5 The Development of a New Cloning Strategy and the Demonstration of its Feasibility*

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

5.1 Overview

During the course of our experiments constructing the Pro_mGlu_n diblock in Chapter 3, it became evident that a different approach was necessary to attain our goal of producing high molecular weight surface-active PAA's. The limiting factor was that the conventional cloning techniques used did not permit us to construct the long DNA sequences that were required to encode for a higher molecular weight tail block. Thus, we could not produce a PAA diblock with a sufficiently long tail block for optimum brush extension. Therefore, alternative cloning strategies were investigated.

In the literature, several techniques have recently been developed for the production of synthetic structural protein polymers.[1,2,3,4] We considered each technique to determine if it could be applied to the production of our PAA design. Although these techniques were successful for their specific applications, we found they were still inadequate for our purposes. Most of these techniques relied on a limited pool of restriction enzymes, inefficient reactions (*e.g.* head-to-tail self-ligation of short DNA blocks), and required numerous experimental steps. These methods were also relatively complicated and provided a low level of control and flexibility. The attempt to apply one of these methods towards the construction of a gene encoding for a high molecular weight PAA diblock is presented in Section 5.3.

After several failed attempts to apply these techniques, we concluded that we must develop a new cloning strategy. Section 5.4 presents the development of a new universal cloning technique for the streamlined production of surface-active PAA's that enables a high level of control over polymer composition and molecular weight, while also being relatively simple and flexible. Chapter 6 will present how we applied this method to the production of a unique surface-active PAA designed to form brush layers on aluminum oxide surfaces. A manuscript submitted to *Biomacromolecules* based partly on the work presented in this chapter is located in Appendix A.

[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).

5.2 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*

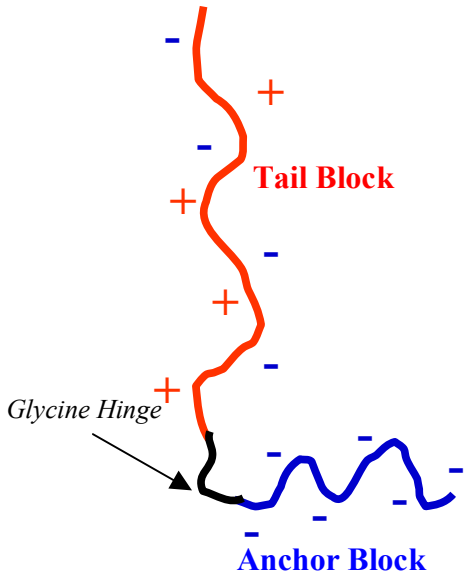

Figure 5.1 Schematic of new brush-forming PAA design.

For the application of alternative cloning strategies, we chose to alter our PAA design. In Chapter 3, a PAA diblock was designed to have the composition Pro_mGlu_n. For all subsequent work, we altered our tail and anchor block compositions. The reason for this change will be discussed in more depth in Chapter 6. For now, it suffices to say that our brush-forming PAA construct will consist of a zwitterionic tail block and an acidic anchor block. Briefly, the zwitterionic tail block will consist of an equal number of acidic and basic amino acids (resulting in a net zero charge near pH7) and the anchor block will remain acidic, but will include a glycine 'hinge' region (which should promote tail extension). A schematic of the new PAA diblock design is given in Figure 5.1. For convenience, we will use DNA sequences based on this PAA design during the investigation of alternative cloning strategies, while the theory behind this PAA design and its production/characterization will be discussed in depth in Chapter 6.

5.3 Application of the Cloning Strategy Developed by McMillan *et al.* for the Production of High Molecular Weight Brush-Forming PAA's

McMillan *et al.* developed a cloning method for the production of elastin-like structural protein polymers.[5] Their method centers on the use of DNA 'monomers' that have complimentary overhangs (or 'sticky ends') at the 5' and 3' ends. Because of these sticky ends, the DNA monomers can be self-ligated together in a head-to-tail manner, forming DNA 'concatemers'. We applied this method in order to ligate together short DNA monomers encoding for zwitterionic PAA sequences. The goal was to construct a

long genetic sequence encoding for a high molecular weight tail block. The desired amino acid sequence of the zwitterionic PAA module is given in Figure 5.2. From this PAA sequence, the corresponding DNA sequence was constructed using the genetic code (Figure 5.3).

Zwitterionic PAA Sequence:

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

Figure 5.2 Amino acid sequence and charge structure of the PAA sequence encoded by the DNA Monomer. G = Gly; S = Ser; E = Glu; D = Asp; K = Lys; R = Arg.

DNA Monomer Sequence:

5' – agac ctcttcg |**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** | cgaaga gattg – 3'

Figure 5.3 DNA monomer sequence (bold) encoding for the PAA sequence given in Figure 5.2. Underlined base pairs represent *Eam*1104I recognition sites and "|" represents cut site.

The DNA monomer was constructed synthetically by the hybridization of 2 complimentary synthetic oligonucleotides. The resulting DNA monomer (see Figure 5.3) contained restriction enzyme sites at the 5' and 3' ends. The restriction enzyme *Eam*1104I was used for its unique ability to cut outside of its recognition sequence. This is an appealing property since the enzyme does not leave extraneous nucleotides that would result in unwanted amino acids in the final PAA sequence. Digestion with *Eam*1104I, results in DNA monomers with complimentary 5' and 3' ends. These monomers were self-ligated using T4 DNA Ligase, resulting in concatemers of increasing size. When run on an agarose electrophoresis gel, the range of concatemers appeared as a 'ladder'. A schematic of this method as well as a gel showing the ligation results are located in Figure 5.4.

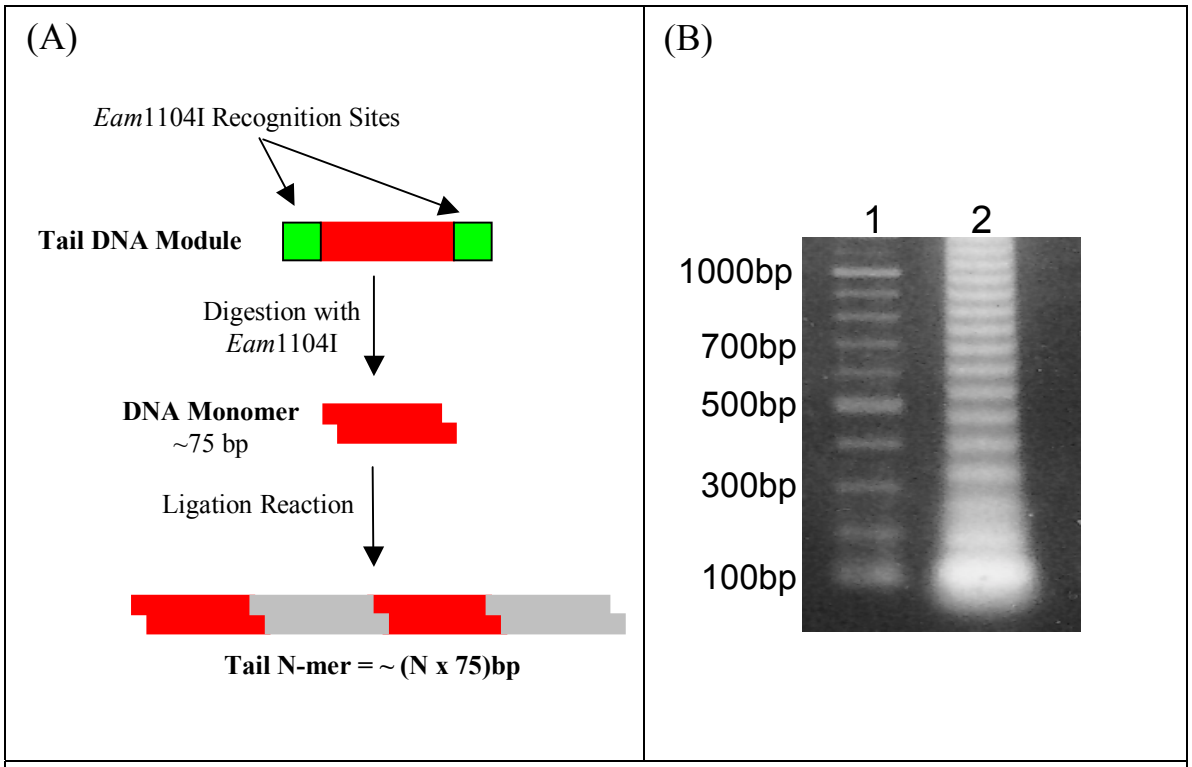

Figure 5.4 Application of the strategy developed by McMillan *et al.* (A) Schematic of the cloning method developed by McMillan *et al.* as applied to the ligation of our DNA modules; (B) agarose electrophoresis gel analysis of the self-ligation of the tail monomers (1% agarose, visualized with ethidium bromide): Lane 1, 100 bp DNA ladder (NEB); Lane 2, ligation reaction.

A DNA module encoding for an acidic anchor block was constructed by hybridizing two synthetic oligonucleotides in the same manner as the tail DNA monomer. However, for the anchor block, an *Eam*1104I site was only engineered into the 5' end. After digestion of the resulting double stranded anchor DNA module with *Eam*1104I, the anchor module was incubated with a mixture of tail concatemers and ligated to their 3' end. Next, a short DNA linker encoding for a chemical cleavage site was ligated onto the 5' end in the same manner. A schematic of the final construct is given in Figure 5.5.

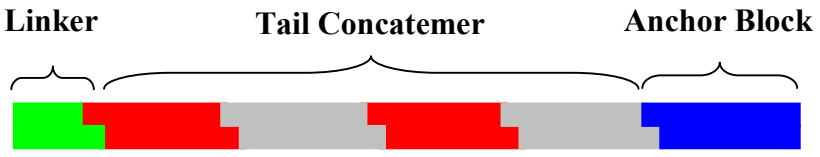

Figure 5.5 Schematic of DNA encoding for PAA diblock. Linker containing chemical cleavage site (green), zwitterionic tail concatemer (red/grey) and anchor block (blue).

Finally, we attempted to clone the DNA construct seen in Figure 5.5 into a plasmid that would express it with an amino-terminal poly-histidine fusion tag (pCR®T7/NT-TOPO®). However, we were unable to obtain any recombinant plasmids containing the 'Linker-Tail-Anchor' construct with more than one or two tail DNA monomers. Even when only the high molecular weight tail concatemers (500-1000 bp) were included in the ligation reactions, we could not isolate recombinant plasmids containing sequences encoding for longer tail blocks.

We, along with the research group of Won and Barron,[4] subsequently learned that there were inherent problems in adapting the method developed by McMillan *et al.* to the construction of DNA molecules longer than 400-700 bp. Won and Barron showed that these long DNA molecules with complimentary sticky ends readily circularize before insertion into a recipient plasmid. Won and Barron observed this circularization at greater than 6 repeats of a 63 bp monomer, or ~380 bp. Therefore, we would expect our concatemers to circularize at greater than 4 repeats of our ~75 bp tail DNA monomer.

Consequently, we determined that the method developed by McMillan *et al.* was unsuitable for reaching our goal of obtaining a high molecular weight PAA with a defined composition. Won and Barron proposed their own method of building long repeats of short DNA sequences using an engineered intermediate cloning vector.[4] For this method, short DNA sequences were ligated together in succession. Once a gene consisting of a high number of repeats was obtained, the entire gene was cloned into a separate expression vector. We attempted to apply this method as well (data not shown). However, it became apparent that this method also lacked the flexibility we were looking for in a cloning strategy. After the final gene was cloned into the expression vector, it could not be easily altered without re-starting the gene assembly process from the beginning. In addition, the cloning steps of the Won and Barron method were numerous and time consuming. The limitations of these two methods led us to develop an improved and more flexible cloning technique for producing PAA's for brush-forming applications.

5.4 Specific Aim #2: *Develop a new cloning strategy to meet the need for a simple and universal method for constructing the genes required for producing specifically designed, high molecular weight brush-forming PAA's in genetically engineered E. coli*

After several attempts using existing cloning strategies, it became evident that a suitable cloning strategy was not currently available for building long DNA sequences encoding for high molecular weight surface-active PAA's. Most of the newer cloning strategies developed for producing synthetic structural protein polymers were very successful, however, only for their specific applications. Therefore, we set out to develop an improved cloning strategy that would be effective in producing surface-active PAA's, yet also be easier, quicker and more flexible. Being able to readily modify the PAA within the expression vector facilitates the rapid investigation of how these modifications affect the resulting PAA product. The following sections discuss our new cloning strategy and present preliminary experimental results that confirm its feasibility.

5.4.1 Features of the New Cloning Strategy

For the task of producing high molecular weight surface-active PAA's we developed a cloning strategy with the goal of incorporating the following features:

1) **Modular** – Small DNA molecules can be synthesized, assembled, and inserted directly into an *expression vector* in *modules*. Thus, an intermediate cloning vector is not needed. These modules of DNA would encode for modules of the final PAA.

2) **Sequentially-modifiable** – The DNA modules can be *sequentially inserted* directly into the *expression vector*. That is, the vector can be re-opened and additional DNA modules can be inserted. The new insertion points can be either at the 5' or 3' end of the previously inserted DNA module. Thus, the PAA composition can be sequentially modified by adding to either its amino- or carboxy-terminus without having to re-start the gene assembly process from the beginning.

3) **Minimal Amino Acid Requirement** – The composition of the resulting PAA should be minimally impacted by the restriction enzymes and cloning method used.

4) **Fusion Tag / Removal** – To aid in PAA purification and detection, there should be the option to include a fusion tag (*e.g.* poly-histidine, HP-Thioredoxin). Additionally, a method to remove the fusion tag after the fusion is expressed should be included. Since removal by enzymatic cleavage is often expensive and inefficient, a chemical

cleavage site (*e.g.* hydroxylamine, CNBr) can be inserted into the junction between the fusion tag and PAA.

5) **Versatility and Transferability** - The cloning strategy should be simple, using only commercially available expression vectors (for prokaryotic, yeast and mammalian expression systems) and molecular biology reagents, and should be easily transferable to other academic laboratories that have basic molecular biology and protein production expertise.

5.4.2 Overview of Strategy

We introduce a new cloning strategy for the production of synthetic polypeptides that is more flexible and will require fewer intermediate experimental steps than those currently used. The strategy that we developed can be summarized as follows:

1. The first recombinant DNA module (DNA Module #1) is inserted via TOPO® TA cloning into an unmodified expression vector (*e.g.* pCR®T7/NT-TOPO®, pBAD/Thio-TOPO®). TOPO® TA cloning is simple and rapid (5 minutes).

2. DNA Module #1 contains a **unique** *Sfo*I restriction enzyme site (5'-*ggc*|*gcc*-3'). This *Sfo*I site can be located anywhere within DNA Module #1. Since this is where the subsequent blocks will be inserted, it will likely be located at either the 5' or 3' end. The DNA sequence 5'-*ggc gcc*-3' encodes for Gly-Ala.

3. The new expression vector containing the DNA Module #1 is then digested with *Sfo*I, and DNA Module #2 is inserted via blunt-end ligation. An appropriate double restriction digest or PCR can be used after each insertion to confirm insert size and orientation. The only requirements for DNA Module #2 are that there are no internal *Sfo*I sites, and the DNA encoding DNA Module #2 has a *ggc* codon (Gly) at the 3' end (**Strategy #1**) or a *gcc* codon (Ala) at the 5' end (**Strategy #2**) so that the unique *Sfo*I site is regenerated for subsequent insertions. The recombinant PAA gene can be expressed, or it can be modified further.

4. Step 3 is repeated until the gene encoding for the desired PAA is complete. As with any cloning method, the DNA should be sequenced before proceeding to expression.

Schematics illustrating the two strategies are shown in Figure 5.6 (**Strategy #1**) and Figure 5.7 (**Strategy #2**). The subtle differences in these strategies (see Step 3 above) allow for interesting experimental possibilities. **Strategy #1**, for example, can be used when one wishes to sequentially insert modules into the growing gene that modify the C-terminus of the growing PAA. Alternatively, **Strategy #2** can be used when one has a known '*cap domain*' (Module #1 in Figure 5.7) and wishes to insert a varying number of

identical modules as 'linker blocks' (Module #3 in Figure 5.7) between it and the module at the C-terminus of the PAA (Module #2 in Figure 5.7). The effect of the length of the 'linker' region (varying the number of modules inserted between #1 and #2) can then be easily investigated.

The amino acid requirement of a Gly-Ala dipeptide sequence is minimal since glycine and alanine possess the two smallest amino acid side chains. Because of their size, they are often found in loop structures in naturally occurring proteins. These loops can impart structural flexibility between the blocks. This is a desirable feature for solvated tail blocks or for the tail-anchor block hinge region. If Gly-Ala sequences are required to occur within any of the modules, the degeneracy of the genetic code can be used to ensure that these do not contain an *Sfo*I recognition site (*e.g.* 5'-*gga gca*-3' encodes for Gly-Ala but will not be digested by *Sfo*I).

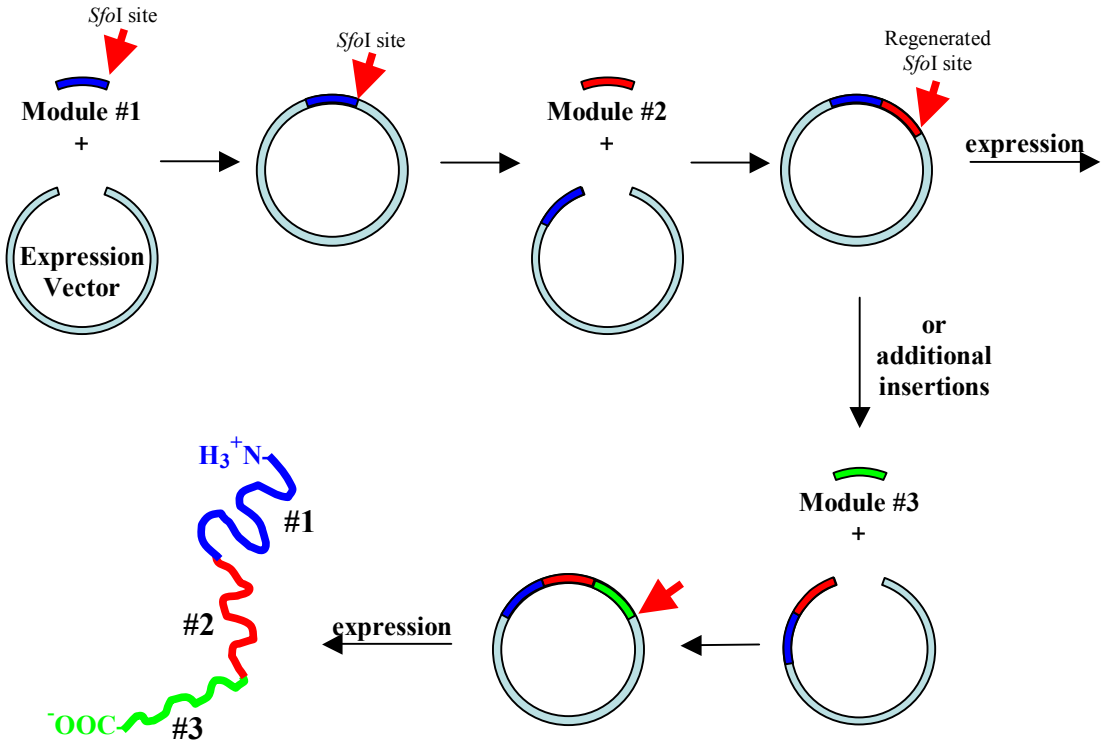

Figure 5.6 Outline of Cloning Strategy #1.

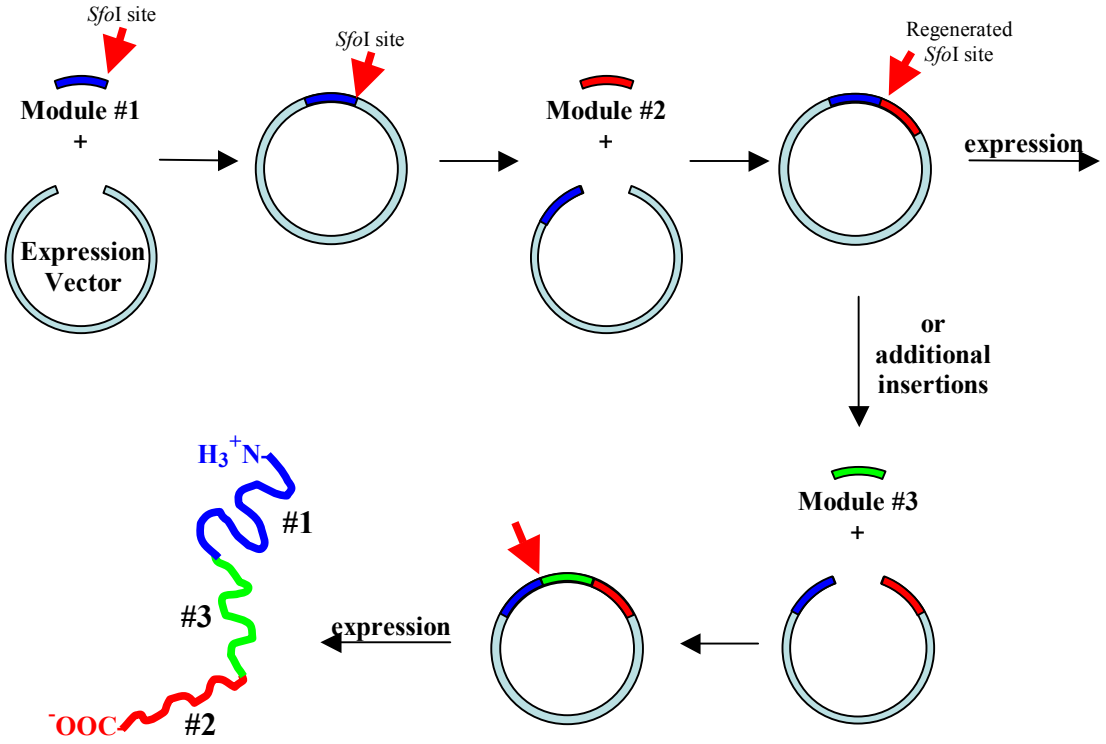

Figure 5.7 Outline of Cloning Strategy #2.

5.4.3 Demonstrating Feasibility of Specific Aim #2

Preliminary experiments were conducted initially to demonstrate the feasibility of the new cloning method outlined above. We accomplished this by using a variation of **Strategy #2** to sequentially insert short DNA modules into a commercially available expression vector. The success of these preliminary experiments confirmed the feasibility of our strategy. For convenience, we used DNA sequences based on the PAA design introduced in Section 5.2.

Choosing Expression Vector

We began by identifying commercially available prokaryotic expression vectors that were suitable for our cloning strategy. It was of great importance that these vectors did not contain any internal *Sfo*I recognition sites. It was also desired that the vectors express the PAA as a fusion protein with a tag that would facilitate purification and detection. We found that Invitrogen's pCR®T7/NT-TOPO® and pBAD/Thio-TOPO® expression vectors

met our criteria. These vectors are used to express recombinant proteins with N-terminal polyhistidine (His_6) and HP-Thioredoxin (Trx) tags, respectively, in *E. coli*. These tags help simplify purification and offer epitopes for immunodetection by Western blot with commercially available antibodies. For the feasibility studies, we chose to use the pCR®T7/NT-TOPO® expression vector.

It should be noted that the cloning strategy outlined above is also compatible with commerically available expression vectors used with the methylotrophic yeast *Pichia pastoris* (pPICZ, pPICZα, pPIC6, pPIC6α, pPICZ-E, pPICZα-E, pGAPZ, and pGAPZα) and certain mammalian cells (pSecTag2 and pSecTag2/Hygro). Therefore, expression systems other than *E. coli* could conceivably be used to produce sequentially modified proteins using this versatile cloning strategy.

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

Next, the unique *Sfo*I restriction site was introduced into the vector by inserting an initial DNA module via TOPO® TA cloning. The restriction site allows for the "opening" of the vector for all subsequent insertions of DNA modules. The sequence of the initial DNA module used here is given in Figure 5.8 and encodes for the PAA sequence given in Figure 5.9. This PAA sequence was designed as an acidic anchor block consisting of a random sequence of glutamate and aspartate residues.

The unique *Sfo*I recognition site (5' *ggc*|*gcc* 3') was incorporated into the module sequence and was placed towards the 5' end. This allowed for subsequent insertions of DNA modules *upstream* of this module (the Anchor DNA 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 added at the 3' end of the module. The STOP codon signals to the protein translation machinery of the host organism to stop making the polypeptide strand at that point.

Anchor DNA Module:

5'- aac ggc | gcc ggc gga ggc ggt gat gac gat gag gat gaa gat gag gaa gat gac gag gac gac gag gat gcc tga -3'

Figure 5.8 Anchor DNA Module sequence. Underlined base pairs represent *Sfo*I recognition site and "|" represents cut site.

Anchor Block PAA Sequence:

N G A G G G G D D D E D E D E E D D E D D E D A *

Figure 5.9 Amino acid sequence encoded by Anchor DNA Module. "NG" represents Asn-Gly hydroxylamine cleavage site; 'blue' residues represent acidic residues, E=Glu, D=Asp; "GA" represents residues encoded by *Sfo*I cloning site; and "*" represents STOP codon.

The Anchor DNA Module 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. The A-overhangs are necessary for the TOPO® TA cloning step. A schematic showing the construction of the module and its insertion into the expression vector is shown in Figure 5.10. Once inserted, orientation of the module is confirmed by double restriction enzyme digest using *Xba*I and *Sfo*I. The resulting vector, denoted as pHis$_6$-A, now contains the unique *Sfo*I site required for the application of the new cloning strategy to sequentially insert additional DNA modules. The only requirement is that the *Sfo*I site is regenerated after each insertion.

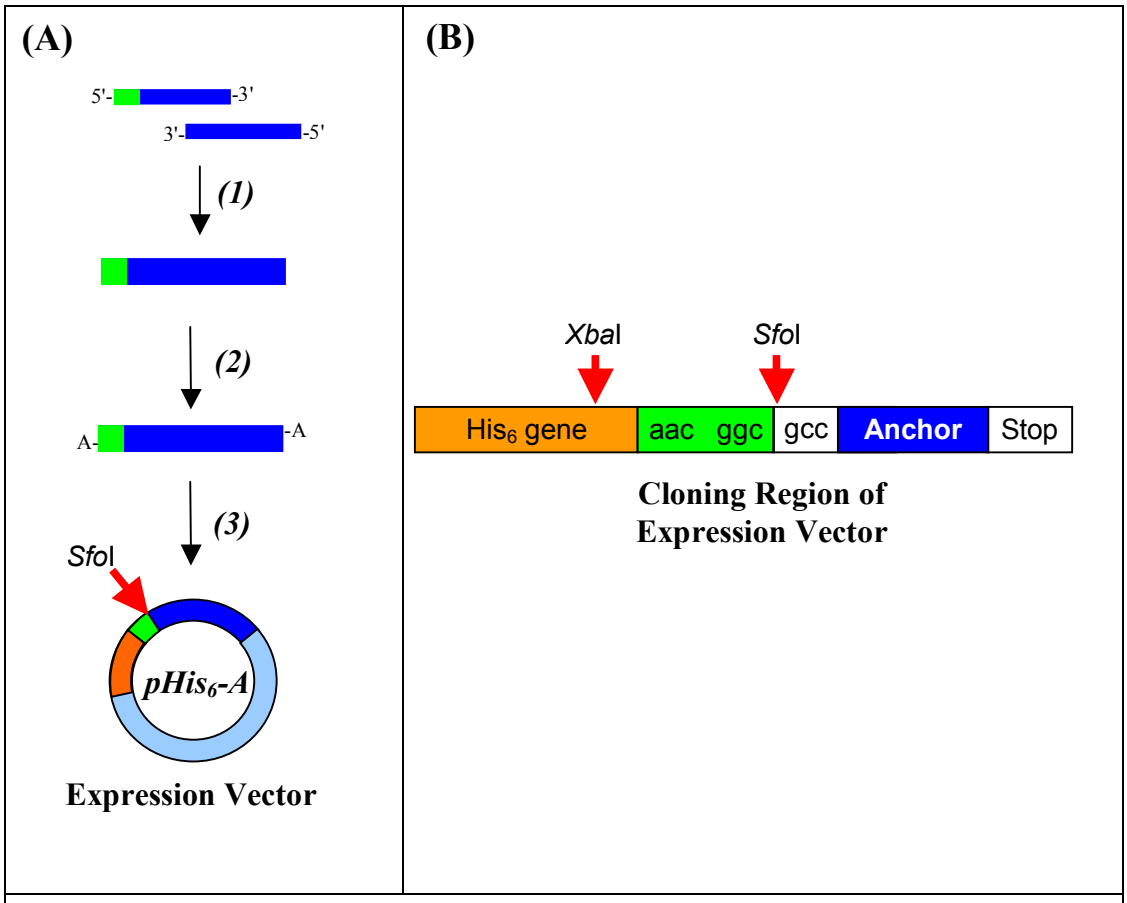

Figure 5.10 Insertion of initial DNA module into the expression vector. (A) Construction of Anchor DNA Module and insertion into expression vector. (1) Hybridization and fill-in of single-stranded synthetic oligonucleotides; (2) Addition of 3' A-overhangs with *Taq* DNA Polymerase; (3) TOPO® TA cloning into pCR® T7/NT-TOPO®. (B) Schematic showing the cloning region including the His_6-tag gene (orange), Asn-Gly dipeptide sequence for hydroxylamine cleavage (green), Anchor DNA Module (blue) and RE recognition sites (red arrows). *Xba*I site is present in the vector and is used to check for insert orientation.

Construction of DNA Modules and Their Insertion into the Expression Vector

In order to confirm the feasibility of our new strategy, we sequentially inserted DNA modules encoding for zwitterionic PAA sequences directly into the expression vector, $pHis_6$-A. Three different DNA modules, encoding for 3 different zwitterionic PAA sequences, were used. The PAA sequences encoded by these DNA modules are given in Figure 5.11. Each DNA module was constructed by annealing two overlapping synthetic oligonucleotides, as with the Anchor DNA Module. However, once the strands were filled in with Sequenase™, each module was PCR amplified using synthetic DNA primers and *PfuTurbo*® DNA Polymerase. This resulted in double-stranded DNA modules with blunt ends (see *Materials and Methods* for DNA sequences of

oligonucleotides and primers). The 5' ends of each module were phosphorylated with T4 polynucleotide kinase prior to insertion into the linearized expression vector via blunt end ligation (Figure 5.12).

Zwitterionic PAA Sequence - A:

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 Sequence - B:

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 Sequence - C:

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

Figure 5.11 Amino acid sequences and charge structure of the three zwitterionic PAA blocks used during the feasibility experiments. G = Gly; S = Ser; E = Glu; D = Asp; K = Lys; R = Arg.

Linearization of the expression vector was accomplished by digesting it with *Sfo*I, which cuts the plasmid DNA at the unique (5' *ggc*|*gcc* 3') site, leaving blunt ends. To prevent self-ligation of the vector during the insertion step, the 5' ends of the linearized vector were dephosphorylated using CIP. The preparation of the linearized plasmid is shown in Figure 5.12.

Finally, the DNA module and the linearized vector were ligated together using T4 DNA Ligase. A 3:1 insert-to-vector molar ratio was used during the ligation reactions. The reactions were incubated for 24 hrs at 14°C and then transformed into One Shot® TOP10 Chemically Competent *E. coli* and selected on LB Agar containing ampicillin. Colonies were analyzed for the presence of the recombinant vector containing the new insert. The entire process of inserting a DNA module into the linearized expression vector is summarized in Figure 5.12.

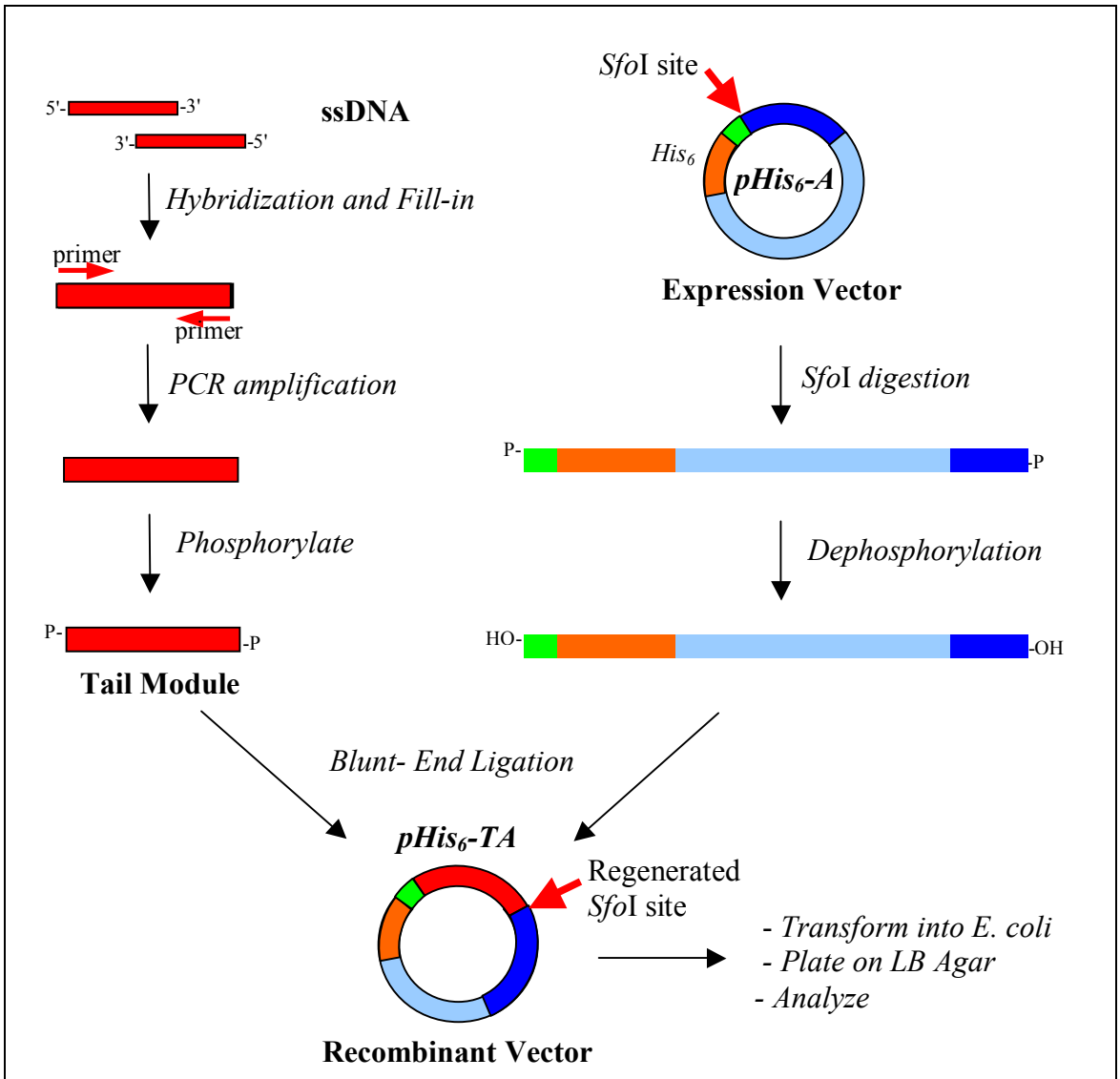

Figure 5.12 Summary of the blunt-end ligation of a DNA module into a linearized expression vector.

The cloning process used in the feasibility studies is summarized in Figure 5.13. Three tail DNA modules were sequentially inserted in the order of Tail A, B then C, generating the following recombinant vectors: pHis$_6$-TA (1 tail block module and 1 anchor block module), pHis$_6$-T$_2$A (2 tail block modules and 1 anchor block module) and pHis$_6$-T$_3$A (3 tail block modules and 1 anchor block module).

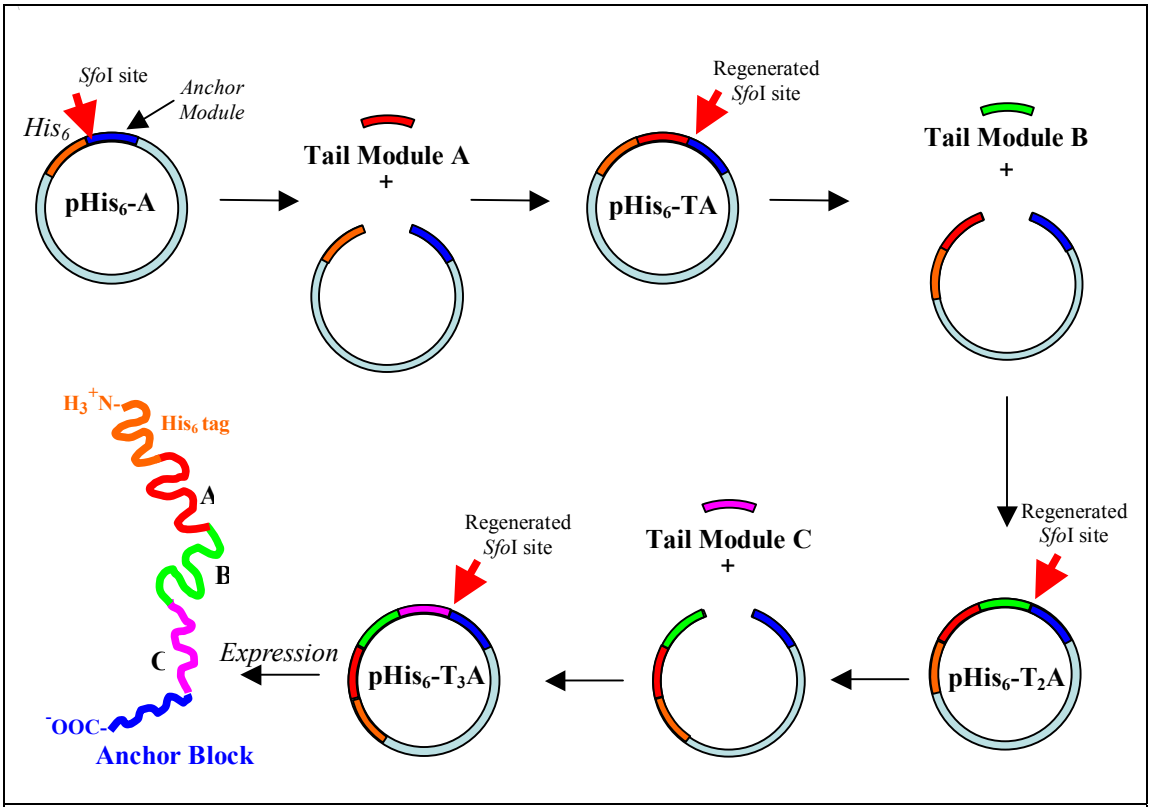

Figure 5.13 Outline of procedure used for feasibility studies (based on Cloning Strategy #2).

Analysis of the Insert Size Using Colony PCR

During blunt-end ligation, it is possible for multiple copies of the modules to be inserted. Even though we minimized this event by optimizing the insert:vector ratio used during the ligation reaction, it was still necessary to check the size of the insert. This can be easily accomplished by colony PCR after transformation, using primers (T7 Forward and pRSET Reverse) designed to hybridize upstream and downstream of the insert. Therefore, after each module was inserted, the PCR product increased by the size of that module (Figure 5.14A). Figure 5.15 shows an agarose gel of the PCR products obtained after the insertion of each tail DNA module (~80 bp in size).

Analysis of the Insert Orientation Using Restriction Enzymes

It is also necessary to confirm the orientation of the insert because blunt-end ligations are non-directional. This was accomplished by double restriction digest. As can be seen in Figure 5.14B, the restriction enzyme *Xba*I cuts within the original vector

sequence upstream of the insertion site (within the His_6 gene). The location of the *Sfo*I site depends on the orientation of the inserted module. In the case shown in Figure 5.14B, if inserted correctly, the restriction fragment size will be proportional to the size of the module inserted. On the other hand, if inserted incorrectly, the location of the *Sfo*I site will change, resulting in a restriction fragment of a different size.

The combination of the colony PCR and double restriction digest analyses were used to identify the clones that contained properly inserted DNA modules. Positive recombinant vectors were then isolated and submitted for DNA sequencing to confirm the fidelity of the cloning procedure (see Section 5.7). Stock cultures containing each recombinant vector were made in 15% glycerol and stored at -80°C.

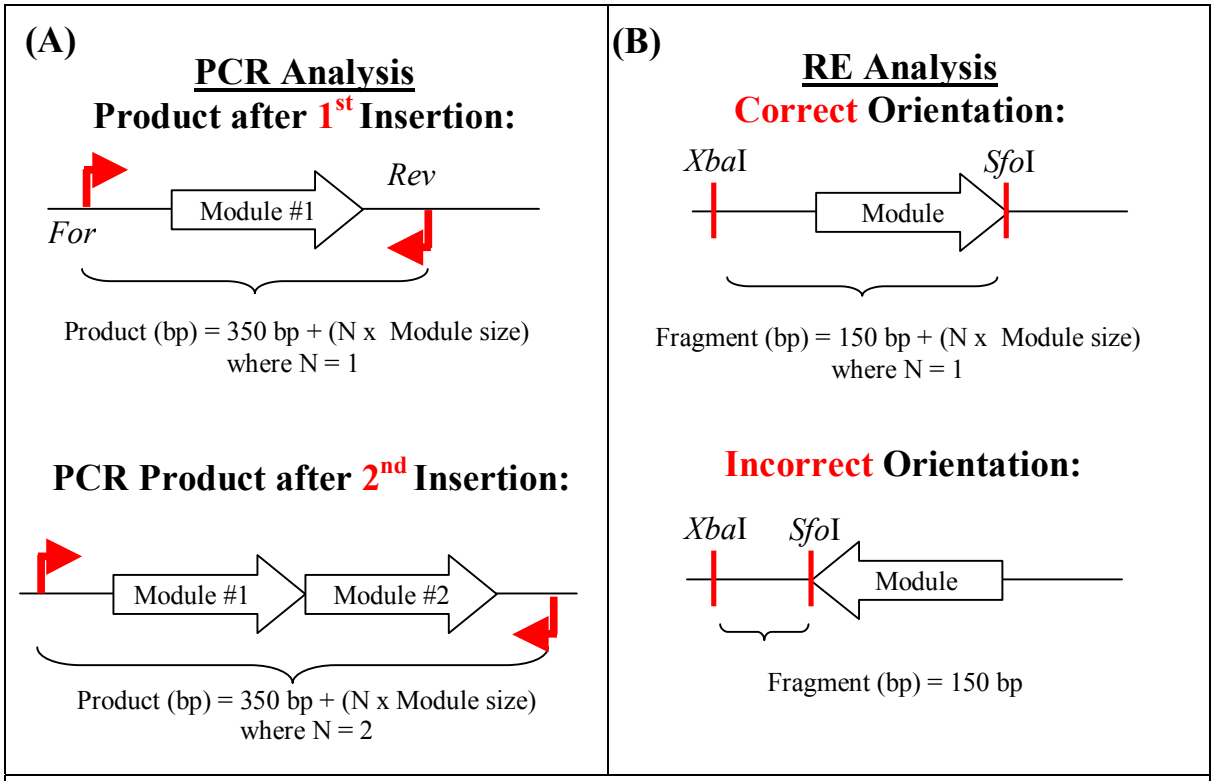

Figure 5.14 Schematic showing the methods used for analyzing transformant colonies after each insertion. (A) PCR analysis of insert size: PCR product increases in size with insertion of modules between flanking primer sites. (B) Restriction enzyme (RE) analysis of insert orientation: if orientation is correct, fragment will increase is size with additional insertions.

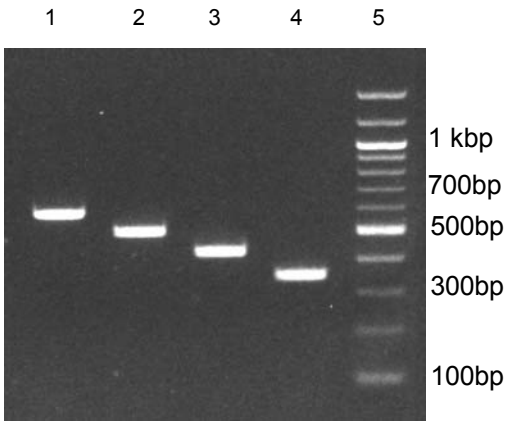

Figure 5.15 Analysis of insert size using PCR. Agarose gel electrophoresis (1.2% agarose, visualized with ethidium bromide) of PCR analysis conduct after each insertion. Lane 1, $pHis_6$-T_3A; Lane 2, $pHis_6$-T_2A; Lane 3, $pHis_6$-TA; Lane 4, $pHis_6$-A; and Lane 5, 100 bp DNA Ladder (NEB). The size of the PCR product is (350 + N(80))bp where N is the number of tail block modules

Analysis of Blunt-End Ligation Efficiency

Blunt-end ligation is considered less efficient when compared to cohesive-end ligation. However, we have shown that we were able to optimize the blunt-end ligation reaction conditions so that we achieved relatively good efficiency. Table 5.1 shows the efficiencies for the 4 sequential DNA module insertions discussed above, including the TOPO® TA cloning step. The efficiencies presented are based on the number of colonies with correct insertion size *and* orientation out of the number of colonies analyzed after each step (>20 colonies).

Step	Insertion Method	Efficiency
Vector + Anchor Module	TOPO® TA cloning	20%
+ 1st Tail Module	Blunt-end ligation	25%
+ 2nd Tail Module	Blunt-end ligation	30%
+ 3rd Tail Module	Blunt-end ligation	20%

Table 5.1 Efficiencies of the insertion steps made during the feasibility experiments using TOPO® TA cloning and blunt-end ligation. Efficiency was calculated by the number of isolated colonies having the correct insert size *and* orientation divided by the number of transformant colonies analyzed (>20).

Results of the Feasibility Study

Using our new cloning strategy we were able to successfully make several sequential insertions directly into an unmodified, commercially available expression vector. After DNA sequencing analysis, we have confirmed the construction of a PAA gene enoding for the composition: $(Tail)_3$-(Anchor). The PAA sequence encoded by the $pHis_6$-T_3A vector is given in Figure 5.16. We conclude that the successful results presented here confirm the feasibility of our new cloning strategy.

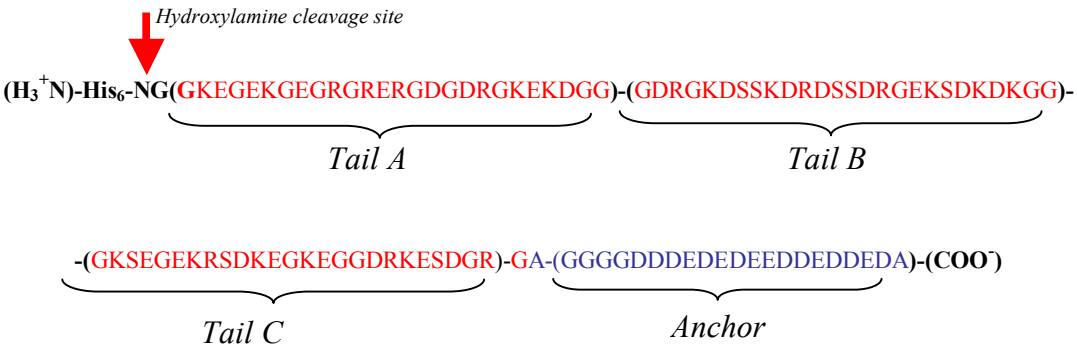

Figure 5.16 The amino acid sequence of the PAA encoded by the $pHis_6$-T_3A vector that was constructed during the feasibility experiments. The PAA is expressed with an N-terminal fusion tag (His_6) and a hydroxylamine cleavage site.

5.5 Conclusions

Prior work has indicated that there was a need for a universal cloning strategy for the production of highly specified, high molecular weight PAA's. In this chapter, we have presented the development of such a method. The successful preliminary experiments presented here confirm the feasibility and indicate the potential effectiveness of this technique. We have shown that this method enables one to specify the exact monomer sequence of a PAA, as well as molecular weight, functionality and secondary structure. The following chapter will discuss the application of the new cloning strategy to the production of a unique PAA construct designed to form brush layers on aluminum oxide surfaces.

5.6 Materials and Methods

Materials. One Shot® TOP10 Chemically Competent *E. coli* strain, expression plasmid pCR® T7/NT-TOPO®, eLONGase® Enzyme mix, LB Broth Base, UltraPure™ agarose, UltraPure™ ethidium bromide solution and T7/pRSET sequencing primers were purchased from Invitrogen Corp. (Carlsbad, CA). *Taq* PCR Core Kit, QIAprep® Spin Miniprep, QIAquick® Gel Extraction/PCR Purification Kits and PAGE-purified synthetic oligonucleotides were purchased from Qiagen, Inc. (Valencia, CA). Sequenase™ Version 2.0 DNA Polymerase was obtained from USB, Inc. (Cleveland, OH). *PfuTurbo*® DNA Polymerase was obtained from Stratagene (La Jolla, CA). The restriction endonuclease *Eam*1104I was purchased from Fermentas, Inc. (Hanover, MD). All other restriction enzymes, DNA modifying enzymes, T4 DNA Ligase, 100 bp DNA and 1 kbp DNA Ladders were purchased from New England Biolabs (Beverly, MA). Synthetic DNA primers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). LB Agar and glycerol were obtained from Sigma (St. Louis, MO). Ampicillin was purchased from VWR International (So. Plainfield, NJ).

General Methods. The procedures for the manipulation of DNA, transformation and cell growth were adapted from published literature[6, 7] 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). Agarose gel images were captured using a Polaroid GelCam. DNA concentrations were calculated from the optical density (OD) obtained using a Milton Roy Spectronic 1201 UV spectrophotometer (Ivyland, PA).

Method Developed by McMillan *et al.*

Construction of DNA Monomers. Two single-stranded synthetic oligonucleotides with complimentary 3' ends were purchased (Table 5.2). To construct the DNA monomer, 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 that the product size was ~100 base-pairs, as desired (data not shown). The product was then used as a template for PCR amplification to manufacture sufficient amounts of the DNA module. PCR was conducted using eLONGage® enzyme mix and the primers given in Table 5.2. Prior to the addition of template DNA, 8-methoxypsoralen was added to the PCR mixture followed by exposure to UV light for 5 min. 8-methoxypsoralen intercalates into contaminating DNA, rendering it harmless in the subsequent PCR reaction.[8] PCR was performed with 40 cycles at 95°C for 1 min (denaturing), 65°C for 1 min (annealing), and 68°C for 1 min (elongation). Large amounts of the double-stranded DNA module were obtained. Reactions were pooled together and the DNA was precipitated using an ethanol/sodium acetate procedure. Precipitated DNA was then re-suspended in sterile water then purified on an agarose electrophoresis gel and extracted into sterile water using the QIAquick® Gel Extraction kit. The DNA module was then digested overnight (~16 hrs) with *Eam*1104I at 37°C. *Eam*1104I was deactivated by heating the mixture to 65°C for 20 min and the monomer was purified using QIAquick® PCR purification kit and dried down using a Savant SpeedVac.

DNA Module		T_A
Sense	5'- agac ctcttcg | ggc aaa gag ggc gag aag ggg gag ggc cg**t ggc cgc gag cgc ggc gat g**-3'	67°C
Antisense	5'- caat ctcttcg | gcc acc gtc ttt ttc ctt acc acg atc gc**c atc gcc gcg ctc gcg gcc a**-3'	
Forward Primer	5'- agtcg **ctcttcg ggc aaa gag ggc gag** -3'	65°C
Reverse Primer	5'- tcagg **ctcttcg gcc acc gtc ttt ttc c** -3'	

Table 5.2 Synthetic oligonucleotide and primer sequences used to construct the DNA module for applying the method developed by McMillan *et al.* Bold-face represents bases involved in hybridization. Underlined bases represent *Eam*1104I recognition sites.

Generation of Tail Concatemers. The DNA monomer was re-suspended in 1X Ligase buffer and T4 DNA Ligase was added. The ligation reaction was incubated at 16°C for 24 hrs. Reactions were then pooled together and the DNA concatemers was purified using the QIAquick® PCR purification kit and then dried down. Agarose electrophoresis showed concatemer formation (Figure 5.4B).

Feasibility Studies for New Cloning Strategy

Insertion of Initial DNA Module. Two overlapping, single-stranded, synthetic oligonucleotides (Table 5.3) were used to construct the Anchor DNA Module. 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 pCR®T7/NT-TOPO®. The entire cloning reaction was transformed into One Shot® TOP10 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 5 mL 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 *Xba*I and *Sfo*I were used to check the orientation of the Anchor DNA Module. The double digestion reaction was incubated overnight (~16 hrs) at 37°C. A sample of the reaction was loaded onto an agarose electrophoresis gel and visualized with ethidium bromide (data not shown). Correct insert orientation yielded a digest fragment (bp) = 150 bp. Incorrect orientation yielded a digest fragment (bp) = 150 bp + (Module Size), where the Anchor DNA Module size is ~70 bp. Plasmids from positive clones were submitted for DNA sequencing (Cleveland Genomics, Inc.) using T7/pRSET primers to validate the gene sequence and orientation. Stocks of positive clones were made in 15% glycerol and stored at -80°C.

Anchor Module		T_A
Anchor Sense	5'-aac ggc | gcc ggc gga ggc ggt gat gac gat **gag gat gaa gat gag gaa gat**-3'	55°C
Anchor Antisense	5'- tca gcc tca ggc atc ctc gtc gtc ctc gtc **atc ttc ctc atc ttc atc ctc**-3'	

Table 5.3 Synthetic oligonucleotides used to construct Anchor DNA Module. Bold-face represents bases involved in hybridization. Underlined bases represent *Sfo*I recognition site.

Generation of Blunt-Ended DNA Modules. Two overlapping, single-stranded, synthetic oligonucleotides (Table 5.4) were used to construct each of the three DNA Modules (Module A, B and C). 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 electrophoresis showed that the product size was ~80 base-pairs, as desired (data not shown). The product was then used as a template for PCR amplification to manufacture sufficient amounts of each DNA module. PCR was conducted for each module using *PfuTurbo*® DNA polymerase and the corresponding amplification primers (Table 5.4). 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 each double-stranded, blunt-ended DNA module were obtained. The reactions for 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. This process is summarized in (Figure 5.12).

Tail Module A		**T_A**
A Sense	5'-ggc aaa gag ggc gag aag ggg gag ggc cgt **ggc cgc gag cgc ggc gatg**-3'	65°C
A Antisense	5'-gcc acc gtc ttt ttc ctt acc acg atc gc**c atc gcc gcg ctc gcg gcc**-3'	
A-For	5'- ggc aaa gag ggc gag aag gg- 3'	59°C
A-Rev	5'- gcc acc gtc ttt ttc ctt acc -3'	

Tail Module B		**T_A**
B Sense	5'- ggc gac cgt ggc aaa gac agc agc aag gat **cgt gac agc tct gac cgt ggt**-3'	63°C
B Antisense	5'- gcc gcc ctt gtc ctt gtc gct ttt ttc **acc acg gtc aga gct gtc acg** -3'	
B-For	5'- ggc gac cgt ggc aaa gac ag -3'	61°C
B-Rev	5'- gcc gcc ctt gtc ctt gtc gc -3'	

Tail Module C		**T_A**
C Sense	5'-ggc aag agc gaa ggc gag aag cgc agc g**at aaa gag ggt aaa gag ggc ggt**-3'	63°C
C Antisense	5'- gcc gcg acc atc gct ctc ttt gcg gtc **acc gcc ctc ttt acc ctc ttt at**-'3	
C-For	5'- ggc aag agc gaa ggc gag aag -3'	61°C
C-Rev	5'- gcc gcg acc atc gct ctc tt-3'	

Table 5.4 Synthetic oligonucleotide and primer sequences used to construct Tail DNA Modules A, B and C. Bold-face represents bases involved in hybridization.

*Linearization of Vector with Sfo*I. The expression vector was linearized by digesting with *Sfo*I, which leaves blunt ends. Plasmid 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. This process is summarized in (Figure 5.12).

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 DNA module, 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 One Shot® TOP10 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) from each transformation 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 5.5), 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). The samples from each PCR reaction were analyzed using agarose gel electrophoresis for the correct size insert (data not shown). The size of the PCR product (bp) = 350 bp + (N x (Module Size)), where N = number of inserts and for these experiments the module size was ~80 bp.

Primers		**T_A**
T7 Forward	5'- taa tac gac tca cta tag gg -3'	50°C
pRSET Reverse	5'- tag tta ttg ctc agc ggt gg -3'	

Table 5.5 Synthetic DNA primers used to analyze insertion size. PCR Product Size (bp) = 350 bp + N(80) where N is the number of insertions of 80bp Modules into the vector, pHis$_6$-A.

Analysis of Insert Orientation – Double restriction digest. Clones containing an insert of correct size were used to inoculate 5 mL of LB Broth containing 100 µg/mL ampicillin. Cultures were incubated overnight (~16 hrs) at 37°C with shaking (200 rpm). Plasmid was isolated from each using the QIAprep® Spin Miniprep kit. Plasmid was re-suspended in 40 µL of sterile water to a concentration of ~100 ng/µL. Restriction enzyme double digests (*Sfo*I/*Xba*I) were conducted to check the orientation of the insert. The 15 µL digestion 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). For properly inserted modules, the size of the restriction fragment (bp) = 150 bp + (N x (Module Size)), where N = number of inserts and for these experiments the module size was ~80 bp. During incorrect orientation, the restriction fragment (bp) = 150 bp. Plasmids from colonies containing the correct insert size and insert orientation were submitted for DNA sequencing (Cleveland Genomics, Inc.) using T7/pRSET primers to validate the gene sequence and orientation. Stocks of positive clones were made in 15% glycerol and stored at -80°C.

5.7 Supplemental Information

5.7.1 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:	pHis$_6$-A
Sequencing primer used:	T7 Forward

```
  1  catcatcatc atcatcatgg tatggctagc atgactggtg gacagcaaat
 51  gggtcgggat ctgtacgacg atgacgataa ggatccaacc cttaacggcg
101  ccggcggagg cggtgatgac gatgaggatg aagatgagga agatgacgag
151  gacgacgagg atgcctgagg ctgaaagggc gaattcgaag cttgatccgg
201  ctgctaacaa agcccgaaag gaagctgagt tggctgctgc caccgctgag
```

His$_6$	bases 1-18
Asn-Gly	bases 94-99
***Sfo*I site**	bases 97-102
Anchor	bases 103-165
Stop codon	bases 166-168

Table 5.6 DNA sequencing results for pHis$_6$-A.

Sequencing results for:	pHis$_6$-TA
Sequencing primer used:	T7 Forward

```
  1  catcatcatc atcatcatgg tatggctagc atgactggtg gacagcaaat
 51  gggtcgggat ctgtacgacg atgacgataa ggatccaacc cttaacggcg
101  gcaaagaggg cgagaagggg gagggccgtg gccgcgagcg cggcgatggc
151  gatcgtggta aggaaaaaga cggtggcgcc ggcggaggcg gtgatgacga
201  tgaggatgaa gatgaggaag atgacgagga cgacgaggat gcctgaggct
251  gaaagggcga attcgaagct tgatccggct gctaacaaag cccgaaagga
301  agctgagttg gctgctgcca ccgctgagca ataactagca taacccctg
```

His$_6$	bases 1-18
Asn-Gly	bases 94-99
Tail (1)	bases 100-174
***Sfo*I site**	bases 175-180
Anchor	bases 181-243
Stop codon	bases 244-246

Table 5.7 DNA sequencing results for pHis$_6$-TA.

Sequencing results for:	$pHis_6$-T_2A
Sequencing primer used:	T7 Forward

```
   1  catcatcatc atcatcatgg tatggctagc atgactggtg gacagcaaat
  51  gggtcgggat ctgtacgacg atgacgataa ggatccaacc cttaacggcg
 101  gcaaagaggg cgagaagggg gagggccgtg gccgcgagcg cggcgatggc
 151  gatcgtggta aggaaaaaga cggtggcggc gaccgtggca aagacagcag
 201  caaggatcgt gacagctctg accgtggtga aaaaagcgac aaggacaagg
 251  gcggcgccgg cggaggcggt gatgacgatg aggatgaaga tgaggaagat
 301  gacgaggacg acgaggatgc ctgaggctga aagggcgaat tcgaagcttg
 351  atccggctgc taacaaagcc cgaaaggaag ctgagttggc tgctgccacc
 401  gctgagcaat aactagcata acccttgggg cctctaaaac gggtcttgaa
```

His_6	bases 1-18
Asn-Gly	bases 94-99
Tail (1)	bases 100-177
Tail (2)	bases 178-252
***Sfo*I site**	bases 253-258
Anchor	bases 259-321
Stop codon	bases 322-324

Table 5.8 DNA sequencing results for $pHis_6$-T_2A.

Sequencing results for:	pHis$_6$-T$_3$A
Sequencing primer used:	T7 Forward

```
   1  catcatcatc atcatcatgg tatggctagc atgactggtg gacagcaaat
  51  gggtcgggat ctgtacgacg atgacgataa ggatccaacc cttaacggcg
 101  gcaaagaggg cgagaagggg gagggccgtg gccgcgagcg cggcgatggc
 151  gatcgtggta aggaaaaaga cggtggcggc gaccgtggca aagacagcag
 201  caaggatcgt gacagctctg accgtggtga aaaaagcgac aaggacaagg
 251  gcggcggcaa gagcgaaggc gagaagcgca gcgataaaga gggtaaagag
 301  ggcggtgacc gcaaagagag cgatggtcgc ggcgccggcg gaggcggtga
 351  tgacgatgag gatgaagatg aggaagatga cgaggacgac gaggatgcct
 401  gaggctgaaa gggcgaattc gaagcttgat ccggctgcta acaaagcccg
 451  aaaggaagct gagttggctg ctgccaccgc tgagcaataa ctagcataac
 501  ccttggggcc tctaaaacgg gtcttgaagg ggtttttttg cttgaaanga
```

His$_6$	bases 1-18
Asn-Gly	bases 94-99
Tail (1)	bases 100-177
Tail (2)	bases 178-255
Tail (3)	bases 256-330
***Sfo*I site**	bases 331-336
Anchor	bases 337-399
Stop codon	bases 400-402

Table 5.9 DNA sequencing results for pHis$_6$-T$_3$A.

5.8 References

[1] Cappello J, Crissman J, Dorman M, Mikolajczak M, Textor G, Marquet M and Ferrari F. "Genetic engineering of structural protein polymers". *Biotechnol. Progr.* (1990); 6: 198-202.

[2] McMillan RA, Lee TAT and Conticello VP. "Rapid assembly of synthetic genes encoding protein polymers". *Macromolecules.* (1999); 32: 3643-3648.

[3] Meyer DE and Chilkoti A. "Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system". *Biomacromolecules*. (2002); 3: 357-367.

[4] Won JI and Barron AE. "A new cloning method for the preparation of long repetitive polypeptides without a sequence requirement". *Macromolecules*. (2002); 35: 8281-8287.

[5] McMillan RA, Lee TAT and Conticello VP. "Rapid assembly of synthetic genes encoding protein polymers". *Macromolecules.* (1999); 32: 3643-3648.

[6] Sambrook J, Fritsch EF and Maniatis T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, (1992).

[7] Ausubel FM. *Current Protocols in Molecular Biology*. J. Wiley: New York, NY, (1997).

[8] Meier, A. "Elimination of contaminating DNA within polymerase chain reaction reagents – implications for a general-approach to detection of uncultured pathogens". *J. Clin. Microbiol*. (1993); 31(3): 646-652.