It may not desorb and ionize completely during the MALDI-TOF MS analysis. However, further analysis of the digest sample indicated that the T_4A product was not present at all. HPLC analysis of the digest reaction only resulted in one peak (at 214 and 280 nm), most likely due to the Trx fragment (data not shown). This suggests that the T_4A product has been degraded. There is evidence in the literature of non-specific proteolysis by enterokinase.[13] This may be occurring here. After the EKMax™ cleaves the Trx tag at the specified site, it possibly continues to degrade the T_4A product. This may explain why a T_4A band does not appear in the SDS-PAGE gel in Figure 6.20A, or in the MS and HPLC analyses. As a result, fusion tag removal via enzymatic digestion was abandoned. In the next section, we turn to a chemical cleavage method to remove the Trx.

6.7 Hydroxylamine Chemical Cleavage

The hydroxylamine cleavage reaction was first described by Bornstein and Balian[14] and its effectiveness as a method to cleave fusion proteins was first demonstrated by Canova-Davis *et al.*[15] Cleavage of the TrxT_4A product was carried out in a buffer containing 2 M hydroxylamineHCl, 2 M guanidineHCl and 0.2 M K_2CO_3 at pH 9.0. The guanidine was included to act as a denaturant, which has been shown to improve yields.[16] The reaction mixture was incubated at 45°C for 20 hrs. To stop the reaction, 3vol. of a 2% TFA solution was added. The cleavage reaction was monitored by SDS-PAGE analysis (Figure 6.22).

The gel in Figure 6.22 shows the progression of the cleavage reaction. The intensity of the TrxT_4A band decreases as several bands <14.4 kDa appear. Trx contains one

internal Asn-Gly bond and additional bonds that are mildly susceptible to cleavage by hydroxylamine.[14,17] The location of these bonds are shown in Table 6.3. Assuming 100% cleavage of these bonds, 6 fragments were expected, including the T_4A product. The sequences and sizes of these fragments are given in Table 6.4. It is unclear whether any of the bands in Figure 6.22 represent the T_4A product because of the polypeptide's unknown staining properties. Therefore, we turned to HPLC and mass spectrometry in order to determine if the desired PAA product was formed during the chemical cleavage reaction.

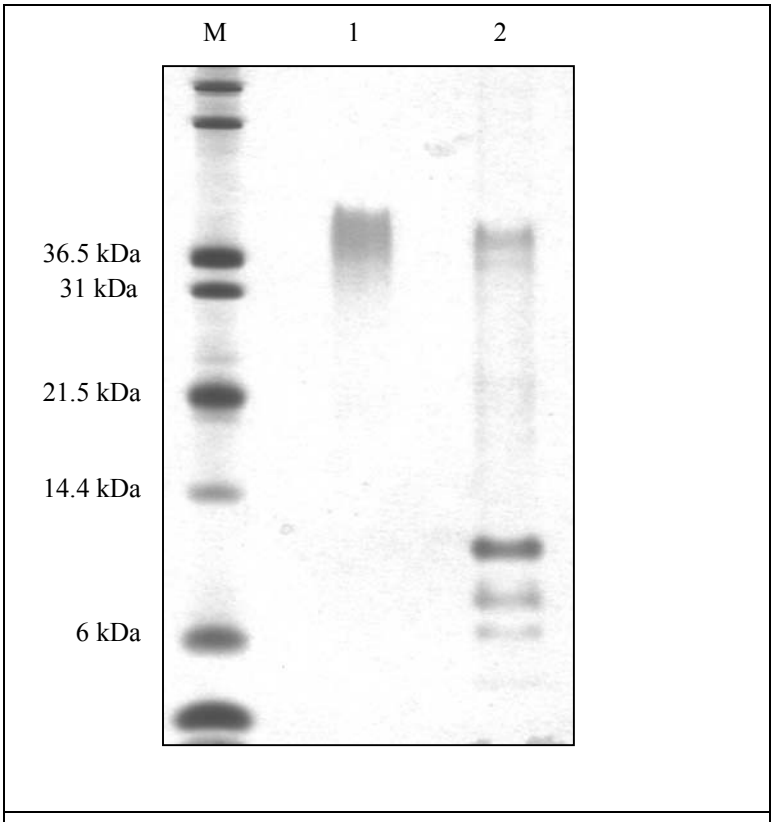


Figure 6.22 SDS-PAGE analysis of the chemical removal of Trx using hydroxylamine. (NuPAGE® Novex 10% Bis-Tris polyacrylamide gels visualized by staining with SilverQuest™ Silver Staining Kit). Lane M, Mark12™ Unstained Standard; Lane 1, $TrxT_4A$ in control buffer (t=20 hrs); Lane 2, $TrxT_4A$ in cleavage buffer (t=20 hrs).

MGSD↓KIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCKMIAPILDGIADGYQG
KLTVAKLNIDHNPGTAPKYGIRGIPTLLLFKN↓GEVAATKVGALSKGQLKEFLDA
N↓LAGSGSGDDDD↓KLALN↓GKSEGEKRSDKEGKEGGDRKESDGRGGKEGEKG
EGRGRERGDGDRGKEKDGGGDRGKDSSKDRDSSDRGEKSDKDKGGGESRDKG
GEKRDSDRSERSDRSSERGAGGGGEEEEEEEEEEA

Table 6.3 Amino acid sequence of the TrxT$_4$A product and the location of the peptide bonds susceptible to cleavage by hydroxylamine (↓).

Peptide Fragment	Amino Acid Sequence	$[M+H]^+$ MW_{avg}
T$_4$A	GKSEGEKRSDKEGKEGGDRKESDGRGGKEGEKGEGRGRER GDGDRGKEKDGGGDRGKDSSKDRDSSDRGEKSDKDKGGGE SRDKGGEKRDSDRSERSDRSSERGAGGGGEEEEEEEEEEA	12826.01
Trx Fragment #1	MGSD	409.43
Trx Fragment #2	KIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCKMIAPILDG IADGYQGKLTVAKLNIDHNPGTAPKYGIRGIPTLLLFKN	8834.29
Trx Fragment #3	GEVAATKVGALSKGQLKEFLDAN	2347.67
Trx Fragment #4	LAGSGSGDDDD	1008.93
Trx Fragment #5	KLALN	558.70

Table 6.4 Peptide fragments resulting from 100% cleavage of TrxT$_4$A by hydroxylamine. Molecular masses were calculated using PeptideMass software.

6.7.1 MALDI-TOF MS Analysis of the Chemical Cleavage Reaction

Mass spectrometry was used to confirm the presence of the T$_4$A product after the Trx tag was removed by hydroxylamine chemical cleavage. TrxT$_4$A was cleaved for 4 hrs at 45°C in the hydroxylamine reaction buffer, then stopped with 3vol. of 2% TFA and frozen. Samples were then thawed and dialyzed against 20 mM Tris pH8.0. A sample of the dialyzed product then was analyzed by MALDI-TOF Mass Spectrometry using an Applied Biosystems Voyager DE-Pro in linear positive ion detection mode. The results are shown below in Figure 6.23.

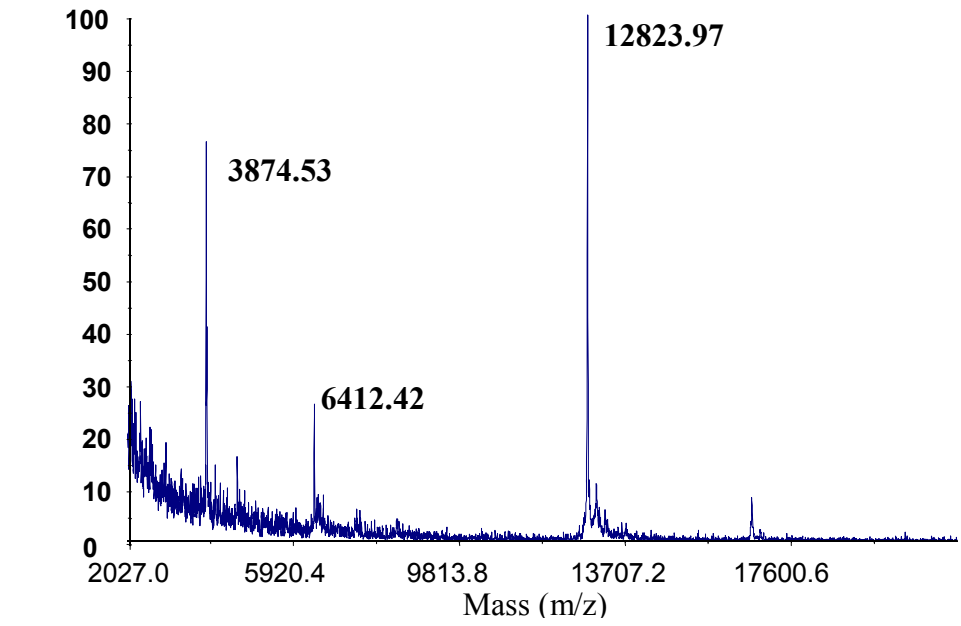


Figure 6.23 Results of MALDI-TOF MS analysis of the hydroxylamine cleavage reaction (t = 4 hrs). Dialyzed reaction samples were diluted 1:1 with Sinapic Acid (Fluka) matrix solution (10 mg/mL Sinapic Acid in 50% acetonitrile/0.05% TFA). The plates were spotted with 1 uL of sample/matrix, and allowed to air dry. A minimum of 400 shots were taken to obtain each spectrum.

A strong signal was obtained at 12,823.35 m/z, along with the corresponding doubly charged ion at 6,412 m/z (Figure 6.23). This was in good agreement with the calculated $[M+H]^+$ m/z of 12,826.01 m/z (MW_{avg}).

The MS analysis of the hydroxylamine cleavage reaction confirms the presence of the T_4A product. This once again validates hydroxylamine chemical cleavage as an efficient method for the removal of fusion tags. More importantly, these results indicate that we have succeeded in producing a specifically designed PAA. The final step is to isolate the product, which will be discussed in the next section.

Identity	MALDI Results	Predicted[a] $[M+H]^+$, MW_{avg}	% Diff[b]
T_4A Product	**12,823.97 m/z**	**12,826.01 m/z**	**0.02%**

Table 6.5 Results from MALDI-TOF MS analysis of the hydroxylamine cleavage reaction.

[a] Molecular weight calculated using PeptideMass software
[b] Difference relative to predicted value

6.8 Isolation of the T_4A Product Using HP-IEC

High performance ion exchange chromatography (HP-IEC) was employed to isolate the T_4A product from the other peptide fragments produced during the hydroxylamine chemical cleavage reaction. Dialyzed reactions samples were loaded onto a Vydac DEAE 301 VHP column. Separation was obtained by using an increasing salt gradient and fractions were collected. A chromatogram showing the elution profiles at 214 nm and 280 nm is shown in Figure 6.24.

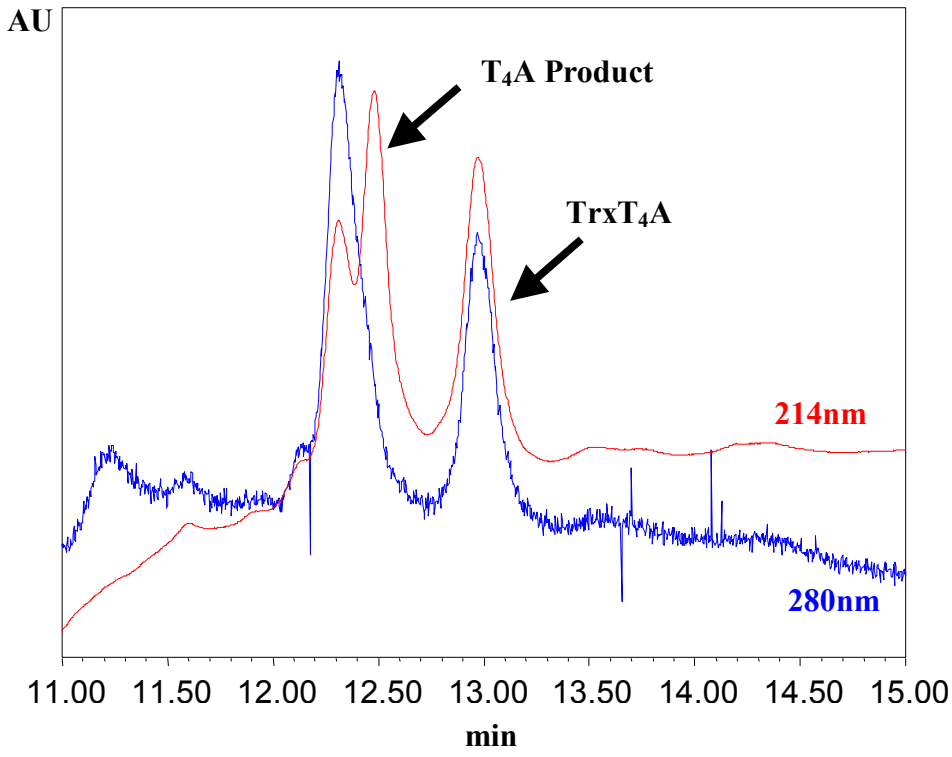


Figure 6.24 Chromatogram showing the HP-IEC purification of the T_4A product from the hydroxylamine cleavage reaction. Sample: Hydroxylamine digest reaction (t=4 hrs) of $TrxT_4A$ stopped with 3vol 2% TFA and frozen at -80°C. Sample was then dialyzed against 20 mM Tris pH8.0. Amount loaded: 100 µL. Column: Vydac DEAE 301VHP with Vydac CPF10 prefilter. 5 µm, 900 Å, 7.5 mm ID x 50 mm L. Conditions: 24°C, load at 0.5 mL/min., UV absorbance at 214 nm (top trace) and 280 nm (bottom trace). Mobile phase: A = 7.5 mM Tris-HCl in 25% acetonitrile, pH 8.0. B = 500 mM NaCl in Buffer A. Gradient: Linear 0 to 32%B over 8 min at 1 mL/min. Then held at 100%B for 4 min. Collection: 0-25 min, 1 min/tube. (Note: y-axis is normalized AU).

Although the T_4A product does not contain amino acids that absorb UV at 280 nm (Trp, Tyr and Cys), due to the high absorptive energy of the peptide bond, it will absorb strongly at 214 nm. Therefore, we can identify the T_4A product by comparing the chromatograms seen in Figure 6.24 and looking for peaks that appear in the 214 nm trace

(top) but not in the 280 nm trace (bottom). Using this method, the peak eluted at ~12.5 min becomes a prime candidate for containing the T_4A product. This peak was collected and submitted for MS analysis to determine if it contained the T_4A product. This is discussed in the next section. It also should be noted that in Figure 6.24 we also see a peak eluted at ~13 min that has absorbance at 214 and 280 nm. From the information acquired in Section 6.5.3, we may assume that this peak corresponds to the presence of $TrxT_4A$ in the sample. This suggests that after 4 hrs, the hydroxylamine reaction did not reach completion and some fusion product remained in the reaction sample uncleaved.

6.8.1 MALDI-TOF MS Analysis of the Isolated T_4A Product

Mass spectrometry was used to confirm the presence of the T_4A product in the fraction collected above. A sample of the HP-IEC fraction was analyzed by MALDI-TOF Mass Spectrometry using an Applied Biosystems Voyager DE-Pro in linear positive ion detection mode. The results are shown below in Figure 6.25.

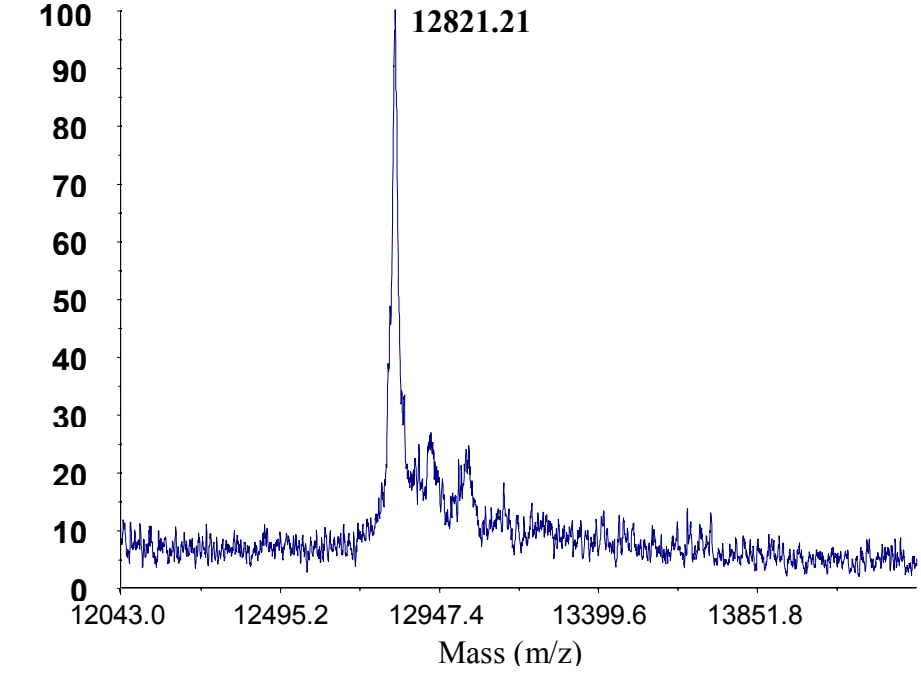


Figure 6.25 Results of MALDI-TOF MS analysis of the isolated T_4A product. HPLC fraction was dried down and reconstituted with deionized water. The sample was then mixed 1:1 with Sinapic Acid (Fluka) matrix solution (10 mg/mL Sinapic Acid in 50% acetonitrile/0.05% TFA). The plates were spotted with 1 uL of sample/matrix, and allowed to air dry. A minimum of 400 shots were taken to obtain each spectrum.

Identity	MALDI Results	Predicted[a] $[M+H]^+$, MW_{avg}	% Diff[b]
T_4A Product	12,821.21 m/z	12,826.01 m/z	0.04%

Table 6.6 Results from MALDI-TOF MS analysis of the isolated T_4A product.

[a] Molecular weight calculated using PeptideMass software
[b] Difference relative to predicted value

A strong signal was obtained at 12,821.21 m/z (Figure 6.25). This was in good agreement with the calculated $[M+H]^+$ m/z of 12,826.01 m/z (MW_{avg}). Therefore, the MS analysis of the HP-IEC fraction confirmed that we have located the peak corresponding to the T_4A product. However, by inspecting Figure 6.24, it can be seen that slightly more method development is necessary to resolve the T_4A peak from surrounding peaks completely. This is required in order to collect a sufficiently pure T_4A product sample for further analyses. However, we were unable to accomplish this due to reasons that will be discussed in the next section.

6.8.2 Instability of the Zwitterionic PAA Block

During the process of producing and analyzing the T_4A product described above, it became evident that the zwitterionic PAA block may be inherently unstable. While we were able to successfully collect the data presented in this chapter, several factors led us to this conclusion. We began to question the stability of the zwitterionic PAA block after noticing that samples seemed susceptible to degradation after long term storage at -80°C and multiple freeze/thaw cycles. Initially, purified samples of the $TrxT_4A$ product behaved as expected during HPLC and MS analysis. However, subsequent analyses (after being stored at -80°C and multiple freeze/thaw cycles) resulted in significantly different results (Figure 6.26).

The top and bottom panels of Figure 6.26 show the HP-IEC and MS analysis, respectively, of the purified $TrxT_4A$ product after storage at -80°C and multiple freeze/thaw cycles. Initially, during HP-IEC, it was found that $TrxT_4A$ elutes at ~13 min (Figure 6.18). In the chromatogram shown in Figure 6.26, several smaller peaks appear in the 10-12.5 min range but a peak corresponding to $TrxT_4A$ does not appear (outlined box). During MS analysis, we expect there to be peak at 25,891.35 m/z (Figure 6.19), however, no peak appears at this m/z in the spectrum below (outlined box). Several

additional peaks corresponding to smaller peptide fragments do appear that were not present initially. The difference between the initial results and the HPLC and MS analyses shown here suggest that, after storage and multiple freeze/thaw cycles, the $TrxT_4A$ product is being degraded.

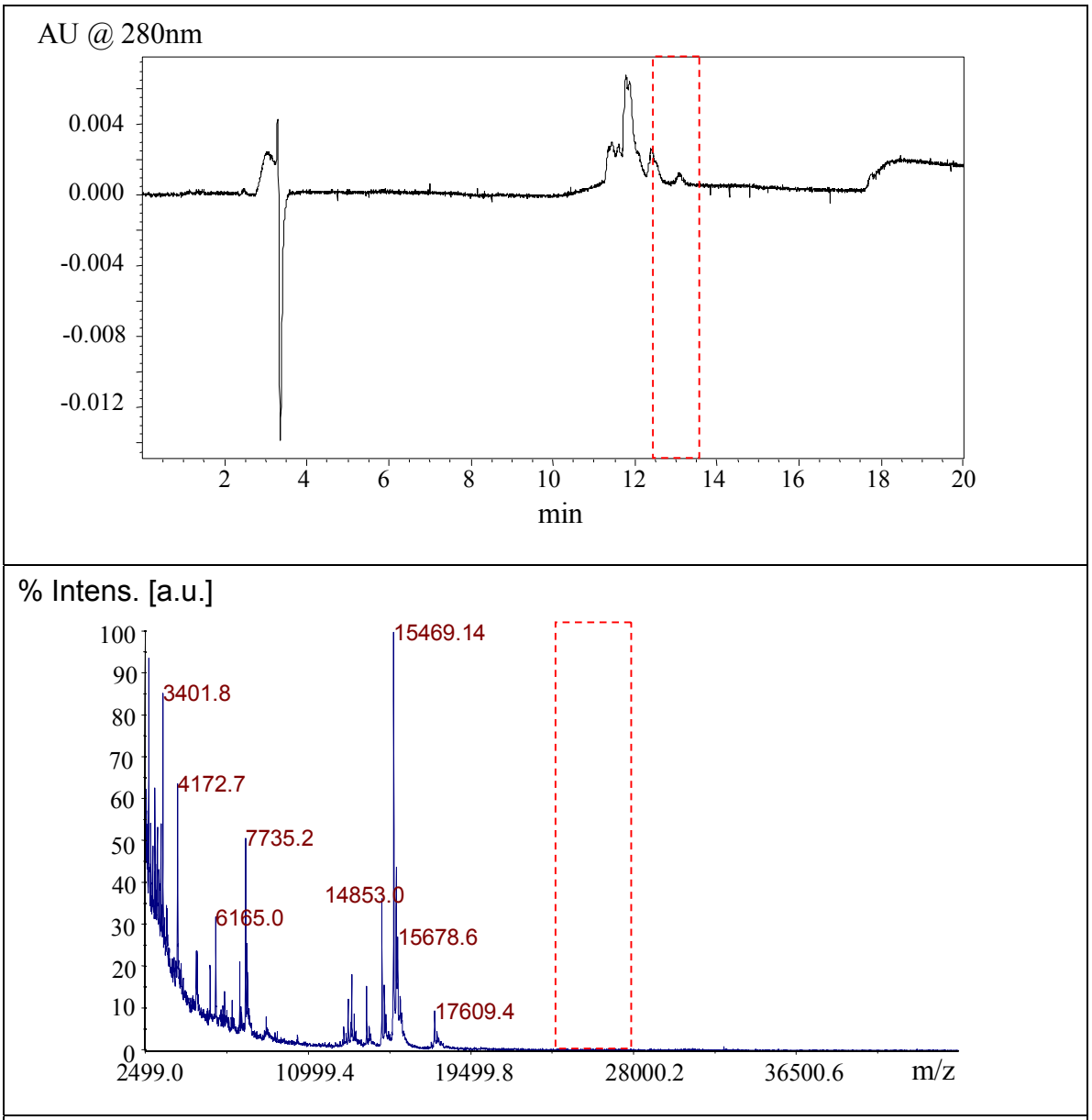


Figure 6.26 HP-IEC and MALDI-TOF MS analysis results suggesting degradation of $TrxT_4A$ product. (Top Panel) HP-IEC analysis of the purified $TrxT_4A$ product after storage at -80°C and multiple freeze/thaw cycles. The column and chromatographic conditions used were the same as in Figure 6.18. Outlined box indicates where $TrxT_4A$ should elute. (Bottom Panel) MALDI-TOF MS analysis of the purified $TrxT_4A$ product after storage at -80°C and multiple freeze/thaw cycles. The sample preparation and analysis was conducted in same manner as all other MS analyses included in this chapter. Outlined box indicates where $TrxT_4A$ peak should occur.

In light of this evidence, we may revisit our conclusions drawn from the enzymatic removal of Trx from the T_4A product (Section 6.6). EKMax™ was used to enzymatically cleave off the Trx tag, but according to MS analysis, no T_4A product appeared after the reaction. We originally attributed this to possible non-specific proteolysis of the product by EKMax™. This may in fact be still true. However, the T_4A product may not have been detected due to degradation as a result of an inherent instability.

The exact cause of the zwitterionic PAA block instability is unknown at this time. It is possible that the T_4A product is susceptible to proteolysis by enzymes that it encounters during its production and isolation. There are many naturally occurring proteases that catalyze the cleavage of peptide bonds at or near Arg/Lys residues (tryspin, endoproteases Lys-C and Arg-C) and Glu/Asp residues (endoproteases Asp-N and Glu-C). These enzymes are serine endoproteases, which is a class of proteases that are common in biological systems. It is quite possible that a protease of this type (native to *E. coli*) may be present during the processing of the $TrxT_4A$ product. If so, it may be able to catalyze the cleavage of peptide bonds within the zwitterionic PAA sequence, due to its abundance of Lys, Arg, Glu and Asp residues. To avoid this problem, protease inhibitors may need to be added to all buffers used during the processing of constructs that contain the zwitterionic PAA block. For example, benzamidine and phenylmethylsulfonyl fluoride (PMSF) are commonly used (at 1 mM) to act as competitive inhibitors of serine proteases.

In addition to proteolysis, it is possible the zwitterionic sequence is inherently labile to some process during its relatively routine processing. For example, the zwitterionic sequence may be labile to freeze/thaw cycles. Given this possibility, it may unfortunately be impossible to produce sufficient quantities of T_4A product for future studies.

6.9 Conclusions

The work presented in this chapter describes the design, production and partial isolation of a unique PAA. The completion of this work was significant because it confirmed the feasibility of the following:

- Applying the new cloning strategy to the construction of the long DNA sequences necessary to produce high MW PAA's with a specifically designed composition.
- The biosynthetic production of a unique PAA specifically designed for brush-forming applications.
- Using hydroxylamine cleavage as a simple method for the removal of fusion tags.

As stated, for the production of optimized, brush-forming PAA's, a new cloning strategy was necessary. In Chapter 5, a new and universal strategy was introduced. In this chapter, it was employed for the production of a specifically designed, high MW PAA diblock consisting of a short acidic anchor block and a long zwitterionic tail block. The utility of the strategy was shown through its ability to exactly control the PAA monomer sequence and MW.

Biosynthetic production and partial isolation of the desired T_4A product was successful. However, evidence suggests that the zwitterionic PAA block may be inherently unstable. This is an unfortunate complication that jeopardizes the practicality of this PAA design. Further analysis will be required to elucidate the cause of this apparent instability. Obviously, if unstable, this PAA would not be suitable for brush-forming applications. This is also unfortunate since we considered this zwitterionic PAA to potentially be a universal tail block for metal oxide systems.

Even with these complications, the successful biosynthetic production and partial isolation of the T_4A product can be viewed as a significant accomplishment. We have shown that we have developed an effective strategy for the design, cloning, expression and purification of specifically designed, high MW brush-forming PAA's. The methodology presented can also be universally applied towards many other PAA designs and applications, paving the way for breakthroughs in fundamental and applied studies of self-assembly phenomena.

6.10 Materials and Methods

Materials. pBAD/TOPO® ThioFusion™ Expression Kit, One Shot® TOP10 Chemically Competent *E. coli* strain, OneShot® MAX Efficiency® DH5αT_1^R Chemically Competent *E. coli* strain, EKMax™, NuPAGE® Novex 10% Bis-Tris polyacrylamide gels, LB Broth Base, UltraPure™ agarose, UltraPure™ ethidium bromide solution, SeeBlue® Plus2 Pre-Stained Standard, Mark12™ Unstained Standard, Colloidal Blue Staining Kit, SilverQuest™ Silver Staining Kit, Anti-Thio™ antibody and Trx Forward/pBAD Reverse sequencing primers were purchased from Invitrogen Corp. (Carlsbad, CA). Sequenase™ Version 2.0 DNA polymerase was obtained from USB Corporation (Cleveland, OH). *Taq* PCR Core Kit, PAGE-purified synthetic oligonucleotides, QIAprep® Spin Miniprep, QIAquick® Gel Extraction/PCR Purification and Ni-NTA Spin Kits were purchased from Qiagen, Inc. (Valencia, CA). *PfuTurbo*® DNA polymerase was purchased from Stratagene (La Jolla, CA). LB Agar, CelLytic™ B Bacterial Cell Lysis Extraction Reagent, L-(+)-arabinose, and Goat Anti-Mouse HRP-conjugate were obtained from Sigma (St. Louis, MO). Immun-Blot™ PVDF membrane was purchased from Biorad (Hercules, CA). Ampicillin and 0.2μm pre-sterilized syringe filters were purchased from VWR International (So. Plainfield, NJ). Metal Enhanced DAB Substrate Kit and Slide-A-Lyzer® Dialysis Cassettes (3.5kDa molecular weight cutoff) were purchased from Pierce (Rockford, IL). All restriction enzymes, Alkaline Phosphatase, Calf Intestinal (CIP), T4 Polynucleotide Kinase, T4 DNA Ligase, 100 bp DNA and 1 kbp DNA Ladders were purchased from New England Biolabs (Beverly, MA). DNA primers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA).

General Methods. The procedures for the manipulation of DNA, transformation, cell growth, product expression and analysis were adapted from published literature[35,36] or from instructions provided by product manufacturers. All reagents for the manipulation of DNA were sterile and DNAse/RNAse free. Enzymatic manipulations of DNA were conducted in reagent buffers supplied by the manufacturer. PCR amplification and DNA extension reactions were performed in an Omnigene thermal cycler from Hybaid (United Kingdom). Automated DNA sequence analysis was performed on a Perkin-Elmer ABI Prism model 377 DNA Sequencer at Cleveland Genomics, Inc. (Cleveland, OH). Cells were lysed using a Tekmar Sonic Disruptor with a microtip sonicator (Mason, OH). Product analysis was conducted by SDS-PAGE using pre-cast NuPAGE Novex 10% Bis-Tris polyacrylamide gels on an Xcell Surelock Mini-Cell apparatus from Invitrogen Corp. The gels were visualized using silver staining or Coomassie Blue staining. Agarose and polyacrylamide gel images were captured using a Polaroid GelCam and Microtek ScanMaker X6EL scanner, respectively. Polypeptide product and DNA concentrations were calculated from the optical density (OD) obtained using a Milton Roy Spectronic 1201 UV spectrophotometer (Ivyland, PA).

New Cloning Strategy

Insertion of Initial DNA Module. Two overlapping, single-stranded, synthetic oligonucleotides (Table 6.7) were used to construct Tail DNA Module #1. To construct the module, the sense and antisense oligonucleotides were mixed and then heated to 95°C. The mixture was cooled down slowly to T_A°C (5°C below the estimated melting temperature (T_m) of the complimentary bases) and held for 5 minutes to anneal the strands. A solution of dNTPs was added, followed by the addition of Sequenase™ to begin the "fill-in" reaction (37°C for 30 min). 3' A-overhangs were added to the DNA module by adding 1 U of *Taq* DNA polymerase and incubating the mixture at 72°C (15 min). Agarose gel electrophoresis showed that the product size was ~80 base-pairs, as desired (data not shown). The module was then TOPO® TA cloned into the expression vector pBAD/Thio-TOPO®. The entire cloning reaction was then transformed into OneShot® MAX Efficiency® DH5αT_1^R Chemically Competent *E. coli* and selected on LB Agar containing 100 μg/mL ampicillin after overnight incubation (~16 hrs) at 37°C. Colonies were chosen and used to inoculate 5mL of LB Broth containing 100 μg/mL ampicillin. Cultures were incubated with shaking (~200 rpm) overnight (~16 hrs) at 37°C. 3 mL of each culture was used to isolate plasmid using the QIAprep® Spin Miniprep kit. Plasmid was eluted from the spin columns into 40 μL of sterile water to a concentration of ~100 ng/μL. Restriction enzymes *Nco*I and *Sfo*I were used to check the orientation of the Tail DNA Module #1. The double digest reactions were incubated overnight (~16 hrs) at 37°C. A sample of each reaction was loaded onto an agarose electrophoresis gel and visualized with ethidium bromide (data not shown). Correct orientation yielded a digest fragment (bp) = ~370 bp + (size of Tail DNA Module #1) = ~450 bp. Incorrect orientation yielded a digest fragment (bp) = ~370 bp. Plasmids from positive clones were submitted for

DNA sequencing (Cleveland Genomics, Inc.) using Trx Forward/pBAD Reverse primers to validate the gene sequence and orientation. Stocks of positive clones were made in 15% glycerol and stored at -80°C. This vector was denoted as $pTrxT_1$.

Tail DNA Module #1		**T_A**
Sense	5'- aac ggc aag agc gaa ggc gag aag cgc agc g**at aaa gag ggt aaa gag ggc ggt** -3'	63°C
Antisense	5'- tca ggc gcc gcg acc atc gct ctc ttt gcg gtc **acc gcc ctc ttt acc ctc ttt at**-'3	

Table 6.7 Synthetic oligonucleotides used to construct Tail DNA Module #1. Bold-face represents bases involved in hybridization. Underlined bases represent *Sfo*I recognition site.

Generation of Blunt-Ended DNA Dimer Module (Tail DNA Modules #2+#3). A dimer module consisting of Tail DNA Modules #2 and #3 (Figure 6.10) was constructed via PCR amplification from an existing recombinant vector, $pHis_6$-T_2A (see Chapter 5). PCR was conducted using *PfuTurbo*® DNA polymerase and the corresponding amplification primers (Table 6.8). PCR was performed with 35 cycles at 95°C for 30 sec (denaturing), 60°C for 30 sec (annealing), and 72°C for 50 sec (elongation). Large amounts of the double-stranded, blunt-ended DNA dimer were obtained. PCR reactions were pooled together and the DNA was purified using the QIAquick® PCR purification kit and dried down using a Savant SpeedVac. DNA was then re-suspended in T4 DNA Ligase Buffer and the 5' ends were phosphorylated with T4 polynucleotide kinase. DNA was purified by agarose gel electrophoresis and extracted into sterile water using the QIAquick® Gel Extraction kit. DNA concentration was determined by optical density (OD) at 260 nm. The final DNA sequence of the dimer module is located in Table 6.8.

Tail DNA Modules #2+#3		**T_A**
DNA Sequence	5'- ggc aaa gag ggc gag aag ggg gag ggc cgt ggc cgc gag cgc ggc gat ggc gat cgt ggt aag gaa aaa gac ggt ggc ggc gac cgt ggc aaa gac agc agc aag gat cgt gac agc tct gac cgt ggt gaa aaa agc gac aag gac aag ggc ggc -3'	---
Tail2 Forward	5'- ggc aaa gag ggc gag aag gg- 3'	60°C
Tail3 Reverse	5'- gcc gcc ctt gtc ctt gtc gc -3'	

Table 6.8 DNA sequence of the dimer module constructed using PCR amplification with primers Tail2 Forward and Tail3 Reverse and the template vector $pHis_6$-T_2A.

Generation of Blunt-Ended Tail DNA Module #4 and Anchor DNA Module. Two overlapping, single-stranded, synthetic oligonucleotides were used to construct Tail DNA Module #4 (Table 6.9) and Anchor DNA Module (Table 6.10). For each module, the sense and antisense oligonucleotides were mixed and then heated to 95°C. The mixture was cooled down slowly to T_A°C (5°C below the estimated melting temperature (T_m) of the complimentary bases) and held for 5 minutes to anneal the strands. A solution of dNTPs was added, followed by the addition of Sequenase™ to begin the "fill-in" reaction (37°C for 30 min). Agarose gel electrophoresis showed the correct product size (data not shown). The product was then used as a template for PCR amplification to manufacture sufficient amounts of each DNA module. The PCR was conducted for each module using *PfuTurbo*® DNA polymerase and the corresponding amplification primers (Table 6.9 and Table 6.10). PCR was performed with 35 cycles at 95°C for 30 sec (denaturing), T_A°C for 30 sec (annealing), and 72°C for 50 sec (elongation). Large amounts of each double-stranded, blunt-ended DNA module were obtained. PCR reactions of each module were pooled together and the DNA was purified using the QIAquick® PCR purification kit and dried down using a Savant SpeedVac. DNA was then re-suspended in T4 DNA Ligase Buffer and the 5' ends were phosphorylated

with T4 polynucleotide kinase. DNA was purified by agarose gel electrophoresis and extracted into sterile water using the QIAquick® Gel Extraction kit. DNA concentration was determined by optical density (OD) at 260 nm.

Tail DNA Module #4		**T_A**
Tail4 Sense	5'-ggc gaa tct cgc gac aag ggc ggt gaa aaa **cgt gat tct gat cgc tct gaa c** -3'	61°C
Tail4 Antisense	5'- gcc gcg ttc gct gct gcg atc gct ac**g ttc aga gcg atc aga atc acg** -3'	
Tail4 Forward	5'- ggc gaa tct cgc gac aag gg -3'	61°C
Tail4 Reverse	5'- gcc gcg ttc gct gct gcg a -3'	

Table 6.9 Synthetic oligonucleotide and primer sequences used to construct Tail DNA Module #4. Bold-face represents bases involved in hybridization.

Anchor DNA Module		**T_A**
Anchor Sense	5'- gcc ggc gga ggc **ggt gaa gag gag gag gaa gag** -3'	61°C
Anchor Antisense	5'- tca gcc tca ggc ttc ctc ttc ttc **ctc ttc ctc ctc ctc ttc acc** -3'	
Anchor Forward	5'- gcc ggc gga ggc ggt g -3'	53°C
Anchor Reverse	5'- tca gcc tca ggc ttc ctc -3'	

Table 6.10 Synthetic oligonucleotide and primer sequences used to construct Anchor DNA Module. Bold-face represents bases involved in hybridization.

*Linearization of Vector with Sfo*I. The expression vector was re-suspended in 10X NEB Buffer 2, *Sfo*I (1 unit per 1 µg DNA) and sterile water to a concentration of 0.5 µg DNA/10 µL and incubated overnight (~16 hrs) at 37°C. The restriction enzyme was deactivated by heating to 65°C for 20 min. The 5'ends of the linearized vector were dephosphorylated to prevent self-ligation by adding 0.5 U of Alkaline Phosphatase, Calf Intestinal (CIP) per µg DNA. The reaction was incubated at 37°C for 4 hrs and then terminated by heating to 75°C for 10 min. The reactions were loaded onto an agarose gel for purification. The linear, dephosphorylated vector was extracted into sterile water using the QIAquick® Gel Extraction kit. DNA concentration was determined by optical density (OD) at 260 nm.

Blunt-end Ligation of Modules into Expression Vector. The blunt-end ligation procedure was adapted from a protocol obtained from Life Technologies. Blunt-end ligations were conducted at an insert-to-vector molar ratio of 3:1 in a 20 µl reaction mixture. This mixture consisted of 10X ligase buffer, 40-60 fmol of linearized vector, 120-180 fmol of insert DNA, 1 Weiss unit of T4 DNA Ligase and sterile water. The ligation mixture was incubated for ~24 hrs at 14°C. The entire ligation reaction was then transformed into OneShot® MAX Efficiency® DH5αT_1^R Chemically Competent *E. coli* and selected on LB Agar containing 100 µg/mL ampicillin after overnight incubation (~16 hrs) at 37°C.

Analysis of Insert Size – Colony PCR. Colonies (>20) were chosen and used to inoculate 50 µL of LB Broth containing 100 µg/mL ampicillin. Cultures were incubated for 3 hrs at 37°C. 1 µL of each culture was directly added to a 20 µL PCR reaction mixture containing 10X PCR buffer, dNTPs, synthetic primers (Table 6.11), sterile water and *Taq* DNA polymerase. PCR was performed with 20 cycles at 95°C for 45 sec (denaturing), 50°C for 30 sec (annealing), and 72°C for 30 sec (elongation). Samples from each PCR

reaction were analyzed by agarose gel electrophoresis (data not shown). The size of the PCR product (bp) = 186 bp + (N x (Module Size)), where N = number of inserts.

Primers		**T_A**
Trx Forward	5'- ttc ctc gac gct aac ctg -3'	50°C
pBADRev2 Reverse	5'- ctc atc cgc caa aac agc c -3'	

Table 6.11 Synthetic DNA primers used to analyze insert size. Bold-face represents bases involved in hybridization. PCR Product Size (bp) = 186 bp + N(module size) where N is the number of insertions into the vector.

Analysis of Insert Orientation – Colony PCR. Colonies containing the correct size insert were then analyzed by Colony PCR to check for insert orientation. 1 µL of each culture was directly added to a 20 µL PCR reaction mixture containing 10X PCR buffer, dNTPs, synthetic primers, sterile water and *Taq* DNA polymerase. PCR was performed with 20 cycles at 95°C for 45 sec (denaturing), 53°C for 30 sec (annealing), and 72°C for 30 sec (elongation). The synthetic primers used are given in Table 6.12. For each insertion, Thio Forward was combined with the reverse primer corresponding to the module inserted. For example, after inserting Tail DNA Module #3, Thio Forward and Tail3 Reverse were used. Samples from each PCR reaction were analyzed by agarose gel electrophoresis (data not shown). Correct insert orientation yielded a PCR product (bp) = 350 bp + (N x (Module Size)), where N = number of inserts. Plasmids from positive clones were submitted for DNA sequencing (Cleveland Genomics, Inc.) using Trx Forward/pBAD Reverse primers to validate the gene sequence and orientation. Stocks of positive clones were made in 15% glycerol and stored at -80°C.

Primers		**T_A**
Thio Forward	5'- tct gat aaa att att cat ctg act -3'	53°C
Tail1 Reverse	5'- tca ggc gcc gcg acc atc gct ctc -3’	
Tail2 Reverse	5’- gcc acc gtc ttt ttc ctt acc -3’	
Tail3 Reverse	5’- gcc gcc ctt gtc ctt gtc gc -3’	
Tail4 Reverse	5’- gcc gcg ttc gct gct gcg a -3’	
Anchor Reverse	5’- tca gcc tca ggc ttc ctc -3’	

Table 6.12 Synthetic DNA primers used to analyze insert orientation. PCR Product Size (bp) = 350 bp + N(module size) where N is the number of insertions into the vector.

Pilot Product Expression. Prior to expression, the recombinant plasmids were transformed into the expression strain, One Shot® TOP10 Chemically Competent *E. coli*. The transformation mixture was spread on LB agar plates containing 100 µg/mL ampicillin and incubated overnight at 37°C. A single colony was then used to inoculate 5 mL of LB medium containing 100 µg/mL ampicillin and incubated overnight at 37°C with shaking (225-250 rpm). The next day, 10 mL of LB medium containing 100 µg/mL ampicillin was inoculated with 500 µL of the overnight culture and incubated at 37°C with shaking (225-250 rpm) to midlog phase (OD_{600} = ~0.5). Expression was induced with the addition of L-(+)-arabinose to a final concentration of 0.02%. A 500 µL aliquot was removed, and the cells were pelleted by centrifugation at maximum speed in a microcentrifuge tube for 30 sec. The supernatant was removed and the pellet was stored at -20°C. This was denoted as the zero time point. Expression was continued for 5-6 hrs, then cells were harvested by centrifugation as above. Each pellet was re-suspended in 125 µL CelLytic™ B Bacterial

Cell Lysis Extraction Reagent and the cells were lysed repeating a freeze(-80°C)/thaw(42°C)/sonicate(10 sec) cycle 4 times. The cell debris was pelleted by centrifugation at maximum speed for 15 min. The soluble cell extract was removed. The protein content of the extracts was analyzed by SDS-PAGE and visualized using Coomassie Blue staining.

$TrxT_4A$ Larger Scale Expression. In a 250 mL culture flask, 50 mL of LB Broth containing 100 µg/mL ampicillin was inoculated with stock TOP10 bacterial culture containing the recombinant vector $pTrxT_4A$. The culture was incubated overnight at 37°C with shaking at 200 rpm. The culture was pelleted in a 50 mL centrifuge tube at 3000 X *g* at 4°C. The supernatant was removed and the pellet was re-suspended in 10 mL of fresh LB media. Four 1 L culture flasks with 250 mL of fresh LB media containing 100 µg/mL ampicillin were inoculated with 2 mL of this cell suspension. The culture was incubated at 37°C with shaking at 200 rpm until mid-log phase (OD_{600} ~ 0.5). Expression of $TrxT_4A$ product was induced with the addition of L-arabinose to a final concentration of 0.02%, and the cultures were incubated for an additional 5-6 hrs at 30°C. The cells were harvested by centrifuging at 3000 X *g* at 4°C, and the cell pellets were stored at -80°C.

$TrxT_4A$ Purification – Ni-NTA Spin Kit. This procedure was adapted from the Ni-NTA spin kit product manual. ~1 g cell pellet was re-suspended in 7 mL of 1X IMAC Binding Buffer (20 mM sodium phosphate, 500 mM NaCl pH 7.8). Cells were lysed by repeating a freeze(-80°C)/thaw(42°C)/sonicate(10 sec) cycle 4 times. The cell debris was pelleted by centrifugation at maximum speed for 15 min. The soluble cell extract was collected. 600 µL of extract was loaded per spin column. The column was centrifuged at 700 X *g* for 2 min. The flow-though was collected. The spin column was washed with 60 µL IMAC Wash Buffer (20 mM sodium phosphate, 500 mM NaCl pH 6.0) and then centrifuged at 700 X *g* for 2 min. This was repeated with another 600 µL of IMAC Wash Buffer. The $TrxT_4A$ product was eluted in two washes with 200 µL IMAC Elution Buffer (20 mM sodium phosphate, 500 mM NaCl pH 4.0). After centrifugation at 700 X *g* for 2 min, the eluate was collected and was adjusted to pH~7 with 1 M Tris pH8.0. The eluate was then dialyzed against water using Slide-A-Lyzer® Dialysis Cassettes (3.5 kDa molecular weight cutoff) at 4°C for 24 hrs. The samples were collected and product concentration was estimated by measuring the OD at 280 nm (assuming an extinction coefficient of 0.548 L/g·cm calculated based on the amino acid sequence of the product[18]). Samples were stored at -80°C. When required, protein was concentrated by drying down using a Savant SpeedVac, then re-suspended the sample to the desired concentration.

Enzymatic Cleavage. EKMax™ was used to remove the HP-Thioredoxin fusion partner. The procedure was supplied by the manufacture. 5 µg of $TrxT_4A$ product was re-suspended in 20 µL digest buffer containing 0.4 U of EKMax™. Reactions were incubated at 37°C for 0-8 hrs. The digestion reactions were monitored by SDS-PAGE and visualized by staining with silver or Colloidal Coomassie Blue.

Hydroxylamine Chemical Cleavage. The chemical cleavage procedure was adapted from published literature.[19] $TrxT_4A$ product was re-suspended in Reaction Buffer (2 M hydroxylamineHCl, 2 M guanidineHCl, 0.2 M K_2CO_3 pH 9.0) to a concentration 0.1-10 mg/mL. The reaction mixture was incubated at 45°C for 4-20 hrs. To stop the reaction, the reaction components were dialyzed away using Slide-A-Lyzer® Dialysis Cassettes (3.5kDa molecular weight cutoff) against deionized water at 4°C for 24 hrs or by the addition of 3vol of 2% TFA. The stopped reaction samples were collected and stored at -80°C. A control reaction was run simultaneously exactly as above without the inclusion of hydroxylamineHCl in the reaction buffer. The cleavage reactions were monitored by SDS-PAGE and visualized by staining with silver or Colloidal Coomassie Blue. For HPLC and MS analysis, samples were dialyzed against 20 mM Tris, pH 8.0 using Slide-A-Lyzer® Dialysis Cassettes (3.5 kDa molecular weight cutoff).

HPLC Purification. HPLC analysis of all protein samples was obtained on a Waters Alliance® HPLC system with: cooled autosampler chamber and column heater; a Waters 2996 photodiode array detector; and a Waters Fraction Collector. High performance ion exchange chromatography (HP-IEC) was performed on a Vydac DEAE 301VHP column (DEAE stationary phase, 5 µm particle size, 900 Å pore size, L = 5 cm, ID = 7.5 mm, column temperature = 24°C) with a Vydac CPF10 column prefilter. The mobile phases used were A: 7.5 mM Tris-HCl in 25% Tris-HCl, pH 8.0 in 25% acetonitrile, and B: 500 mM NaCl in Buffer A. Gradient programs are given in individual chromatograms. Fractions were collected

at 1 min intervals, frozen, then vacuum dried in a Labconco Centrivap. Dried samples were reconstituted in water prior to further analysis by MALD-TOF MS.

MALDI-TOF Mass Spectrometry. Samples were analyzed by MALDI-TOF Mass Spectrometry using an Applied Biosystems Voyager DE Pro or a Bruker OmniFlex (Nebraska Center for Mass Spectrometry at the University of Nebraska-Lincoln). Spectra were acquired manually in linear positive ion detection mode, with an accelerating voltage of 25 kV and a 300 nsec extraction delay time. A minimum of 400 shots per samples were used to obtain each final spectrum. Spectra y-axis represents %Intensity normalized to strongest peak. Spectra are presented without any smoothing. Before analysis, samples were mixed 1:1 with Sinapic Acid (Fluka) matrix solution (10 mg/mL Sinapic Acid in 50% acetonitrile/0.05% TFA). 1 µL of sample/matrix was spotted onto a 100-well stainless steel sample plate (ABI, Framingham, MA) and dried in ambient conditions. The instrument was calibrated using the following protein standards (Sigma, MS-CAL1): bovine insulin, equine Cytochrome C, equine apomyoglobin and rabbit muscle aldolase.

6.11 Supplemental Information

6.11.1 DNA Sequencing Results

The results from the DNA sequencing (conducted by Cleveland Genomics, Inc.) of the plasmids relevant to the work presented in this chapter are given below. The one-letter nucleotide abbreviations are used and the sequences are written in order of 5' → 3'.

Sequencing results for:	pTrx-T_1
Sequencing primer used:	Trx Forward (Invitrogen)

```
  1  cctttnnncn ggatcggtga tgacgatgac aagctcgccc ttaacggcaa
 51  gagcgaaggc gagaagcgca gcgataaaga gggtaaagag ggcggtgacc
101  gcaaagagag cgatggtcgc ggcgcctgaa agggcgagct tgaaggtaag
151  cctatcccta accctctcct cggtctcgat tctacgcgta ccggtcatca
201  tcaccatcac cattgagttt aaacggtctc cagcttggct gttttggcgg
251  atgagagaag attttcagcc tgatacagat taaatcagaa cgcagaagcg
```

Asn-Gly	bases 43-48
Tail (1)	bases 49-120
***Sfo*I site**	bases 121-126
Stop codon	bases 127-129

Table 6.13 DNA sequencing results for pTrx-T_1.

Sequencing results for:	pTrx-T_3
Sequencing primer used:	pBADRev2 (custom – see Table 6.11)

```
   1  gaaataaatt ttgtttaaac tttaaggaag gngatataca ataccccatg
  51  gggatctgat taaaattatt cattttgant tgaatgattc tttttganac
 101  tgatgtaact taaggcagat ggtgcaattc ttggttgatt tntggcacac
 151  tggtgcggtc cgngcaaaat gatcgctccg attttggatg aaatcgctga
 201  cgaatatcag ggcaaactga ccgttgcaaa actgaacatc gatcacaacc
 251  cgggcactgc gccgaaatat ggcatccgtg gtatcccgac tctgctgctg
 301  ttcaaaaacg gtgaagtggc ggcnaccaaa gtgggtgcac tgtntaaagg
 351  tcagttgaaa gagttcctcg acgctaacct ggccggctnt ggatccggtg
 401  atgacgatga caagctcgcc cttaacggca agagcgaagg cgagaagcgc
 451  agcgataaag agggtaaaga gggcggtgac cgcaaagaga gcgatggtcg
 501  cggcggcaaa gagggcgaga agggggaggg ccgtggccgc gagcgcggcg
 551  atggcgatcg tggtaaggaa aaagacggtg gcggcgaccg tggcaaagac
 601  agcagcaagg atcgtgacag ctctgaccgt ggtgaaaaaa gcgacaagga
 651  caagggcggc gcctgaaagg gcgagcttga aggtaagcct atccctaacc
 701  ctctcntcgg tctcgattct acgcgtaccg gtcatcatca ccatcaccat
 751  tggttnaacc gtctccnnna agnnnnnttt ttccct
```

Asn-Gly	bases 424-429
Tail (1)	bases 430-504
Tail (2)	bases 505-582
Tail (3)	bases 583-657
***Sfo*I site**	bases 658-663
Stop codon	bases 664-666

Table 6.14 DNA sequencing results for pTrx-T_3. (DNA sequences shown are reverse compliment of raw data).

Sequencing results for:	pTrx-T_4
Sequencing primer used:	pBADRev2 (custom – see Table 6.11)

```
  1  tgggtgcact gtctaaaggt cagttgaaag agttcctcga cgctaacctg
 51  gccggctctg gatccggtga tgacgatgac aagctcgccc ttaacggcaa
101  gagcgaaggc gagaagcgca gcgataaaga gggtaaagag ggcggtgacc
151  gcaaagagag cgatggtcgc ggcggcaaag agggcgagaa gggggagggc
201  cgtggccgcg agcgcggcga tggcgatcgt ggtaaggaaa aagacggtgg
251  cggcgaccgt ggcaaagaca gcagcaagga tcgtgacagc tctgaccgtg
301  gtgaaaaaag cgacaaggac aagggcggcg gcgaatctcg cgacaagggc
351  ggtgaaaaac gtgattctga tcgctctgaa cgtagcgatc gcagcagcga
401  acgcggcgcc tgaaagggcg agcttgaagg taagcctatc cctaaccctc
451  tcctcggtct cgattctacg cgtaccggtc atcatcacca tcaccattga
501  gtttaacgtc tccgaaagg
```

Asn-Gly	bases 93-98
Tail (1)	bases 99-173
Tail (2)	bases 174-251
Tail (3)	bases 252-329
Tail (4)	bases 330-404
***Sfo*I site**	bases 405-410
Stop codon	bases 411-413

Table 6.15 DNA sequencing results for pTrx-T_4. (DNA sequences shown are reverse compliment of raw data).

Sequencing results for:	pTrx-T_4A
Sequencing primer used:	pBADRev2 (custom – see Table 6.11)

```
  1  ggcgccaacc caaagtgggt gcactgtcta aaggtcagtt gaaagagttc
 51  ctcgacgcta acctggccgg ctctggatcc ggtgatgacg atgacaagct
101  cgcccttaac ggcaagagcg aaggcgagaa gcgcagcgat aaagagggta
151  aagagggcgg tgaccgcaaa gagagcgatg gtcgcggcgg caaagagggc
201  gagaaggggg agggccgtgg ccgcgagcgc ggcgatggcg atcgtggtaa
251  ggaaaaagac ggtggcggcg accgtggcaa agacagcagc aaggatcgtg
301  acagctctga ccgtggtgaa aaaagcgaca aggacaaggg cggcggcgaa
351  tctcgcgaca agggcggtga aaaacgtgat tctgatcgct ctgaacgtag
401  cgatcgcagc agcgaacgcg gcgccggcgg aggcggtgaa gaggaggagg
451  aagaggaaga agaggaagcc tga
```

Asn-Gly	bases 108-113
Tail (1)	bases 114-188
Tail (2)	bases 189-266
Tail (3)	bases 267-344
Tail (4)	bases 345-419
***Sfo*I site**	bases 420-425
Anchor	bases 426-470
Stop codon	bases 471-473

Table 6.16 DNA sequencing results for pTrx-T_4A. (DNA sequences shown are reverse compliment of raw data).

6.11.2 Amino Acid Sequences

The amino acid sequences of the polypeptides relevant to the work presented in this chapter are given in Table 6.17. The one-letter amino acid abbreviations are used and the sequences are written in order of amino terminus → carboxy terminus.

Polypeptide	Amino Acid Sequence
Trx-T_1	MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGP CKMIAPILDGIADGYQGKLTVAKLNIDHNPGTAPKYG IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA GSGSGDDDDKLALNGKSEGEKRSDKEGKEGGDRKES DGRGA
Trx-T_3	MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGP CKMIAPILDGIADGYQGKLTVAKLNIDHNPGTAPKYG IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA GSGSGDDDDKLALNGKSEGEKRSDKEGKEGGDRKES DGRGGKEGEKGEGRGRERGDGDRGKEKDGGGDRG KDSSKDRDSSDRGEKSDKDKGGA
Trx-T_4	MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGP CKMIAPILDGIADGYQGKLTVAKLNIDHNPGTAPKYG IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA GSGSGDDDDKLALNGKSEGEKRSDKEGKEGGDRKES DGRGGKEGEKGEGRGRERGDGDRGKEKDGGGDRG KDSSKDRDSSDRGEKSDKDKGGGESRDKGGEKRDSD RSERSDRSSERGA
Trx-T_4A	MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGP CKMIAPILDGIADGYQGKLTVAKLNIDHNPGTAPKYG IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA GSGSGDDDDKLALNGKSEGEKRSDKEGKEGGDRKES DGRGGKEGEKGEGRGRERGDGDRGKEKDGGGDRG KDSSKDRDSSDRGEKSDKDKGGGESRDKGGEKRDSD RSERSDRSSERGAGGGGEEEEEEEEEA

Table 6.17 Amino acid sequences of polypeptides relevant to Chapter 6.

6.12 References

[1] Krsmanovic JL. Ph.D. Thesis, Virginia Polytechnic Institute & State University, Virginia, USA, 2003.

[2] Robinson CR and Sauer RT. "Optimizing the stability of single-chain proteins by linker length and composition mutagenesis". *Proc. Natl. Acad. Sci. USA*. (1998); 95: 5929-2934.

[3] Pinaud F, King D, Moore H-P and Weiss S. "Bioactivation and cell targeting of semiconductor CdSe/ZnS nanocrystals with phytochelatin-related peptides". *J. Amer. Chem. Soc*. (2004); 126: 6115-6123.

[4] Jonsson B, Khan MO, Askesson T and Woodward CE. "Modulation of colloidal forces with polyampholytes". *Langmuir*. (2002); 18(4): 1426-1432.

[5] Zhulina EB, Dobrynin AV and Rubinstein M. "Adsorption of polyampholytes at charged spherical particles". *J. Phys. Chem. B.* (2001); 105(37): 8917-8930.

[6] Dobrynin AV, Rubinstein M and Joanny JF. "Polyampholyte solutions between charged surfaces: DeBye-Huckel theory". *J. Chem. Phys.* (1998); 109(20): 9172-9176.

[7] McGrath KP, Tirrell DA, Kawai M, Mason TL and Fournier M. "Chemical and biosynthetic approaches to the production of novel polypeptide materials ". *Biotechnol. Progr.* (1990); 6: 188-192.

[8] Trinh T, Jessee J, Bloom FR and Hirsch V. "STBL2 : An *Escherichia coli* strain for the stable propagation of retroviral clones and direct repeat sequences". *Focus*. (1994); 16: 78-80.

[9] Marques CM and Joanny JF. "Block Copolymer Adsorption in a Nonselective Solvent". *Macromolecules.* (1989); 22: 1454-58.

[10] Wu DT, Yokoyama A and Setterquist RL. "An Experimental Study on the Effect of Adsorbing and Non-adsorbing Block Sizes on Diblock Copolymer Adsorption". *Polym. J.* (1991); 23(5): 709-714.

[11] Miyada CG, Stoltzfus L and Wilcox G. "Regulation of the *ara*C gene of *Escherichia coli*: catabolite repression, autoregulation and effect on *ara*BAD expression". *Proc. Natl. Acad. Sci.* (1984); 81: 4120-4124.

[12] Lee N, Francklyn C, and Hamilton EP. "Arabinose-induced binding of AraC Protein to $ara I_2$ activates the *ara*BAD operon promoter". *Proc. Natl. Acad. Sci.* (1987); 84: 8814-8818.

[13] Stoll VS, Manohar AV, Gillon W, Macfarlane ELA, Hynes RC and Pai EF. "A thioredoxin fusion protein of VanH, a D-lactate dehydrogenase from *Enterococcus faecium*: cloning, expression, purification, kinetic analysis and crystallization". *Protein Sci.* (1998); 7(5): 1147-1155.

[14] Bornstein, P. and Balian, G. "Cleavage at Asn-Gly bonds". *Method. Enzymol.* (1977); 47: 132-145.

[15] Canova-Davis E, Eng M, Mukku V, Reifsnyder DH, Olson CV, and Ling VT. "Chemical heterogeneity as a result of hydroxylamine cleavage of a fusion protein of human insulin-like growth factor I". *Biochem. J.* (1992); 278: 207-213.

[16] Kwong MY and Harris RJ. "Identification of succinimide sites in proteins by N-terminal sequence analysis after alkaline hydroxylamine cleavage". *Protein Sci.* (1994); 3: 147-149.

[17] Hiller Y, Bayer EA and Wilchek M. "Studies on the biotin-binding site of avidin. Minimized fragments that bind biotin". *Biochem. J.* (1991); 278: 573-585.

[18] Gill SC, Von Hippel PH. "Calculation of protein extinction coefficient from amino acid sequence data". *Anal. Biochem.* (1989); 182(2): 319-326.

[19] Walker, JM. *The Protein Protocols Handbook*. Humana Press Inc., Totowa, NJ, (1996).