

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Chemical & Biomolecular Engineering Theses,
Dissertations, & Student Research

Chemical and Biomolecular Engineering,
Department of

4-2011

DE NOVO GENE SYNTHESIS BY RAPID POLYMERASE CHAIN ASSEMBLY COUPLED WITH IMMUNOAFFINITY PURIFICATION: A NOVEL PROCESS AND WORKSTATION

Joel R. TerMaat

University of Nebraska-Lincoln, jtermaat@nebrwesleyan.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/chemengtheses>



Part of the [Biochemical and Biomolecular Engineering Commons](#)

TerMaat, Joel R., "DE NOVO GENE SYNTHESIS BY RAPID POLYMERASE CHAIN ASSEMBLY COUPLED WITH IMMUNOAFFINITY PURIFICATION: A NOVEL PROCESS AND WORKSTATION" (2011). *Chemical & Biomolecular Engineering Theses, Dissertations, & Student Research*. 8.
<https://digitalcommons.unl.edu/chemengtheses/8>

This Article is brought to you for free and open access by the Chemical and Biomolecular Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Chemical & Biomolecular Engineering Theses, Dissertations, & Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

DE NOVO GENE SYNTHESIS BY RAPID POLYMERASE CHAIN ASSEMBLY
COUPLED WITH IMMUNOAFFINITY PURIFICATION:
A NOVEL PROCESS AND WORKSTATION

by

Joel R. TerMaat

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Engineering

(Chemical & Biomolecular Engineering)

Under the Supervision of Professors Anuradha Subramanian and David S. Hage

Lincoln, Nebraska

April, 2011

DE NOVO GENE SYNTHESIS BY RAPID POLYMERASE CHAIN ASSEMBLY
COUPLED WITH IMMUNOAFFINITY PURIFICATION:
A NOVEL PROCESS AND WORKSTATION

Joel R. TerMaat, Ph.D.

University of Nebraska, 2011

Advisors: Anuradha Subramanian and David S. Hage

The *de novo* synthesis of genes is emerging as a powerful tool in biotechnology. The ability to synthesize genes of any desired sequence opens the door to seemingly unlimited research possibilities. Major advances have been made recently in *de novo* gene synthesis, with Polymerase Chain Assembly (PCA) routinely used to construct functional sequences from short single-stranded oligonucleotides. However, current PCA techniques are lacking in speed and fidelity. Additionally, substantial undesired reactants/products are present in the final reaction. A novel process and accompanying workstation has been developed that incorporates rapid PCA synthesis coupled with subsequent affinity purification of the synthesis mixture. The system enables fast and accurate PCA synthesis and isolation of the full length DNA of interest.

In the synthesis step, the desired sequence is assembled and PCR amplified in a fast thermocycler to generate a high yield of product with minimal runtime and errors. A traditional 2-step PCA-PCR approach is utilized to assemble and amplify the full-length gene. Alternatively, integration of PCA and PCR into a single rapid reaction is also employed, working reliably up to about 1 kb. For the synthesis of genes longer than 1.5 kb, a convergent rapid synthesis strategy is proposed in which the full-length sequence is assembled by a series of synthesis steps from smaller fragments. In this work, a variety of genes ranging from 600 bp up to 3.8 kb in length are synthesized by rapid PCA techniques.

The second section of the workstation employs two affinity columns to isolate the desired full-length product from shorter unwanted reactants/products inherent in the PCA

reaction. During PCR amplification, labels are incorporated into the desired product on both ends via PCR primers. Undesired products contain only one of these labels, or no label at all. The first column interacts with one of the labels to partially purify the mixture. The intermediate product is then subsequently purified via the second column to isolate the full-length sequence. An initial prototype workstation was developed, while the 2nd generation instrument consisted of process refinements using two antibody columns for immunoaffinity purification.

DEDICATION

To my wife Amelita—without your patience and unwavering support it would have been impossible to achieve my goals. To my two sons, Dean & Layne, for bringing a joy into my life I could never have imagined possible.

AUTHOR'S ACKNOWLEDGEMENTS

A special thanks to Dr. Subramanian, whose guidance challenged me to think critically and to remain persistent in the pursuit of understanding the unknown.

GRANT INFORMATION

This work was supported by the National Institutes of Health (R21 RR022860 and R33 RR022860)

Table of Contents

1	Background and Motivation for the Gene Synthesis Workstation	1
1.1	The Polymerase Chain Reaction (PCR)	3
1.2	Gene Synthesis Techniques.....	5
1.3	PCA Synthesis Using a Rapid Thermocycler	8
1.4	Application of Affinity Purification to PCA Gene Synthesis	10
1.5	Introduction to the Workstation Process	12
1.6	Developmental Objectives	14
1.7	References	15
2	Proof-of-Concept for the Workstation	18
2.1	Introduction	18
2.1.1	Rapid PCR-based synthesis	18
2.1.2	Purification of synthetic genes.....	20
2.2	Materials & Methods.....	21
2.2.1	Design of Oligonucleotides.....	21
2.2.2	Polymerase Chain Assembly Conditions.....	21
2.2.3	PCR Amplification Conditions	22
2.2.4	Labeling of synthesis products.....	23
2.2.5	BrdU and Iminobiotin affinity columns.....	24
2.2.6	Affinity Purification Hardware	25
2.2.7	Purification Methods.....	28
2.2.8	Gel Electrophoresis	29

2.2.9	Cloning and Sequencing	30
2.3	Results & Discussion	30
2.3.1	929 bp puc19 B-lactamase sequence	31
2.3.2	612 bp endothelial protein C receptor sequence	33
2.3.3	615 bp puc19 ori sequence.....	35
2.3.4	1548 bp human Thrombomodulin sequence.....	37
2.3.5	Synthesis Rates, Cloning, and Sequencing	38
2.4	Conclusion.....	40
2.5	References	41
2.6	Publication and Author Contributions.....	42
3	Gene synthesis by integrated polymerase chain assembly and PCR amplification ..	44
3.1	Introduction to the concept of integrated PCA-PCR.....	44
3.2	Materials & Methods.....	48
3.3	Results & Discussion	50
3.3.1	Number of Cycles	50
3.3.2	Effect of Oligo Concentration.....	51
3.3.3	Effect of Primer Concentration	54
3.3.4	Limitations arising from number of oligonucleotides	55
3.4	Conclusion.....	60
3.5	References	61
3.6	Publication and Author Contributions.....	62
4	Improvements to the Gene Synthesis Workstation	64
4.1	Introduction	64

4.1.1	Improvements to the 2-step PCA process	69
4.1.2	Investigation of target-ligand affinity systems	69
4.1.3	Affinity purification process	72
4.2	Materials & Methods.....	73
4.2.1	Optimization of 2-step PCA-PCR.....	73
4.2.2	PCR testing of labels.....	74
4.2.3	Affinity columns	75
4.2.4	Batch purification testing.....	76
4.2.5	Flow-thru affinity purification hardware	78
4.2.6	Flow-thru immunoaffinity purification.....	80
4.3	Results & Discussion	83
4.3.1	Optimization of 2-step PCA-PCR.....	83
4.3.2	Labeled primer testing	87
4.3.3	Batch purification with high salt elution.....	88
4.3.4	Batch purification with free label elution buffer.....	89
4.3.5	Column Regeneration	91
4.3.6	Non-specific binding testing	92
4.3.7	Selection & Serial batch testing.....	93
4.3.8	Flow-thru purification.....	96
4.4	Conclusions	98
4.5	References	99
4.6	Author Contributions.....	102
5	Convergent synthesis by rapid PCA in the workstation	103

5.1	Introduction	103
5.1.1	Convergent strategy for longer genes	103
5.1.2	Application of purification to convergent-based synthetic genes.....	108
5.1.3	Convergent synthesis of a 3.8kb mWAP-hPC sequence	109
5.2	Materials & Methods.....	111
5.2.1	mWAP-hPC: Sequence information and design of oligonucleotides	111
5.2.2	mWAP-hPC convergent synthesis.....	112
5.2.3	Flow-through Immunoaffinity Purification	113
5.2.4	Batch immunoaffinity purification	115
5.2.5	Gel Electrophoresis & Purification.....	116
5.2.6	Cloning & Sequencing.....	117
5.3	Results & Discussion	117
5.3.1	mWAP-hPC fragment synthesis	117
5.3.2	Flow-through immunoaffinity purification.....	119
5.3.3	2 nd round stitching to construct the 3.8kb sequence	123
5.3.4	Follow-up batch testing.....	128
5.3.5	mWAP-hPC sequencing results.....	132
5.4	Summary & Conclusions	134
5.5	References	137
5.6	Author Contributions.....	139
6	Summary, Analysis and Future Work.....	140
6.1	Rapid PCA-based gene synthesis	140
6.1.1	Summary	140

6.1.2	Impact of assembly oligo length	142
6.1.3	Oligo design flexibility with longer assembly oligos	143
6.1.4	Economic considerations	145
6.1.5	Error Frequency Considerations	146
6.1.6	Future Work with rapid PCA.....	148
6.2	Affinity purification of synthetic genes	150
6.2.1	Summary	150
6.2.2	Impact of labeling DNA.....	151
6.2.3	Affinity purification yield	152
6.2.4	Economic considerations	153
6.2.5	Future work with generic labeled primers	154
6.2.6	Future work with FITC detection	155
6.2.7	Future work with other affinity columns	156
6.2.8	Broader applications of purification of labeled DNA.....	157
6.3	Closing thoughts.....	158
6.4	References	159
	Appendix 5A- mWAP-hPC sequence information.....	162
5A.1	Sequence information	162
5A.2	Oligo Hybridization Map.....	163
5A.3	Sequencing Results for Gene (1+2).....	166
5A.4	Sequencing Results for Gene (3+4).....	172
	Appendix 5B- Synthesis of chimeric gene sequences	178
5B.1	Introduction	178

5B.2	Chimeric genes: Design of sequence and assembly oligonucleotides.....	181
5B.3	PCA synthesis of chimeric genes	183
5B.4	Cloning & Sequencing.....	184
5B.5	Results & Discussion.....	185
5B.6	Conclusions	187
5B.7	References	188
5B.8	Contributions	189
Appendix 5C: Details of Chimeric Sequences.....		190
5C.1	Chimera #1 Beginning Design	190
5C.2	Chimera#1 Resulting Sequence.....	191
5C.3	Chimera #1 Oligo Design Hybridization Map	192
5C.4	Chimera #2 Starting original nucleotide sequence	193
5C.5	Resulting Forward Translation of Chimera #2	194
5C.6	Chimera #2 Oligo Design Hybridization Map	195

Table of Figures

Figure 1-1 Simplified schematic of one PCR cycle.....	5
Figure 1-2 Illustration of the PCA process for a simple 4 oligo assembly.....	7
Figure 1-3 Picture of the PCRJet rapid thermocycler.....	9
Figure 1-4 Simple illustration of affinity chromatography.....	11
Figure 1-5 Schematic of the rapid gene synthesis workstation.....	13
Figure 2-1 Picture of the PCRJet rapid thermocycler.....	19
Figure 2-2 Schematic of the prototype affinity chromatography station.....	20
Figure 2-3 Picture of the Prototype workstation using a vertical breadboard.	25
Figure 2-4 PCA, PCR and Affinity Purification of the pUC19 β -lactamase gene.	31
Figure 2-5 PCA, PCR and affinity purification of the Endothelial Protein C Receptor gene.....	33
Figure 2-6 PCA, PCR and affinity purification of the pUC19 Origin of Replication gene.	35
Figure 2-7 PCA, PCR and affinity purification of the human Thrombomodulin gene. .	37
Figure 2-8 EcoRI restriction enzyme digest of 10 clones.....	39
Figure 3-1 Traditional 2 step PCA/PCR and Integrated PCA-PCR vs. # of cycles.....	51
Figure 3-2 Effect of oligo concentration on Integrated PCA-PCR.....	52
Figure 3-3 Effect of primer concentration on Integrated PCA-PCR.	55
Figure 3-4 Traditional 2 step PCA/PCR and Integrated PCA-PCR vs. # of oligos using the TM-1 oligo set.....	57

Figure 3-5. Integrated PCA-PCR products after 37 cycles using a modified Barnes and Frawley Protocol.	58
Figure 4-1 Schematic of the next generation affinity section.	73
Figure 4-2 Picture of the new affinity purification system.	79
Figure 4-3 PCA oligo concentration – PCR aliquot amount reaction sets for the 640 bp chimera “A” and 1548 bp TM-1 genes.	84
Figure 4-4 PCR inhibition testing using labeled and regular primers.	87
Figure 4-5 Initial batch testing of the five labels with high salt in the elution buffer with the 640 bp chimera A sequence.	88
Figure 4-6 Batch testing of the five labels with label in the elution buffer for the 640 bp chimera A sequence.	90
Figure 4-7 Batch testing of the columns after glycine regeneration using the 640 bp chimera A sequence.	91
Figure 4-8 Batch testing of the five columns with unlabelled 640 bp chimera A product.	93
Figure 4-9 Serial batch purification of the labeled 1.4 kb full chimera gene using anti-FITC, followed by anti-DIG.	95
Figure 4-10 Initial purification of labeled 1.4 kb full chimera using the next generation workstation.	97
Figure 5-1 Schematic of convergent synthesis of a long gene.	107
Figure 5-2 Different workstation processes for long gene synthesis from fragments. .	109
Figure 5-3 PCR products of the different mWAP-hPC fragments generated by rapid 2-step PCA-PCR.	118

Figure 5-4 Flow-through immunoaffinity purification of a 4-fragment (~950 bp) mixture.	120
Figure 5-5 Flow-through purification of a 2-fragment (~1.9 kb) mixture (A) and the subsequent Amicon concentration (B).	121
Figure 5-6 Flow-through purification of the full-length labeled 3.8kb direct stitching product.	122
Figure 5-7 Direct stitching of unpurified labeled PCR products from either 4 fragments (lane 1) or 2 fragments (lane 2).	124
Figure 5-8 Effect of fragment template concentration on stitching.	125
Figure 5-9 Stitching comparison of unpurified vs. gel-purified using 4-fragments.	127
Figure 5-10 Comparison of the effect of DNA length on chromatographic binding.	129
Figure 5-11 Effect of incubation time on binding amount at room temperature and 37°C.	131
Figure 5-12 Batch testing with long incubation times [(1+2) and (3+4)] with serial purification with FITC then digoxigenin, and subsequent stitching (Stitch, lane 1) of the purified products.	132
Figure 6-1 Oligo map for an example 600 bp sequence employing 100-mers with gaps allowed.	144
Figure 6-2 Fraction clones correct vs. gene length at different error frequencies as estimated by the Poisson distribution.	148
Figure 6-3 The addition of dummy 5' and 3' flanker sequences allows for generic labeled primers to be used for many different synthetic genes.	155
Figure 6-4 Illustration of an integrated Ab-FITC/streptavidin column.	157

1 Background and Motivation for the Gene Synthesis Workstation

Ever since the double-helix structure of DNA was first described by Watson and Crick (1953), scientists have wondered what the possibilities might be with the knowledge of genetic instructions. Can DNA sequences be modified or created a priori to produce a desired function, even unnatural? By knowing the desired output, the ability to synthesize genes by programming a genetic code opens the door to possibilities limited only by man's imagination and cunning.

While many DNA syntheses rely on modification of existing natural sequences to alter functionality, the concept of *de novo* synthesis of gene sequences no longer constrains genetic coding to those based in nature. In some views a blessing, technical challenges still abound for facile synthesis of gene sequences, knowingly or unknowingly good or bad. Additionally, the moral and ethical implications of gene synthesis cannot be ignored, especially as scientists move forward with creating ever longer gene sequences. Just recently, Gibson and others (2008) synthesized a 582,970bp genome for *Mycoplasma genitalium*, the smallest free-living organism.

Despite a rapidly growing gene synthesis biotechnology sector, fast and accurate synthesis of moderate gene fragments is still in relative infancy. High error rates are just one of many difficulties that plague DNA synthesis. Since one error in a synthesized gene sequence can destroy the desired functionality of the expressed protein, a high fidelity process is needed that minimizes post synthesis manipulation to correct the error

or reduces the burden of screening error-riddled products to isolate an error-free sequence.

The primary motivation for this work is to develop a novel workstation platform called a Gene Synthesis Workstation for rapid and high-fidelity synthesis of gene sequences of moderate length. At the core of the workstation is a rapid thermocycler, which facilitates expedited synthesis and reduction of introduced errors in the synthesis process. Since gene synthesis within a thermocycler typically leads to a myriad of products, it is usually necessary to take the polymerase chain reaction (PCR) end-product through a purification step to increase the ratio of desired product to undesired products prior to further processing. During the polymerase chain assembly (PCA) process, only a few copies of the full length gene are constructed against a large background of shorter oligonucleotides. Traditionally, PCA is followed by primer-mediated PCR amplification to selectively amplify the full length product. However, undesired oligonucleotides from the assembly are still significantly present in the PCR product as some are linearly amplified.

A secondary aspect of the workstation is to provide a novel alternative chromatography based purification system to isolate the desired full-length gene resulting from the PCR-based gene synthesis. Typically, gel purification is employed which consists of conducting gel electrophoresis to separate the PCR product mixture by molecular weight. The desired product band (section of gel at a certain molecular weight) is manually cut from the gel and then a kit is used to purify the product from the gel slice. However,

band cutting and clean-up can be time consuming due to the manual nature of the technique. For the synthesis of longer DNA, multiple DNA constructs need to be assembled, purified, and ligated. Manual separation of tens or even hundreds of each DNA constructs (each with different lengths) can be tedious and manually intensive. In the alternative chromatography section, synthesis products are purified through a serial arrangement of two columns which exploit binding of artificial labels that have been incorporated into the synthesis products.

The chromatography section of the workstation is not necessarily meant to completely replace gel purification for the synthesis of all genes. Rather, it is meant as an option available to researchers conducting gene synthesis. This is even true for the specifics of the PCA synthesis described herein. Given the infinite diversity of sequences and lengths that one may desire to synthesize, it is clear that one single technique may not be superior to others for all cases. Specific techniques or variations or combinations thereof may be optimal for any given sequence. While the rapid thermocycler PCA synthesis can be viewed as generally superior in terms of thermal damage in comparison to slow thermocyclers, it and the chromatographic purification should be viewed as one of many options. Molecular biologists should select from all available tools depending on the specifics of the gene sequence to be synthesized.

1.1 The Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) is one of the most valuable techniques available to molecular biologists. Discovered by Kary Mullis in 1983 (Rabinow 1996), PCR creates billions of copies of target DNA amplicon from just a few starting templates using

thermal cycling. At the core of the process is the polymerase enzyme. Double-stranded DNA (dsDNA) is first denatured into single stranded DNA (ssDNA) within a thermocycler. Typically two primers, i.e. short synthetic oligonucleotides of about 20 base pairs, are employed which target specific sequences of the template. When the temperature is lowered in the thermocycler (to about 60°C depending on the melting temperature of the primers employed) after denaturation, the primers anneal to the complementary template sequences. Polymerase enzymes then bind to the 3' end of the annealed primers. Deoxyribonucleotides (dNTP's) are present within the PCR mixture, and are thereby added to the 3' end of the primer complementary to the template. Typically, the thermocycler temperature is then raised to around 72°C for optimal extension activity of the polymerase. At the end of the temperature cycle is the construction of two new dsDNA. When the cycling is repeated over multiple cycles, the constructed amplicons are exponentially amplified, theoretically doubling every cycle. A simplified schematic of one PCR cycle is illustrated in Figure 1-1. One method for detection of the PCR products is agarose gel electrophoresis, in which the molecular weight of the DNA amplicon is compared to a reference ladder.

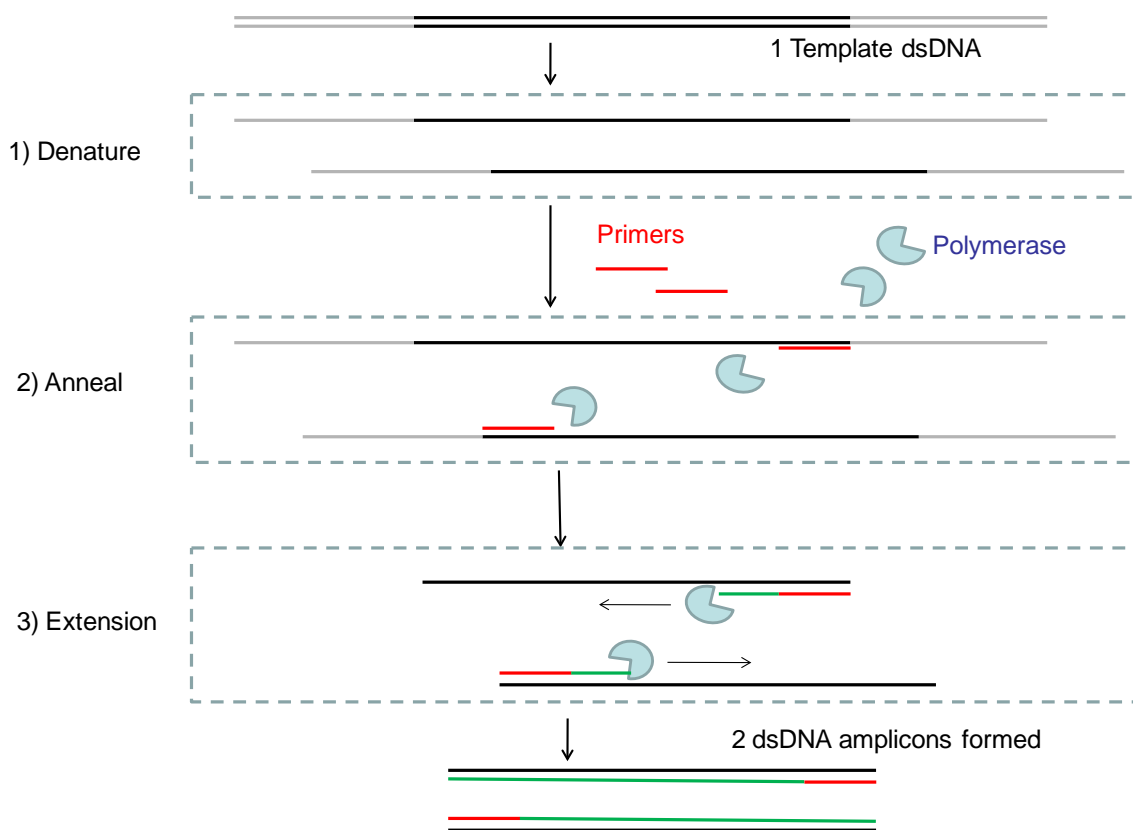


Figure 1-1 Simplified schematic of one PCR cycle.

Refer to text for description. The template DNA at the beginning of the process may either be starting DNA template (exterior sequence shaded in grey and removed for clarity) or amplicon constructed in a previous cycle.

1.2 Gene Synthesis Techniques

A plethora of synthesis options are available which rely on basic molecular biology techniques (see Czar and others (2009) for an in-depth review). Three of the most common techniques are phosphoramidite synthesis, ligase chain reaction (LCR), and polymerase chain assembly (PCA). Automated phosphoramidite synthesizers routinely create short (~15-100 bp) nucleotide sequences because elongation efficiencies and self-complementarity lower the yields of longer sequences. Integrated DNA Technologies

(Coralville, IA) commercially offers “ultramers”--synthetic oligonucleotides up to 200bp in length (Allen and others 2009). While phosphoramidite synthesis is the method of choice for synthesis of short ssDNA sequences, such as PCR primers, it is impractical for synthesis of longer dsDNA sequences exhibiting biological functionality. In this respect, phosphoramidite synthesis is utilized to construct the building blocks for LCR and PCA. In the LCR technique, a mixture of oligos are denatured and annealed together under stringent conditions. A DNA ligase is present to ligate the oligos into dsDNA. LCR is limited in its usage due to the slow reaction rate, and PCA is still typically required for final synthesis assembly.

The first description of the polymerase chain assembly (PCA) method to create longer synthetic genes from overlapping short oligonucleotides is commonly attributed to Stemmer and others (1995). In essence, it is a variation of the PCR process in which a thermo-stable DNA polymerase is used to “stitch” together shorter oligonucleotides. Overlapping oligonucleotides are subjected to PCR thermocycling and polymerase extension to form the gene of interest. Stemmer and others (1995) assembled a 1.1kb β -lactamase encoding gene and a 2.7 kb plasmid from short 40mer oligonucleotides. PCA or variations thereof have been frequently used to synthesize genes from a few hundred to many kilobases (Mehta and others 1997; Smith and others 2003; Gao and others 2004; Mamedov and others 2007). PCA also is a means to create synthon precursors that can be utilized to create even larger constructs (Kodumal and others 2004). Recently, Gibson and others (2008) synthesized a 582,970 bp *Mycoplasma genitalium* genomes

using recombination techniques from 5-7kb overlapping “cassettes” assembled from $\sim 10^4$ oligonucleotides of ~ 50 bp in length each.

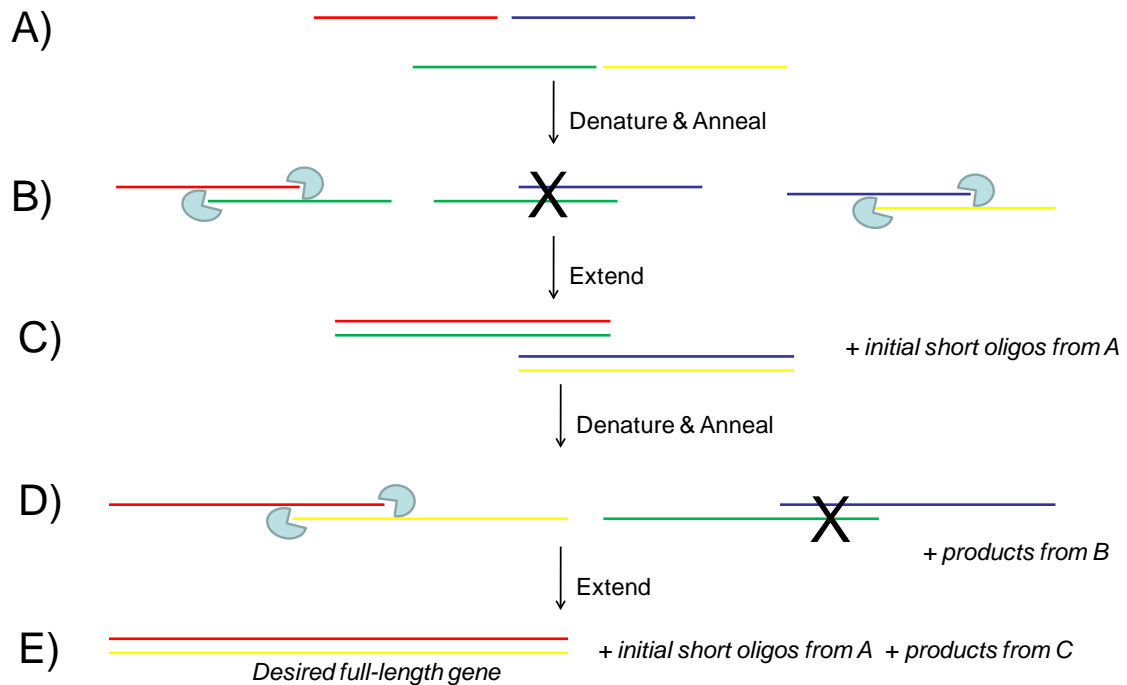


Figure 1-2 Illustration of the PCA process for a simple 4 oligo assembly.

“X” denotes oligo hybridizations that are unproductive.

An illustration of the PCA process is shown in Figure 1-2. During the PCA process, a mixture of oligos is subjected to the same temperature cycling as in PCR amplification. Neighboring oligos that are complementary to one another hybridize during the annealing step. The polymerase then extends the oligos to completion. Note that the polymerase only is effective in extending from the 3' end, meaning that many of the hybridizations that occur are unproductive. The newly formed dsDNA is then denatured into ssDNA. During the next annealing step, these longer products may now hybridize to new neighboring sequences. After repeated cycling, the oligos continually increase in length

until the full-length dsDNA is constructed. As the maximum number of full-length products is limited by the amount of outermost oligos and the process is fairly inefficient, a relatively small amount of full-length product is obtained against a large background of unwanted products. These unwanted products from inner oligos are obviously a necessary evil—the starting oligos must be present for PCA but they become a nuisance once the full-length product is constructed. Typically, PCA is followed by PCR amplification of the full-length gene to increase its amount relative to unwanted products. In a separate reaction using an aliquot of the PCA product, PCR primers are employed to exponentially amplify the full-length gene.

1.3 PCA Synthesis Using a Rapid Thermocycler

As noted previously, the cornerstone of the workstation is the utilization of a rapid thermocycler to perform PCA. Almost all prior PCA research has used “heat-block” thermocyclers with slow ramp rates, with typical run-times for PCA synthesis around 1 to 3 hours. Similar run-times are needed when follow-up PCR is done to selectively amplify the full-length gene. The development herein is based upon the initial work of Mamedov and others (2007), in which rapid PCA was conducted in the PCRJet[®] (Quintanar and Nelson 2002) to construct the endothelial protein-C receptor and thrombin receptor genes. The PCRJet[®] is a rapid thermocycler which exploits high heat transfer rates by forcing heated or room temperature compressed air over small diameter glass capillaries. Total runtimes in the PCRJet[®] are decreased to about 10-20 minutes for each PCA and PCR step.



Figure 1-3 Picture of the PCRJet rapid thermocycler

The PCRJet is used for the PCA and PCR steps in the workstation.

While a beneficial reduction in run-times using a faster thermocycler is obvious, the real importance of rapid PCA lies in the minimization of errors produced in the assembly process. By minimizing the times spent at elevated temperatures, thermal damage to the DNA is minimized. Pienaar and others (2006) demonstrate that errors due to thermal damage primarily arises from exposure of ssDNA to elevated temperatures via A+G depurination, oxidative damage of guanine to 8-oxoG and cytosine deamination to uracil. With standard thermocyclers, typical hold times at denaturation are 10 to 30 seconds, not including the cumulative time to reach the set-point is large due to the slow ramp rates of these instruments. In the compressed air thermocycler, the time spent at elevated temperatures is on the order of a few seconds. Thus, PCA with a rapid thermocycler is used as a basis for this work as it provides superior synthesis of products with high fidelity and functionality. The thermostable DNA polymerase from *Thermococcus kodakaraensis* (KOD) is also employed in this work. KOD polymerase is attractive for

its fast extension rate and proof-reading activity in which it may correct base pair errors (Griep and others 2006).

1.4 Application of Affinity Purification to PCA Gene Synthesis

Affinity chromatography, especially immunoaffinity chromatography, has grown considerably in use over the past several decades. The first application of affinity chromatography is commonly attributed to Starkenstein (1910) while the first antibody-antigen immunocapture separations were developed in the 1930's (d'Allesandro and Sofia 1935; Landsteiner and van der Scheer 1936). Briefly, affinity chromatography is a separations technique which relies on the attractive (affinity for) intermolecular forces between a ligand and receptor. The target molecule, in this case the full-length synthesized gene, may be isolated if it possesses a target that is preferentially captured by the introduced ligand. Immunoaffinity chromatography relies on the affinity between antibody and antigen. Due to the nature of antibody-antigen binding, immunoaffinity purification is in general highly specific.

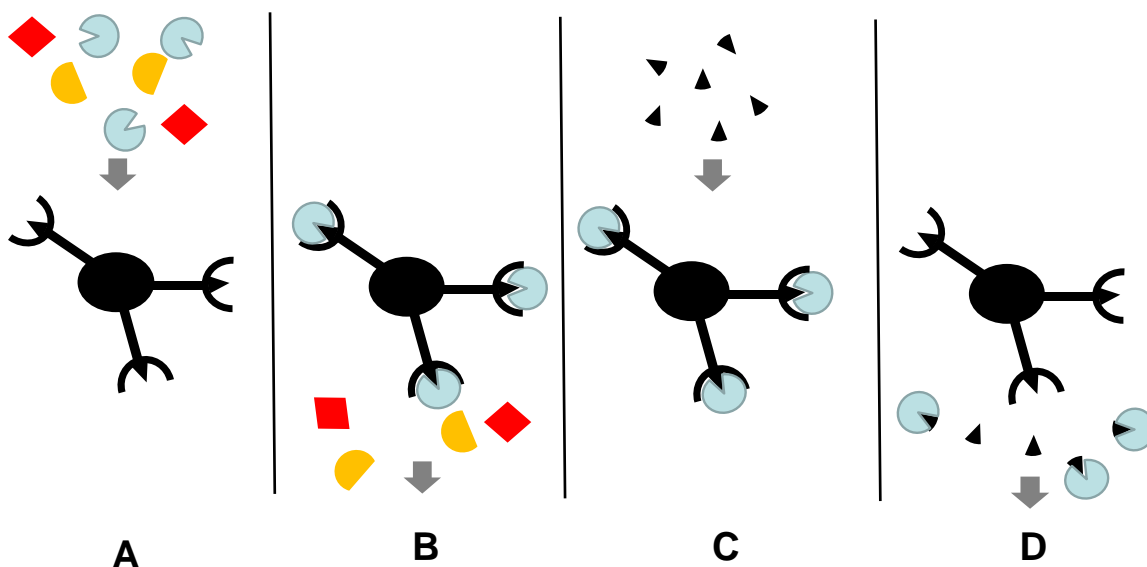


Figure 1-4 Simple illustration of affinity chromatography.

A) A mixture of compounds is introduced to immobilized ligands. B) Target molecules bind specifically to ligand sites while others are washed through. C) An elution buffer is applied to the immobilized ligand. D) Target molecules are eluted from the immobilized ligand.

In standard PCA synthesis, there are not unique parts of the full-length gene that can be readily exploited by affinity separations. The labels must be incorporated into the full-length gene by adding labels to the PCR primers, attached at the 5' end. After the full-length gene has been assembled, new primer-derived copies of the full-length gene (existing as double-stranded DNA, dsDNA) that are formed during the PCR amplification will possess a label on each end. Partially extended PCA products that are linearly amplified will possess one of the labels, but not both. Thus, the two-label full-length gene is unique to the mixture, and can thereby be isolated by using a two-column serial purification.

There are several benefits of the automated purification over manual gel purification. First, the procedure has the potential to be faster than gel purification (i.e. < 30 minutes compared to ~1-2 hours). Second, one need not be concerned with introducing synthetic DNA errors due to UV damage as in gel purification, as any aliquots to gels need not be directly used for cloning. Additionally, the automation of purification may reduce the labor burden thereby increasing the throughput potential for multiple gene synthesis by PCA. Lastly, affinity purification may be more effective than gel purification in isolating the desired full-length gene as gel purification offers fairly low resolution. One may expect a better quality purified fraction with affinity separation, thereby aiding cloning efforts, as it is very likely that positive clones will possess the full-length insert and thereby reduce the amount of screening needed.

1.5 Introduction to the Workstation Process

To advance the gene synthesis field, a novel process and accompanying workstation has been developed that incorporates rapid PCA synthesis coupled with affinity purification. The system enables fast and accurate PCA synthesis and isolation of the full length DNA of interest, enabling facile cloning and expression. In the synthesis step, the desired sequence is PCA assembled and PCR amplified in a fast thermocycler to generate a high yield of product with minimal runtime and errors. A traditional 2-step PCA-PCR approach is utilized to assemble and amplify the full-length gene. Alternatively, integration of PCA and PCR into a single rapid reaction is also employed, working reliably up to about 1 kb. The second step in the PCR workstation employs two columns

to isolate the desired full-length product from shorter unwanted reactants/products. During PCR amplification, labels are incorporated into the dsDNA product on both ends. The first column interacts with one of the labels to partially purify the mixture. The intermediate product mixture is then subsequently purified further via the second column. A schematic of the workstation process is shown in Figure 1-5.

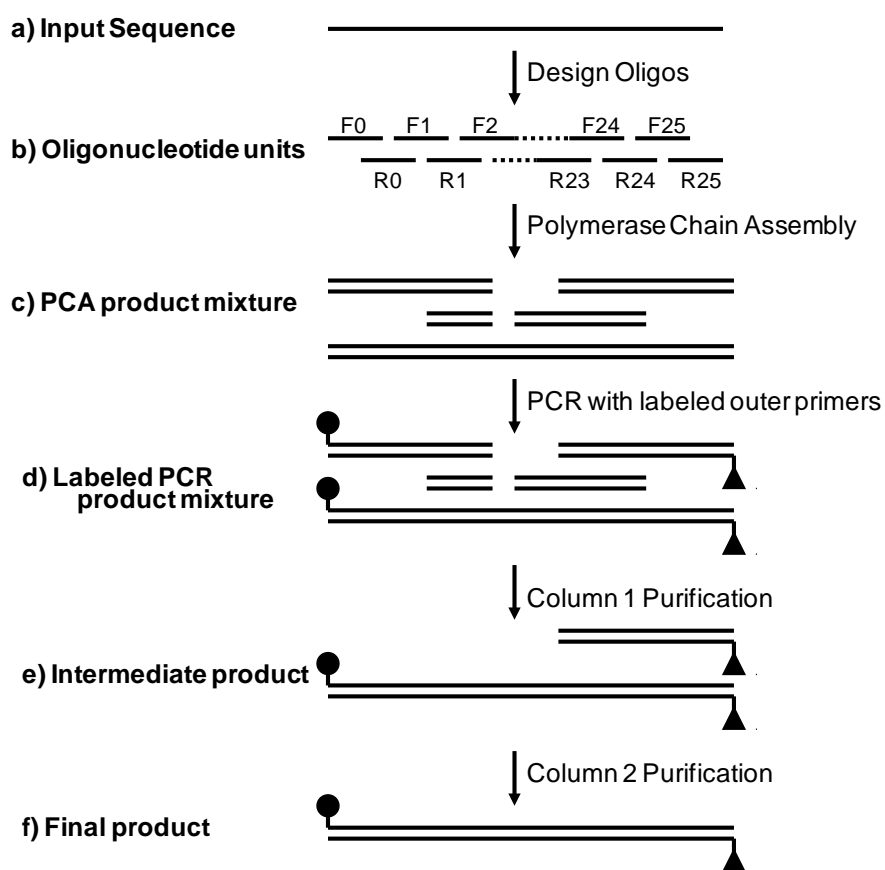


Figure 1-5 Schematic of the rapid gene synthesis workstation

A serial two-column purification is employed after gene synthesis.

1.6 Developmental Objectives

While others have broken ground in the field of PCR-based gene synthesis, the field is still searching for improved methods. The speed and high fidelity that can be achieved by rapid thermocycling has been described (Mamedov and others 2007), but has by and large gone unappreciated. New demonstrations of gene synthesis by rapid PCA, along with the construction of even longer gene sequences by the technique, are of great importance to wider adoption. The objectives of this work are outlined below:

1. Demonstrate the robustness and repeatability of rapid PCA gene synthesis for a variety of different genes. As the specific sequences of each gene play a significant impact on synthesis success, it is of critical importance that rapid PCA is proven as a robust technique.
2. Develop proof-of-concept for affinity purification platform and method using two columns in series to isolate genes synthesized by rapid PCA. The proof-of-concept represents the first ever demonstration of the process.
3. Develop a technique that combines PCA assembly and PCR amplification into a single step procedure to even further reduce thermocycling runtimes and the labor requirements to produce the desired gene in high yield.
4. Further develop the chromatography section with improved performance and characteristics based upon the proof-of-concept findings. Immunoaffinity purification using two antibody columns are preferred.
5. Develop a novel rapid PCA method that enables synthesis of longer gene sequences. The longest gene synthesized by Mamedov and others (2007) using

rapid PCA was a 1,548 bp thrombomodulin gene from a pool of 77 oligonucleotides.

The process and workstation steps are demonstrated for a variety of synthetic genes. In Chapter 2, four different genes were synthesized in an initial proof-of-concept workstation. Chapter 3 investigates a novel technique that integrates PCA and PCR steps in a single rapid thermocycling process. Chapter 4 presents development of a next generation instrument with refined processes and hardware. In Chapter 5, a long 3.8 kb gene sequence was synthesized to challenge the next generation workstation and demonstrate a convergent synthesis approach. To construct longer genes that are otherwise not amenable to direct synthesis, the sequence may be segmented into smaller fragments that are synthesized in parallel in 1st round PCA-PCR. 2nd round integrated PCA-PCR is performed to join the fragments into the full-length gene. The process is demonstrated for the synthesis of the 3.8 kb gene sequence by utilizing either 2 or 4 smaller fragments. Provided in an appendix, chimeric gene sequences integrating two encoding regions for two different proteins were explored. Chapter 6 provides an analysis of the research and direction for future work.

1.7 References

S.D. Allen, T.M. Luebke, and S.D. Rose. Ultramers- The longest oligonucleotides available with mass spectrometry. Taken from http://www.idtdna.com/home/Ultramers_Technical_Report_IDT.pdf (accessed January 12, 2009)

- Czar, M.J., J.C. Anderson, J.S. Bader, J. Peccoud. Gene Synthesis demystified. *Trends Biotechnol.* **27**: 63-72 (2009)
- d'Allesandro, G. and F. Sofia. The adsorption of antibodies from the sera of syphilitics and tuberculosis patients. *Z. Immunitats.* **84**: 237-250 (1935)
- Gao, X., P. Yo, A. Keith, T. Ragan, and T. Harris. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**: e143-e164 (2004)
- Gibson, D.G., et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220 (2008)
- Griep, M., C.A. Kotera, R.M. Nelson, and H.J. Viljoen. Kinetics of the DNA polymerase *pyrococcus kodakaraensis*. *Chem. Eng. Sci.* **61**: 3885–3892. (2006)
- Kodumal, S.J., K.G. Patel, R. Reid, H.G. Menzella, M. Welch, and D.V. Santi. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 5573-5578 (2004)
- Landsteiner, K. and J. van der Scheer. Cross reactions of immune sera to azoproteins. *J. Exp. Med.* **63**: 325-339 (1936)
- Mamedov, T.G., N.V. Padhye, H. Viljoen, and A. Subramanian. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387 (2007)
- Mehta, D.V., R.J. DiGate, D.L. Banville, and R.D. Guiles. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. *Prot. Expr. Purif.* **11**: 86-94 (1997)

- Pienaar, E., M. Theron, M. Nelson, and H.J. Viljoen, H.J., A quantitative model of error accumulation during PCR amplification. *Comput. Biol. Chem.* 30, 102-111. (2006)
- Quintanar, A. and R.M. Nelson. A process and apparatus for high-speed amplification of DNA. U.S. Patent No. 6,472,186. (2002)
- Rabinow, P. Making PCR: A Story of Biotechnology. University of Chicago Press. ISBN 0-226-70147-6. (1996)
- Smith, H.O., C.A. Hutchison, C. Pfannkoch, and J.C. Venter. Generating a synthetic genome by whole genome assembly: Φ X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 100: 15440-15445 (2003)
- Starkenstein, E. Ferment action and the influence upon it of neutral salts. *Biochem. Z.* **24**: 210-218 (1910)
- Stemmer, W.P.C., A. Cramer, K.D. Ha, T.M. Brennan, and H.L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*. **164**: 49-53 (1995)
- Watson, J.D. and F.H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 171: 737-8

2 Proof-of-Concept for the Workstation

To demonstrate the rapid PCA synthesis coupled with affinity purification process described in Chapter 1, a proof-of-concept workstation was developed. The system enables fast and accurate PCA synthesis, ending with isolation of the full length DNA of interest. In the synthesis section, a rapid 2-step PCA assembly with subsequent PCR amplification technique is used to generate a high yield of product with minimal runtime and errors. During PCR amplification, labels are incorporated into the dsDNA product on both ends. The second step in the PCR workstation employs two affinity columns to isolate the desired full-length product from shorter unwanted reactants/products. The first column interacts with one of the labels to partially purify the mixture. The intermediate product mixture is then subsequently purified further via the second column. Four genes were constructed and purified using the workstation to demonstrate the approach.

2.1 Introduction

2.1.1 Rapid PCR-based synthesis

For a background into gene synthesis by PCR, please refer to Chapter 1. The first section of the workstation is a rapid compressed-air thermocycler, the PCRJet[®] (Quintanar and Nelson 2002; Whitney 2004). The rapid thermocycler is utilized for the PCA and PCR steps of the workstation process. The PCRJet[®] accommodates up to eight 1.5 mm outer diameter glass capillaries (part # 04929292001, Roche Applied Science, Indianapolis, IN). Each capillary can hold up to 25 µl of reaction mixture.



Figure 2-1 Picture of the PCRJet rapid thermocycler

The PCRJet® (see Figure 2-1) is one of the fastest PCR thermocyclers built to date. Briefly, the instrument forces heated or room temperature compressed air at high velocity over the glass capillaries. The high heat transfer of the forced convection is coupled to the fast equilibrium characteristics of the small diameter capillaries. This results in heating and cooling rates in excess of $10^{\circ}\text{C}/\text{sec}$ with only a few seconds of hold time required at the set-point temperature for the PCR mixture to reach the desired temperature. In conjunction with the fast thermocycler, KOD polymerase is employed for its particularly suitable high fidelity and fast extension rates (Griep and others 2006) to give very short runtimes of about 10 minutes.

To test the robustness of the prototype workstation, four different genes were selected. The details of the four genes and the primer sequences used are listed in Table 2-1.

Gene	Accession number	Size (bp)	GC content	# of Oligos	Primer pair sequences
EPCR-1	L35545	612	59%	35	5-TTTTGTAGCCAAGACGCCTC-3 5-CGAAGTGTAGGAGCGGCTT-3
Thrombomodulin	BC035602	1548	42%	77	5-ATCTCGAGAAAAGACATCATC-3 5-TAGCGGCCGCTTAAGAATG-3
puC-19 β -lactamase	L09137	929	48%	52	5-GTAACTTGGTCTGACAGTTAC-3 5-TGAGACAATAACCCTGATAAATG-3
puC-19 ori	L09137	615	54%	36	5-CGCGTTGCTGGCGTTTTTC-3 5-AAAGGATCTTCTTGAGATCCT-3

Table 2-1 Genes synthesized in the proof-of-concept workstation.

2.1.2 Purification of synthetic genes

The second section of the synthesis workstation is tasked with purification of the reaction products to isolate the desired full-length dsDNA of interest. A schematic of the prototype workstation is shown in Figure 2-2.

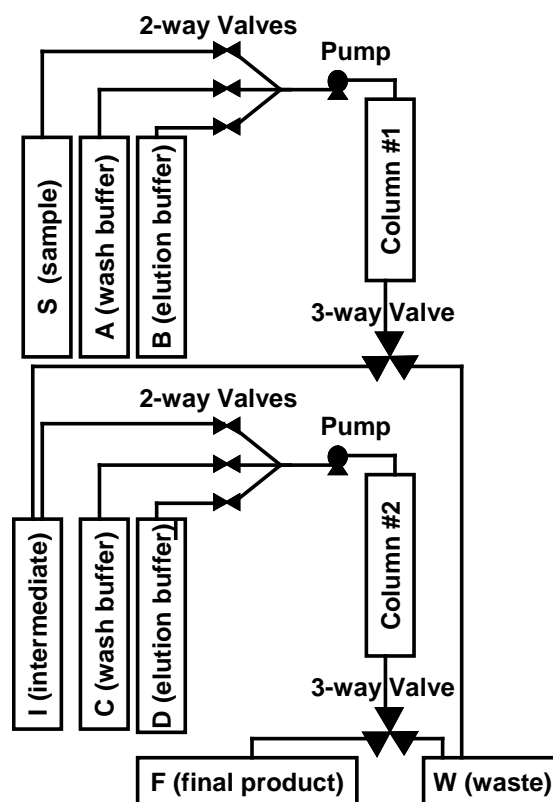


Figure 2-2 Schematic of the prototype affinity chromatography station.

One commonly used interaction is the streptavidin-biotin (or biotin analog) system (Hofman and others 1980; Fudem-Goldin and Orr 1990; Hirsch and others 2002). Iminobiotin is an attractive analog due to the reversible binding and elution characteristics at modified pH values. Additionally, the thymidine analog bromodeoxyuridine (BrdU) antibody-antigen systems also find common use in immunoaffinity separations, especially for in vivo labeling (Haider and others 1997; Kobayashi and others 1998; Urbach and others 1999; Artursson and Jansson 2003; Walters and Field 2006). Thus these two labels are employed in the prototype workstation as demonstration of the concept.

2.2 Materials & Methods

2.2.1 Design of Oligonucleotides

The Gene2Oligo computer program of Rouillard and others (2004) was used to design assembly oligonucleotides and optimize melting temperatures of the hybridization units. Oligonucleotides at either 50 or 100 μ M in nuclease-free water were obtained from Integrated DNA Technologies (Coralville, IA).

2.2.2 Polymerase Chain Assembly Conditions

The synthesis of the genes was performed in accordance with the work of Mamedov and others (2007). Equimolar PCA was carried out in 25 μ L reaction volumes as listed in Table 2-2.

Reagent	β -lactamase	pUC19ori	Thrombomodulin	EPCR-1
dNTPs	200 μ M	200 μ M	200 μ M	200 μ M
MgSO ₄	4mM	4mM	5mM	4mM
Non-acetylated BSA	0.4mg / mL	0.4mg / mL	2mg / mL	0.4mg / mL
KOD hot start polymerase	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction
KOD Reaction Buffer	1X	1X	1X	1X

PCR Protocol	β-lactamase		pUC19ori		Thrombomodulin		EPCR-1	
	Temp (°C)	Time (s)	Temp (°C)	Time (s)	Temp (°C)	Time (s)	Temp (°C)	Time (s)
Hot start	94	30	94	30	94	30	92	30
Denaturation	94	2	94	2	92	2	92	1
Annealing	56	10	56	10	60	15	68	4
Elongation	72	10	72	10	72	7		
# of Cycles	30		30		30		30	
Final extension	72	25	72	25	72	20	-	-

Table 2-2 Polymerase Chain Assembly conditions.

2.2.3 PCR Amplification Conditions

PCR amplification using regular or labeled primers was carried out in 25 μ L reaction volumes as listed in Table 2-3. PCR was performed on aliquots of either the assembly or affinity purification products.

PCR Protocol	β -lactamase		pUC19ori		h-TM		EPCR-1 ¹	
PCR Protocol	Temp (°C)	Time (s)	Temp (°C)	Time (s)	Temp (°C)	Time (s)	Temp (°C)	Time (s)
Hot start	94	30	94	30	92	30	92	30
Denaturing	94	2	94	2	92	2	92	1
Annealing	55	3	55 / 56 ³	3	60	3 / 2 ⁴	60	2
Elongation	72	10	72	10	72	15	72	4
# of Cycles	30		30		40		35	
Final extension	72	25	72	25	72	20	72	20

Reagent	β -lactamase	pUC19ori	h-TM	EPCR-1 ¹
dNTPs	200 μ M	200 μ M	200 μ M	200 μ M
Synthetic DNA	0.5 μ L / 25 μ L reaction	0.5 μ L / 25 μ L reaction	2 μ L / 25 μ L reaction	0.7 μ L / 25 μ L reaction
MgSO ₄	4mM	4mM	5mM	4mM
Non-acetylated BSA	0.4mg / mL	0.4mg / mL	0.4 – 2.0 mg / mL ⁴	0.4mg / mL
KOD hot start polymerase	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction	0.5U / 25 μ L reaction
KOD Reaction Buffer	1X	1X	1X	1X
Forward primer	0.6 μ M / 0.7 μ M ²	0.5 μ M - 0.7 μ M ³	0.6 μ M	0.7 μ M
Reverse primer	0.6 μ M / 0.7 μ M ²	0.5 μ M - 0.7 μ M ³	0.6 μ M	0.7 μ M

Table 2-3 PCR amplification conditions.

- 1: PCR amplification of purified product was performed using 1 μ L template with pH adjusted to 7 by adding 1.06 μ L of Tris base.
- 2: Unlabelled PCR of PCA product: 0.6 μ M each. Labelled PCR of PCA product or unlabelled PCR of purified product: 0.7 μ M each
- 3: Unlabeled amplification of PCA product was done using 0.7 μ M of each primer at 55°C annealing temperature. Labeled amplification of PCA product was done using 0.5 μ M of each primer at 56°C annealing temperature. Unlabeled amplification of purified product was done using 0.6 μ M of each primer at 56°C annealing temperature.
- 4: BSA concentrations were 2.0 and 0.4 mg/mL for labeled and unlabeled amplifications, respectively. Annealing times were 3 and 2 seconds for labeled and unlabeled amplifications, respectively.

2.2.4 Labeling of synthesis products

The forward and reverse primers used for PCR amplification were labeled with iminobiotin and BrdU labels attached to their 5' ends, respectively. During the PCR amplification step of PCA synthesized product, the polymerase-extended primers contain

the labels. The labeled primers were obtained from IBA GmbH (Göttingen, Germany), while regular primers were obtained from Integrated DNA Technologies (Coralville, IA).

2.2.5 BrdU and Iminobiotin affinity columns

Chicken polyclonal antibody against BrdU (anti-BrdU Pab) was purchased from Immunology Consultants Laboratory, Inc. (Newberg, OR). Activated NHS-Sepharose resin was purchased from Sigma Chemical Company (St. Louis, MO). The anti-BrdU-Pab was coupled to the NHS-Sepharose resin per the protocol supplied by the manufacturer. Briefly, the resin as supplied was drained and washed with 1 M ice-cold HCl and decanted. The resin was subsequently washed with approximately 10 column volumes of 0.1 M sodium bicarbonate, 0.5 M NaCl, pH 8.0 buffer (ligand coupling buffer) prior to antibody coupling. The target antibody was dissolved in 0.1 M sodium bicarbonate, 0.5 M NaCl, pH 8.0 buffer (ligand coupling buffer), mixed with decanted resin at a volume ratio of 3:1 and placed on an end-to-end rotator in a cold room (4°C). Upon completion of the antibody coupling step, the supernatant was drained and excess reactive groups on the resin were blocked with 1M ethanolamine, pH 8.0 for 1-hour at room temperature. Upon completion of the blocking step, the resin was sequentially washed with 0.1 M sodium bicarbonate, 0.05 M Tris-HCl, 0.5 M NaCl, pH 8.0 and 10 mM Tris-HCl, 50 mM NaCl, pH 7.4. Unless used immediately, the antibody coupled gel was stored at a neutral pH in the presence of 0.05% NaN₃ at 2°C to 8°C.

For the iminobiotin affinity capture, its well-known interaction with streptavidin was employed. Ready to use 1ml HiTrap Streptavidin HP columns were purchased from GE Healthcare (part # 17-5112-01, Piscataway, NJ).

2.2.6 Affinity Purification Hardware

The hardware layout of the affinity workstation is in accordance with Figure 2-2. The affinity chromatography section is comprised of pumps, valves, tubing, columns, reagent / buffer vessels and the necessary electronics and software to control all of these components. The components were assembled on a vertical breadboard using perforated sheet metal to maximize flexibility and different arrangements for prototyping as needed. A picture of the prototype workstation is shown in Figure 2-3.



Figure 2-3 Picture of the Prototype workstation using a vertical breadboard.

The components follow that of the schematic in Figure 2-2. The power supply and custom printed circuit board are located on the right.

The key hardware components of the system are described briefly below:

Pumps. There are two pumps present in the prototype affinity section, one pump for each column. The displacement diaphragm pumps (model # 120SP2410-4TE, Bio-Chem Valve Inc., Boonton, NJ) are factory calibrated to 10 μ L per pump stroke (at zero back-pressure) at a rate of up to 2 Hz (1.2ml per minute max). The actual flow rates of the pumps were calibrated to account for the tubing lengths and backpressure of the system. Higher flow rate pumps are also available. Bio-Chem Valve components were chosen for the chemical inertness of the internal components and attractive pricing compared to other pump designs.

2-way valves. The station has six normally closed 2-way valves. These micro isolation valves (model # 038T2S24-32-4, Bio-Chem Valve Inc.) provide a compact method to direct fluid flow.

3-way valves. There are two 3-way isolation valves in the station (model # 075T3MP42-32, Bio-Chem Valve Inc.). Each 3-way valve selects the direction for the fluid to flow downstream of the columns – either directed towards the next stage or towards a waste container.

Tubing & Fittings. The pumps and valves are connected with chemically inert PTFE tubing (part # PTFE1030, Western Analytical Products, Wildomar, CA). This replaceable and sterilizable tubing has a 0.030” inner diameter and 1/16”

outer diameter. Various PTFE and PEEK fittings and connectors connect the workstation components (Western Analytical Products, Wildomar, CA).

Reagent vessels. Several reagent vessels are used by the station for tubing priming, column washing and pH control. The buffer vessels were 50 mL centrifuge tubes (model # 2553, CLPdirect, San Diego, CA), while the feed and intermediate vessels were 5ml tubes, both as a matter of convenience.

Affinity columns. A 1 mL HiTrap Streptavidin HP column (part # 17-5112-01, GE Healthcare, Piscataway, NJ) was used for the iminobiotin capture. A custom anti-BrdU Pab (Immunology Consultants Laboratory, Inc., Newberg, OR) was coupled to NHS-sepharose resin (Sigma Chemical Company, St. Louis, MO). The resin was prepared and packed in a Tri-corn 5/20 glass column (part # 28-4064-08, GE Healthcare, Piscataway, NJ).

Power supply. A 24V DC power supply capable of delivering 60W is used to power the pumps and valves (model # MAP55-1024, Power-One, Camarillo, CA).

Controller card. A custom printed circuit board (PCB) was built to trigger the pumps and valves (MIS Engineering, Lincoln, NE). This PCB allows for either manual control with push buttons or with an external computer control via 5 VDC inputs.

The manual control option was used to prime the system; all tubing and affinity purification columns were filled with the proper initial buffer. Automated control was used to then perform the pump and valve control for the actual affinity purification. A KPCI-3107 board (Keithley Instruments) was interfaced to the controller card to run the automated protocol. A simple Visual Basic 6.0 software interface was written by Dr. Scott Whitney to control valve settings and operating the pumps at each programmed step by sending digital outputs to the custom printed circuit board.

2.2.7 Purification Methods

Typical purification procedures are described here. Unless otherwise specified, all experimentation was performed following these protocols.

- The binding buffer **A** for the anti-BrdU column was 10mM Tris-HCl, 50mM NaCl, pH 7.4.
- The elution buffer **B** for the anti-BrdU column was 20mM Tris-HCl, 500mM NaCl, pH 7.5.
- The binding buffer **C** for the Streptavidin column was 50mM ammonium carbonate, 500mM NaCl, pH 10.0.
- The elution buffer **D** for the Streptavidin column was 50mM ammonium acetate, 500mM NaCl, pH 4.0.

The affinity workstation was initialized by priming lines with the respective solutions. The intermediate vessel was preloaded with 3 mL of 1x binding buffer (C) and the sample vessel loaded with 200 μ L of labeled PCR product mixture and 1200 μ L sterile

distilled water. The high amount of feed product was used to overcome dilution and gel sensitivity considerations; smaller feed amounts (i.e. 25 μ L) may also be employed. Pumping was set at 1 Hz for an approximate flow rate of 0.6 mL/min. The software was programmed to run the following sequence:

- Equilibrate columns and tubing with respective binding buffers **A** and **C**
- Load in sample mixture **S** to anti-BrdU column
- Wash BrdU column with binding buffer **A**
- Elute intermediate product from BrdU column with elution buffer **B**
- Load intermediate mix **I** to streptavidin column
- Wash streptavidin column with binding buffer **C**
- Elute final product **F** from streptavidin column with elution buffer **D**

To adequately characterize the purification characteristics, the protocol for the four genes was performed in a semi-automated fashion. Samples for binding, washing, and elution fractions for both columns were collected manually in approximately 200 μ l increments from the respective outlet tubing into 1.5 ml microcentrifuge tubes for gel analysis.

To fully demonstrate the automated protocol, purification with the β -lactamase synthetic gene was conducted once the protocol was sufficiently optimized. The automated protocol was completed in 33 minutes at the conservative 0.6 ml / min flow-rates.

2.2.8 Gel Electrophoresis

For the electrophoresis gels, typically 8 μ L of each PCA or PCR sample or 30 μ L of each affinity chromatography fraction were run on 1.0% agarose gel stained with ethidium bromide at 10V/cm in TAE buffer.

2.2.9 Cloning and Sequencing

The synthetic genes were gel-purified using a Gel Elute Extraction Kit (Qiagen Inc., Valencia, CA). Purified DNA was ligated to TOPO vector using Zero Blunt[®] TOPO[®] PCR Cloning Kit (Invitrogen, Carlsbad, CA). The ligated plasmid was transformed into competent cells of *E. coli* strain TOP-10 (Invitrogen); ten positive clones were selected. Purified plasmids from selected clones were sequenced at the University of Nebraska-Lincoln sequencing facility (CEQ8000 DNA Sequencer, Beckman Coulter, Fullerton, CA) using vector and gene specific primers.

2.3 Results & Discussion

The application of the prototype workstation to the synthesis and purification of β -lactamase, pUC19ori, TM and EPCR-1 are elaborated below. The four genes were successfully assembled from 52, 36, 77 and 35 synthetic oligonucleotides, respectively. In other testing with unlabelled product, some non-specific binding was observed with the anti-BrdU column (results not shown), suggesting that the full-length gene purity in the final effluent may be partially compromised if undesired iminobiotin-labelled products are non-specifically bound to the anti-BrdU column in the first stage. However, the four labeled genes tested did not show any significant background present.

2.3.1 929 bp puc19 B-lactamase sequence

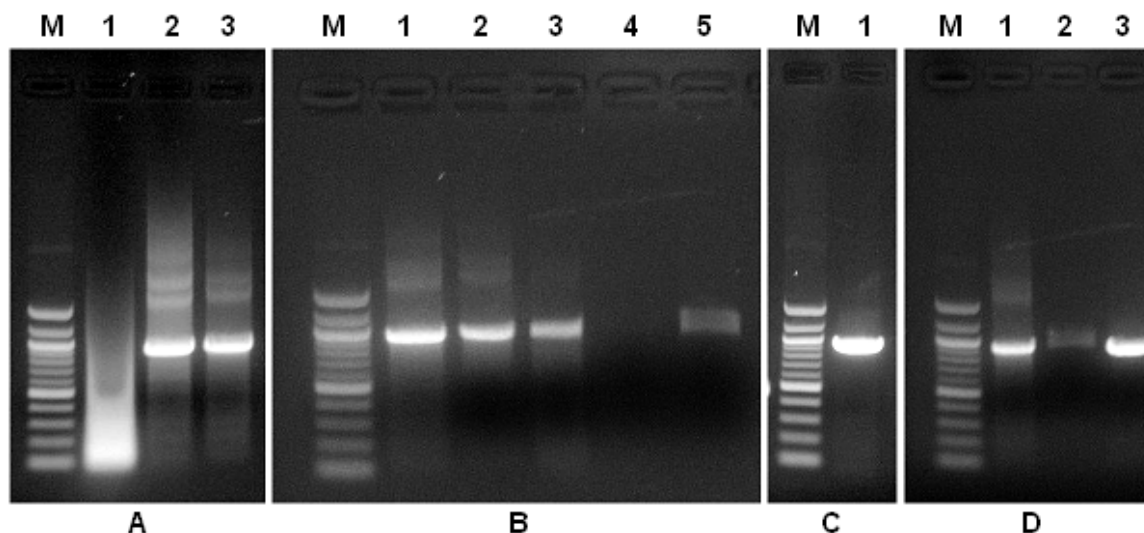


Figure 2-4 PCA, PCR and Affinity Purification of the pUC19 β -lactamase gene.

A: PCA & PCR products- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: PCA product; lane 2: PCR product with unlabeled primers; lane 3: PCR product with labeled primers.

B: Semi-automated Affinity Purification- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: labeled PCR product (8 μ L); lane 2: wash sample from BrdU column (30 μ L); lane 3: intermediate elution product from BrdU column (30 μ L); lane 4: wash sample from streptavidin column (30 μ L); lane 5: final elution product from streptavidin column (30 μ L).

C: PCR of Affinity Purification Product- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: 929bp PCR of the affinity purification product.

D: Fully Automated Affinity purification- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: PCR product with labeled primers (feed); lane 2: Final Elution Product; lane 3: 929bp PCR of the Affinity Purification Product.

The assembly product and PCR amplification (with both regular and labeled primers) of the assembly product is shown in Figure 2-4A. The success of the PCA is confirmed by the presence of the correct size full length bands from the PCR amplifications. Note that a significant smear is present in the PCR steps due to a complex mixture of 1) carryover and linear amplification of short assembly oligonucleotides and 2) duplex formation of different length single stranded DNA with complementary sequence regions.

Samples from wash/elution fractions of the affinity purification protocol are Figure 2-4B. As expected, the intermediate elution product from the anti-BrdU column has both the full-length product present along with undesired shorter products that possess the BrdU label but not the iminobiotin label. Subsequent purification of this intermediate through the streptavidin column results in only the full length product with both labels in the elution fraction. The absence of visible bands in the streptavidin wash step is attributed to the high dilution of the intermediate product. The slight upward shift of the final product is due to the high salt present in the elution buffer. The purity of the affinity purification product and its correct size was confirmed by subsequent PCR (Figure 2-4C). Sampling of the affinity purification steps interrupted the protocol and resulted in a semi-automated process.

To demonstrate the robustness of the workstation, the protocol was run in fully automated mode—sampling was only done at the final elution step. The eluted product and its PCR amplification of the automated procedure are shown in Figure 2-4D. The success of the automated procedure provides evidence that the affinity purification approach may be

used as an alternative to gel purification. Gel purification typically requires about 2 hours to conduct gel separation and then recover the desired band from the gel slice. In contrast, the fully automated protocol required 33 minutes for completion. The comparison is not necessarily clear-cut as the affinity purification is at higher cost and additional processing steps may be needed depending on downstream objectives. Yet, the additional advantage is that the affinity chromatography does not expose the synthesized DNA to UV damage.

2.3.2 612 bp endothelial protein C receptor sequence

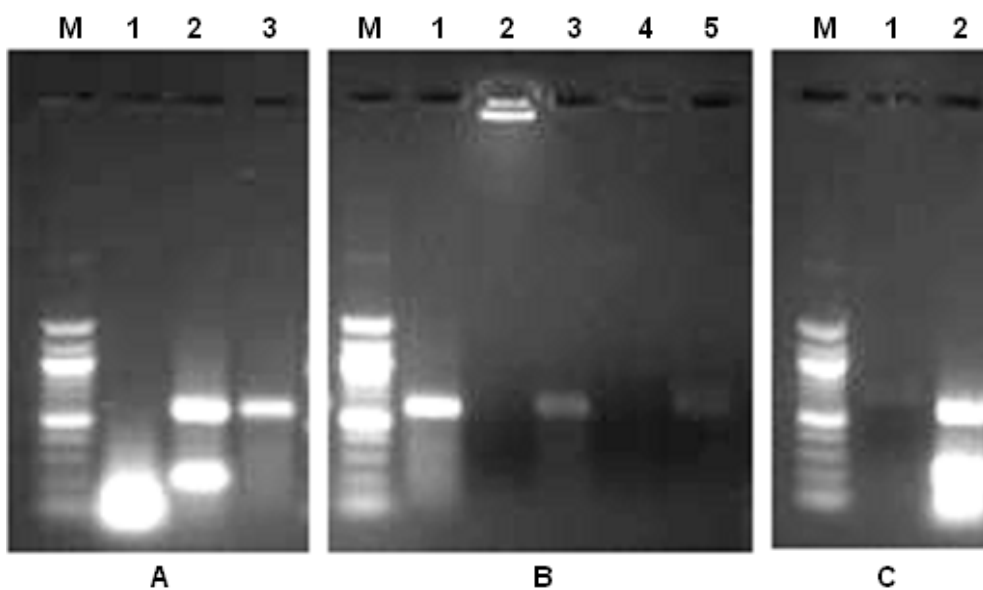


Figure 2-5 PCA, PCR and affinity purification of the Endothelial Protein C Receptor gene.

A: PCA & PCR products- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: PCA product; lane 2: PCR product with unlabeled primers; lane 3: PCR product with labeled primers.

B: Semi-automated Affinity Purification- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: labeled PCR product (8 μ L); lane 2: wash sample from BrdU column (30 μ L); lane 3: intermediate elution product from BrdU column (30 μ L); lane 4: wash sample from streptavidin column (30 μ L); lane 5: final elution product from streptavidin column (30 μ L).

C: PCR of Affinity purification Product- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: final elution product; lane 2: PCR of the affinity purification product.

The assembly product and PCR amplification (with both regular and labeled primers) of the assembly product is shown in Figure 2-5A. The success of the PCA is confirmed by the presence of the correct size full length bands from the PCR amplifications. A secondary non-specific band was obtained with regular primers (Figure 2-5A, lane 2), despite optimization attempts. As with the other genes, a significant smear is present in the PCR steps.

Samples from wash/elution fractions of the affinity purification are shown in Figure 2-5B. As expected, the intermediate elution product from the anti-BrdU column has both the full-length product present along with undesired shorter products that possess the BrdU label but should not contain the iminobiotin label. Subsequent purification of this intermediate through the streptavidin column results in the full length product with both labels in the elution fraction. The absence of visible bands in the streptavidin wash step is attributed to the high dilution of the intermediate product. The slight upward shift of the final product is due to the high salt of the elution buffer. The anti-BrdU wash gel lane

(Figure 2-5B, lane 2) was corrupted—other gels have shown this product to look similar to the labeled PCR product lane (results not shown).

The purity of the affinity purification product and its correct size was confirmed by subsequent PCR with regular primers (Figure 2-5C). The non-specific amplification of undesired products is attributed to PCR and not the purification performance (compare to Figure 2-5A, lane 2).

2.3.3 615 bp puc19 ori sequence

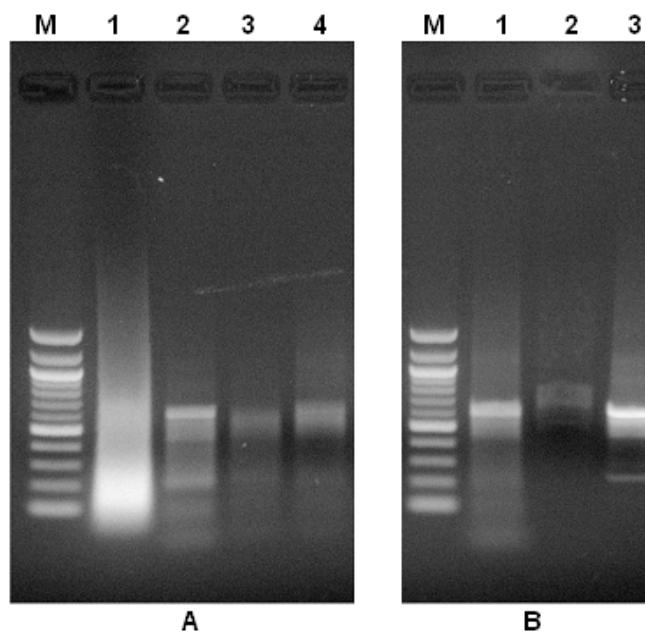


Figure 2-6 PCA, PCR and affinity purification of the pUC19 Origin of Replication gene.

A: PCA & PCR products- Lanes (M) 100bp DNA ladder (NEbiolabs); lane 1: PCA product; lane 2: PCR product with unlabeled primers; lane 3: PCR product with labeled primers; lane 4: PCR product with labeled primers (new reverse primer).

B: Semi-automated Affinity purification- Lanes (M) 100bp DNA ladder (NEbiolabs);

lane 1: Iminobiotin-labeled PCR product (8 μ L); lane 2: Elution product from streptavidin column (30 μ L); lane 3: Concentrated elution product.

The assembly product and PCR amplification (with both regular and labeled primers) of the assembly product is shown in Figure 2-6A. The success of the PCA is confirmed by the presence of the correct size full length bands from the PCR amplification with regular primers (Figure 2-6A, lane 2). Efficient PCR with labeled primers was problematic (Figure 2-6A, lane 3). The first labeled primer set yielded a smeared band at approximately the correct size, and the poor yield was unattractive for purification testing. PCR with mixing of labeled/unlabeled primers showed correct amplification with both forward primers, but only in combination with the regular reverse primer (results not shown). This suggested that the reverse BrdU-labeled primer might be the culprit; a replacement was ordered from the supplier. However, the result was similar using the new primer (Figure 2-6A, lane 4).

Due to the difficulties with the BrdU-labeled primer, a single column purification with streptavidin was employed to partially purify the gene. First, the PCA product was amplified using the iminobiotin-labeled forward primer and the regular reverse primer (Figure 2-6B, lane 1). The affinity purification workstation and software were reconfigured to run only the streptavidin column section. The purified elution product is shown in Figure 2-6B, lane 2. Note the slight upward shift due to high salt as seen with the other genes.

With only one column used in the purification, it is expected that impurities will still be present—those which possess the labeled 5' end but may or may not be the desired full-length product. The product should be similar to the intermediate in the tandem purification scheme. To reveal the presence of impurities (along with confirming the correct length by gel mobility), a purification kit was used on the elution fraction per manufacturer's protocol (Cat # A9281, Promega Corporation, Madison, WI). The concentrated elution product is shown in Figure 2-6B, lane 3. As anticipated, a background of undesired products still exists after the single column purification.

2.3.4 1548 bp human Thrombomodulin sequence

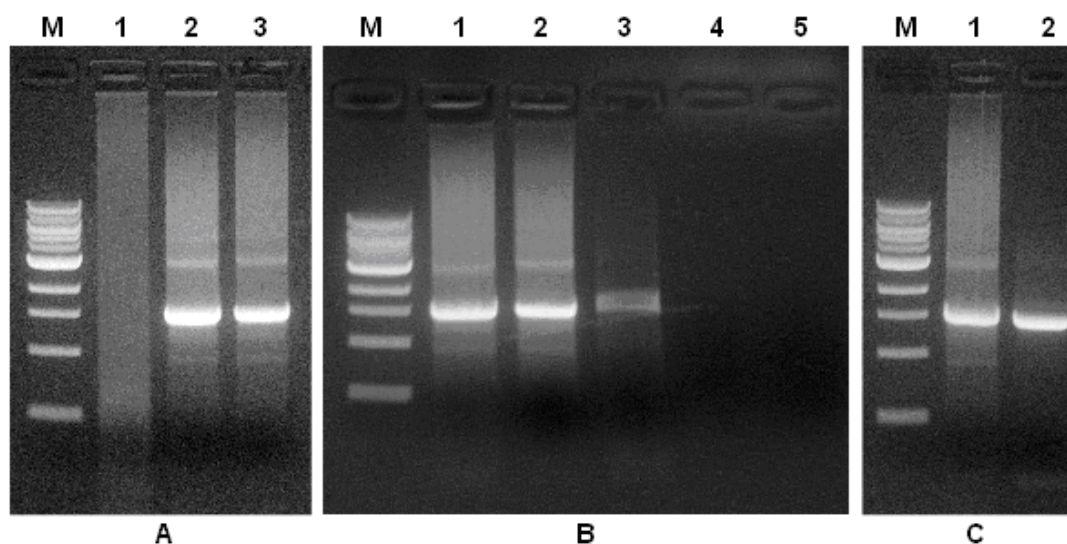


Figure 2-7 PCA, PCR and affinity purification of the human Thrombomodulin gene.

A: PCA & PCR products- Lanes (M) 1kb DNA ladder (NEBiolabs); lane 1: PCA product; lane 2: PCR product with unlabeled primers; lane 3: PCR product with labeled primers.

B: Semi-automated Affinity Purification- Lanes (M) 1kb DNA ladder (NEBiolabs); lane 1: labeled PCR product (8 μ L); lane 2: wash sample from BrdU column (30 μ L); lane 3: intermediate elution product from anti-BrdU column (30 μ L); lane 4: wash sample from

streptavidin column (30 μ L); lane 5: final elution product from streptavidin column (30 μ L).

C: PCR of Affinity purification Product- Lanes (M) 1kb DNA ladder (NEbiolabs); lane 1: PCR product with labeled primers (feed); lane 2: PCR of the affinity purification product. The assembly product and PCR amplification (with both regular and labeled primers) of the assembly product is shown in Figure 2-7A. The success of the PCA is confirmed by the presence of the correct size full length bands from the PCR amplifications. Note that a significant smear is present in the PCR steps due to a complex mixture of 1) carryover and linear amplification of short assembly oligonucleotides and 2) duplex formation of different length single stranded DNA with complementary sequence regions.

Samples from wash/elution fractions of the affinity purification protocol are shown in Figure 2-7B. As expected, the intermediate elution product from the anti-BrdU column has both the full-length product present along with undesired shorter products that possess the BrdU label but not the iminobiotin label. The absence of visible bands in the streptavidin wash step is attributed to the high dilution of the intermediate product. The absence of any visible product in the streptavidin elution sample is likely attributable to low binding efficiency of the labels and the detection sensitivity of the gel. However, the presence of desired product was confirmed by subsequent PCR (Figure 2-7C) on the elution fraction.

2.3.5 Synthesis Rates, Cloning, and Sequencing

	puC-19 β -lactamase	puC-19 ori	h-TM	EPCR-1
Product Size (bp)	929	615	1548	612
# of oligonucleotides	52	36	77	35
Assembly Rate (bp/hr)	4,355	4,085	6,057	6,994
Total errors in ten clones sequenced	8	0	8	0
Errors/kb	0.86	NIL	0.52	NIL

Table 2-4 Summary of the four assembled genes.

Assembly times and error rates for the four genes are summarized in Table 2-4. All four genes were synthesized at rates that exceed 4,000 bp/hr. Note the difference in the synthesis rates of the pUC19 origin of replication and the EPCR gene, both of which are of comparable length. Due to the differences in their oligonucleotide sets (from which they are assembled), the pUC19 origin of replication required a lengthier assembly protocol than the EPCR gene. All error rates for the four different genes were below 1 error/kb, calculated after sequencing of ten clones each. The EcoRI digests of ten clones obtained from the purified β -lactamase product are shown in Figure 2-8. The gel indicates that the full-length gene insert was present in all clones. The presence of the full-length insert in all ten clones was confirmed by sequencing.

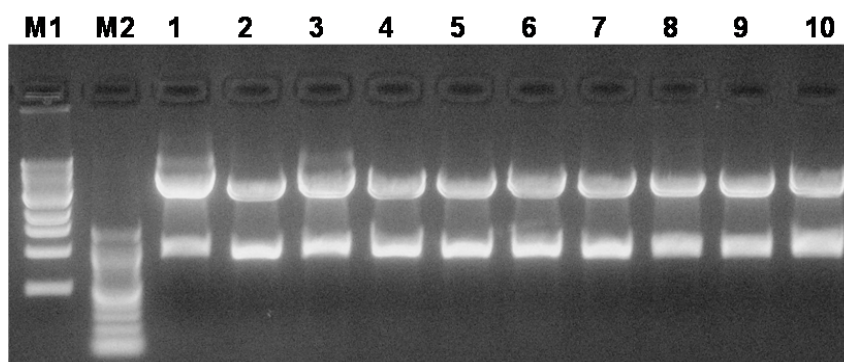


Figure 2-8 EcoRI restriction enzyme digest of 10 clones.

Lane M1: 1kb DNA ladder (NEbiolabs); Lane M2: 100bp DNA ladder (NEbiolabs); lane 1-10: EcoRI digest of ten clones (10 μ L each)

2.4 Conclusion

The primary objective of demonstrating feasibility of the synthesis and purification platform has been met. Four different genes were constructed rapidly and with high fidelity in the PCRJet[®] thermocycler. A fully automated purification was performed on the synthesized β -lactamase gene, indicating that the serial affinity purification approach may serve as an alternative to gel purification. Tempered success was obtained for the three other genes tested. Thrombomodulin also performed rather well, with the exception of low yield in the final effluent that was below the gel detection limit. PCR amplification of the final effluent clearly demonstrated that the full-length TM gene was present. For pUC19 ori and EPCR, difficulties arose in PCR amplification, thus limiting the clarity of purification results obtained.

However promising, process refinements are recommended such that the robustness of the platform in all respects is increased. The up-front thermocycler hardware is well-developed, yet the PCA and PCR reaction conditions may be further optimized for specificity, yield, and runtimes. While the initial affinity purification system served its purpose, a next generation should seek to address present concerns, both in hardware and processes. For example, in testing the purification setup with unlabelled PCR products, the anti-BrdU column did exhibit some non-specific binding (results not shown). If some of these non-specific products contain the iminobiotin label, they will be captured by the streptavidin column and eluted in the desired final effluent. Some impurities in the final effluent are not necessarily lethal to the success of the process. Rather, the objective with

any purification technique is to increase the relative signal of the desired product against undesired products.

2.5 References

- Artursson, V. and J.K. Jansson. Use of bromodeoxyuridine immunocapture to identify active bacteria associated with arbuscular mycorrhizial hyphae. *Appl. Environ. Microbiol.* **69**: 6208-6215 (2003).
- Fudem-Goldin, B. and G. Orr. Methods in Enzymology. Vol. 184. M. Wilchek and E. Bayer, Eds. Academic Press, New York, (1990) pp. 167–173.
- Griep, M., C.A. Kotera, R.M. Nelson, and H.J. Viljoen. Kinetics of the DNA polymerase *pyrococcus kodakaraensis*. *Chem. Eng. Sci.* **61**: 3885–3892. (2006)
- Haider, S.R., G. Juan, F. Traganos, and Z. Darzynkiewicz. Immunoseparation and immunodetection of nucleic acids labeled with halogenated nucleotides. *Exp. Cell Res.* **234**: 498-506 (1997).
- Hirsch, J.D., L. Eslamizar, B.J. Filanoski, N. Malekzadeh, R.P. Haugland, J.M. Beechem, and R.P. Haugland. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**: 343-357 (2002).
- Hofmann, K., S.W. Wood, C.C. Brinton, J.A. Montibeller, and F.M. Finn. Iminobiotin affinity columns and their application to retrieval of streptavidin. *Proc. Natl. Acad. Sci. U.S.A.* **77**: 4666-4668 (1980).
- Kobayashi, T., T. Rein, and M.L. Depamphilis. Identification of primary initiation sites for DNA replication in the hamster dihydrofolate reductase gene initiation zone. *Mol. Cell. Biol.* **18**: 3266-3277 (1998).

- Mamedov, T.G., N.V. Padhye, H. Viljoen, and A. Subramanian. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387 (2007).
- Quintanar, A. and R.M. Nelson, 2002. A process and apparatus for high-speed amplification of DNA. U.S. Patent No. 6,472,186. (2002)
- Rouillard, J.M., W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**: W176-W180 (2004).
- Urbach, E., K.L. Vergin, and S.J. Giovannoni. Immunochemical detection and isolation of DNA from metabolically active bacteria. *Appl. Environ. Microbiol.* **65**: 1207–1213 (1999).
- Walters, S.P. and K.G. Field. Persistence and growth of fecal bacteroidales assessed by bromodeoxyuridine immunocapture. *Appl. Environ. Microbiol.* **72**: 4532-4549 (2006).
- Whitney, S.E. Analysis of rapid thermocycling for the polymerase chain reaction. *ETD collection for University of Nebraska - Lincoln*. Paper AAI3131568. (2004)

2.6 Publication and Author Contributions

A substantial portion of this work was published under:

- TerMaat, J.R., T.G. Mamedov, E. Pienaar, S.E. Whitney, and A. Subramanian. Automated two-column purification of Iminobiotin and BrdU labeled PCR products for rapid cloning: application to genes synthesized by Polymerase Chain Assembly. *J. Chromatogr. Sci.* **48**: 120-4. (2010)

Author Contributions:

J.R. TerMaat was primarily responsible for development and testing of the affinity workstation.

T.G. Mamedov was primarily involved in the design of gene sequences, coupling of the anti-BrdU antibody to sepharose resin, and batch testing of the affinity purification.

E. Pienaar was primarily responsible for conducting PCA synthesis and PCR amplification reactions for the four genes.

S.E. Whitney developed the Visual Basic software to control the pump and valve settings to operate the workstation.

A. Subramanian led the scientific direction of the research.

3 Gene synthesis by integrated polymerase chain assembly and PCR amplification

In traditional PCA, equimolar concentrations of single stranded DNA oligonucleotides are repeatedly hybridized and extended by a polymerase enzyme into longer dsDNA constructs, with relatively few full-length sequences being assembled. Thus, traditional PCA is followed by a second primer-mediated PCR reaction to amplify the desired full-length sequence to useful, detectable quantities. Integration of assembly and primer-mediated amplification steps into a single reaction using a high-speed thermocycler is shown to produce similar results. For the integrated technique, the effects of oligo concentration, primer concentration, and number of oligonucleotides are explored. The technique is successfully demonstrated for the synthesis of two genes encoding EPCR-1 (653 bp) and pUC19 β -lactamase (929 bp) in under 20 minutes. However, rapid integrated PCA-PCR was found to be problematic when attempted with the TM-1 gene (1509 bp). Partial oligonucleotide sets of TM-1 could be assembled and amplified simultaneously, indicating that the technique may be limited to a maximum number of oligonucleotides due to competitive annealing and competition for primers.

3.1 Introduction to the concept of integrated PCA-PCR

PCA is a variation of the PCR process in which many oligonucleotides (“oligos”) are repeatedly hybridized, extended by a polymerase, and denatured. The desired gene sequence is segmented into short phosphoramidite synthesized oligo sequences of ~15 to 100 bp in length. The oligo design is typically aided by an oligo design software program (Rouillard and others 2004). Using PCR thermocycling, the oligos are hybridized to one another and undergo polymerase chain extension to construct longer

dsDNA. After repeated thermal cycling, the desired full-length sequence is assembled. In traditional equimolar PCA, relatively little full-length dsDNA is completely assembled during the PCA process. Thus, PCA is usually followed by a second primer-mediated PCR amplification of the full length template to produce useful (and detectable) quantities of the synthetic gene.

The ability to conduct PCA and primer-mediated amplification simultaneously is highly attractive due to reduced labor, reagents, and instrument time. A more subtle advantage of the integrated technique is that a reduction of synthetic errors attributable to thermal damage can be logically assumed from fewer thermal cycles (Pienaar and others 2006). Chen and others (1994) described PCR based *de novo* gene synthesis of a 779 bp bacteriorhodopsin gene. Chen and others (1994) employed purified 70 to 100-mers with ~20 bp overlap between neighboring oligos, along with short outer oligos (primers). In single reactions, they assembled gene fragments from either 4 or 12 oligos, with outer primers providing for PCR amplification. Barnes and Frawley (2003) used approximately picomole amounts of each 40-mer per 100 μ L reaction volume (i.e. ~1 nM to 10 nM concentration each oligo). Outer primers were present at 200 nM each. Using integrated PCA-PCR, Barnes and Frawley (2003) constructed a 731 bp gene and a 2187 bp gene. However, their recommended thermal cycler parameters were denaturation at 93°C for 1 minute and annealing/extension at ~55°C for 20 minutes over 20-25 cycles. Given these long annealing times, the Barnes and Frawley method yields runtimes of about 7-10 hours. Tian and others (2004) coined the term PAM (polymerase assembly multiplexing) for spiked outer primers combined with the oligo pool to perform

integrated PCA-PCR. Using PAM (2 μ L oligo mixture, 400 nM each primer) coupled with a highly discriminating oligonucleotide selection method, 21 ribosomal genes were synthesized, cloned, and sequenced. Selected error-free genes were combined using sequential PAM reactions to yield a 14.6 kb operon.

Gao and others (2004) presents a unique thermodynamically balanced inside-out (TBIO) synthesis scheme to conduct PCA with gradient oligo concentrations. Sense strand oligos are used only in the C-terminal half and anti-sense strand oligos are only used in the N-terminal half with the innermost oligos overlapping. The oligo concentrations were increased step-wise from 40 nM (innermost oligos) to 200 nM (outermost oligos). TBIO proceeds from the “inside-out” with each consecutive cycle growing strands outwards from the low concentration innermost oligos to the outermost oligos. This is in contrast to traditional PCA in which assembly occurs throughout the oligos in every cycle. However, this TBIO technique was limited to a maximum of six oligo pairs in a single PCR and thereby required repetitive PCR steps to synthesize longer genes.

A rapid integrated PCA-PCR technique has been developed using short annealing times (~10 seconds), a high-speed thermocycler, and fast KOD polymerase. Using the rapid technique, genes can be assembled and amplified in high yield in under 20 minutes. Wu and others (2006) explored key parameters in a conventional PCR thermocycler for their one-step simplified gene synthesis approach. Wu and others (2006) explored key parameters in a conventional PCR thermocycler for their one-step simplified gene synthesis approach (integrated PCA-PCR). Polymerase selection, oligo concentration,

and ratio of outer primers were tested for three different genes (209, 777, and 936 bp). Similarly, we have focused our efforts on exploring several key variables that are important to the development of a rapid technique—the number of thermal cycles, the concentration of oligonucleotides, and the concentration of outer primers. As noted by Wu and others (2006), KOD polymerase performs better than Taq or Pfu enzymes for integrated protocols. Also, the fast extension rate of KOD (Greip and others 2006) is well-suited to rapid thermocycling, and is thus used in our experiments. In instances where the rapid protocol failed, the Barnes and Frawley (2003) protocol was attempted using a conventional block cycler.

First, oligo concentration strongly influences integration of assembly and amplification. PCA usually employs high concentrations (25 nM to 100 nM) of each oligo for assembly, whilst followup PCR on an aliquot has a lower oligo background (~5 nM). Oligo concentrations must be high enough to enable efficient (and feasible time-scale) PCA. However, PCA oligos/shorter products are complementary to the full-length template and some compete for the primer pool during PCR amplification. So, one must at the same time provide a low oligo background concentration that will minimally interfere with amplification efficiency. Reconciling the oligo concentrations between the two steps is paramount to integrating PCA-PCR. Second, the concentration of primers (or outer oligos) affects the outcome of the integrated technique. Third, as the number of oligos present in the reaction mixture increases both the assembly and PCR amplification become more difficult. While there is no theoretical limit on the assembly or

amplification process, there is a practical limit on the number of oligos that can be used robustly for the technique.

To investigate the integrated PCA-PCR technique, three genes were selected as representative candidates: a 653 bp gene containing the endothelial protein C receptor (EPCR) sequence, a 929 bp gene sequence encoding the pUC19 β -lactamase gene, and a 1509 bp gene sequence encoding the thrombomodulin-1 (TM-1) gene. All three of these genes have been previously shown to work well under the traditional two-step approach (Mamedov and others 2007; TerMaat and others 2010).

3.2 Materials & Methods

Oligonucleotides spanning the templates coding the EPCR gene, pUC19 β -lactamase gene, and TM-1 gene were designed from cDNA sequences (GenBank accession nos. L35545, L09137, and BC035602, respectively). The Gene2Oligo computer program of Rouillard and others (2004) was used to design the assembly oligos. Oligos and primers were obtained from Integrated DNA Technologies (Coralville, IA) at 50 μ M in nuclease-free water.

Thermocycling protocols were performed in a rapid compressed air PCRJet[®] thermocycler (Quintanar and Nelson 2002) using glass capillaries. All reaction volumes were 25 μ l with the following reagents fixed: 200 μ M of each dNTP, 5 mM MgSO₄, 400 μ g/ml non-acetylated BSA, and 0.5 unit KOD hot-start polymerase from Novagen (Madison, WI) in 1x manufacturer's buffer. All thermocycling protocols had fixed

conditions of a 30 s hot start at 94°C, denaturation of 94°C for 2 s, extension of 72°C for 10 s, and a final extension at 72°C for 15 s. The number of cycles and annealing conditions were varied. *Traditional PCA*: 20 or 30 cycles with annealing conditions of 56°C for 10 s. Each reaction contained 0.1 µM of each oligonucleotide. *Followup PCR*: 20 or 35 cycles with annealing conditions of 56°C for 3 s, and each reaction contained 1 µl PCA product template and 0.7 µM of each primer. All reaction products were electrophoresed on 1% agarose gels stained with ethidium bromide using 6 µl of each product.

For the integrated PCA-PCR experiments, a variety of thermocycling conditions and reaction compositions were used. The standard conditions for the integrated PCA-PCR experiments were 40 cycles with annealing conditions of 56°C for 10 s. Each standard reaction contained 10 nM of each oligonucleotide and 0.7 µM of each primer. For oligo concentration experiments, 1 nM to 100 nM of each oligo were employed, with an additional 1 nM sample with 30 s annealing time. For primer concentration experiments, 0 µM to 1.5 µM each primer were employed. To investigate the number of integrated PCA-PCR cycles, 20 to 50 cycles were used. For integrated PCA-PCR experiments with the TM-1 oligo set, the number of oligos present in each reaction was varied from 46 to 76 oligos. The thermocycling conditions were 70 cycles with annealing at 62°C for 15 s. Each reaction contained 5 nM of each oligonucleotide along with an additional 700 nM each outer oligo.

The protocol described by Barnes and Frawley (2003) was attempted with the β -lactamase and TM-1 oligo sets. KlenTaq1 Polymerase and Rockstart buffer from DNA Polymerase Technologies (St. Louis, MO) were employed in the reaction mixture along with 1.3M betaine (Sigma Aldrich) and (20, 10, or 5) nM of each oligo. A conventional block thermocycler was used to execute the 20 minute annealing time protocol.

3.3 Results & Discussion

3.3.1 Number of Cycles

Figure 3-1 shows integrated PCA-PCR of EPCR (Figure3-1A) and β -lactamase (Figure 3-1B) using 20 to 50 cycles. The left lanes of each gel show the equimolar PCA and followup PCR, respectively, for the traditional two-step approach with 30 cycles of PCA and 35 cycles of PCR. The right lanes of Figure 1b show the equimolar PCA and follow-up PCR for β -lactamase with 20 cycles each. As expected, the amplification of the desired product increased up to a maximum along with the number of thermal cycles. With the use of a rapid thermocycler, the 40 cycle protocol was completed in 19.2 minutes, and was used for further reactions. However, one can see that a high yield of product can be obtained in 30 cycles (14.5 minutes).

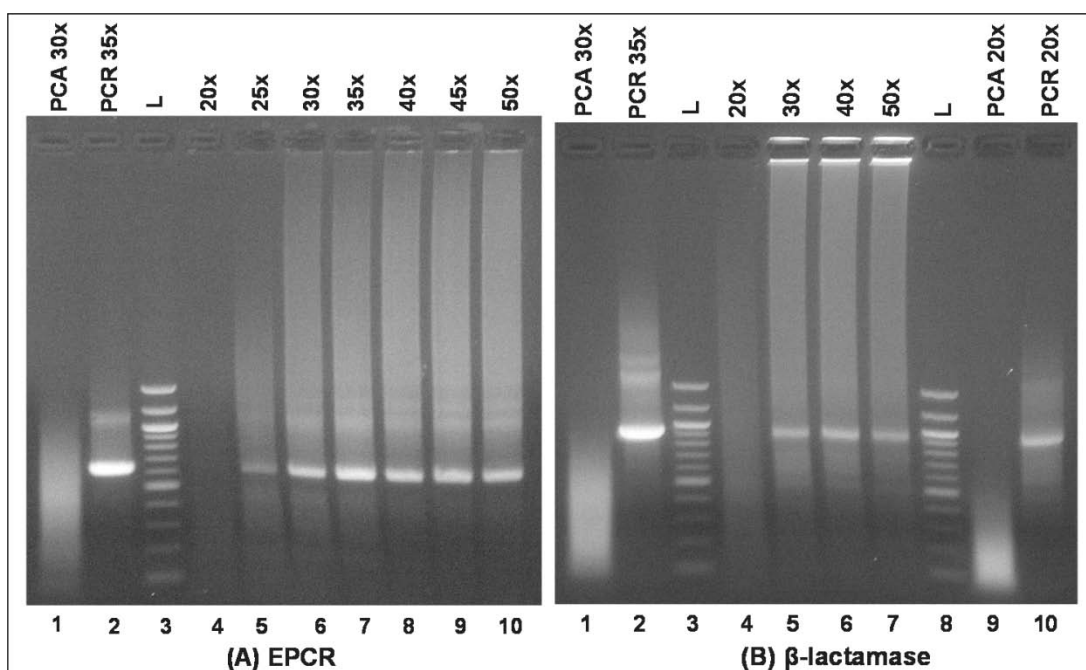


Figure 3-1 Traditional 2 step PCA/PCR and Integrated PCA-PCR vs. # of cycles.

(A) EPCR gene. lane 1: Equimolar PCA with 30 cycles; lane 2: Followup PCR with 35 cycles; lanes 4 thru 10: Integrated PCA-PCR vs. # of cycles. (B) β -lactamase gene. lane 1: Equimolar PCA with 30 cycles; lane 2: Followup PCR with 35 cycles; lanes 4 thru 7: Integrated PCA-PCR vs. # of cycles; lane 9: Equimolar PCA with 20 cycles; lane 10: Followup PCR with 20 cycles. Lanes (L): 100bp DNA ladder (NEbiolabs).

3.3.2 Effect of Oligo Concentration

The integrated technique using 1 nM to 100 nM each oligo was tested for both EPCR (Figure 3-2A) and β -lactamase (Figure 3-2B). The right lane of each gel represents a 1 nM of each oligo sample with the annealing time increased to 30 seconds. For both genes, 100 nM did not yield any distinguishable product from the background. For 25 nM reactions, EPCR showed a band at the correct molecular weight while β -lactamase

did not. Both genes had strong product amplification at 10 nM and 5 nM. The 1 nM EPCR sample with 30 second annealing had an extremely faint band, while 10 second annealing product did not. The reverse situation was observed for 1 nM β -lactamase samples.

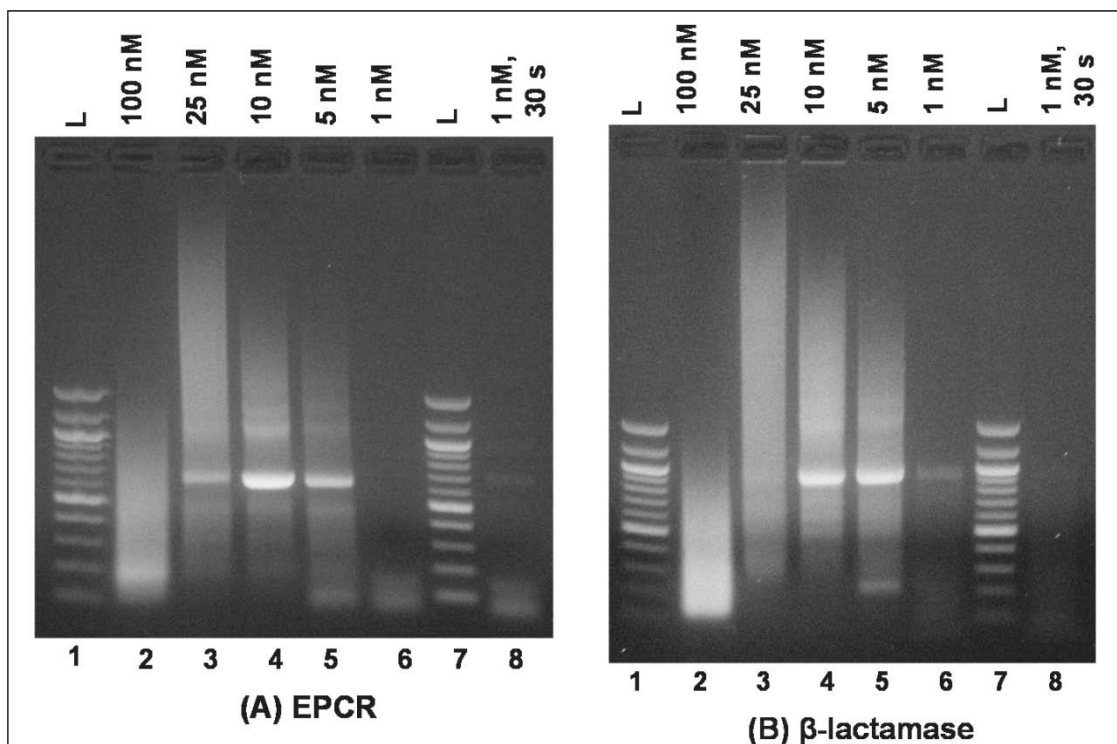


Figure 3-2 Effect of oligo concentration on Integrated PCA-PCR.

(A) EPCR gene. (B) β -lactamase gene. Lanes 2 thru 6: (100, 25, 10, 5, and 1) nM of each oligo, respectively. Lane 8: 1 nM of each oligo with 30 second annealing time.

Lanes (L): 100bp DNA ladder (NEbiolabs).

The concentration of oligonucleotides is paramount to simultaneous assembly (PCA) and amplification (PCR) in a single reaction. As can be seen in Figure 3-2, the oligo concentration of 100 nM typically employed in the traditional two-step PCA and followup PCR is unsuitable for use in the integrated approach. While 25 nM of each

oligo did yield product in the case of the 653 bp EPCR gene, it did not in the case of the longer 929 bp β -lactamase sequence. Only when the oligo concentrations were about 5 to 10 nM did the integrated technique show high yield of the desired product in a single step. This is in good agreement with the findings of Barnes and Frawley (2003) with their long annealing time integrated protocol. Interestingly, Wu and others (2006) was able to obtain the desired product at higher concentrations up to 100 nM, but only for short genes with few oligos. At 5 nM, Wu and others (2006) had only the 209 bp gene construct visible product. In exploring oligo concentrations, both the conventional and rapid techniques became less forgiving as the number of oligos present increased. 5 nM to 10 nM concentration of each oligo is recommended as close to optimal for rapid integrated PCA-PCR.

Further reduction of oligo concentrations to 1 nM resulted in little to no product. This indicates that the optimum of around 5 to 10 nM drops off sharply. In essence, proper assembly is intermittent at 1 nM or lower concentrations and is not advisable. There is no theoretical minimum on the necessary oligo concentrations for PCA per se, as it is conceivable that only one strand of each oligo would be needed. However, the oligos become so dilute within the reaction mixture that the probability of correct hybridization at each cycle becomes miniscule. This is exacerbated by the fact that, upon later cycles, these sparse intermediate products must hybridize. Thus, the propagation of assembly is contingent upon intermediate product concentrations that are considerably less than 1 nM. As intermediate assembly products are at extremely low (or non-existent)

concentrations, even the longest of annealing times may be insufficient to promote continued assembly.

3.3.3 Effect of Primer Concentration

The effect of primer concentrations (0 to 1.5 μM) is shown in Figure 3-3 (EPCR: Figure 3-3A, β -lactamase: Figure 3-3B). Wu and others (2006) specified a primer to assembly oligo ratio, but the impression is that absolute primer amounts may have more weight to integrated PCA-PCR techniques than a primer ratio. Note that primers can be replaced by simply spiking the outer assembly oligos (with consideration given to optimal annealing temperatures). The maximum amount of full-length product attainable clearly depends on the amounts of the outer primers. With no primers present, no distinct band at the expected size was visible for either gene while primer concentrations of 0.1 μM did produce weak amplification. 0.4 μM to 0.7 μM of each primer provided good amplification of the product. High primer concentrations of 1 μM or more also produced good results. Just as in regular PCR, high primer concentrations can cause increased probability of misannealing and thereby inhibit the desired amplification. Thus, concentrations of primers or outer oligos of about 0.4 to 0.7 μM each is recommended.

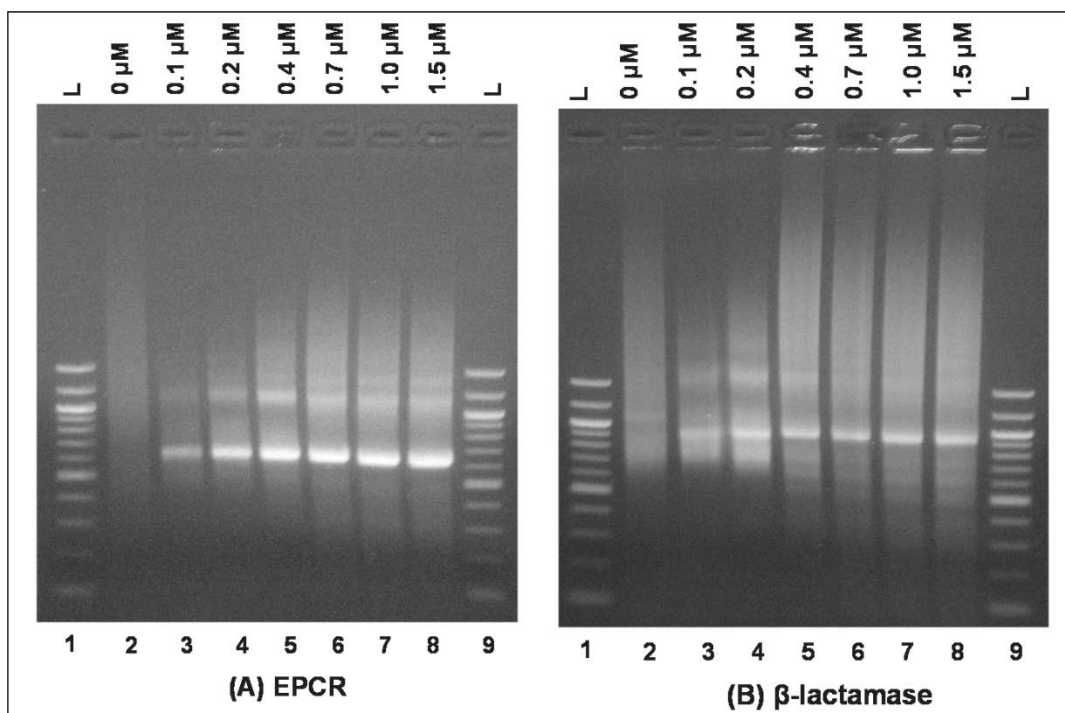


Figure 3-3 Effect of primer concentration on Integrated PCA-PCR.

(A) EPCR gene. (B) β -lactamase gene. Lanes 2 thru 6: (0 (none), 0.1, 0.2, 0.4, 0.7, 1.0, and 1.5) μ M each primer, respectively. Lanes (L): 100bp DNA ladder (NEbiolabs).

3.3.4 Limitations arising from number of oligonucleotides

When rapid integrated PCA-PCR was attempted on TM-1 (76 oligos, 1509 bp), no desired amplification product was visible on the gel (results not shown). Numerous adjustments to TM-1 oligo concentrations, primer concentrations, and increased number of cycles did not alter the outcome (results not shown). We hypothesized that integrated PCA-PCR may be practically limited to a maximum number of oligos. Thus, the number of oligos used within the reaction was decreased to investigate the point at which the integrated technique failed.

Figure 3-4A shows the resulting products using the TM-1 oligo set for 46 to 76 oligos. The 76 oligo reactions are the full-length 1509 bp gene, while fewer oligo reactions result in shorter products (partial gene sequence). All oligos were set at a concentration of 5 nM. An additional 700 nM each outer oligo employed in lieu of primers. The optimal annealing conditions were found to be 62°C for 15 s since longer oligos were used. The number of cycles was increased to 70 to maximize yield. As can be seen from the figure, the amount of amplified product decreases as the number of oligos increases and integrated PCA-PCR fails to produce a distinct product at about 60 to 66 oligos. The decreased band intensity for 56 oligos is due to a compromise in annealing temperatures (not optimal) for the oligo set. Figure 3-4B shows the traditional equimolar PCA and follow-up PCR products using the full TM-1 oligo set. The desired band at ~1509 bp is visible after completion of the two step approach.

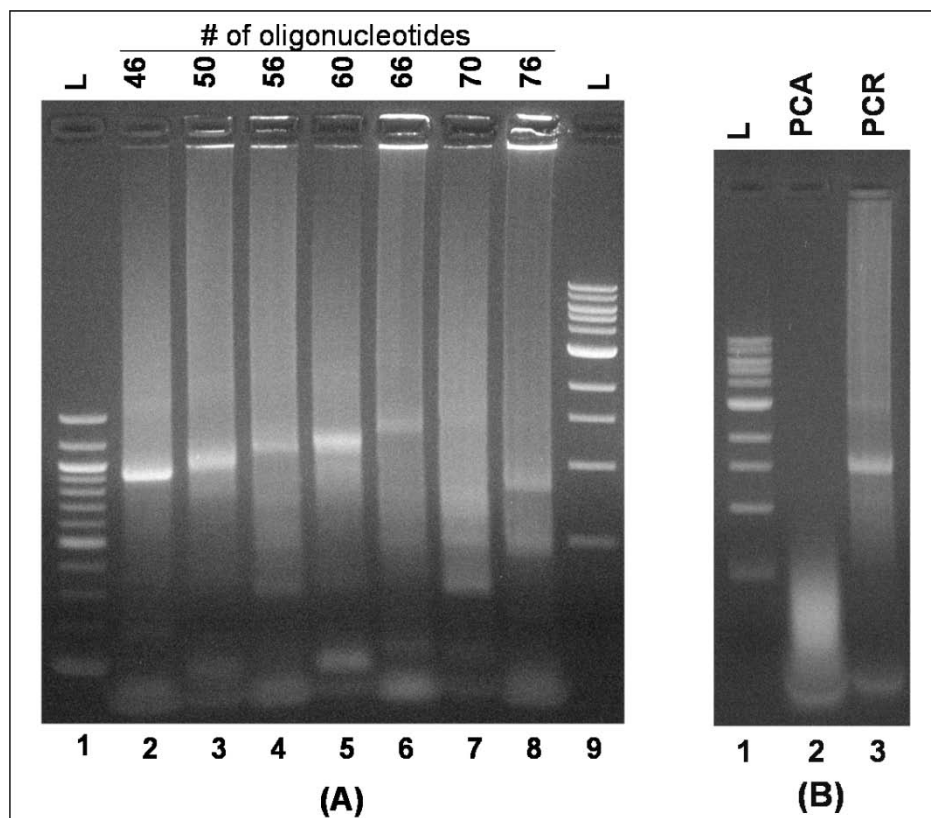


Figure 3-4 Traditional 2 step PCA/PCR and Integrated PCA-PCR vs. # of oligos using the TM-1 oligo set.

(A) Integrated PCA-PCR using different # of oligos from the TM-1 oligo set. Lanes 2 thru 8: 46, 50, 56, 60, 66, 70, and 76 total oligos, respectively. (B) Equimolar PCA (lane 2) with follow-up PCR (lane 3) of the full TM-1 gene 76 oligo set. Lanes (L): 100bp DNA ladder or 1kb DNA ladder (NEbiolabs).

The use of integrated PCA-PCR to assemble longer genes is possible given the findings of Barnes and Frawley (2003). To see if the TM-1 gene could be assembled with longer annealing times, we increased the annealing times from 10 seconds up to 14 minutes in the compressed air thermocycler. In no case was the integrated PCA-PCR technique successful for the TM-1 gene. Negative results (not shown) were also found for another

76 oligo set that was available in our lab. Follow-up PCR was performed and demonstrated that the oligos had assembled full-length templates, but never reached an efficient amplification phase. The exact integrated PCA-PCR protocol described by Barnes and Frawley (2003) was then attempted with the β -lactamase and TM-1 oligo sets using a conventional block thermocycler. In Figure 3-5, a discernable band was visible after 25 and 37 cycles for β -lactamase with 5 nM each oligo. However, none of our experiments with the TM-1 oligo set were successful even with the same reaction conditions. Followup PCR of the 25 cycle products revealed that β -lactamase was indeed fully assembled but TM-1 was not (results not shown). While we certainly cannot rule out the possibility of success as we did not use PAGE-purified oligos or KlenTaqLA polymerase mix, the TM-1 gene was not simultaneously assembled and amplified in any of our reactions.

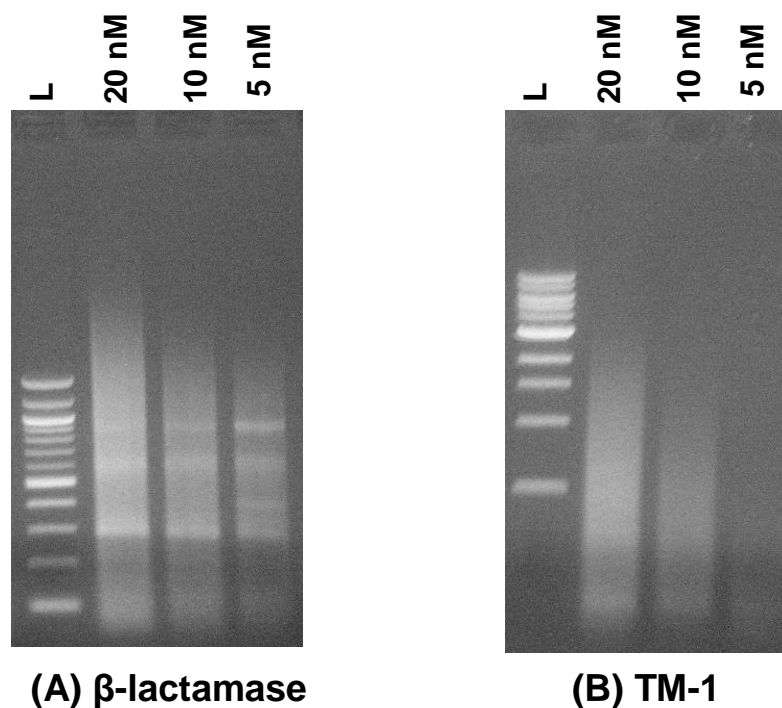


Figure 3-5. Integrated PCA-PCR products after 37 cycles using a modified Barnes and Frawley Protocol.

(A) β -lactamase using either 20, 10, or 5 nM each oligo. (B) TM-1 using either 20, 10, or 5 nM each oligo. L: 100bp or 1kb ladder (New England Biolabs)

Logically, both PCA and PCR amplification become more difficult as the number of oligos and gene length increase. Given the failure of the integrated technique on TM-1 but the success of the 2-step approach, it is fair to say that the 2-step approach is more robust. In general, one would expect such a result, as two separate steps facilitate the use of different oligo concentrations. In equimolar PCA, higher oligo concentrations (25 nM to 100 nM) allow for efficient assembly even for longer gene sequences with a large number of oligos. Follow-up PCR on an aliquot of the PCA product is performed with a low concentration of background oligos (~5 nM) due to the dilution of the aliquot. This dichotomy of oligo concentrations is not possible with the integrated approach.

To allow for efficient amplification, the background oligos should be present in minimal concentrations. All background oligos are complementary to the full-length PCR amplification templates. Thus, these oligos may inhibit polymerase extension by annealing to the template downstream of the primer. Experiments using Vent polymerase (New England Biolabs) with strand displacement activity did not result in successful amplification of TM-1 (results not shown). Additionally, inner oligos that have been extended to completion on their 3' end will compete for primers and be linearly amplified. As the number of oligos in the reaction mixture increase, so does the primer competition and upstream annealing. For example, 26 possible forward oligos are present for β -lactamase. In contrast, 38 forward oligos are present in the full TM-1 set. As an additional complication, the formation of DNA duplex pairs and gel analysis

inferences will depend on the relative oligo pool at the last thermal cycle. A distinct band may not be visible even though a significant amount of the full-length top strand and bottom strands may be present. It is expected that using 100-mers as described by Chen and others (1994) instead of 40-mers would likely allow for longer genes to be assembled and amplified with an integrated technique. However, PAGE purification of assembly oligos would be needed to limit unwanted deletion errors in the synthesized product.

3.4 Conclusion

Integration of PCA assembly and primer-mediated PCR amplification into a single reaction using rapid thermocycling has been demonstrated. Unlike the traditional 2-step synthesis approach, the concentration of initial oligos must be lowered with an optimum about 5 nM to 10 nM each oligo. Standard PCR reaction and primer conditions may be used. The rapid integrated PCA-PCR technique was successful for two genes, a 653bp EPCR gene (35 oligos) and a 929 bp β -lactamase gene (52 oligos). Annealing times of 10 seconds were found to be sufficient for the technique. Both genes were assembled and amplified in a single reaction, requiring less than 20 minutes total (for 40 cycles) of rapid thermocycling to obtain the full-length genes in high yield.

As the number of oligos present within the reaction increases, one must reconcile the dichotomy of optimal oligo concentrations. Lower concentrations favor PCR amplification, but hinder efficient oligo assembly due to extremely dilute intermediate products. Higher concentrations promote assembly but inhibit PCR amplification of constructed full-length templates by upstream annealing of complementary oligos and

competition for primers. For the 1509 bp TM-1 gene synthesis with 76 oligos, we were unable to obtain a distinct gel electrophoresis product. The long annealing protocol of Barnes and Frawley (2003) was equally poor for TM-1 but successful for β -lactamase synthesis. Using partial oligo sets for TM-1, it was determined that about 60 oligos is the maximum number that can be routinely used for rapid integrated PCA-PCR. This corresponds to about 1200 bp if ~40mers are used. While there is likely some dependence on the specific TM-1 gene sequence, some insights can be drawn from the number of oligo results. Above 60 oligos, results may be unsatisfactory or optimization too laborious and the two-step approach preferable.

3.5 References

- Barnes, W.M. and E.R. Frawley. Stream-lined Gene Assembly PCR. In: Dieffenbach, C.W., Dveksler, G.S. (Eds.), *PCR Primer: A laboratory manual*. Cold Spring Harbor Laboratory Press. New York. pp. 475-481. (2003)
- Chen, G.Q., I. Choi, B. Ramachandran, and J.E. Gouaux. Total gene synthesis: novel single-step and convergent strategies applied to the construction of a 779 base pair bacteriorhodopsin gene. *J. Am. Chem. Soc.* **116**: 8799-8800. (1994)
- Gao, X., P. Yo, A. Keith, T.J. Ragan, and T.K. Harris. Thermodynamically Balanced Inside-Out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**: e143-e164. (2004)

- Mamedov, T.G., N.V. Padhye, H. Viljoen, A. Subramanian. Rational *de novo* gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387. (2007)
- Pienaar, E., M. Theron, M. Nelson, and H.J. Viljoen, H.J., A quantitative model of error accumulation during PCR amplification. *Comput. Biol. Chem.* 30, 102-111. (2006)
- Quintanar, A., and R.M. Nelson. A process and apparatus for high-speed amplification of DNA. U.S. Patent No. 6,472,186. (2002)
- Rouillard, J.M., W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**: W176-W180 (2004).
- TerMaat, J.R., T.G. Mamedov, E. Pienaar, S.E. Whitney, and A. Subramanian. Automated two-column purification of Iminobiotin and BrdU labeled PCR products for rapid cloning: application to genes synthesized by Polymerase Chain Assembly. *J. Chromatogr. Sci.* **48**: 120-4. (2010)
- Tian, J., H. Gong, N. Sheng, X. Zhou, and E. Gulari, X. Gao, and G. Church. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature.* **432**: 1050-1054. (2004)
- Wu, G., J.B. Wolf, A.F. Ibrahim, S. Vadasz, M. Gunasinghe, and S.J. Freeland. 2006. Simplified gene synthesis: a one step approach to PCR-based gene construction. *J. Biotechnol.* **124**: 496-503. (2006)

3.6 Publication and Author Contributions

A substantial portion of this work was published under:

TerMaat, J.R., E. Pienaar, S.E. Whitney, T.G. Mamedov, and A. Subramanian. Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler. *J. Microbiol. Methods* **79**: 295-300. (2009)

Author Contributions:

J.R. TerMaat was primarily responsible for development of the rapid integrated PCA-PCR technique.

E. Pienaar conducted additional investigation into the rapid integrated PCA-PCR technique.

S.E. Whitney provided computer modeling that aided in the development of the technique.

T.G. Mamedov was primarily responsible for the design of the sequences for the genes employed in this work.

A. Subramanian led the scientific direction of the research.

4 Improvements to the Gene Synthesis Workstation

A next generation rapid PCA-based gene synthesis and affinity purification platform has been developed that generates improved yield and specificity from both sections of the workstation. While feasibility of combining rapid PCA synthesis and subsequent two-column affinity purification of synthesized products was conceptually demonstrated previously in Chapter 2, several deficiencies and opportunities were identified. In prior syntheses, the PCA oligo concentration and the amount of PCA product aliquot used for subsequent PCR were determined to require additional optimization as genes longer than 1 kb presented some difficulties in synthesis. In prior proof-of-concept purification, some non-specific binding of the BrdU system was observed, in addition to unattractive intermediate pH adjustment and generally low levels of isolated full-length construct. In this work, five different target-ligand systems (BrdU, desthiobiotin, digoxigenin, FITC, and TRITC) were explored to select two of these systems that convey improved levels of yield and specificity in a desirable process. Digoxigenin and FITC were selected among the five for use in a simpler yet improved flow-through immunoaffinity process that directly applies first column effluent to the second column. A 1.4 kb gene was synthesized and purified in the next generation workstation as a representative test sequence.

4.1 Introduction

Polymerase Chain Assembly (PCA) to synthesize genes from a pool of oligonucleotides was first described by Stemmer and others in 1995. In brief, short overlapping oligonucleotides are hybridized together, extended by a DNA polymerase, and then denatured in a PCR thermocycler—the process is repeated over many cycles to construct

the full-length double-stranded DNA of interest. The technique of PCA has been exploited by other researchers to construct a variety of sequences ranging from a few hundred bases up to many kilobases in length (Mehta and others 1997; Smith and others 2003; Gao and others 2004; Kodumal and others 2004). PCA was a primary technique to create 5-7 kb sequences utilized by Gibson and others (2008) to construct a 582,970 bp *Mycoplasma genitalium* genome using recombination techniques.

Mamedov and others (2007) applied the use of a rapid compressed air thermocycler (Quintanar and Nelson 2002, Whitney 2004) to facilitate extremely short runtimes with high fidelity. The rapid PCA technique was further expanded in Chapter 2 (also see TerMaat and others 2010) to include affinity purification of the synthesized genes. In Chapter 2, serial two-column purification was employed to selectively capture the full-length synthetic gene. BrdU and iminobiotin labels were incorporated into the gene via amplification primers, thereby facilitating selective capture of labeled DNA in the affinity columns. These two affinity systems have been separately employed for protein purification (Haider and others 1997; Urbach and others 1999; Hirsch and others 2002; Walters and Field 2006). However, the application of these systems in a serial purification arrangement for isolation of synthetic genes was not performed by others to our knowledge. In Chapter 2, the rapid synthesis and subsequent purification of genes was investigated using four genes ranging from 612 to 1548 bp.

While combined rapid PCA and affinity purification of synthetic genes was conceptually demonstrated in Chapter 2, the results presented therein show that the proof-of-concept

process suffered from several short-comings. The objective of this work was to develop a next generation synthesis and purification platform that addresses deficiencies and unknowns arising from the proof of concept research. The foundation of the rapid PCA synthesis technique was relatively well-developed, yet the optimal conditions of PCA oligo concentration and subsequent PCR template concentration were not fully understood. Prior work indicated that as the gene to be synthesized increased in length (typically greater than 1 kb), the synthesis became very sensitive to these two variables. However, optimal conditions for these parameters were unknown and the rationale behind the effects not clearly understood; thus further investigation was undertaken in this work.

For the purification process, the performance characteristics of the initial anti-BrdU and Streptavidin ligands, while demonstrating feasibility, were not considered optimal with respect to the yield and specificity. The BrdU purification did demonstrate some non-specific binding of unlabelled DNA, thereby resulting in some possible impurities being present in the isolated product. Due to the binding and elution conditions of the two columns in the proof-of-concept workstation, an intermediate adjustment step added some complexity to the system. Lastly, the yield of isolated product in the final effluent was also identified for improvement. For example, the yield of isolated product after affinity purification was below the gel detection limit for the TM-1 gene tested therein. Thus, a considerable portion of the development herein focused on selection and testing of different purification systems for use in the next generation platform. A total of five different affinity based systems were compared to select the two top candidates for

further testing. Additionally, biospecific elution conditions enabled testing of a modified purification process in which effluent from the first column is directly applied to the second column.

The scope of this research is summarized in the following key objectives:

- Elucidate general insight into the optimal PCA oligo concentration and PCR template aliquot conditions for rapid PCR to be implemented in the workstation.
- Identify five different affinity purification systems as candidates for implementation in the purification section of the workstation.
- Screen the label candidates for any potential inhibition during incorporation of the labels into amplicons during PCR.
- Conduct batch testing of label-ligand systems to compare yield and specificity.
- Determine appropriate elution conditions for the systems, either non-specific salt elution or biospecific elution using free labels in the elution buffer.
- Implement two selected candidates in an improved flow-through process with accompanying instrumentation.
- Validate the new system for the purification of a 1.4 kb gene sequence.

For this work, a 1.4 kb sequence was primarily employed as a model gene. This gene sequence was a chimeric gene encoding for both EPCR and TM-2 regions in a single

sequence joined by a myc linker sequence. By utilizing both inner and outer primers, a total of three different amplicons could be generated—a 640 bp chimera “A” fragment (primarily the EPCR section), a 753 bp chimera “B” fragment (primarily the TM-2 section), and a 1377 bp (herein 1.4 kb) full chimeric gene sequence. Please refer to Appendices 5B and 5C for further details on the chimera sequence, noting that all references in this chapter of “chimera” or its fragments refer to a modified chimera #2 gene that lacks the His-6 tag (see hybridization map in Appendix 5C.6 for the modified sequence). Note that the A and B fragments overlap each other by 16 bp, thus reconciling the gene lengths. The details of the chimera and its fragments employed in this work are given in Table 4-1.

Description	Length	# oligos	Forward Primer (5'-->3')	Reverse Primer (5'-->3')
Full-length chimera	1377 bp	77	ATCTCGAGAAAAGATTTGTAG	TAGCGCCGCTTACTTACC
Chimera A fragment	640 bp	35	ATCTCGAGAAAAGATTTGTAG	ACGGCGCTATTCAGATCC
Chimera B fragment	753 bp	42	ATCTGAATAGCGCCGTCG	TAGCGCCGCTTACTTACC

Table 4-1 Details of the chimera gene sequence and its fragments.

For investigation of PCA parameters, the chimera “A” fragment was tested along with the TM-1 sequence described in Chapter 2. In PCR inhibition tests, all three chimera amplicons were tested (A and B with only one labeled primer, full-length with two labeled primers to elucidate any detrimental interaction). Comparative testing of label systems by batch purification was performed using the chimera “A” fragment for consistency. Serial purification, both batch and flow-through, using two labels was performed with the full-length chimera as no inner labeled primers were procured.

4.1.1 Improvements to the 2-step PCA process

Two of the most critical parameters in the 2-step PCA-PCR process are the oligo concentration used in PCA, and the amount of PCA product used in subsequent PCR. In prior research (Mamedov and others 2007; TerMaat and others 2010), typically smaller genes were assembled with 100 nM each oligo with 1 μ l of this PCA product used as template for PCR. However, longer synthetic genes were assembled using ~25 nM each oligo with 2 μ l used as template (Mamedov and others 2007). While the assemblies may have approximately the same total oligo concentrations, it begs several questions: Is there a general rule of thumb for an optimal oligo concentration? What is the sensitivity of the 2-step process to oligo concentrations? Is there an optimal aliquot amount for follow-up PCR? Do the answers to these questions depend on the gene length (or # of oligos)? It was postulated that, as the gene length (and thereby # of oligos) increases, the process will become more sensitive to the point of failure, for similar reasons as found in integrated PCA-PCR. To provide some preliminary insight into these questions, a brief set of experiments was devised in which the PCA oligo concentration, and the subsequent PCA product aliquot for PCR, were varied for two genes: the 640 bp chimera A fragment and the 1548 bp TM-1 gene sequence from Chapter 2.

4.1.2 Investigation of target-ligand affinity systems

The concept of affinity purification as demonstrated in Chapter 2 is not specific to the BrdU and iminobiotin systems employed therein. Several deficiencies existed in which correction would be desirable, most notably non-specific binding found with the BrdU antibodies. Additionally, demonstration of the process with alternative purification

systems supports greater generalization of the workstation. The label-antibody systems that could potentially be employed in the system are seemingly limitless. Antibodies against all sorts of molecules are theoretically an option, as evidenced by the plethora of options commercially available. The type of antibody (polyclonal or monoclonal), the originating host (goat, sheep, chicken, mouse, rabbit, etc), as well as the specific supplier compound the issue. Lastly, other affinity ligand-target systems can also be considered for use in the workstation. However, utilization of antibodies is particularly advantageous due to their high specificity in general. As exhaustive analysis of all possible schemes was practically impossible, a highly focused effort was undertaken.

In total, five different candidates were selected for initial testing based upon a prior work and a survey of literature. The original BrdU antibody (goat polyclonal antibody, Immunology Consultants) used in Chapter 2 was retained for testing. For the streptavidin affinity separation, desthiobiotin was used as a replacement for iminobiotin with the HP Streptavidin resin (GE Healthcare, Piscataway, NJ). Desthiobiotin has shown to be eluted using competitive binding with biotin in the elution buffer, whereas biotin binding is essentially irreversible and iminobiotin purification typically employs pH adjustment (Hirsch and others 2002). Three other candidates were also selected: Digoxigenin (DIG), Fluorescein isothiocyanate (FITC), and tetramethyl isothiocyanate (TRITC). Digoxigenin is a steroid found in the *Digitalis* plants and is commonly employed as a hapten (Höltke and Kessler 1990). FITC and TRITC are commonly used fluorescent labels that may be attached to proteins (Moser and Hage 2006; Hermanson 2008). For Digoxigenin, FITC, and TRITC, a search of commercial suppliers for antibodies was

conducted; cost, purity, host, and amount were weighed in the selection. All of the procured antibodies used were immunoaffinity purified against the antigen, except for the anti-TRITC which was protein A purified. The details of the labels and employed ligands are given in Table 4-2.

Label	CAS #	Ligand Source	Ligand Supplier	Supplier Part #
BrdU	59-14-3	chicken polyclonal	Immunology Consultants	CBDU-65A-Z
Desthiobiotin	533-48-2	HP Streptavidin	GE Healthcare	17-5112-01
Digoxigenin	1672-46-4	goat polyclonal	Genetex	GTX73152
FITC	27072-45-3	goat polyclonal	Bethyl Labs	A150-112A
TRITC	95197-95-8	mouse monoclonal	Rockland Inc.	200-301-246

Table 4-2 Label and ligand details.

To evaluate the labels in an effort to select only two for testing in the next generation workstation, performance of labeled primers in the PCR process (inhibitory or not) was the first screening step. Incorporating labels into the oligonucleotide primers for PCR amplification of the full-length synthetic gene template is necessary to provide targets for selective purification of the full-length synthetic gene from the oligo pool resulting from the PCA process. PCR inhibition due to the labeling of primers was not expected because of the relatively small size of the molecules and the success already achieved with BrdU and iminobiotin in the proof-of-concept. However, overlooking this due diligence step could have had onerous consequences if indeed the selected labels were inhibitory to PCR. The five candidates were then subjected to a number of batch elution tests to provide comparative measures of 1) the efficacy of binding, and 2) elution characteristics under salt and free label in solution conditions. The total cost consideration was based on cost of the antibody, the cost of the primer labeling, and the

cost of the label for inclusion in elution buffers. Lastly, the buffer compatibility between selected label purification systems was also not overlooked since whatever two candidates are chosen must work synergistically in the serial purification.

4.1.3 Affinity purification process

Another aspect of the research was to refine the affinity purification section. After selection of digoxigenin and FITC as the two candidates and demonstration of product elution from competitive binding with free labels, a revised hardware schematic was developed for the serial immunoaffinity process. The biospecific elution by free labels facilitated a simplification of the workstation. The first column (anti-FITC) product effluent was directly applied to the second column (anti-digoxigenin) as it was not anticipated that free FITC label would interfere with the binding due to the specificity of antibody based separations. A schematic of the hardware is shown in Figure 4-1.

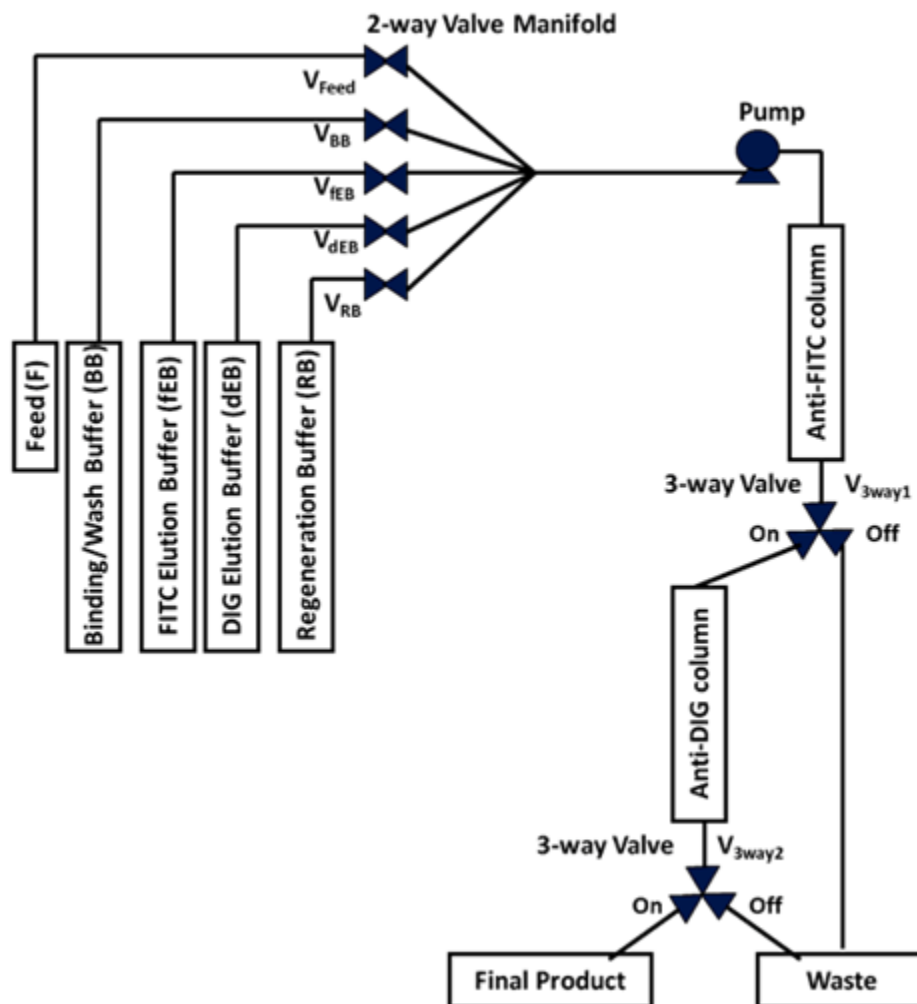


Figure 4-1 Schematic of the next generation affinity section.

4.2 Materials & Methods

4.2.1 Optimization of 2-step PCA-PCR

The TM-1 oligo set (77 oligos, 1548bp) from Chapter 2 and the chimera A oligo set (35 oligos, 640bp) were used as model systems. Oligo concentrations in the PCA step were either 12.5, 25, 50, or 100 nM each oligo. For each of the PCA products, the aliquot for PCR amplification was either 0.25, 0.5, 1, 2, or 4 μ l. Each 25 μ l PCA and PCR reaction contained the standard reagents of 200 μ M each dNTP, 5 mM MgSO_4 , 400 μ g/ml non-

acetylated BSA, and 0.5 U KOD polymerase in 1x manufacturer's reaction buffer. For PCR amplification, 0.7 μ M each primer was used. PCA thermocycling conditions in the PCRJet for both genes were: Hot Start of 94°C for 30 sec; 30 cycles of [94°C for 2 sec, 56°C for 10 sec, 72°C for 10 sec]; and final extension of 72°C for 15 sec. PCR conditions were identical except for 1) 40 cycles with 56°C for 3 sec annealing was used for TM-1, and 2) 30 cycles with 58°C for 3 sec annealing was used for chimera A. PCR products were analyzed by 1% gel electrophoresis stained with ethidium bromide using 10 μ l each product and 12 μ l reference ladder.

4.2.2 PCR testing of labels

The chimeric gene sequence of EPCR-myc-TM2 was used as the model sequence. One advantage of this chimeric gene sequence is that inner primers were previously shown to be robust to amplify shorter products. Thus, three total amplicons could be tested: 1) 640 bp chimera "A" (primarily the EPCR region) with labeled/regular outer forward primer and regular inner reverse primer, 2) 753 bp chimera "B" (primarily the TM-2 region) with labeled/regular outer reverse regular/label primer and regular inner forward primer, or 3) 1.4 kb full chimera with both labeled or regular outer primers to verify no interaction between labels existed. Outer primers labeled on the 5' end with the respective label or unmodified were obtained from IBA GmbH (Göttingen, Germany). The labeled primers from IBA GmbH were HPLC-purified post-labeling to ensure a high percentage (>95%) of the primers were in fact labeled. Both forward and reverse outer primers were labeled with each of the five candidates. Regular inner primers were procured from Integrated DNA Technologies (Coralville, IA).

First, equimolar PCA was performed to create the three templates for the PCR testing. Each 25 μ l PCA reaction contained 200 μ M each dNTP, 5 mM MgSO_4 , 400 μ g/ml non-acetylated BSA, and 0.5 U KOD polymerase in 1x manufacturer's reaction buffer. The chimera "A" and chimera "B" assembly utilized 100 nM each oligo, whilst the full chimera was at 25 nM. PCA thermocycling conditions in the PCRJet were: Hot Start of 94°C for 30 sec; 30 cycles of [94°C for 2 sec, 56°C for 10 sec, 72°C for 10 sec]; and final extension of 72°C for 25 sec.

As outlined above, the possible combinations of labeled and regular primers were explored for each of the three amplicons. For chimera "A" and chimera "B", 1 μ l of the PCA product was used as template for PCR, whilst 2 μ l was taken for the full chimera amplification. 0.7 μ M each primer was employed in each PCR reaction. Otherwise the other reaction components for PCR were the same as for PCA. PCR conditions in the PCRJet were: Hot Start of 94°C for 30 sec; 35 cycles of [94°C for 2 sec, 58°C for 3 sec, 72°C for 10 sec]; and final extension of 72°C for 25 sec. 1% gel electrophoresis stained with ethidium bromide was performed loading 10 μ l of each PCR product and 12 μ l molecular weight DNA ladders.

4.2.3 Affinity columns

The BrdU antibody and HP Streptavidin ligands from Chapter 2 were retained. The three new antibodies (for DIG, FITC, and TRITC) were coupled to N-Hydroxysuccinimidyl-Sepharose(R) 4 Fast Flow (Sigma Aldrich, 4% highly cross-linked beaded agarose, part # H8280) to a concentration of 0.5 mg per ml. Briefly, the sepharose resin-isopropanol

suspension was centrifuged, the isopropanol removed, and then washed with 10 resin volumes of ice-cold 1 mM HCl. Antibody was placed into 0.1M NaHCO₃, 0.5M NaCl, pH 8.0, and then added to the resin. Coupling was allowed to proceed overnight at 4°C on an end-to-end rotator. After centrifuging and removing the supernatant, any remaining coupling sites were blocked by washing twice in 5 resin volumes of 1.0M ethanolamine, pH 8.3, and then once in 5 resin volumes of 0.2M Tris-HCl, pH 8.5, with each step at 1 hour at room temp on an end to end rotator. The resin was then equilibrated in protein binding buffer (PBB) by washing five times in 5 resin volumes of 10mM Tris-HCl, 50mM NaCl, pH 7.4. The coupled resins were stored at 4°C in 0.05% sodium azide.

4.2.4 Batch purification testing

To investigate the binding and elution characteristics, the chimera “A” (EPCR) fragment was used as the initial test gene. 60 µl of labeled PCR product (with respective label) was mixed with 180 µl of PBB (10mM Tris-HCl, 50mM NaCl, pH 7.4; application buffer), and 40 µl of this mixture removed as a feed sample. 100 µl of each resin was placed in a microcentrifuge tube. After adding the feed to the resin, the 2:1 solution:resin was placed on an end-to-end rotator at room temperature. After binding, the tube was centrifuged and the supernatant removed (same procedure for wash and elution steps). The resin was washed with 200 µl PBB. Elution was attempted using 200 µl aliquots of elution buffer with a 4 hour incubation time. 30 µl of each fraction was analyzed by 1% agarose gel electrophoresis with 12 µl of 100bp DNA ladder (New England Biolabs) for reference. Due to the higher detection sensitivity, gels were stained with GelRed dye (Biotium, Hayward, CA, part # 41001) rather than the typical ethidium bromide.

Initial testing with high salt elution: Binding was allowed to proceed overnight. Three washes were performed. One elution fraction with 20 mM Tris-HCl, 500 mM NaCl, pH 7.4 using was performed with 4 hour elution incubation time.

Testing with free label in elution buffer: Elution buffers with free labels were prepared with the label at half-maximum solubility, 500 mM NaCl. 1% DMSO (Sigma Aldrich) was included to increase the solubility of the labels. All free labels were obtained from Sigma Aldrich (part numbers: BrdU- B5002; biotin - B4501; DIG – D9026; FITC - F3651; TRITC – 87918). Binding of Feed was reduced to 1 hour. 2 washes and 2 elution steps were employed with incubation times for elution at 15 minutes each step.

Regeneration testing- The exact resin aliquots used for free label elution testing of labeled PCR product were regenerated with 500 µl glycine, pH 2.2 for 10 minutes at room temperature. The exact same protocol used for free label elution testing was then carried out.

Non-specific binding testing: Fresh resins were utilized under the same protocol as the labeled PCR product with free label elution buffer. Regular PCR product devoid of any labels was used in the feed.

Serial testing with digoxigenin and FITC: The full chimera product labeled with digoxigenin and FITC (forward and reverse primers, respectively and arbitrarily) was

used for serial batch testing. 90 μ l of this product was mixed with 270 μ l PBB; the additional amount was used in an effort to overcome detection deficiencies especially after a serial purification. Binding to fresh anti-FITC resin (100 μ l) was allowed to proceed for 1 hour, then washed three times, then eluted three times with FITC free label elution buffer (15 minutes incubation each elution step). The first two FITC elution fractions were combined (after saving 30 μ l each as samples for gel analysis) and an equivalent amount of water added to adjust salt concentration. This intermediate feed was incubated with fresh anti-DIG resin (100 μ l) for 1 hour, then washed three times, then eluted three times with DIG free label buffer (15 minutes incubation each elution step).

4.2.5 Flow-thru affinity purification hardware

A new enclosure was constructed that was approximately 15" wide x 15" tall x 10" deep. The electronics (power supply, circuit board, and wiring harnesses) were located in the rear of the unit and isolated from any fluid paths by way of a middle divider. The schematic of pumps, valves, columns, fittings, tubing, and vessels according to the layout were mounted onto the divider wall. A picture of the workstation is shown in Figure 4-2. Please reference the schematic shown in Figure 4-1.



Figure 4-2 Picture of the new affinity purification system.

Many of the same components employed in the Chapter 2 workstation were retained in the next generation workstation. Please refer to Chapter 2 for the details of the pumps, valves, tubing, custom circuit board, power supply, and buffer vessels. Omnifit 3mm id x 2.5cm precision glass columns (Western Analytical Products, part # 006-CC-03-02-FF) were employed for both columns. A digital input/output terminal board (Keithley Instruments) and power entry module were mounted at the back wall to allow for the system electronics to be completely hidden. Molex mini-fit pins and wall-mount connectors mounted on the middle divider were employed within wiring harnesses to tidy up electrical wiring. A custom-built outer enclosure was constructed from 1/16" aluminum with hinged plexi-glass front doors, framed with 1/2" aluminum bar stock. The

middle divider was employed to separate electronic components from the chromatography process.

All buffers were placed proximal each other along with the feed. The fluids flow through a valve manifold consisting of 2-way valves that may be opened during each stage. A multi-port union consolidates the outputs into a single line. A pump is activated to pull the selected fluid(s) through valves that are open at each stage, and the fluid is pumped through the first column. Three way valves are present at the effluent end of each column to direct the flow to waste, the second column, or the final effluent. Prior to direct application of desired FITC effluent to the anti-digoxigenin column, it is necessary to divert feed loading and wash fractions that contain unwanted digoxigenin-labeled product to waste. Therefore, a three-way valve was required to be present in this configuration between the FITC and digoxigenin columns.

4.2.6 Flow-thru immunoaffinity purification

Initial testing of the flow-thru immunoaffinity workstation was done with the labeled full chimera sequence. Anti-FITC and anti-digoxigenin antibody resins were each packed into separate Omnifit glass columns with a bed volume of 170 μ l. This amount was determined to be sufficient, as it represents at least a ten-fold excess of antibody binding sites to labeled product. Estimating the molecular weight of the igG antibodies at 150 kDa and a concentration of 0.5 mg/ml, this yields approximately 567 picomoles. In the PCR reaction using 0.7 μ M of labeled primer and 100 μ l total product as feed, the maximum amount of label load to the first column is 70 picomoles. Each igG molecule

has the capacity to bind two antigens in the two Fab fragments. Note that the label binding burden of the second column is substantially less due to wash-thru of labels and inefficiencies of first column binding. However, both columns used the same bed volume for simplicity. Sufficient bed volume binding capacity is also justified empirically from scale-up of batch testing in which 100 μ l of resin was used against 60 μ l of PCR product.

The buffer vessels in the workstation were filled with the following buffers:

- Binding/Wash Buffer (BB): 10mM Tris-HCl, 50mM NaCl, pH 7.4
- FITC Elution Buffer (fEB): FITC label at max solubility in water with 1% DMSO, 100mM NaCl
- Digoxigenin Elution buffer (dEB): DIG label at max solubility in water with 1% DMSO, 250mM NaCl
- Regeneration Buffer (RB): 0.1M phosphate buffer, pH 2.5

Note the sixth vessel was not productively used yet filled with water for priming of the tube. Prior to installation of the columns, all lines were primed via the software interface. The respective buffer vessel fluid was primed up to the 7-port union. Binding buffer was primed into the feed line and all lines downstream of the union. After installation, the columns were equilibrated with approximately 2 ml of binding buffer. With the system initialized, the following protocol was used. Please refer to the hardware schematic shown in Figure 4-1 and the picture of the workstation in Figure 4-2. Effluent samples of approximately 200 μ l were collected.

1. Load Feed to FITC column. Mixed 100 μ l PCR product with 300 μ l BB in sample Feed microcentrifuge tube. 40 μ l was saved as a Feed aliquot for gel analysis. With V_{Feed} on, pumping set at 20 μ l /min (200ms on / 29800ms off per stroke). When ~20ul left, added 40 μ l BB to tube and continued pumping. The first 200 μ l was not collected. Collected samples #1-3 from V_{3way1} waste.
2. Wash FITC column. With V_{BB} on, pumping set at 0.3 ml /min (200ms on / 1800 ms off per stroke) for a total of 1.2 ml. Collected samples #4-9 from V_{3way1} waste.
3. Elute FITC column/Bind DIG column. With V_{fEB} on and V_{3way1} on, pumping set at 20 μ l /min (200ms on / 29800ms off per stroke) for a total of 1.2 ml. Collected samples #10-15 from V_{3way2} waste.
4. Wash DIG column. With V_{BB} on and V_{3way1} on, pumping set at 0.3 ml /min (200ms on / 1800 ms off per stroke) for a total of 1.2 ml. Collected samples #16-21 from V_{3way2} waste.
5. Elute DIG column. With V_{dEB} on, V_{3way1} on, and V_{3way2} on, pumping set at 40 μ l / min (200ms on / 14800 ms off per stroke) for a total of 1.6 ml. Collected samples #22-29 from V_{3way2} collect.
6. Flush DIG column. With V_{dEB} on and V_{3way1} on, pumping set at 0.3 ml / min (200ms on / 1800 ms off per stroke) for a total of 1.2 ml.
7. Regenerate columns. With V_{RB} on, pumping set at 0.3 ml / min (200ms on / 1800 ms off per stroke) for a total of 3ml. Repeated for 3ml with V_{3way1} on.

8. Equilibrate columns. With V_{BB} on, pumping set at 0.6 ml / min (200ms on / 800 ms off per stroke) for a total of 6 ml. Varied V_{3way1} and V_{3way2} to clear lines.

30 μ l of each sample was analyzed by 1% gel electrophoresis using GelRed stain with 12 μ l of 1 kb ladder.

4.3 Results & Discussion

4.3.1 Optimization of 2-step PCA-PCR

The results of PCA oligo concentration coupled with PCR aliquot are shown in Figure 4-3. For the short 640 bp gene, the reactions were insensitive to the variables—high yield of product was obtained for all samples. However, for the longer 1548 bp gene, the yield of product became very sensitive to the variables and their interaction. A desired product band was distinguishable for only a few of the combinations in this case. For the 1548 bp gene, [25 nM / 2 μ l] and [50 nM / 1 μ l] yielded the best products, while samples to the right (next level of PCR aliquot amount) of these optimal values gave strong bands with slightly reduced yields.

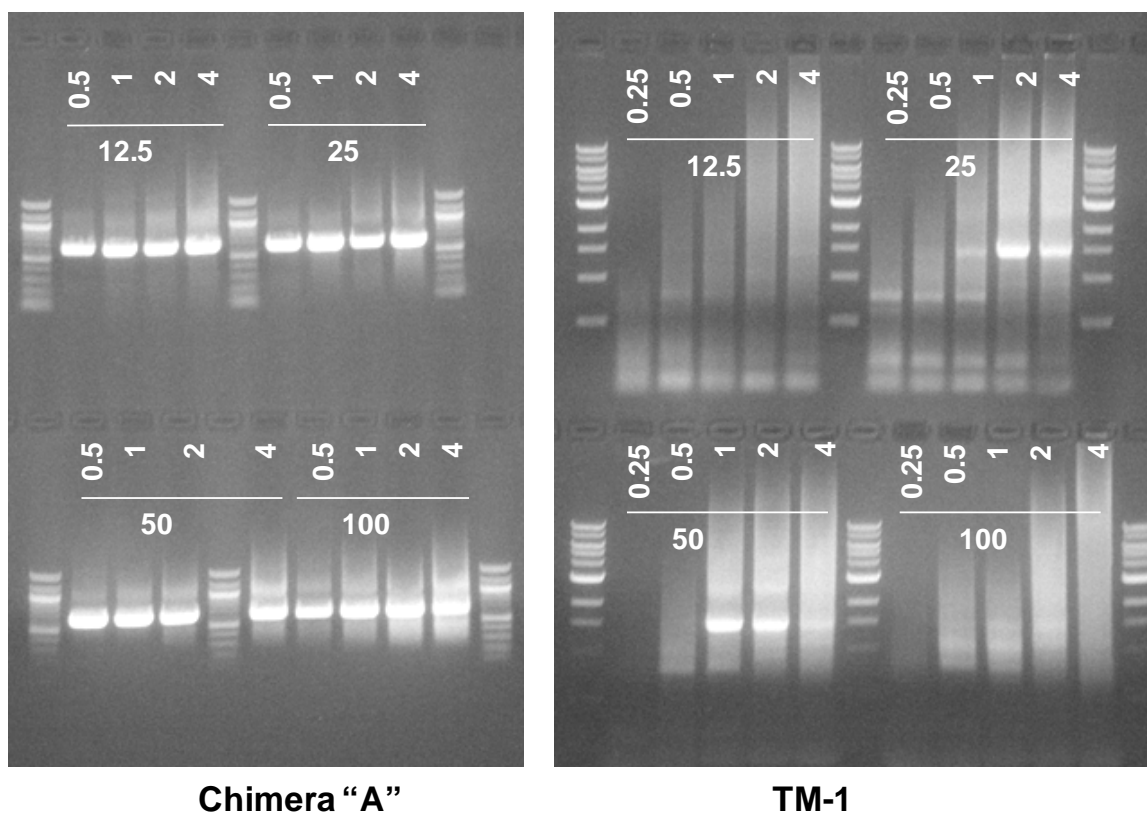


Figure 4-3 PCA oligo concentration – PCR aliquot amount reaction sets for the 640 bp chimera "A" and 1548 bp TM-1 genes.

PCA was performed at either 12.5, 25, 50 or 100 nM each oligo. Aliquots of 0.25, 0.5, 1, 2, or 4 μ l were used for PCR.

As expected, the window of acceptable oligo concentrations and PCR aliquots closes as the gene length increases. Of course, the specific genes employed in the experimental set highly influence the results. Thus, a quantitative analysis of gel electrophoresis yields does not provide intrinsic benefit from a design aspect. Note that normalization of the PCR aliquot to an effective oligo concentration may be more apt. In the TM-1 set for example, the most distinct product band for each PCA oligo concentration can be seen at an effective PCR concentration of 2 nM.

The significance of oligo concentration to integrated PCA-PCR as detailed in Chapter (also see TerMaat and others 2009) supports that these parameters are also of high importance to the 2-step process. In integrated PCA-PCR, one must balance the oligo concentration—it must be sufficient for assembly yet low enough so that exponential amplification is not greatly inhibited. In the 2-step process, one has considerable control over the oligo concentrations during each of the separate PCA and PCR reactions. As noted in Chapter 3, high PCA oligo concentrations are reduced by only taking an aliquot of the PCA product for follow-up PCR.

Synthesis of shorter genes that do not require a substantial number of oligos (35 in the case of chimera A) do not require a high initial concentration of each oligo for the reaction to progress through intermediate products to the full-length gene. Thus, shorter genes may be assembled over a wide range of assembly oligo concentrations. Since in these PCA reactions a considerable number of full-length templates are formed, the PCR aliquot amount is not critical to ensure a sufficient number of templates are present for amplification.

Under this rationale, one would expect preference for high oligo concentrations and a relatively large PCR aliquot for longer genes. However, other aspects of the reaction must be considered in the case of a large pool of oligos. For PCA, high oligo concentrations to promote assembly through intermediate products are counter-acted by inefficiencies arising from the high total oligo concentration. For example, the PCA

reaction may become “poisoned” by the sheer number of oligos present at 100 nM due to increased probability of misannealing events, or even polymerase sequestration (Kainz 2000). In the case of TM-1 at 100 nM, the total oligo concentration was 7.7 μ M. In contrast, chimera A at 100 nM was 3.5 μ M total oligo concentration. At the other spectrum end, low concentrations of assembly oligos are insufficient for the assembly to progress through intermediate products. As the number of oligos in a reaction increases, the number of intermediate actions that must occur increases along with more dilute levels of these intermediates. A second aspect that must be considered non-trivial with longer gene synthesis is the PCR aliquot amount. It is necessary to get a sufficient number of full-length templates into the PCR amplification reaction (theoretically one template). At high assembly oligo concentrations, amplification may be inhibited by too large of an aliquot due to carryover of undesired oligos that are complementary to the template. Too small of an aliquot in any case will not provide the template required for amplification. Taking both of these points into consideration, it is understandable why the assembly oligo concentration coupled with PCR aliquot amount must be balanced for longer gene synthesis.

From the experiments, some general rules of thumb can be gleaned. In prior research, shorter genes (< 1 kb) were assembled at 100 nM each oligo. While the facile nature of these short genes tolerates this concentration, it is in many ways unnecessarily excessive. Rather than rely on this forgiveness, it is generally recommended that lower oligo concentrations be employed even for short genes. Therefore, for short and long genes alike, the PCA concentration / PCR aliquot combinations of [25 nM / 2 μ l] or [50 nM / 1

μl] are recommended. These findings are relied upon for future work found in Chapter 5. Genes up to ~1.9 kb were successfully constructed under these conditions. Unsuccessful attempts at a 2.7 kb gene reaffirmed the difficulties that arise to such an extent that no distinguishable band can be attained regardless of the parameters employed.

4.3.2 Labeled primer testing

The resulting PCR products employing labeled primers are shown in Figure 4-4. As expected, the labeled primers did not affect PCR amplification in the three cases tested (chimera A and B with single labeled primer, full chimera with combination of two labeled primers). In all cases, amplicon yield from a labeled primer was essentially equivalent to PCR amplification with regular primers. Thus, further testing was pursued with all five candidates.

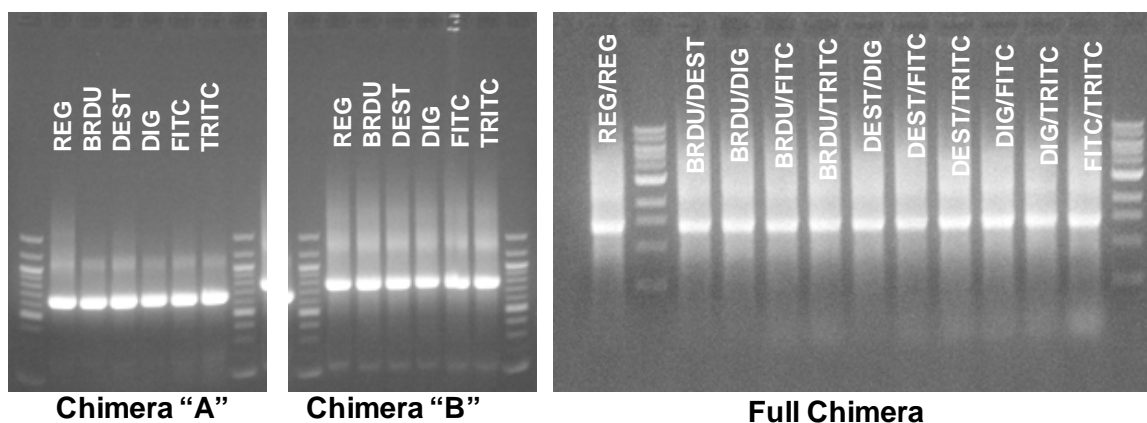


Figure 4-4 PCR inhibition testing using labeled and regular primers.

REG: regular, DEST: desthiobiotin, DIG: digoxigenin. Chimera A: 640 bp; Chimera B: 753 bp; Full Chimera: 1.4 kb.

4.3.3 Batch purification with high salt elution

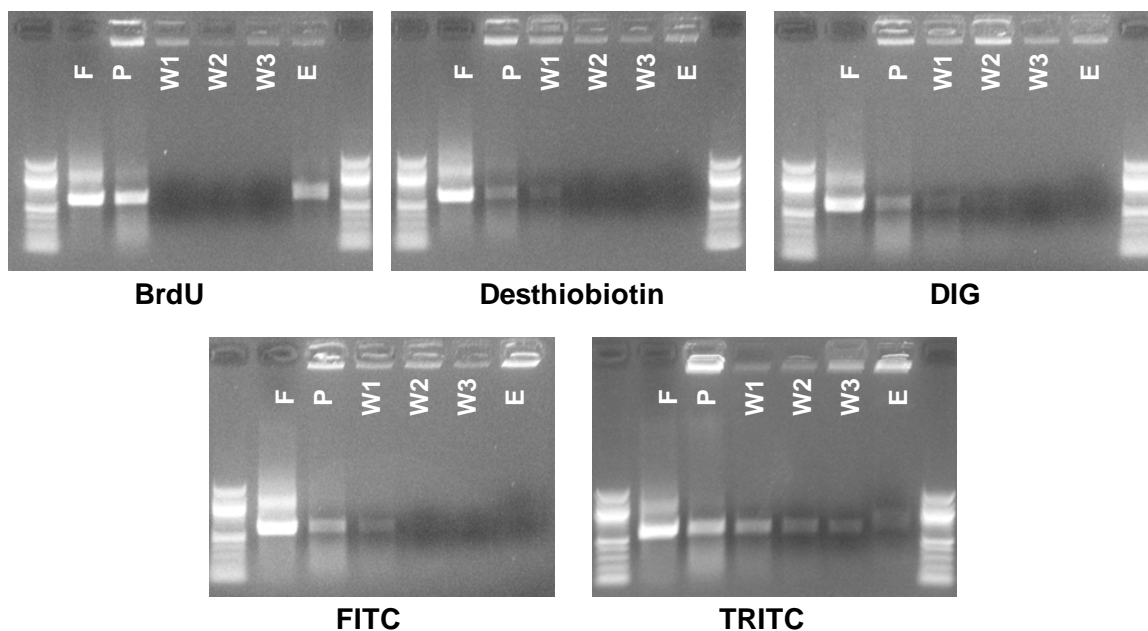


Figure 4-5 Initial batch testing of the five labels with high salt in the elution buffer with the 640 bp chimera A sequence.

F: Feed, P: post-binding supernatant, W: Wash step, E: Elution.

The results from the initial conservative protocol with long binding and elution incubation times reveal several key findings. In comparing the feed and post-bind supernatant samples, the amount of labeled product binding to each resin can be inferred. Of course, it also serves to verify that coupling of the ligands was successful. In this case, anti-streptavidin, anti-digoxigenin, and anti-FITC resins demonstrated fairly high quantities of binding. Anti-BrdU and anti-TRITC resins did not demonstrate such high levels. Although imperfection exists in removing supernatants, the products visible in the wash steps are indicative of bleed-off of the bound product and weaker binding. In this respect, anti-TRITC showed that product slowly was lost over wash steps. Lastly and most importantly for the objective, the presence and amount of product in the elution

fraction revealed the success of the tested 500 mM NaCl elution buffer. The higher salt was successful for BrdU elution and TRITC elution (the latter muddled due to product bleed-off). However, the higher salt elution failed to elute streptavidin, digoxigenin, and FITC labeled-products, implying that the binding affinity of these ligand-target systems was relatively strong by comparison.

In the gels, one will notice that the product band is shifted up (appears to be longer length) in the elution fraction in comparison to the other samples and reference ladder. This mobility shift is a consequence of the higher salt concentration employed in the elution fraction. This phenomena is also observed in other samples in this work in which the salt concentration was altered.

4.3.4 Batch purification with free label elution buffer

The anticipated direction from the beginning of the candidate selection was the use of free labels in the elution buffer for biospecific elution conditions. This reverse-role competitive binding mechanism employed was believed to be a preferred method, especially as it may facilitate a simplified workstation layout via direct application of first column effluent to the second column. The results of the testing are shown in Figure 4-6.

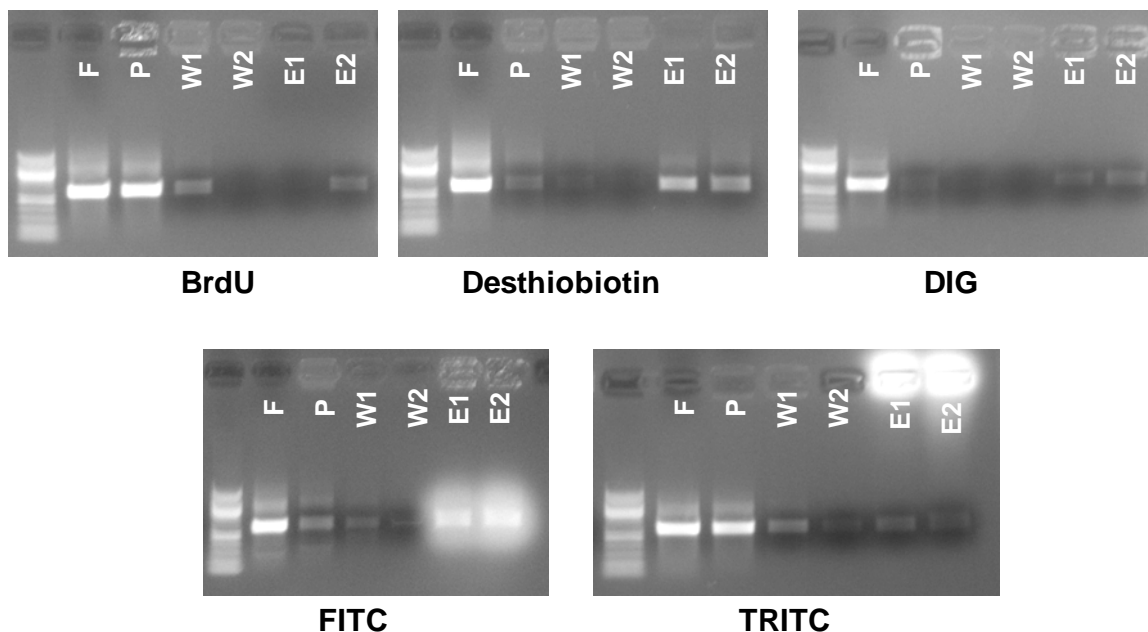


Figure 4-6 Batch testing of the five labels with label in the elution buffer for the 640 bp chimera A sequence.

F: Feed, P: post-binding supernatant, W: Wash step, E: Elution step.

The results demonstrate that the label elution buffers have succeeded in displacing the bound labeled DNA. For all 5 systems, product bands were visible in the elution fractions upon application of the elution buffer. In most of the cases, product eluted over both fractions indicating that the accumulation of free label competition gradually displaces product. In looking at the post-bind supernatant samples in comparison to the feed, once again desthiobiotin, DIG, and FITC bound significantly with the reduction to 1 hour incubation. Additional elution steps would have likely released more product in these three cases as the mass balance of all the samples indicate that some product may still be bound. This is especially true for DIG in which elution was gradual over the two elution fractions. Fluorescent background signals are present in FITC and TRITC

elutions; the FITC background may overlap bands whilst the TRITC label does not migrate in the gel.

4.3.5 Column Regeneration

The objective was to regenerate the column for repeated use by releasing any bound free label. Carry-over labeled DNA was highly unlikely as the resins remained incubated with some elution buffer in between runs. The identical batch purification procedure was then performed to determine the efficacy of the glycine regeneration. The purification results after glycine regeneration are shown in Figure 4-7.

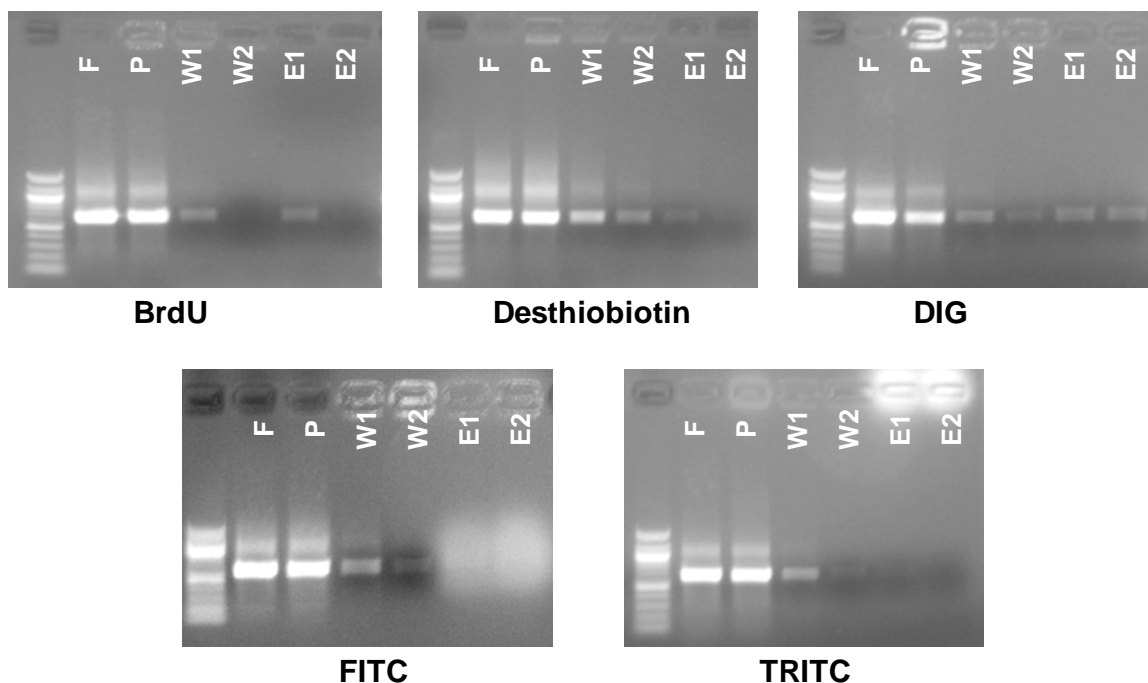


Figure 4-7 Batch testing of the columns after glycine regeneration using the 640 bp chimera A sequence.

F: Feed, P: post-binding supernatant, W: Wash step, E: Elution step.

As can be seen from the gels and in direct comparison to fresh resin testing done in 4.3.4, binding of the product was greatly reduced in all cases. BrdU, digoxigenin, and FITC exhibited very slight binding activity as evidenced by an increase in product upon elution buffer application. Elution bands for FITC were masked by the free label fluorescence. The performance reduction was either due to the glycine denaturing the antibodies and rendering them inactive, or the glycine failed to remove the bound free labels from the binding sites. As glycine conveys relatively harsh regenerations conditions, an alternative regeneration buffer of pH 2.5, 0.1 M phosphate buffer was used for further column regeneration. This buffer has been shown to release bound fluorescein from anti-FITC antibodies (Clarke and Hage 2001). The low pH used typically results in a conformational change of the ligand, thereby facilitating release of the target.

4.3.6 Non-specific binding testing

In general, it can be assumed that antibody-based binding systems are highly specific for the target. However, this cannot necessarily be assumed to be unquestionably true, especially in cases where polyclonal antibodies are employed. Thus, nonspecific binding using regular (unlabelled) PCR product with fresh resins is shown in Figure 4-8. The anti-BrdU resin was the only antibody to exhibit visible non-specific binding. This is supportive of non-specific binding observed in the initial proof-of-concept testing.

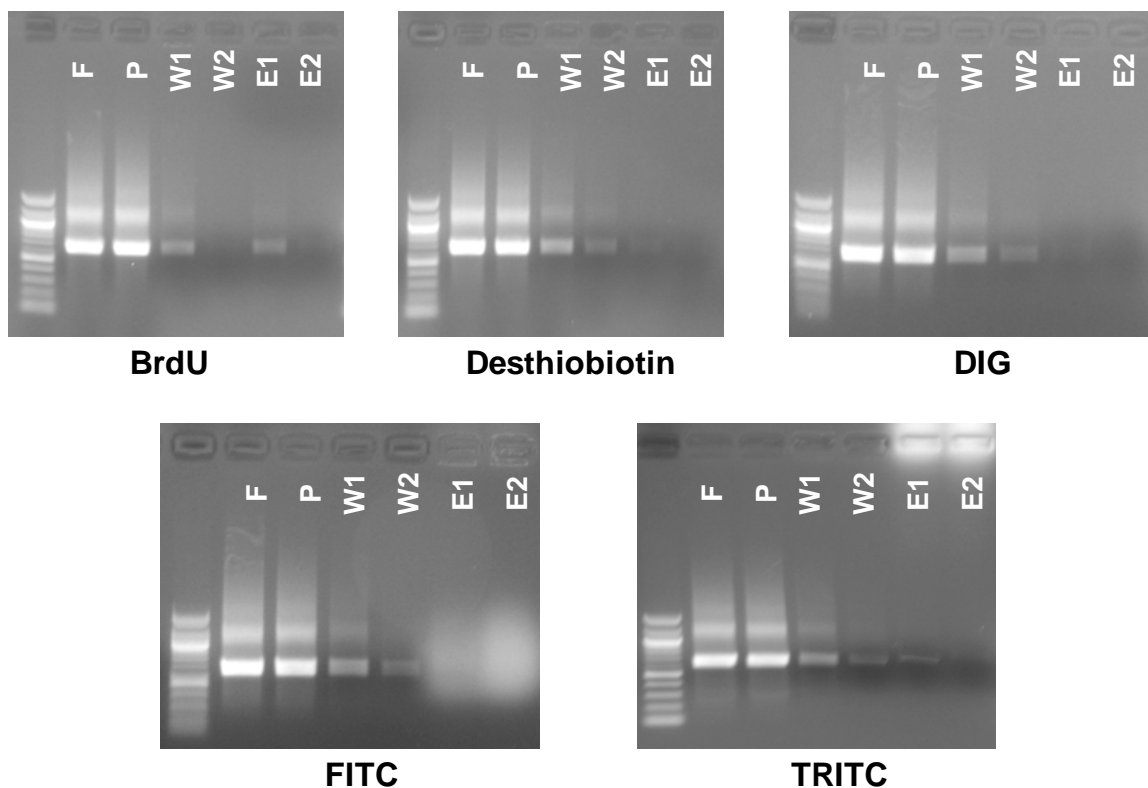


Figure 4-8 Batch testing of the five columns with unlabelled 640 bp chimera A product.

F: Feed, P: post-binding supernatant, W: Wash step, E: Elution step.

4.3.7 Selection & Serial batch testing

After weighing the batch results, anti-digoxigenin and anti-FITC were chosen as the candidates for further exploration. Anti-BrdU was easily eliminated due to its non-specific binding and low binding yield of labeled product. Anti-TRITC was less attractive due to the weak binding characteristics coupled with high costs of antibody and free label for the elution buffer, although optimized isocratic conditions may be feasible. While anti-FITC had very low regeneration with glycine, the proven regeneration buffer (Clarke and Hage 2001) made it, along with anti-DIG, the most promising candidates. The gradual elution of DIG-labeled products under free label elution buffer may be

improved with optimization. Desthiobiotin performed rather well overall and would have been a justifiable choice. However, regeneration was expected to be problematic since the biotin-streptavidin exhibits very tight binding affinity. The use of StrepTactin Sepharose with HABA (2-(4'-hydroxyphenylazo)benzoic acid) elution or regeneration may be a useful technique (Müller and Schmidt 2000) if desthiobiotin labeling is considered in future work.

In this light, the use of iminobiotin may be considered also an attractive option for its pH-based binding/elution protocol and proven regeneration. The iminobiotin pH adjustment may be particularly advantageous as the second stage in a serial purification arrangement with an antibody-based system. There are two considerations for this scenario to be feasible: 1) the labeled product will successfully bind in the first stage under streptavidin elution conditions so that all products not captured by the first column do not bind to the streptavidin (i.e. first column binding buffer at pH 4.0), and 2) the elution buffer from the first column is at the proper binding pH for the streptavidin (i.e. pH 10). If reasonable validity can be achieved regarding these two aspects, then it would be possible to have both columns directly in series, or for that matter even a single column packed with two distinct matrix layers of the respective ligands.

Digressing aside, the initial two labels selected for further testing in the workstation were the digoxigenin and FITC. The next step was to conduct serial purification with the two labels in batch processing. As an arbitrary choice, the forward primer was labeled with digoxigenin, while the reverse primer was labeled with FITC. Initially, digoxigenin was

used as the first purification step, followed by FITC purification. While a sufficient amount of purified product was visible in the final effluent, the yield was unimpressive. A short ~100bp undesired PCR product complicated matters and optimization strategies for PCA and PCR did little to improve the result. As a consequence, FITC purification was performed first followed by digoxigenin. This strategy gave slightly better yield in the final effluent, and was used in further testing.

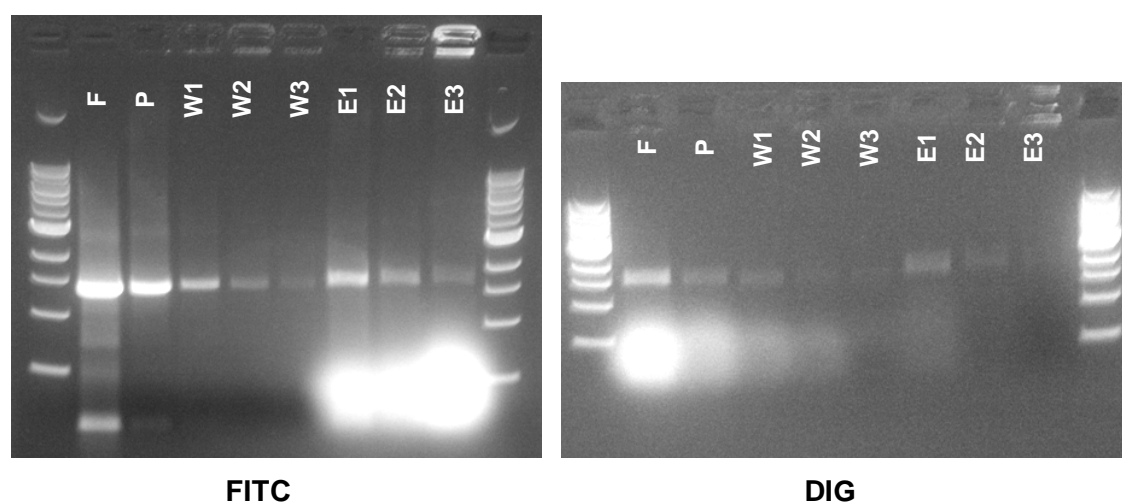


Figure 4-9 Serial batch purification of the labeled 1.4 kb full chimera gene using anti-FITC, followed by anti-DIG.

The first two elution fractions of anti-FITC were combined as feed to the anti-DIG resin.

F: Feed, P: post-binding supernatant, W: Wash step, E: Elution step.

The serial purification demonstrated that the process is successful. Note that the non-specific ~100bp band did compete with the desired product, and bound in greater relative amounts to the FITC column. Strong elution bands were found in the FITC elution products (the lower gel fluorescence is due to the free FITC label in the elution buffer).

For the DIG purification, the feed bound in moderate amount, and clearly yielded the desired band in the final elution fractions. The upward shift in the elution samples is due to the higher salt concentration. On a final note, the elution buffers used herein were at 500 mM NaCl, thus the reason why an intermediate adjustment was done. Other testing (results not shown) has demonstrated that lower salt concentrations are adequate. Thus, 100 mM NaCl in the FITC elution buffer and 250 mM in the DIG elution buffer were employed in further experiments.

4.3.8 Flow-thru purification

A key refinement to the workstation from the initial proof-of-concept was a simplified process. In the initial BrdU and streptavidin arrangement, an intermediate pH adjustment step (and vessel) was utilized in order to adjust BrdU elution product to the proper binding pH of streptavidin. This intermediate step was a conservative approach rather than attempting a high pH in the BrdU elution buffer. With two label-antibody based columns in which free labels are used in the elution buffers, it was deemed unnecessary to have an intermediate adjustment vessel. The rationale is that the presence of free FITC label should have no impact on digoxigenin binding due to the biospecific elution conditions. Thus, it was deemed possible to directly load the anti-FITC elution effluent to the anti-digoxigenin resin.

That said, the previous batch serial testing included a higher salt concentration for anti-FITC elution than the anti-digoxigenin application buffer. While the higher salt would promote sharper elution, the free label carries the primary elution burden. For the FITC

label elution buffer, a salt concentration of 100mM NaCl was determined to strike a balance between elution behavior and sufficient binding to the digoxigenin antibody.

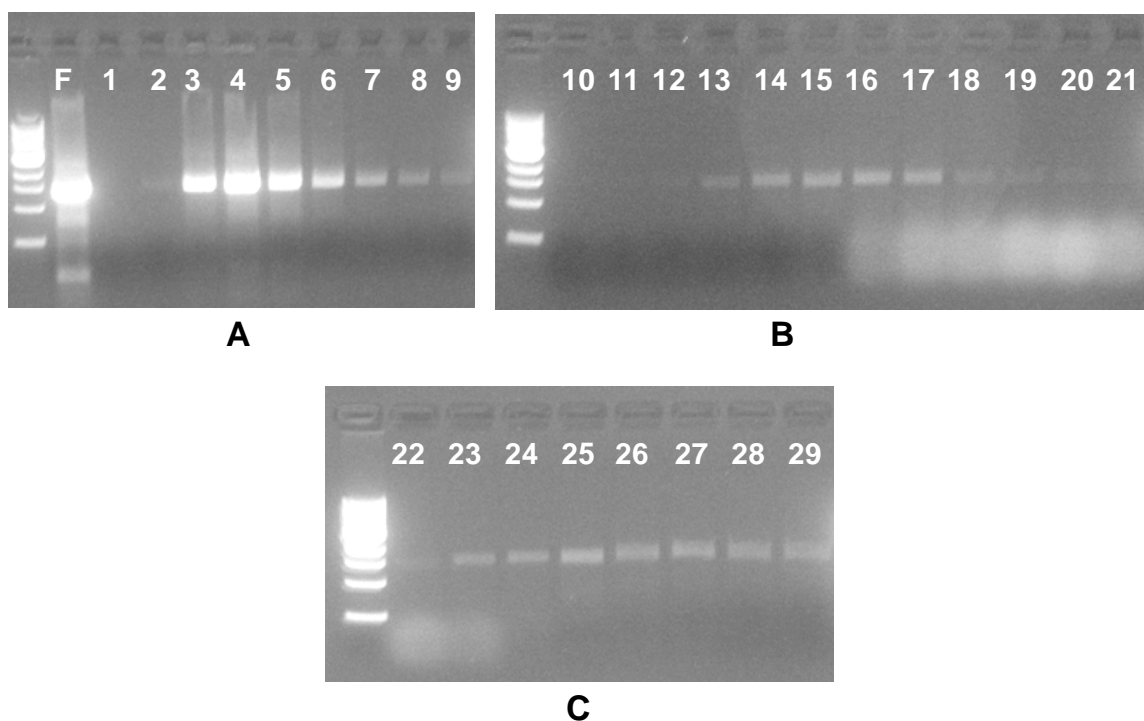


Figure 4-10 Initial purification of labeled 1.4 kb full chimera using the next generation workstation.

A: samples # 1-9. B: samples # 10-21. C: samples # 22-29. Please refer to the protocol for details of the collected samples.

The result shown in Figure 4-10 demonstrate that the initial flow-through immunoaffinity purification has performed well. One can see that binding to the FITC antibody was inefficient, but as expected given the batch testing results with chimera. Upon elution from the anti-FITC column, a significant amount of the full-length product also did not bind to the anti-digoxigenin antibody, but confirms that FITC binding was successful. It can be expected that improved yields from both columns is possible with further process

optimization. However, a sufficient amount of purified product was obtained in the final effluent to be visible by gel electrophoresis. Please refer to Chapter 5 for continued optimization to improve: 1) binding yield of the columns, 2) reduction of bleed-off during column washing, and 3) regeneration of antibody column activity.

4.4 Conclusions

Extending from the proof-of-concept described in Chapter 2, an improved process and accompanying workstation for the rapid synthesis and purification of *de novo* gene sequences have been developed. Within the upfront rapid PCA synthesis section, the PCA oligo concentrations and the PCR aliquots were explored. While assembly of shorter genes was found to be relatively insensitive to these parameters, longer genes (i.e. above 1 kb) are highly susceptible due to increasing complexity stemming from the oligo pool. As logically expected, the interaction of these two parameters played an important role in successful synthesis. As a general guideline, PCA oligo concentrations of 25 to 50 nM each and respective PCR aliquots of 1 to 2 μ l appear to provide the best conditions for routine construction of gene sequences.

An extensive comparison of five different label-ligand systems (BrdU, desthiobiotin, digoxigenin, FITC, and TRITC) was performed in an effort to improve the purification of synthesized genes. All five labels were amenable to incorporation into synthetic DNA via amplification primers, as none of the labels effected any reduction in PCR yield compared to unlabelled primers. Batch purification studies employing a 640 bp gene sequence were performed to compare the binding and elution behavior of labeled DNA to

procured ligands coupled to NHS-sepharose. Of these five candidates, digoxigenin and FITC were selected for further testing in a serial immunopurification arrangement.

Biospecific elution conditions employing competition from free labels enabled a simpler yet improved serial purification process in which the product effluent from the first column could be directly applied to the second column. The serial immunoaffinity purification was first demonstrated in batch mode using a 1.4 kb chimeric gene sequence. The process was then implemented in a new flow-through chromatography platform. The 1.4 kb gene sequence was successfully isolated in the workstation under automated software control of pumps and valves as validation of the method.

In summary, the refined process and workstation disclosed herein with different labels than the proof-of-concept clearly demonstrates the generalization of the technique for the rapid synthesis and purification of gene sequences. Of course, the selected digoxigenin and FITC labels used in the new immunoaffinity platform are by no means purported to be paramount to all other affinity systems. Rather, they represent only improved alternatives to BrdU and iminobiotin. Given the demonstration in this Chapter and Chapter 2, it should be clear that the workstation may be modified to accommodate other affinity systems that one may desire to employ for purification of synthetic DNA.

4.5 References

- Clarke, W., and D.S. Hage. Development of sandwich HPLC microcolumns for analyte adsorption on the millisecond time scale. *Anal. Chem.* **73**: 1366-1373. (2001)
- Gao, X., P. Yo, A. Keith, T. Ragan, and T. Harris. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**: e143-e164 (2004)
- Gibson, D.G., et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220 (2008)
- Haider, S.R., G. Juan, F. Traganos, and Z. Darzynkiewicz. Immunoseparation and immunodetection of nucleic acids labeled with halogenated nucleotides. *Exp. Cell Res.* **234**: 498-506 (1997).
- Hermanson, G.T. Bioconjugate Techniques. Elsevier. Amsterdam, The Netherlands (2008)
- Hirsch, J.D., L. Eslamizar, B.J. Filanoski, N. Malekzadeh, R.P. Haugland, J.M. Beechem, and R.P. Haugland. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**: 343-357 (2002)
- Höltke, H.J. and C. Kessler. Non-radioactive labeling of RNA transcripts in vitro with the hapten digoxigenin (DIG); hybridization and ELISA based detection. *Nucleic Acids Res.* **18**: 5843-5851. (1990)
- Kainz P. The PCR plateau phase—towards an understanding of its limitations. *Biochim. Biophys. Acta.* **1494**: 23-27. (2000)

- Kodumal, S.J., K.G. Patel, R. Reid, H.G. Menzella, M. Welch, and D.V. Santi. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 5573-5578 (2004)
- Mamedov, T.G., N.V. Padhye, H. Viljoen, and A. Subramanian. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387 (2007)
- Mehta, D.V., R.J. DiGate, D.L. Banville, and R.D. Guiles. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. *Prot. Expr. Purif.* **11**: 86-94 (1997)
- Moser, A.C. and D.S. Hage. Chromatographic Immunoassays. In: Hage. D.S. Handbook of Affinity Chromatography. CRC Press. Boca Raton, FL. (2006)
- Müller, H. N. and T.G.M. Schmidt. Chromatography using Strep-tag affinity peptides. In: Kastner, M. (Ed), Protein Liquid Chromatography. Elsevier Science B.V. Amsterdam, The Netherlands. (2000)
- Quintanar, A. and R.M. Nelson, 2002. A process and apparatus for high-speed amplification of DNA. U.S. Patent No. 6,472,186. (2002)
- Smith, H.O., C.A. Hutchison, C. Pfannkock, and J.C. Venter. Generating a synthetic genome by whole genome assembly: Φ X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 15440-15445 (2003)
- Stemmer, W.P.C., A. Cramer, K.D. Ha, T.M. Brennan, and H.L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene.* **164**: 49-53 (1995)

- TerMaat, J.R., E. Pienaar, S.E. Whitney, T.G. Mamedov, and A. Subramanian. Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler. *J. Microbiol. Methods* **79**: 295-300. (2009)
- TerMaat, J.R., T.G. Mamedov, E. Pienaar, S.E. Whitney, and A. Subramanian. Automated two-column purification of Iminobiotin and BrdU labeled PCR products for rapid cloning: application to genes synthesized by Polymerase Chain Assembly. *J. Chromatogr. Sci.* **48**: 120-4. (2010)
- Urbach, E., K.L. Vergin, and S.J. Giovannoni. Immunochemical detection and isolation of DNA from metabolically active bacteria. *Appl. Environ. Microbiol.* **65**: 1207–1213 (1999).
- Walters, S.P. and K.G. Field. Persistence and growth of fecal bacteroidales assessed by bromodeoxyuridine immunocapture. *Appl. Environ. Microbiol.* **72**: 4532-4549 (2006).
- Whitney, S.E. Analysis of rapid thermocycling for the polymerase chain reaction. *ETD collection for University of Nebraska - Lincoln*. Paper AAI3131568. (2004)

4.6 Author Contributions

- J.R. TerMaat was primarily responsible for development and testing of synthesis and affinity workstation.
- T.G. Mamedov conducted sequence and oligo design of the TM-1 and chimeric gene sequences used in this work.
- S.E. Whitney developed the Visual Basic software to control the pump and valve settings to operate the workstation.
- A. Subramanian led the scientific direction of the research.

5 Convergent synthesis by rapid PCA in the workstation

A novel convergent synthesis strategy using rapid thermocycling has been developed, facilitating the construction of longer genes. In the technique, shorter overlapping fragments are first synthesized from approximately 40 bp assembly oligonucleotides using rapid 2-step PCA-PCR. These fragments are then stitched together and the full-length sequence amplified using rapid integrated PCA-PCR. To demonstrate the rapid convergent approach, synthesis of a 3.8kb gene sequence consisting of a mouse whey acidic protein and cDNA for human protein C (mWAP-hPC) was undertaken. Synthesis of the 3.8kb gene was performed by stitching either two fragments (approximately 1.9kb each) or four fragments (approximately 950 bp each). The rapid gene synthesis workstation developed in Chapter 4 was utilized for the parallel synthesis of fragments and convergent stitching. The application of the immunoaffinity purification section of the workstation was also explored to determine if this purification provides any distinct advantages in the convergent strategy. While difficulties arose in the cloning of the synthetic products, the success of the convergent strategy was demonstrated by correct gene sizes and direct sequencing of the fragments. The 3.8 kb gene represents by far the largest gene constructed to date by rapid PCR thermocycling techniques. Provided in the appendix, rapid PCA was also used to construct two novel chimeric gene sequences containing domains for both endothelial protein C receptor and thrombomodulin.

5.1 Introduction

5.1.1 Convergent strategy for longer genes

Polymerase chain assembly (PCA) is a technique for constructing synthetic gene sequences by polymerase extension of a pool of oligonucleotides utilizing a PCR thermocycler (Stemmer and others 1995). Since this first description, PCA or variants thereof have been utilized for the construction of a number of gene sequences ranging from a few hundred to many kilobases (Mehta and others 1997; Smith and others 2003; Gao and others 2004). Due to its facilitation of *de novo* sequence synthesis, PCA has taken root as a primary technique used in the gene synthesis field (Czar and others 2009). The use of a rapid thermocycler to yield short synthesis runtimes and high fidelity was first proposed by Mamedov and others (2007).

A convergent-based synthesis approach using PCA has been achieved by others, but typically ligation and/or cloning strategies were involved in the convergence. Chen and others (1994) demonstrated sequential ligation of three synthetic fragments within a pUC19 vector, in addition to direct PCR-based synthesis of the full-length 770 bp gene. Kodumal and others (2004) used PCR-based synthesis to generate 500-800 bp synthon precursors, ligated these precursors together into ~5 kb components, and used cloning to generate a 32 kb sequence. Recently, Gibson and others (2008) synthesized a 582,970 bp *Mycoplasma genitalium* genomes using recombination techniques from 5-7kb overlapping “cassettes” assembled from $\sim 10^4$ oligonucleotides of ~50bp in length each. To our knowledge, a convergent based approach using PCA (most certainly rapid PCA) in the multiple convergent steps has not been reported.

It is well-known that the larger the gene to be synthesized, the complexity increases with the number of oligos and the probability of success for direct PCA synthesis is greatly reduced. The sheer number of oligos required for the direct synthesis of a long gene may reduce reaction efficiencies, such as via polymerase sequestration (Kainz 2000). The large number of intermediate product reactions needed to progress towards the full-length gene also reduces success rates. Another complication that can arise is that hybridization specificity may be decreased; that is, mis-annealing of oligos may occur as sequences have increased probability of base pair commonality. This is especially apt for genes in which repeat sequences are present. The point at which direct PCA synthesis becomes impossible (or practically so) is dependent on the specific sequence, and could range from 1 kb (for genes with high sequence commonality, such as high GC content) to 5kb or more in which unique oligo hybridization units can be designed.

The technique of rapid integrated PCA-PCR developed in Chapter 3 (also see TerMaat and others 2009) has broader implications than just reducing a two step thermocycling process into one step. It can also facilitate an improved pathway for the synthesis of longer gene fragments. To construct longer genes by polymerase chain assembly, integrated PCA-PCR may not only be used in the first-round synthesis of shorter (<1 kb) fragments, but also as a follow-up to join these fragments into larger constructs. Even though gene synthesis with the technique was problematic in genes greater than 1 kb, the association is indirect. It is believed that the root cause of the difficulties stemmed from the number of oligos. Such an assumption is founded on competitive annealing arguments—complementary internal oligos bind to the full-length template and inhibit

exponential amplification. In the extreme, one can imagine only two oligos, each of 1 kb in length with a hybridization overlap. Rapid integrated PCA-PCR on these two oligos would seem trivial. In the presence of primers, the two oligos would assemble to create 2 kb dsDNA and be thereby exponentially amplified completely unencumbered by competitive annealing. Thus, it is possible to apply rapid integrated PCA-PCR to form larger constructs in certain instances.

For the sake of argument, consider a hypothetical 3kb gene which cannot be constructed by traditional PCA for the above reasons. Of course, direct integrated PCA-PCR will certainly not work where a 2-step PCA with follow up PCR has failed. However, an alternative approach is possible. The full-length sequence can be segmented into smaller fragments that can be facilely created by direct techniques (traditional 2-step or integrated PCA-PCR). In the case of the 3kb gene, it could be divided into two 1.5 kb fragments, three 1 kb fragments, or so on. Neighboring fragments have overlapping hybridization units to facilitate further assembly. After the fragments have been successfully synthesized separately, these fragments can be used as templates for a 2nd round rapid integrated PCA-PCR. A schematic of the process is illustrated in Figure 5-1 for convergent synthesis from a representative 3 fragments.

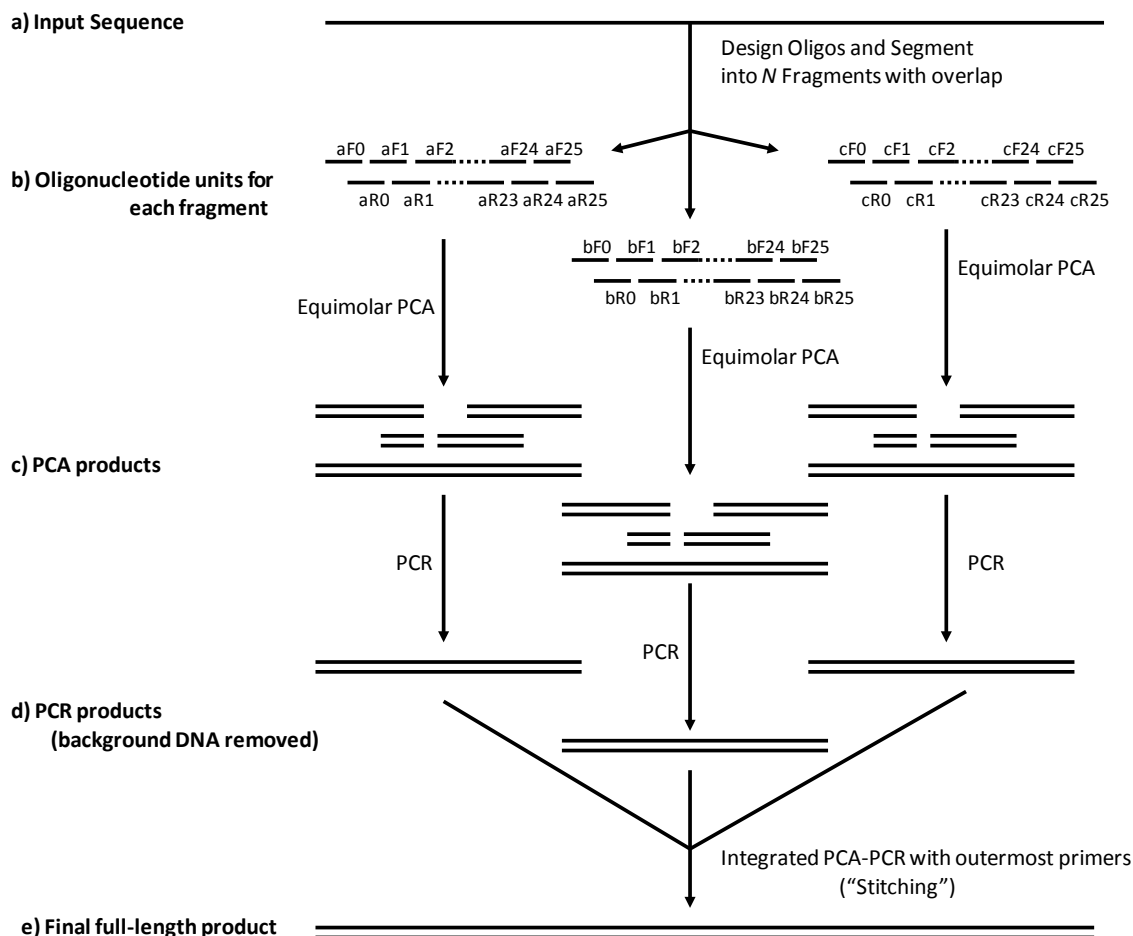


Figure 5-1 Schematic of convergent synthesis of a long gene.

Traditional 2-step PCA-PCR for 1st round synthesis of fragments and rapid integrated PCA-PCR for 2nd round stitching is illustrated.

The advantages of this convergent synthesis strategy are that it enables a reduction in number of oligos in each reaction while also potentially isolating undesired mis-annealing oligo combinations from one another. There are a number of variants possible considering the option to do either equimolar PCA-followup PCR (2-step) or integrated PCA-PCR during the fragment synthesis, and either one for the 2nd round stitching reaction. For the fragment synthesis, the two step process is preferable since in most

cases the fragments need not be shorter than 1kb (see conclusions from Chapter 3) and that a higher yield and relative purity of the fragments is preferred as templates in the stitching reaction. Given that typically only a few fragments are needed in the stitching reaction, integrated PCA-PCR for 2nd round stitching is logical. This convergent synthesis strategy may conceivably be extended further to 3rd, 4th, etc. rounds of synthesis assembly to synthesize even longer constructs, but not explored in this work.

5.1.2 Application of purification to convergent-based synthetic genes

Coupling PCA-based synthesis (traditional or integrated) of fragments with another (or more) PCA stitching steps presents unique opportunities for the application of the affinity workstation developed previously. In Chapter 2 and Chapter 4, the affinity workstation was used as a final step following synthesis of the relatively short genes tested. In Chapter 4 for example, digoxigenin and FITC were selected for use in the immunoaffinity workstation. Incorporation of these two labels into the synthetic gene via amplification primers allowed for isolation of the synthetic gene through a serial immunocapture purification. Both digoxigenin and FITC have also been employed as haptens for purification by others (Höltke and Kessler 1990; Moser and Hage 2006).

For longer genes, the workstation may possibly be applied in a similar fashion after stitching of the full-length gene. As with the shorter genes, a final PCR amplification with regular primers may be required to retain functionality of the gene for further processing (i.e. cloning). Alternatively, the intermediate fragments may each be labeled and purified prior to the final stitching. A comparison of the options is shown in Figure

5-2. Also note that the affinity purification may also be applied to both the intermediate fragments and final full-length sequence.

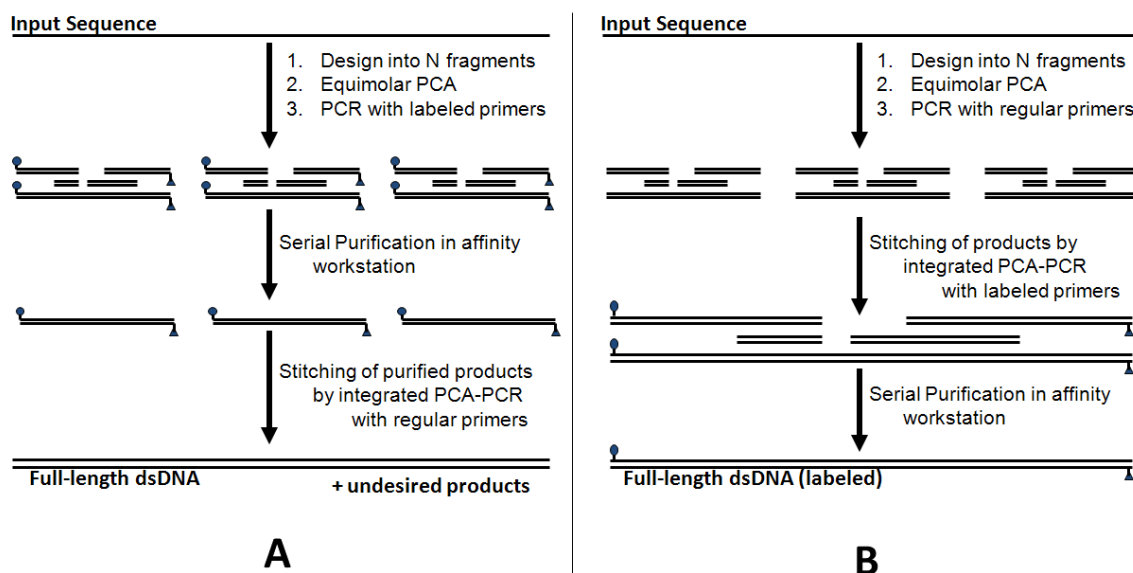


Figure 5-2 Different workstation processes for long gene synthesis from fragments.

(A) Intermediate purification of labeled fragments followed by 2nd round stitching. (B) Direct stitching of fragments using labeled primers followed by purification of the full-length gene.

5.1.3 Convergent synthesis of a 3.8kb mWAP-hPC sequence

As a difficult challenge of the rapid PCA workstation and convergent approach, a 3810 bp sequence was selected. This sequence, graciously provided by William Velander, consists of a 2,424 bp mouse whey acidic protein (mWAP) sequence and a 1386 bp cDNA sequence for human protein C (hPC). mWAP contains regulatory elements that can direct cDNA expression at high levels in pig mammary gland, whilst hPC is a

regulator of hemostasis, making it of great importance to the development of anti-coagulation therapies (Velandar and others 1992). Similar hybrid gene constructs of mWAP-hPC developed by Velandar and others (1992) have demonstrated high recombinant hPC expression levels of 1 g per liter in the milk of transgenic pigs. Initial demonstration of rapid PCA techniques to synthesize a mWAP-hPC sequence is clearly attractive; future variants of the hybrid construct without reliance on existing sequence constraints may be an extension of the convergent approach.

The 3.8 kb gene sequence represents by far the largest sequence synthesized to date by rapid PCA. Direct synthesis of the full-length gene was not possible in the workstation (results not shown)—thus, the convergent strategy was pursued using 2-step rapid PCA-PCR to construct the fragments followed by rapid integrated PCA-PCR. Stitching was tested from either 2 fragments approximately 1.9kb each in length or 4 fragments approximately 950bp each in length. Additionally, the flow-through immunoaffinity purification section of the workstation was utilized as either an intermediate step or final purification step as illustrated in Figure 5-2. The purpose was to determine the benefit, if any, of the purification sequence synthesis and yield. To understand the effects of incubation time and DNA lengths on chromatographic performance with respect to incubation times, temperatures and the effect of DNA length, batch purification studies were also conducted. Additionally, batch purification with overnight incubation of the ~1.9 kb fragments was conducted to maximize yield for stitching.

Provided as appendix, rapid PCA was employed for the *de novo* construction of a chimeric gene sequences encoding for both endothelial protein C receptor (EPCR) and thrombomodulin (TM). These two ~ 1.4 kb gene sequences serve as additional examples of the rapid PCA technique. As a demonstration to the convergent strategy, a chimeric gene was assembled from two fragments. In this chapter, a chimeric gene or its fragments was utilized for investigation of chromatographic behavior where noted.

5.2 Materials & Methods

5.2.1 mWAP-hPC: Sequence information and design of oligonucleotides

The 3810bp sequence is given in Appendix 5A.1. The Gene2Oligo program of Rouillard and others (2004) was employed—the entire sequence was input to generate the complete hybridization map shown in Appendix 5A.2. Modified Gene2Oligo settings of: T_m priority mode, 10 nM each oligo concentration, 1 M Na^+ , and 69 +/- 10°C were required to generate a reasonable program output. A total of 204 oligos comprised the sequence set. The oligo set was then segmented into 2 or 4 fragments as outlined in Table 5-1. Primers were designed for each fragment, with the overlap between fragments simply the corresponding hybridization units between neighboring oligo sets. Assembly oligos and regular primers were obtained from Integrated DNA technologies at 100 μM in nuclease-free water. Labeled primers for each fragment labeled on the 5' end (digoxigenin on forward primer and FITC on reverse primer) were obtained from IBA GmbH (Göttingen, Germany). Note that in certain instances where noted in the text, a modified R4 (R4.1) was used that lacked “TT” on the 3' end.

Fragment	# of oligos	Product Size	Primers	Forward Primer (5'-->3')	Reverse Primer (5'-->3')
1	50	985 bp	F1 / R1	GAATTCTTTCACTGCTAAACAG	GCCCACTGGCTGTTATTC
2	50	936 bp	F2 / R2	GAATAACAGCCAGTGGGC	CCATGTAGAGAGTGTGGG
3	52	1000 bp	F3 / R3	CCGACACTCTCTACATGGAG	TCCTCTAGGCAGTAATGCG
4	52	943 bp	F4 / R4	GCATTACTGCCTAGAGGAGG	CTAAGGTGCCAGCTCTT
(1+2)	100	1903 bp	F1 / R2	GAATTCTTTCACTGCTAAACAG	CCATGTAGAGAGTGTGGG
(3+4)	104	1925 bp	F3 / R4	CCGACACTCTCTACATGGAG	CTAAGGTGCCAGCTCTT
Full-length	204	3810 bp	F1 / R4	GAATTCTTTCACTGCTAAACAG	CTAAGGTGCCAGCTCTT

Table 5-1 Assembly details and primers used in the mWAP-hPC synthesis.

5.2.2 mWAP-hPC convergent synthesis

The first step in the convergent approach was to synthesize the 4 fragments (1, 2, 3, and 4) separately or combine into 2 parts ((1+2) and (3+4)) by 2-step rapid PCA-PCR. 2nd round stitching of fragments by integrated PCA-PCR was performed to construct the full sequence. Each 25 µl reaction contained the standard reagents of 200 µM each dNTP, 5 mM MgSO₄, 400 µg/ml non-acetylated BSA, and 0.5 U hot-start KOD polymerase in 1x manufacturer's reaction buffer. The optimal concentration of assembly oligos in the PCA reaction was determined to be 50 nM each. For PCR amplification and 2nd round stitching by integrated PCA-PCR, 0.7 µM each primer was used in accordance with Table 5-1. 1 µl respective PCA product as template was used for PCR amplification. For the stitching reaction, 1 µl each of the fragment PCR products was employed (i.e. 4 parts: 1 µl each of 1, 2, 3, and 4; 2 parts: 1 µl each of (1+2) and (3+4)) unless otherwise noted. Note that the synthesis of longer fragments (1+2+3), (2+3+4), and (1+2+3+4) was also attempted but not successful. Thermocycling conditions in the PCRJet are given in Table 5-2.

	1,2,3, and 4 fragments		(1+2) and (3+4) fragments		Stitching of full-length sequence
	PCA	PCR	PCA	PCR	
Hot Start	94°C for 30 sec				
# of cycles	30x	35x			45x
Denaturation	94°C for 2 sec				
Annealing	56°C for 10 sec	56°C for 3 sec	56°C for 10 sec	56°C for 3 sec	56°C for 10 sec
Elongation	72°C for 10 sec		72°C for 20 sec		72°C for 30 sec
Final Extension	72°C for 25 sec				72°C for 30 sec

Table 5-2 Thermocycling conditions for the convergent synthesis.

5.2.3 Flow-through Immunoaffinity Purification

A modified protocol to the one presented in Chapter 4 was utilized to purify either a mixture of the 4 fragments, a mixture of the 2 fragments, or the full-length gene. The buffer vessels in the workstation were filled with the following buffers:

- Binding/Wash Buffer (BB): 2.5 mM Tris-HCl, 12.5 mM NaCl, pH 7.4
- FITC Elution Buffer (fEB): FITC label at max solubility in water with 1% DMSO, 100 mM NaCl
- Digoxigenin Elution buffer (dEB): DIG label at max solubility in water with 1% DMSO, 250 mM NaCl
- Regeneration Buffer (RB): 0.1 M phosphate buffer, pH 2.5

Note the sixth vessel was not productively used yet filled with water for priming of the tube. Prior to installation of the columns, all lines were primed via the software interface. The respective buffer vessel fluid was primed up to the 7-port union. Binding buffer was primed into the feed line and all lines downstream of the union. After installation, the columns were equilibrated with approximately 2 ml of binding buffer.

With the system initialized, the following protocol was used. Effluent samples of approximately 250 μ l were collected.

1. Load Feed to FITC column. Mixed 100 μ l labeled PCR product with 400 μ l BB in sample Feed microcentrifuge tube. 40 μ l was saved as a Feed aliquot for gel analysis. With V_{Feed} on, pumping set at 20 μ l / min (200ms on / 29800ms off per stroke). When ~20 μ l left, added 100 μ l BB to tube and continued pumping. When ~20 μ l left again, switched to pure BB microcentrifuge tube for an additional 500 μ l. The first 250 μ l was not collected. Collected samples #1-3 from $V_{3\text{way}1}$ waste.
2. Wash FITC column. With V_{BB} on, pumping set at 40 μ l /min (200 ms on / 14800 ms off per stroke) for a total of 1.75 ml. Collected samples #4-10 from $V_{3\text{way}1}$ waste.
3. Elute FITC column/Bind DIG column. With V_{fEB} on and $V_{3\text{way}1}$ on, pumping set at 20 μ l /min (200 ms on / 29800ms off per stroke) for a total of 1.5 ml. Collected samples #11-16 from $V_{3\text{way}2}$ waste.
4. Wash DIG column. With V_{BB} on and $V_{3\text{way}1}$ on, pumping set at 40 μ l /min (200 ms on / 14800 ms off per stroke) for a total of 1.75 ml. Collected samples #17-23 from $V_{3\text{way}2}$ waste.
5. Elute DIG column. With V_{dEB} on, $V_{3\text{way}1}$ on, and $V_{3\text{way}2}$ on, pumping set at 40 μ l / min (200 ms on / 14800 ms off per stroke) for a total of 1.75 ml. Collected samples #24-31 from $V_{3\text{way}2}$ collect.

Upon completion, columns that were to be re-used were flushed with 2 ml each of the respective elution buffer, regenerated with ~6 ml of regeneration buffer, and then equilibrated with ~6 ml of binding buffer. Where noted, final elution fractions were combined and concentrated using an Amicon DNA purification kit (part # UFC503096, Millipore, Billerica, MA) per manufacturer's instructions with one water wash

5.2.4 Batch immunoaffinity purification

To characterize the binding and elution parameters of the antibody resins, several different batch experiments were carried out in microcentrifuge tubes:

DNA length effects: An equal volume mixture of FITC-labeled chimera #2 "A" PCR product (640 bp) and full-length chimera #2 (1377 bp) was prepared. The 100 µl mixture was mixed with 300 µl binding buffer, 30 µl removed as Feed, and then subjected to binding to anti-FITC resin on an end-to-end rotator at room temperature. 30 µl aliquots of the supernatant were removed after 45 min and 90 min.

Binding incubation time and temperature: 70 µl of digoxigenin labeled chimera #2 "A" PCR product (640 bp) was mixed with 220 µl of binding buffer. After removal of a Feed aliquot, this Feed was subjected to binding to 100 µl of the anti-digoxigenin resin over varied time periods at either room temperature (end-to-end rotator) or in a 37°C incubator (shaking). 30 µl aliquots of the supernatant after centrifugation were removed after 15 min, 30 min, 1 hr, 2 hrs, or 4 hrs.

Overnight serial batch binding: The mWAP-hPC (1+2) and (3+4) PCR products, both labeled with Digoxigenin and FITC, were processed separately. 120 µl of PCR product was mixed with 360 µl binding buffer and added to 150 µl of anti-FITC resin. Binding proceeded overnight on an end-to-end rotator at room temperature. The microcentrifuge tube was centrifuged and the post-bind supernatant removed (same procedure after wash and elution steps). The resin was washed 3 times in 300 µl binding buffer for 10 minutes incubation. Two elution steps were performed using 300 µl FITC elution buffer, with 4 hours incubation each step. The elution fractions were combined as an intermediate Feed, and then the procedure repeated for purification with 125 µl of anti-digoxigenin resin. The final elution fractions were combined and concentrated using an Amicon DNA purification kit (part # UFC503096, Millipore, Billerica, MA) per manufacturer's instructions with one water wash. Stitching of the concentrated products was performed using primers as described in the text.

5.2.5 Gel Electrophoresis & Purification

PCR products were run on a 1% agarose gel stained with ethidium bromide using 10 µl each product and 12 µl 1kb ladder (New England Biolabs). Where gel purification was used, the appropriate band was excised from the gel and processed using a purification kit (Cat # A9281, Promega Corporation, Madison, WI) for cloning and sequencing. Samples obtained from affinity purification fractions were analyzed on a 1% agarose gel stained with GelRed using 30 µl each product and 12 µl 1kb ladder (New England Biolabs).

5.2.6 Cloning & Sequencing

Cloning and sequencing of each of the 2-fragments (gel-purified) was performed by GeneWiz, Inc. (South Plainfield, NJ). However, GeneWiz was unable to produce any colonies using either TA cloning or TOPO blunt-end cloning strategies. To understand what might be lethal to the cloning, direct sequencing of the PCR products was performed. See Appendix 5A.3 and 5A.4 for the sequencing results of gene (1+2) and (3+4), respectively.¹

5.3 Results & Discussion

5.3.1 mWAP-hPC fragment synthesis

The synthesis of the four 950 bp fragments and the two ~1.9 kb fragments by rapid 2-step PCA-PCR was straightforward. Figure 5-3 shows the resulting PCR products for the various fragments. High yield was obtained for all PCR products with a minimal amount of assembly background. Note that the particular products shown in Figure 5-3 were generated using labeled (digoxigenin and FITC) primers—similar results were also obtained using regular primers. The 1.9 kb fragments demonstrated excellent yield and purity given the length of these products. This is indicative of robust assembly oligo design and supportive of optimal conditions used in the synthesis.

¹ The full 3.8 kb gene was first sent to Biomeans, Inc. (Sugar Land, TX) for cloning and sequencing. However, Biomeans failed to provide any results or evidence the work was conducted.

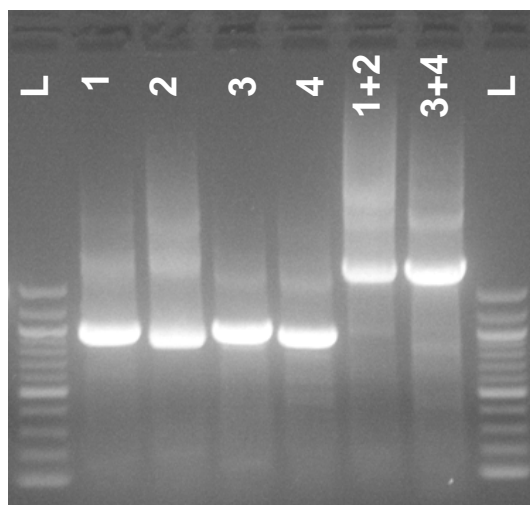


Figure 5-3 PCR products of the different mWAP-hPC fragments generated by rapid 2-step PCA-PCR.

Please refer to Table 5-1 for the respective lengths of the products.

It is important to note that all of similar fragment synthesis reactions were performed in parallel. That is, the ~950 bp fragments were PCA assembled in separate capillaries in a single thermocycler run, and the PCR amplification of these fragments was also performed in a single thermocycling run. Likewise, side-by-side processing was performed for the ~1.9kb fragments. This is a consequence of the entire 204 oligo set having been designed comprehensively and then segregated into fragments with amplification primers also matched for melting temperature. Thus, annealing temperatures were reasonably similar to one another. This parallel processing significantly reduces the instrument time required. Coupled with rapid thermocycling protocols in which each run required only about 20 minutes, the actual thermocycling time for fragment synthesis becomes insignificant compared to other steps in the experimental process such as preparation of oligo and reaction mixes.

5.3.2 Flow-through immunoaffinity purification

The flow-through immunoaffinity workstation was employed to purify labeled intermediate fragments. In line with parallel synthesis in the thermocycler, the feeds to the anti-FITC column were equal volume mixtures of labeled PCR products. Purification fractions for the four ~950 bp fragments are shown in Figure 5-4. The binding to the anti-FITC column was moderate, as evidenced by comparison of the Feed to the pass-through. Upon elution from the FITC column and binding to the anti-digoxigenin column, more pass-through of the full-length product is present. However, a respectable amount of purified product was obtained in the final elution from the anti-digoxigenin column.

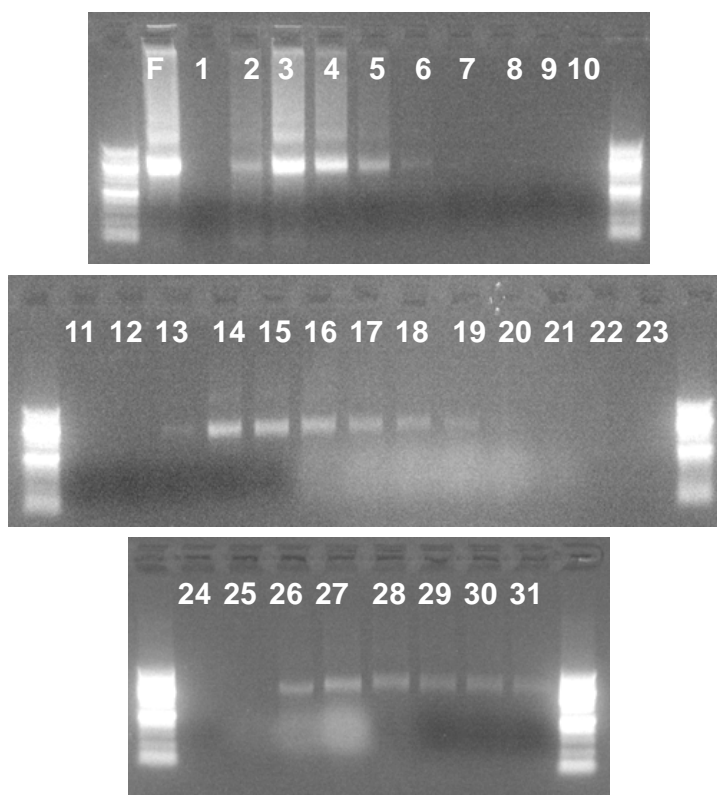


Figure 5-4 Flow-through immunoaffinity purification of a 4-fragment (~950 bp) mixture.

Refer to the materials and methods section for sample details.

Similar chromatographic behavior was obtained for the purification of the two ~1.9kb fragment mixture feed as shown in Figure 5-5A. However, the binding yield at both columns was reduced when compared to the 4-fragment purification. This resulted in even less purified product yield in the final anti-digoxigenin elution. While yield was reduced, faint bands are still visible by gel electrophoresis. Therefore, two final elution fractions (#26 and #27) were combined and processed using an Amicon 30K MW kit to concentrate the samples. 5 μ l of this Amicon product is shown in Figure 5-5B. The

concentration was successful since the intensity was slightly better than the original 30 μ l elution samples in a reduced volume.

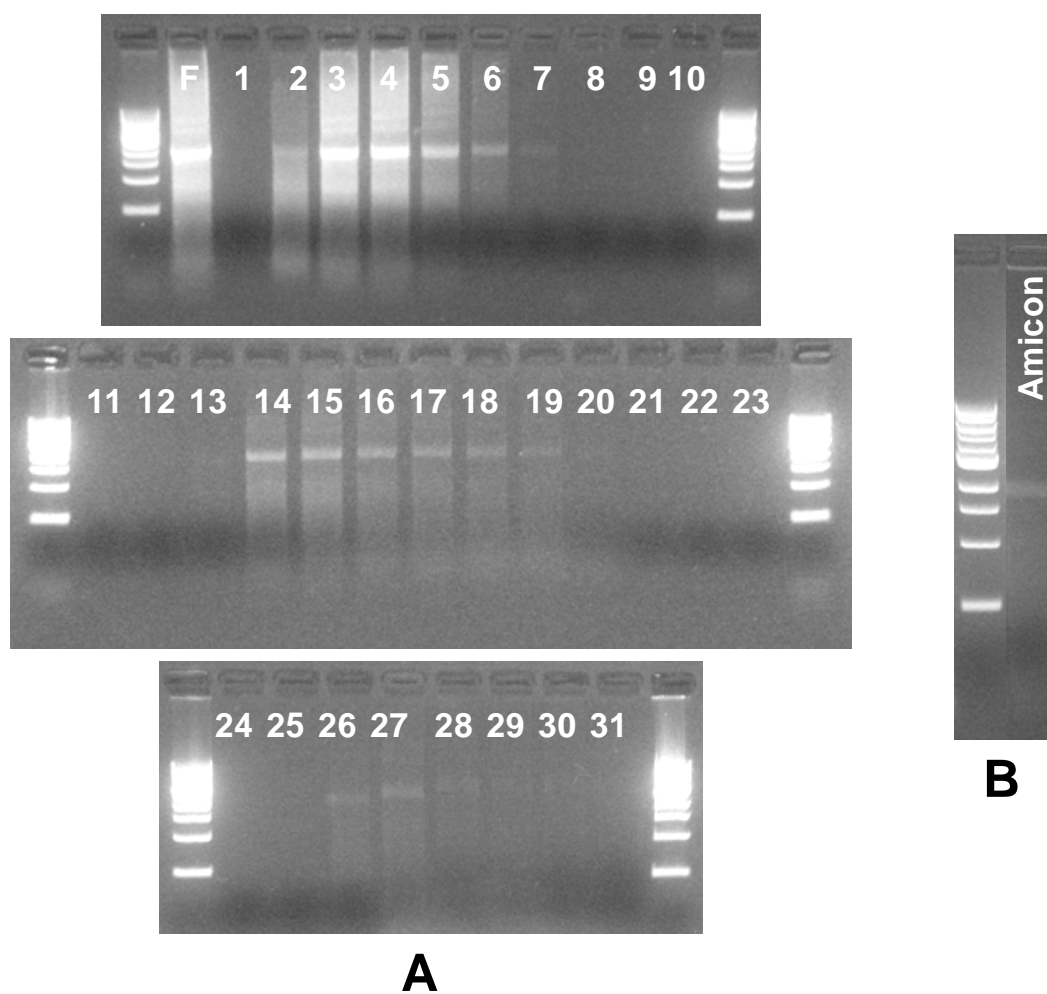


Figure 5-5 Flow-through purification of a 2-fragment (~1.9 kb) mixture (A) and the subsequent Amicon concentration (B).

Refer to the materials and methods section for sample details.

The alternative approach of conducting direct stitching of the 3.8 kb band with labeled primers followed by purification was also tested (Figure 5-6). However, the binding yield was so poor that no product was visible in the final effluent. Insufficient product

for gel detection was also the case even after purification and concentration with the Amicon kit (results not shown). There is some product visible in the FITC elution / digoxigenin bind (samples #14, 15, and 16). Extrapolation from the reduced yield of 1.9 kb fragments compared to 950 bp fragments provided some justification for this result. The differing results from the three different gene lengths insinuate that the binding is perhaps dependent on gene length.

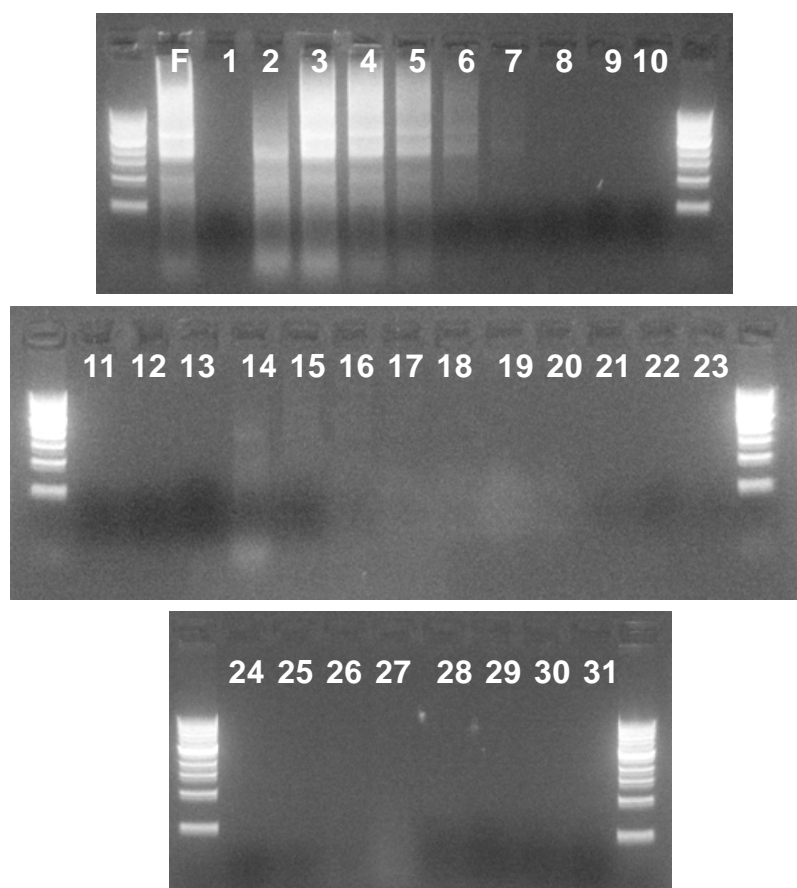


Figure 5-6 Flow-through purification of the full-length labeled 3.8kb direct stitching product.

Refer to the materials and methods section for sample details.

5.3.3 2nd round stitching to construct the 3.8kb sequence

Stitching by 2nd round integrated PCA-PCR from fragments is shown in Figure 5-7. Direct stitching using 1 µl each of unpurified PCR product (from either 4 or 2 fragments) is shown in Figure 5-7. The desired 3.8 kb band was successfully stitched together in both cases. Note that for stitching from 4 fragments, there are incomplete side products at ~1.9 kb (2 of the 4 joined) and ~2.7 kb (3 of the 4 joined) in addition to the 3.8 kb band (4 of the 4 joined). There is also likely remaining ~950 bp template bands (not joined whatsoever), but not visible due to conversion of these products into the larger fragments. Such quantization is an inherent by-product of the stitching from more than 2 fragments. In the stitching from 2 parts, only the 3.8 kb band (2 of the 2 joined) is distinctly visible. The lack of a strong band also at the template length of 1.9 kb also indicates a high conversion efficiency of the stitching. For reference, in other experiments high yield was obtained when conducting stitching on 1+2 and 3+4 fragments to yield ~1.9 kb products (results not shown).

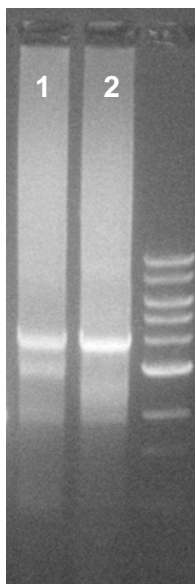


Figure 5-7 Direct stitching of unpurified labeled PCR products from either 4 fragments (lane 1) or 2 fragments (lane 2) to produce the 3.8 kb full-length sequence.

One of the hypotheses of the research was that intermediate immunoaffinity purification of fragments would increase the yield of the stitching reaction. The belief was that the removal of unwanted background DNA that is complementary to the template would reduce competitive annealing and increase the stitching efficiency. However, when stitching was conducted on the flow-through purified fragments (either 4 or 2 fragments), the desired 3.8kb band was never visible on the gel despite exhaustive optimization attempts.

On the surface, this result seems contrary to reason. To understand why purified products were inferior in stitching to the unpurified products, one must examine the other parameters important to the stitching process. Aside from purity of templates, the concentration of templates is of critical importance, supported extensively by traditional

PCA theory. An immense difference exists in the amount of template used in the purified vs. unpurified stitching reactions. Due to the low binding yield at each column stage and the final product dilution over the elution buffer effluent fractions, stitching from purified products employed much lower template concentrations. The concentration of final flow-through effluent product with the Amicon kit improved template concentration, but was still insufficient for stitching.

To demonstrate the importance of template concentration to the stitching reaction, different template concentrations were compared using the 2 fragments. 2 μ l, 1 μ l, 1 μ l of a 10x dilution, and 1 μ l of a 100x dilution, for each of the fragments was employed. The results of the template concentration upon stitching are shown in Figure 5-8. As can be seen from the results, 1 μ l each fragment gave the highest yield while 2 μ l resulted in slightly less yield and conversion of the ~1.9 kb templates. It is clear that dilution of the templates greatly affected the result—the 10x dilution gave only slight yield whereas the 100x dilution completely failed to yield the desired band.

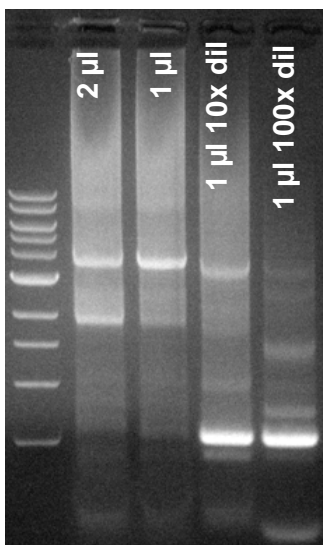


Figure 5-8 Effect of fragment template concentration on stitching of the 3.8 kb gene.

At the lower concentrations, a non-specific band at ~600 bp was generated. This non-specific band was also highly prevalent in stitching reactions from purified products. In a perfect world, PCR amplification is possible from a single template (or in this case, one molecule of each fragment). Yet as any PCR practitioner knows, low-copy (especially single-copy) PCR may have inherent difficulties due to the fragile reactions occurring in the first few cycles and formation of unwanted primer interactions (Chou and others 1992). The problems are exacerbated in the stitching case—during a cycle two very dilute template molecules must anneal through diffusion before undesired products from templates and primers are formed and dominate further cycles. Thus, it is understandable why the lower template concentrations of purified products fail to yield the full-length gene despite having higher purity. To prove that purified products result in better yield than unpurified products when template concentrations are equal, gel-purified and concentrated (to approximate concentration of PCR products) was compared to unpurified products utilizing the 4-fragment stitching. As can be seen in Figure 5-9, a higher yield is obtained for gel-purified fragments. A similar improvement in yield and purity was obtained in stitching gel-purified 2-fragments compared to unpurified templates (results not shown).

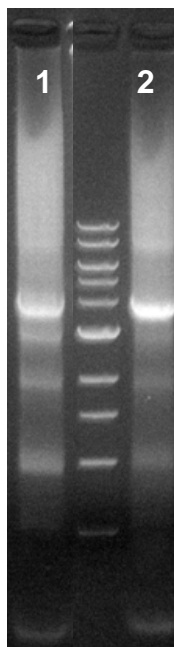


Figure 5-9 3.8 kb stitching comparison of unpurified vs. gel-purified using 4-fragments.

1 μ l each of the unpurified PCR products (lane 1) or 4 μ l of a combined 4-fragment gel purification (lane 2) was employed to give similar template concentrations.

Other testing revealed that the F1/R4 combination necessary to amplify the full-length gene was contributory to the problems at hand. As noted in the materials and methods section, a modified outermost reverse primer (i.e. R4.1) was deemed to give improved amplification. While R4.1 was better than the original R4 at lower template concentrations, it did not give sufficient improvement to give successful stitching from immunoaffinity products. The F1/R3 and F2/R3 primer combination brackets the stitching point of the 2-fragment reaction; stitching using these primer combinations showed marked improvement in the desired band yield (results not shown). Since a shorter product than the full 3810 bp sequence would lack the desired functionality, these

shorter products only served to separate the complications arising from PCR from the base hypothesis.

5.3.4 Follow-up batch testing

Due to the low yield of the flow-through immunoaffinity purification and its subsequent impact on stitching, batch purification testing was performed. In light of the decreased yield with longer length synthetic genes from flow-through testing, further investigation of DNA length effects was undertaken. Using a chimeric gene, direct binding competition using an equal volume mixture of the 640 bp “A” fragment and 1377 bp full-length PCR product was performed. The feed and post-binding supernatants after 45 min and 90 min are shown in Figure 5-10. If length was not a factor, one would expect that the relative ratios of feed bands to post-binding bands would be constant; that is, both products would bind in the same relative amounts. However, the shorter 640 bp band bound in greater amounts than the 1377 bp band, again indicating a binding preference for the shorter gene. Other batch purification experiments with wash and elution steps showed more binding and elution yield for the shorter gene. Thus, it is believed that the gene length does indeed impact the immunoaffinity performance.

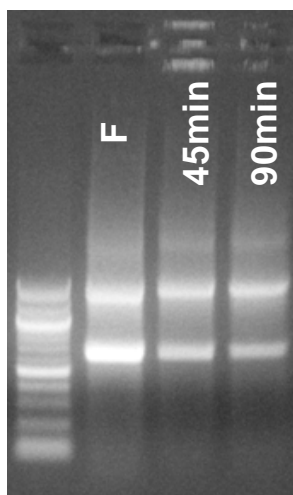


Figure 5-10 Comparison of the effect of DNA length on chromatographic binding.

Competitive binding of a 640 bp chimera “A” fragment and a full-length 1377 bp chimera gene was performed.

The significance of gene length may be rationalized in several ways. At the extreme, the ligand obviously has a higher affinity for a free label than a label with extra attachments. This is clear from the utilization of free labels in the elution buffers to competitively displace bound labeled DNA. The longer the attachment to the free label is, the binding constant is decreased, although the argument could be made that beyond very localized attachments the decrease is negligible. Another consideration is steric hindrance conveyed by the long DNA chain. It is unlikely, especially with longer sequences, that the chain exists in linear form. Rather, the DNA may coil up and may limit the accessibility of the label to the ligand. Even if the label is at times exposed and does bind to the ligand, longer chains may increase the disruption of this bond to disassociate the complex.

Another drawback of the flow-through system is that the residence time may be insufficient for the specific antibody columns employed, even under the slow flow rates employed in the protocol. To investigate, the batch binding over various incubation times are shown in Figure 5-11. It can be seen that most binding occurs initially and decreases at each interval, in line with second order reaction kinetics (or pseudo-first order in the case of a large antibody excess—exact quantitative kinetics plots were not performed on gel intensities). To reach reasonably close to equilibrium in which most of the labeled DNA has bound, around one hour is needed which is not feasible in this flow-through system. To determine if increased temperature would affect binding rates, the incubation time experiment was repeated at 37°C (Figure 5-11). In comparing the binding behavior at the two different temperatures in Figure 5-11, the increased temperature did not have a visible impact on the binding rates. Given the nature of the antibody binding, it is understandable that it may not follow normal Arrhenius rate kinetics and explain why the binding was not improved at 37°C. This is in good agreement with the studies conducted by Lipschultz and others (2002), who found that antigen-antibody docking can be energetically less favorable at elevated temperatures and result in a higher apparent off-rate. Thus, aside from increased residence time and amount of ligand present, there is little leeway to improve the performance of the existing purification systems.

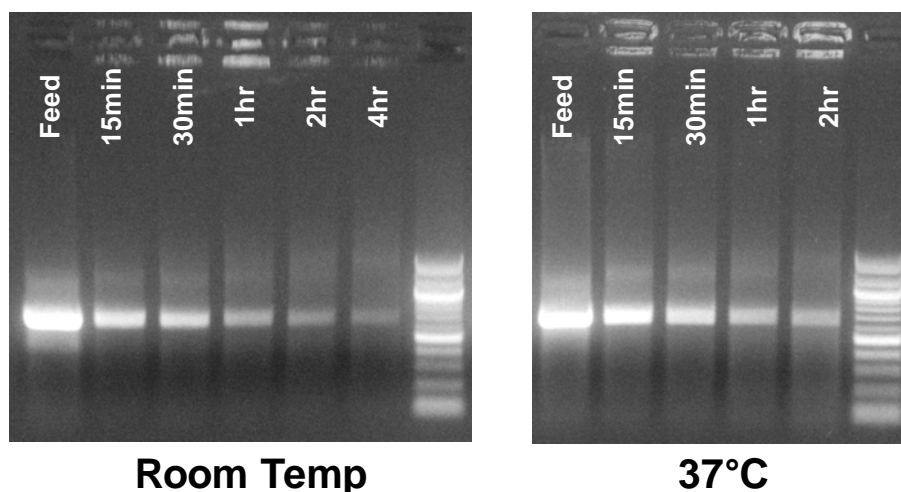


Figure 5-11 Effect of incubation time on binding amount at room temperature and 37°C.

In an attempt to get sufficient yield for stitching, overnight binding and 4 hour elution incubation times were applied to batch serial purification of the (1+2) and (3+4) fragments in separate tubes. The resin amounts used were increase to 150 μ l and 125 μ l, FITC and digoxigenin respectively, in case there was any notion of insufficient binding sites. The purification samples are shown in Figure 5-12. The yield is only marginally improved as even overnight binding of these longer genes did not result in a significant decrease in post-bind samples. This implies that the binding may have reached equilibrium and nothing short of greatly increasing the number of binding sites would alter the outcome. Regardless, after concentrating with the Amicon kit the yield was at least better than in previous tests. Given the slight increase in template concentration, stitching was possible with these immunoaffinity purified products (Figure 5-12 “Stitch”). The stitching product was still inferior to unpurified product stitching due to the decreased template concentration, but at least demonstrated some measure of success.

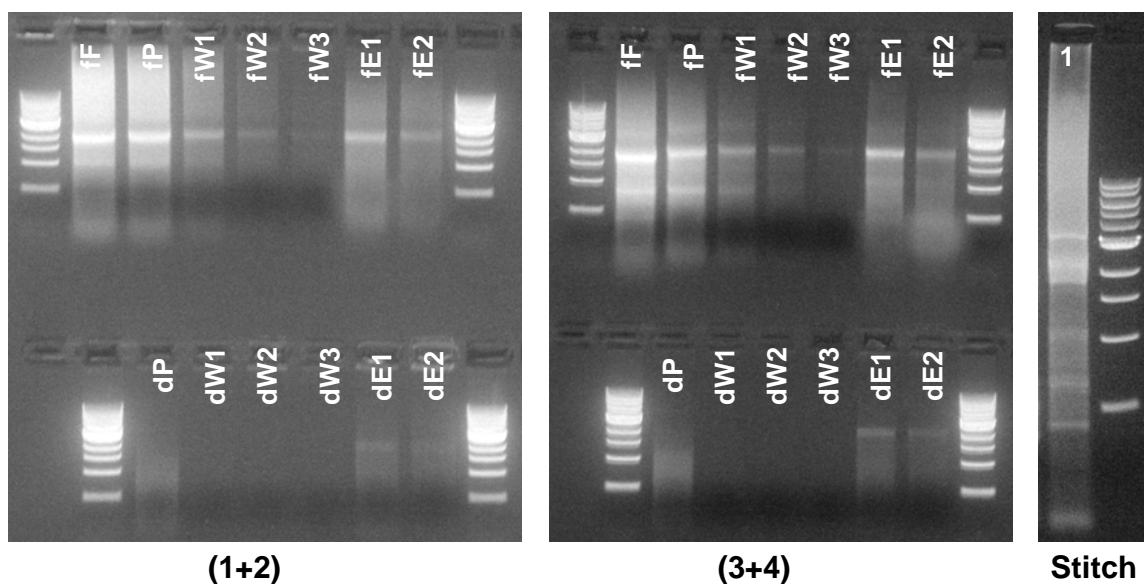


Figure 5-12 Batch testing with long incubation times [(1+2) and (3+4)] with serial purification with FITC then digoxigenin, and subsequent stitching (Stitch, lane 1) of the purified products.

f: FITC; d: digoxigenin; F: feed; P: post-bind supernatant; W: wash (three washes); E: elution (two elution fractions).

5.3.5 mWAP-hPC sequencing results

Difficulties arose in the cloning of gel-purified (1+2) and (3+4) synthetic fragments. While the exact reason for the lack of clones was not determined, the presence of repeat sequences within the genes was identified as one possible cause. As such, direct sequencing of the PCR products was performed to, at the very least, provide a basic understanding of the constructed sequences. The sequencing results for the (1+2) and (3+4) genes are given in Appendix 5A.3 and 5A.4, respectively. It is important to

recognize that the sequencing data is not “clean” due to the residual presence of assembly products and superimposition of peaks.

The sequencing results of gene (3+4) show that the desired gene was successfully constructed. The sequencing data matches the reference sequence very closely, with only a number of base pair positions in question. It can be seen that in nucleotide positions where an error may be present, the improper “G” generated in the data is flanked by repeat “A” nucleotides (indicating confounding sequence signal overlap). Thus, there is a high level of confidence that the (3+4) gene was successfully constructed, barring typical error rates that one would expect from the PCA process.

It is more difficult to draw definite conclusions regarding the accurate synthesis of gene (1+2) from the sequencing data. The synthetic sequence matches the reference sequence by and large over the first 800 bp and from 1300 to 1903 bp. However, the sequencing data appears corrupted in the middle, starting at the presence of a problematic repeat “GAGA...” section in the mWAP region starting at position 760. This repeat region compromised the data from the F2 sequencing primer. Thus, the relative correctness of this middle portion cannot be unequivocally stated.

The correct size construct as determined by gel electrophoresis for not only (1+2) and (3+4) but also for each of the four ~ 950 bp fragments. Additionally, the correct 3.8 kb band was produced in the stitching reactions from both 4 fragments and 2 fragments. The production of the desired sizes in all these cases from a pool of oligos without actual

assembly of the basic desired sequences would defy logic. With these considerations and the current sequencing data, there is strong evidence that the individual fragments and the entire 3810 bp mWAP-hPC was constructed. It is certainly true that errors are likely present. However, the isolation of error-free clones for the entire 3.8 kb gene was not the immediate objective of the work at the onset given typical error rates observed in PCA-based synthesis. It was expected that intermediate screening of fragments would be necessary before stitching the full-length construct, but in this case difficulties in cloning presents a hurdle to overcome in future work. With the failure of the ~ 1.9 kb cloning efforts, the next logical step would be to attempt cloning and sequencing of the four ~ 950 bp fragments. If error-free clones can be isolated for these fragments, stitching of these isolates and subsequent cloning/sequencing of the full-length gene should give definitive proof of success. A further discussion of strategies that may be explored to construct and isolate error-free sequences for longer genes is presented in Chapter 6.

5.4 Summary & Conclusions

A convergent gene synthesis strategy was developed in which rapid PCR thermocycling was employed to construct longer gene sequences not amenable to direct synthesis. In the technique, rapid 2-step PCA-PCR is recommended for the initial synthesis of sequence fragments that can be readily constructed. Each fragment contains hybridization overlap with adjacent sequence fragments such that the fragments can hybridize with each other in the annealing PCR step. Thus, a second stitching reaction by rapid integrated PCA-PCR assembles these fragments with simultaneous amplification of the full-length gene.

The rapid convergent strategy was applied to the synthesis of a 3.8 kb mWAP-hPC gene sequence. Either two (~1.9 kb) or 4 (~950 bp) fragments were easily constructed from short single-stranded assembly oligos in the first round reactions. Direct stitching of these fragments using 1 μ l each as templates successfully generated a band at the desired 3.8 kb length. Due to anticipated error rates, intermediate screening of the two ~1.9 kb fragments was attempted. However, problems arose in cloning of these fragments. Direct sequencing of these fragments (after gel-purification) revealed that, to the exclusion of some errors, the desired fragments were generated. The success of the convergent strategy is given strong supporting evidence from these direct sequencing results and the gel electrophoresis band lengths obtained.

Serial immunoaffinity purification in the workstation was explored for utility in the rapid convergent synthesis technique, both as 1) post-synthesis purification of the full-length gene, and 2) intermediate purification of fragments prior to stitching. With respect to the latter, the hypothesis was that the stitching reaction would benefit from higher purity template fragments. However, a by-product of the purification was a lower template concentration in stitching. For this particular gene, these lower template concentrations proved lethal to stitching success. Only under overnight batch incubation times could enough fragment concentration be achieved after serial purification to successfully generate the 3.8 kb band (and even then inferior to direct stitching). Note that other sequence synthesis may not be susceptible to this issue, and purification may prove

beneficial in other cases. As a basic demonstration, gel-purified fragments at comparable template concentrations to direct PCR products resulted in higher yield.

For flow-through post-purification of the full-length gene, an insufficient yield was obtained to detect by gel electrophoresis. Based on flow-through testing and batch competitive binding testing between DNA of different lengths, it is believed that the immunocapture yield was highly dependent on gene length. Shorter fragments (> 1kb) had relatively acceptable binding levels, whereas binding amounts were substantially reduced for longer products. While the current immunoaffinity purification can be considered acceptable for shorter gene purification, future improvements to the chromatographic performance for longer genes are recommended. Since elevated temperatures did not increase binding, the main parameters for optimization of the existing FITC and digoxigenin systems are ligand concentration, total ligand amount, and residence time in the flow-through system.

An important aspect of the synthesis of longer genes is that many of the processes therein on the fragments can be done in parallel. Multiple reaction capillaries (one for each fragment) may be processed simultaneously during PCA and PCR thermocycling. This would require that all fragments may be acceptably processed under the thermocycling conditions, most notably the annealing temperature. To accomplish this, the melting temperatures of hybridization overlaps are reasonably matched to one another during the oligo design. Likewise, the fragments may be purified simultaneously in the affinity columns; that is, the synthesis reaction products combined in the feed to result in a

purified mixture of the fragments in the final effluent. Such parallel approaches would significantly reduce the time requirements when compared to separate sequential processing.

5.5 References

- Chen, G.Q., I. Choi, B. Ramachandran, and J.E. Gouaux. Total gene synthesis: novel single-step and convergent strategies applied to the construction of a 779 base pair bacteriorhodopsin gene. *J. Am. Chem. Soc.* **116**: 8799-8800. (1994)
- Chou, Q., M. Russell, D.E. Birch, J. Raymond, and W. Bloch. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* **20**: 1717-1723. (1992)
- Czar, M.J., J.C. Anderson, J.S. Bader, J. Peccoud. Gene Synthesis demystified. *Trends Biotechnol.* **27**: 63-72. (2009)
- Gao, X., P. Yo, A. Keith, T. Ragan, and T. Harris. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**: e143-e164 (2004)
- Gibson, D.G., et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220 (2008)
- Höltke, H.J. and C. Kessler. Non-radioactive labeling of RNA transcripts in vitro with the hapten digoxigenin (DIG); hybridization and ELISA based detection. *Nucleic Acids Res.* **18**: 5843-5851. (1990)

- Lipschultz, C.A., A. Yee, S. Mohan, Y. Li, and S.J. Smith-Gill. Temperature differentially affects encounter and docking thermodynamics of antibody--antigen association. *J. Mol. Recognit.* **15**: 44-52. (2002).
- Kainz P. The PCR plateau phase—towards an understanding of its limitations. *Biochim. Biophys. Acta.* **1494**: 23-27. (2000)
- Kodumal, S.J., K.G. Patel, R. Reid, H.G. Menzella, M. Welch, and D.V. Santi. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 5573-5578 (2004)
- Mamedov, T.G., N.V. Padhye, H. Viljoen, and A. Subramanian. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387 (2007)
- Mehta, D.V., R.J. DiGate, D.L. Banville, and R.D. Guiles. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. *Prot. Expr. Purif.* **11**: 86-94 (1997)
- Moser, A.C. and D.S. Hage. Chromatographic Immunoassays. In: Hage. D.S. Handbook of Affinity Chromatography. CRC Press. Boca Raton, FL. (2006)
- Rouillard, J.M., W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**: W176-W180 (2004)
- Smith, H.O., C.A. Hutchison, C. Pfannkock, and J.C. Venter. Generating a synthetic genome by whole genome assembly: Φ X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 15440-15445 (2003)

- Stemmer, W.P.C., A. Cramer, K.D. Ha, T.M. Brennan, and H.L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*. **164**: 49-53 (1995)
- TerMaat, J.R., E. Pienaar, S.E. Whitney, T.G. Mamedov, and A. Subramanian. Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler. *J. Microbiol. Methods* **79**: 295-300. (2009)
- Velander, W.H., J.L. Johnson, R.L. Page, C.G. Russell, A. Subramanian, T.D. Wilkins, F.C. Gwazdauskas, C. Pittius, and W.N. Drohan. High-level expression of a heterologous protein in the milk of transgenic swine using the cDNA encoding human protein C. *Proc. Natl. Acad. Sci. USA*. **89**: 12003-12007. (1992).

5.6 Author Contributions

- J.R. TerMaat was responsible for development and testing of the rapid convergent synthesis and purification.
- S.E. Whitney developed the software to control the pump and valve settings to operate the affinity workstation.
- A. Subramanian led the scientific direction of the research.

6 Summary, Analysis and Future Work

6.1 Rapid PCA-based gene synthesis

6.1.1 Summary

Based upon the initial work of Mamedov and others (2007), the rapid PCA platform and methodologies further developed in this work offer an advantageous pathway for the fast and accurate synthesis of genes. In comparison to PCA with traditional heat-block cyclers (Stemmer and others 1995), PCA synthesis times are reduced from typically from several hours down to less than 30 minutes (for two thermocycling protocols comprising 2-step PCA with follow-up PCR amplification). The reduction in run-times can be considered more than just a speed novelty. As noted by others, fast thermocycling protocols lead to a reduction in error rates attributable to thermal damage (Pienaar and others 2006). Another advantage of rapid PCR over traditional PCR is that the short annealing times used in the former can provide higher fidelity of hybridization/annealing events (Wittwer and Garling 1991). An increase in the accuracy of synthesis implies that less downstream efforts are needed, such as site-directed mutagenesis or cloning/sequencing screening efforts (Kodumal and others 2004). Lastly, rapid PCR presents advantages over longer runtimes when optimization of reaction conditions are required since testing of different conditions can be completed relatively quickly.

While rapid PCA does provide these advantages, it should be noted that PCA with conventional block cyclers has been routinely used for successful synthesis of genes

(Czar and others 2009). Recently, PCA was utilized to construct small building blocks used to assemble the complete 582,970 bp genome for *Mycoplasma genitalium* by recombination techniques (Gibson and others 2008). It was obviously not the objective, nor fiscally conscionable, to compete in a race for length. Currently, PCA is a primary synthesis technique for sequences ranging from several hundred bp to several kb—above this level typically requires pairing with recombination techniques. Thus, the primary concern was to demonstrate rapid PCA within this stratum as a robust and accurate technique.

In this work, rapid PCA was utilized for the synthesis of seven different gene sequences ranging from 612 bp to 3.8 kb in length, not including shorter sequence fragments therein. The variety of sequences and relative ease by which these were assembled demonstrate the robustness of the approach. In most instances, a 2-step PCA with follow-up PCR process was employed. In Chapter 3, a rapid integrated PCA-PCR approach was developed in which genes up to approximately 1 kb in length (starting from ~ 40 bp assembly oligos) could be synthesized in a single reaction. In Chapter 5, a novel convergent approach for the rapid synthesis of longer gene sequences was disclosed. In the convergent technique, shorter sequence fragments are assembled by 2-step rapid PCA-PCR, and then stitched together in a subsequent reaction using rapid integrated PCA-PCR. A 3.8 kb mWAP-hPC gene sequence—the largest constructed to date by rapid thermocycling methods—was stitched from either two or four first-round synthetic fragments.

6.1.2 Impact of assembly oligo length

As part of the research scope for this work, the length of the assembly oligos was limited to roughly 40 bp each. The rationale was that shorter assembly oligos possess fewer errors, as the error rate from automated phosphoramidite chemistry is around 0.3% (McClain and others 1986). Thus, limiting the assembly oligo length aided in preservation of the sequence quality without requiring any extraneous quality control. Even still, Smith and others (2003) ascertained that each 42-mer used in their synthesis contained ~50% truncated species, implying a higher error rate. Additionally, hybridization among oligos was specified so that no gaps were allowed between neighboring oligos on the same strand. This oligo design strategy allowed for some discrimination at annealing to exclude oligos with deletions present, albeit resulting in a highly oligo-dense PCA reaction.

While the short assembly oligo approach with standard desalting purification proved adequate for the most part in this work, revisiting of the assembly oligo length is recommended. In many instances, especially for the chimeric genes, deletions were present in the sequencing data. While stringent PCA hybridization should have limited the incorporation of deletions, the quality of oligo building blocks does impact the end results. Linshiz and others (2008) provide a unique recursive strategy for the synthesis of DNA from error-prone oligos that is akin to the convergent approach. Instead of “one-pot” assembly and cloning, the divide and conquer recursion sought to optimize synthesis steps to generate an error-free clone. However, the recursive technique was only demonstrated for a trivial 768 bp *GFP* sequence. Smith and others (2003) increased the

quality of the oligo pool by first conducting gel purification prior to ligation/PCA to construct a 5,386 bp Φ X174 bacteriophage.

If longer assembly oligos may be employed, it is expected that longer genes may be more readily constructed. For example, a 4 kb gene synthesized from ~40 bp assembly oligos (20 bp hybridization units) would require roughly 200 oligos at the current specification. If 100-mers can be employed, the number of oligos needed would be reduced to approximately 50 oligos if 20 bp overlaps are used. The reservation arising with the latter is that gaps will be present between hybridization sections—there will be no hybridization screening process for elimination of nucleotide deletions in these regions.

Using the trivial equation a 40-
mer synthesis will be 67% full-length, whereas a 100-mer has a mere 36% yield, assuming a 99% efficiency. It is therefore certain that standard desalting with longer oligos should not be used directly due to the majority of oligos being in the truncated form; the oligos must be of the high purity with truncated species removed. Integrated DNA Technologies currently offers DNA oligos up to 200bp in length (“ultramers”, Allen and others 2009) with optional PAGE purification to remove truncated species. If the quality of 100-mers (or even longer) can be assured to a reasonable level, then this approach becomes feasible. Returning to the 4 kb example, a 50-oligo PCA reaction to directly construct the full-length gene would seem within reason.

6.1.3 Oligo design flexibility with longer assembly oligos

Another benefit of utilizing longer assembly oligos is the design flexibility. In the no-gap approach, hybridization units were interdependent. Adjusting the length of one hybridization unit affected the others, resulting in the need for global optimization strategies of the complete sequence hybridization map in order to match melting temperatures (Rouillard and others 2004; Louw and others 2010). Condoning gaps allows for a more flexible design strategy since each hybridization overlap can essentially be designed independently. Nucleotides can be added or subtracted at whim for each hybridization region to obtain desired T_m (melting temperature) values. This is expected to result in a more straightforward design of oligos with improved matching of T_m values. One can proceed in a more linear progression from one hybridization unit to the next. An illustration of an example oligo map is given in Figure 6-1. Note that the oligo lengths need not be exactly the same length, and that typically one oligo (i.e. towards the 3' end) may be much shorter than others dependent on the sequence and desired oligo lengths.

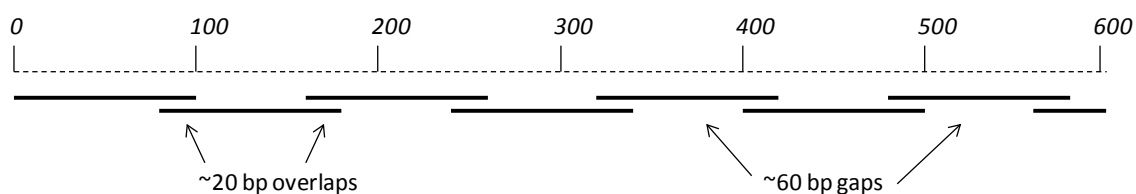


Figure 6-1 Oligo map for an example 600 bp sequence employing 100-mers with gaps allowed.

~20 bp overlaps can be increased or decreased for optimal melting temperatures. Note the short oligo at the 3' end.

6.1.4 Economic considerations

PAGE purification is known to be the method of choice for purifying longer oligos with unit-base resolution, overcoming a key technical hurdle. The other consideration is the economic cost of the longer oligo approach. The current cost of an ultramer from IDT is \$0.75/base on a 4 nmole scale compared to \$0.35/base for standard DNA oligos. The scale is more than sufficient, and even smaller miniaturization of oligo synthesis amount is an outstanding economic barrier for gene synthesis (Tian and other 2004). Note that the presence of gaps significantly reduces the number of total nucleotides synthesized by phosphoramidite chemistry (5,000 bp vs 8,000 bp for a 4 kb sequence). 200-mers would require approximately 4,500 nucleotides. The other economic element is the cost of the PAGE purification—currently retailing at \$105 (for a single oligo, the price of a larger volume order is expected to be discounted significantly). Smith and others (2003) utilized a pooled oligo approach that also could be employed to reduce purification costs. Thus, while a premium will be paid for the longer oligo approach, any cost premium would be entirely justifiable if labor costs are reduced in downstream steps, and most importantly produces the desired construct more efficiently.

The longer assembly oligo approach is not new—IDT offers gene synthesis services using their ultramers. With a number of biotech companies offering gene synthesis, a price war in the gene synthesis industry has been raging (Czar and others 2009). Advertised prices are currently around \$0.50 / bp, usually including insertion into a plasmid vector, cloning and sequencing confirmation. With IDT for example, sequences ranging from 400 bp to 1.5 kb are currently listed at a base price of \$0.55 / bp including

the extras listed. Note that this is less than what one can purchase only the ultramers themselves for at retail (\$0.75 / bp, and not too much more than regular \$0.35 / bp oligos). Due to the vertical integration of processes involved in the synthesis steps by these larger companies and the aggressive pricing, individual researchers may be discouraged that they will incur higher costs to perform the synthesis themselves since the oligos themselves alone are priced comparably to a pre-assembled sequence. However, additional charges are more often than not added for sequence complexity and length, and for many sequences these companies fail to deliver or refuse to quote altogether (Czar and others 2009).

6.1.5 Error Frequency Considerations

In this work, rapid PCA with a proof-reading KOD polymerase (Takagi and others 1997) was demonstrated as a reliable technique for obtaining error-free clones. In Chapter 2, the error rates for all four genes were less than 1 error / kb. This is compared to approximately 2-5 errors / kb usually found in conventional PCA synthesis (Mamedov and others 2007). In the case of the 1.4 kb chimeric genes (see Appendix 5B), error rates were not calculated; yet error-free clones for each of the two variants was found after screening of only ten clones each. Since error rates typically limit conventional PCA synthesis length to 500-800 bp before a screening step is recommended, a small decrease in error frequency can drastically reduce the cloning requirements (Kodumal and others 2004). Next-generation sequencing methods such as emulsion PCR (Marguiles and others 2005) that employ *in vitro* clonal amplification rather than molecular cloning may

also reduce gene synthesis labor costs. While initial investigation into the 3.8 kb failed to provide error-free clones in the fragments tested due to cloning issues, rapid PCA can be viewed as a preferred technique for synthesis of genes.

The importance of gene synthesis error rate reduction cannot be over emphasized. Kodumal and others (2004) conducted an in-depth study of gene synthesis error rates in which over 1 million bp were sequenced. In confirmation with other conventional PCA results, the error frequency was about 0.2% (1 error per 500 bp) or higher. Rapid PCA synthesis in this work (see Chapter 2) and in Mamedov and others (2007) has yielded error frequencies of about 0.1% (1 error per 1 kb). Following Kodumal and others (2004), the error frequency in gene synthesis can be described as a Poisson distribution:

$$P(x, \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

Where λ is the average number of outcomes (i.e. the error frequency, EF), t is the interval (herein the gene length, L), and x is the # of outcomes. Since we are interested in determining the fraction of clones, F , that are error-free ($x=0$), the equation reduces to simply $F = e^{-(EF \times L)}$. A graph of the fraction correct vs. gene length at different error frequencies is shown in Figure 6-2.

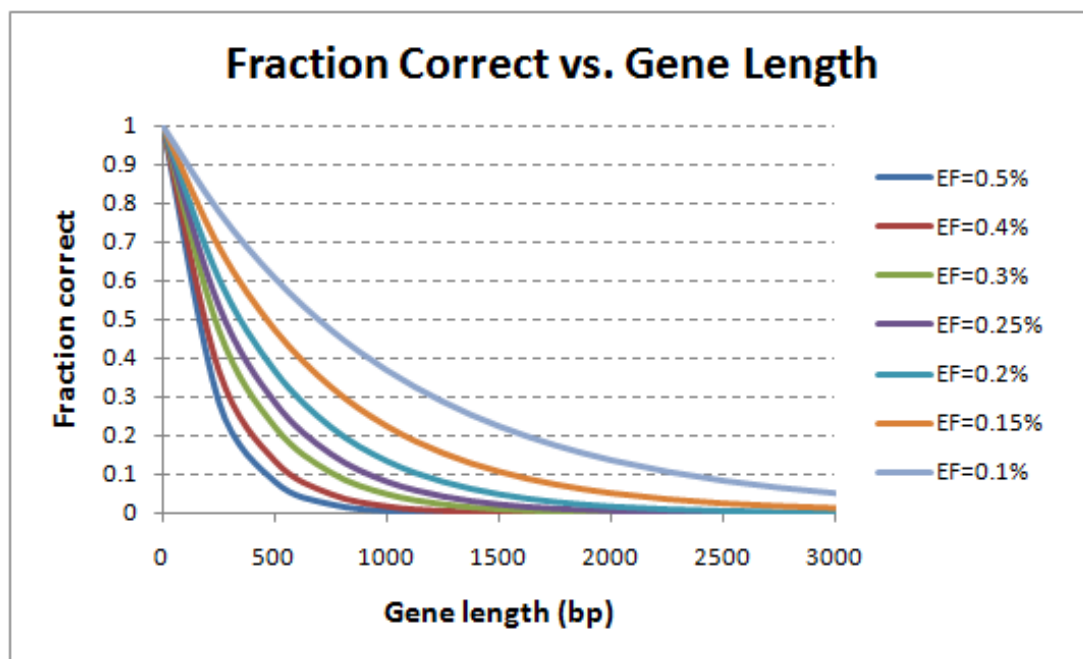


Figure 6-2 Fraction clones correct vs. gene length at different error frequencies as estimated by the Poisson distribution.

The dramatic effect of a reduction in error rates is obvious from Figure 6-2. In conventional PCA with an EF ~ 0.2 , Kodumal and others (2004) limited synthon fragments to 500-800 bp to effectively screen error-free clones from unpurified oligos. With a reduction by rapid PCA techniques to $\sim 0.1\%$, error-free clones can be effectively isolated with minimal screening up to about 1.5 kb with the employed unpurified oligos. In this work, three different genes (TM-1, chimera #1, chimera #2) of ~ 1.4 to 1.5 kb in length yielded error-free clones from just ten picked colonies, supporting the rapid PCA fidelity.

6.1.6 Future Work with rapid PCA

In general, future work with rapid PCA is recommended to focus on the above proposed use of longer assembly oligos. The objectives therein would be three-fold: 1) investigate the robustness of the process steps, particularly with respect to elimination of truncated assembly oligos and deletion error avoidance, 2) demonstrate the capacity to synthesize longer constructs in each PCA reaction for both single and convergent strategies, and 3) statistical comparison of the error rates when employing either rapid or conventional thermocyclers. Such efforts could further cement the superiority of rapid PCA for fast and accurate synthesis of genes many kb in length with minimal cloning burden to isolate an error-free clone.

With regards to the specific gene syntheses presented in this work, there are several areas for future work. For the error-free clones of the two chimeric gene sequences isolated in Appendix 5B, the next step is to conduct protein expression of the chimeric genes and activity assays of the EPCR and TM-2 domains to determine if the chimeric gene yields improved activity. Of course, success of protein expression and particularly protein folding is not a guarantee especially in the case of the “flipped” TM-2 region in chimera #1. The determination of the optimal linker region for synergistic EPCR and TM activity by construction of variants would also serve as an interesting endeavor. In the case of the 3.8 kb gene sequence pursued in Chapter 5, more work is recommended to generate an error-free clone. The initial step would be to attempt cloning of the four ~950 bp part fragments in the hopes that cloning proceeds properly. Given the difficulties in cloning of the longer sequences it may very well turn out that problems arise again, but at the very least provide some direction for the next steps. If troublesome regions become

apparent that cannot be readily fixed by site-directed mutagenesis or the recursive strategy of Linshiz and others (2008), then the recommendation would be to pursue the longer oligo assembly strategy from 100+-mers.

6.2 Affinity purification of synthetic genes

6.2.1 Summary

The companion affinity workstation developed in this work facilitated the purification of full-length synthetic genes from unwanted assembly oligos and incomplete reaction products. Labels were incorporated into the gene during the PCR amplification process by the use of primers labeled on the 5' ends. Since only the full-length gene is expected to contain both labels, the serial immunoaffinity purification resulted in only the full-length product in the final effluent (acknowledging that some non-specificity may exist). The immunoaffinity purification was meant to serve as an alternative to manually-intensive (and possibly error-generating) gel purification typically performed prior to cloning and sequencing.

In Chapter 2, four different genes were tested against a serial affinity arrangement with anti-BrdU and Streptavidin (genes labeled with iminobiotin) columns. A bread-board hardware mock-up was used to implement the serial purification. The 929 bp β -lactamase gene clearly proved the concept, while the 1548 bp TM-1 gene was also successfully purified but at low levels. The short puc19 ori and EPCR genes (612 and

648 bp, respectively) were also successfully purified to some extent, but slight issues with labeled PCR amplification prevented clear analysis of the affinity purifications. Since the anti-BrdU column did demonstrate some non-specificity, in Chapter 4 a next generation platform was developed. FITC and digoxigenin proved to be the best candidates compared to BrdU, TRITC, and desthiobiotin options. The serial anti-FITC and anti-digoxigenin immunoaffinity purification was implemented in an improved hardware and process. This final workstation configuration was validated using a 1.4 kb chimeric gene sequence. In Chapter 5, this immunoaffinity workstation was employed for the purification of mWAP-hPC gene sequences—a mixture of ~950 bp fragments, a mixture of ~1.9 kb fragments, and the full 3.8 kb gene sequence.

6.2.2 Impact of labeling DNA

While the immunoaffinity purification did show isolation of full-length products, there are several considerations for the long-term adoption of this technique. First, one drawback of the affinity purification is that the necessary labels are present in the isolated product. As such, typically another PCR amplification thereon with regular amplification primers was performed for cloning functionality, thereby obviating much of the benefit of the affinity purification. Now, insertion of labeled product into a cloning vector has not been attempted to date, but the ligation reaction is expected to be inhibited by the label presence on the 5' end. Alternatively, shifting of the labels into a middle location of the primer may help or a simple cleavage reaction to remove the labels (i.e. restriction enzyme cutting of the ends) can be envisaged.

6.2.3 Affinity purification yield

Likely the largest short-coming of the current affinity workstation was its relatively low yield. The flow-through purification performed relatively well for genes under 1kb, but as the length of the sequence increased the yield was substantially reduced. Even batch purification of longer gene sequences failed to produce stellar binding. This is indicative of a larger K_d (dissociation constant) as the gene length increased. The other primary parameters for future optimization with the existing FITC and dioxigenin columns to increase the yield are the antibody to label ratio, the ligand concentration, and residence time in the flow-through system. It would be advantageous to decrease the processing time in the workstation given that slow flow rates were employed to date.

One possible explanation for low binding amounts in the affinity purification is the size of the target DNA. In Chapter 5, it appeared that as the gene length increased, the binding amount decreased. Taking the average molecular weight of each base pair at 660 Da, a 1kb gene is approximately 660 kDa which is fairly large by protein purification standards. For example, BSA and fibrinogen are two commonly purified proteins with molecular weights of approximately 67 kDa and 340 kDa, respectively. While the immunoaffinity purification provided fairly good yield of products up to 1kb, the binding amounts were noticeably reduced for the 1.9 kb (~1254 kDa) and no product visible on the gel for the 3.8 kb gene (~2508 kDa). It is believed that steric hindrance reduced the binding as the gene length increased due to limited accessibility of the labels and ligand binding sites. Thus, it is not difficult to rationalize the low yields obtained for longer genes, and thus limiting the recommended maximum gene length in the current

purification system to ~ 1 kb. Of course, this is not to say that higher molecular weight products cannot be successfully purified in high yield after continued optimization of the existing chromatography. Also note that multiplication of yields from the consecutive columns further complicated matters by reduction of overall yield in the final effluent.

As gel electrophoresis was a necessary, but highly unreliable quantitation, detection method in the initial workstation development, calculations of the disassociation constant (K_d) was not worthwhile. Additionally, the impure nature of the synthesis mixtures did not allow for direct K_d measurements with respect to gene length. Given the preliminary implications of gene length on binding, a thorough quantitative investigation of K_d is warranted. It is apparent that the antibodies possessed high affinity for the free label from the use of biospecific competitive binding in the elution step. Tethering of a DNA chain of specific lengths impacts the epitope binding site either directly from the attachment or via steric hindrance. Quantitative assay schemes applied against the free label and labeled DNA of known pure length would be advantageous to fully understand the phenomena.

6.2.4 Economic considerations

Another barrier to adoption of the immunoaffinity purification is the cost-effectiveness of the approach. In this work, primers unique to each synthetic gene were employed at an approximate cost of \$70 each from IBA GmbH (Göttingen, Germany). The average cost of each purification was about \$17 per run, assuming the columns can be acceptably regenerated for three total uses (not including labeled primers or labor costs). The bulk

of the economics is dominated by the cost of the FITC and digoxigenin antibodies. Of course, more column regenerations would significantly amortize the cost per test. For example, 10 column uses before replacement would reduce the cost to around \$5 per test. Additionally, miniaturization of the purification scale, particularly the feed load, would further reduce the amount of antibodies used per test. However, primer labeling still adds unattractive costs to the purification.

6.2.5 Future work with generic labeled primers

A possible solution to reduce purification costs is to tailor the synthesis of genes to use generic labeled primers that can be used for all synthetic genes by adding flanking 5' and 3' sequences that are not specific to the gene. An option is the His-tag-Stop-NotI sequence employed on the 3' end of the chimeric gene sequences (see Appendix 5B). On the 5' end, an additional flanker could consist of the XhoI plus additional outer nucleotides (similar for the 3' primer if the His-tag is not desired). The restriction enzyme sites facilitate direct cleavage of the labels post-purification for insertion into a complementary cloning vector (i.e. pPICZαA for *Pichia pastoris*). An example of the generic flanker regions is given in Figure 6-3 without the His-tag as part of the generic 3' sequence. A variety of generic primer options can be envisioned to select from depending on the desired restriction enzyme sites and avoidance of internal sequence commonality in the dummy sequence flankers—the objective is just to stream-line and reduce costs of the purification.

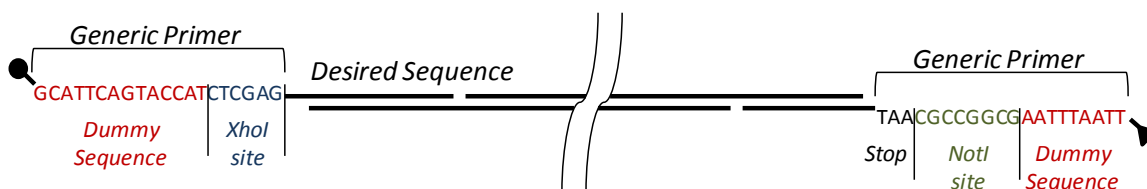


Figure 6-3 The addition of dummy 5' and 3' flanker sequences allows for generic labeled primers to be used for many different synthetic genes.

Example XhoI and NotI sites allow for cleavage of the labels prior to vector insertion.

6.2.6 Future work with FITC detection

To date, approximately 100 μ l of product (combined four separate PCR reaction volumes) in the feed was in part a necessity as gel electrophoresis was the specified means for detection. It is known that absorbance (UV/Vis) detectors are generally limited in sensitivity to about an LOD of 10^{-8} to 10^{-7} M (Moser and Hage 2006), precluding the use of these detectors in the current workstation. With the labeled PCR concentration at 700 nM, it can be expected that the full-length PCR product is only a fraction of this amount. Further dilution upon feed mixing, reduced yield through two columns, and dilution in final effluent, is estimated to put the final concentration at $\sim 10^{-9}$ or less.

One of the reasons that FITC was chosen for use in the final configuration of the workstation was for its fluorescent characteristics and subsequent LOD when used with fluorescent detectors (Moser and Hage 2006). Various types of fluorescence detectors and flow cells capable of detecting the amounts observed in the workstation could thereby be used to track the FITC-labeled products. While laborious sampling of effluent

fractions and gel analysis thereon was performed in the development, only monitoring of the effluent from the digoxigenin column would be necessary to automatically switch the corresponding 3-way valve to collect the final product. A sensitive fluorescent detection scheme would support reduction of feed loads and column sizing, important to the economics of the purification step.

6.2.7 Future work with other affinity columns

Alternatives to the FITC and digoxigenin serial purification may be a worthwhile future endeavor. As noted in Chapter 4, any number of affinity-based separations may prove suitable for purifying hapten-tagged synthetic genes in addition to antibody-antigen systems. Also discussed briefly in Chapter 4, the iminobiotin label with streptavidin binding may be particularly advantageous as the second stage in a serial purification arrangement with an antibody-based system. Unlike Chapter 2, future work would center on testing both columns directly in series or even embodied as a single column packed with two distinct matrix layers of the respective ligands. There are two aspects to this arrangement that need explored: 1) the labeled product will successfully bind in the first stage under streptavidin elution conditions so that all products not captured by the first column do not bind to the streptavidin (i.e. first column binding buffer at pH 4.0), and 2) the elution buffer from the first column is at the proper binding pH for the streptavidin (i.e. pH 10). After sufficient washing of the first stage antibody column, a transition step is likely to be recommended to transition the pH of the streptavidin to 10 prior to FITC elution and product loading to the streptavidin. The process is illustrated in Figure 6-4,

noting that a pH transition step between buffers 1 and 2 and a regeneration step are not shown. Also, the optimal salt concentrations for each buffer are not listed.

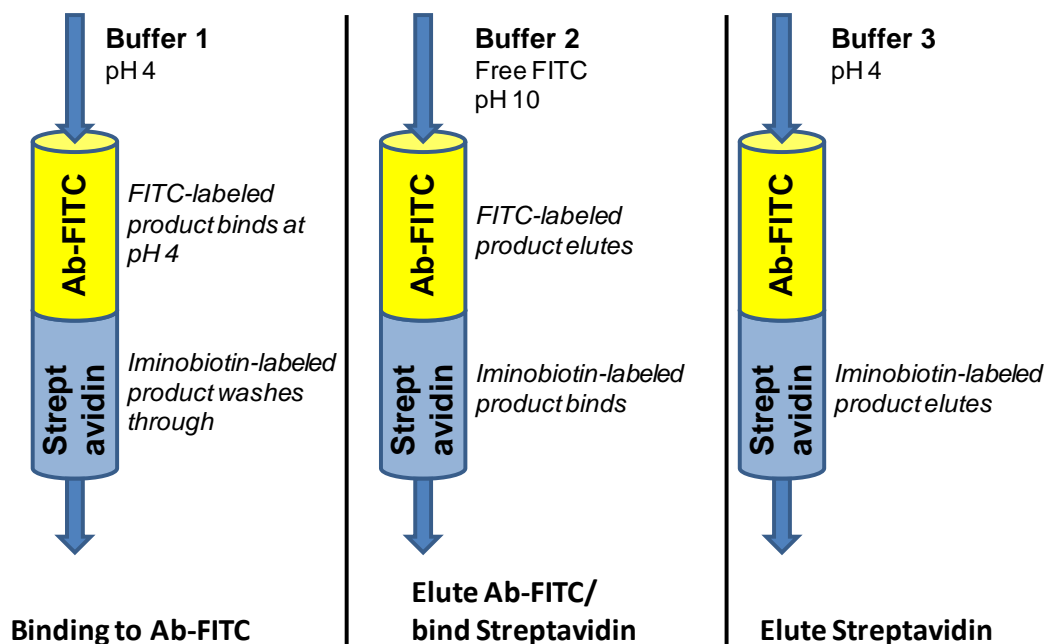


Figure 6-4 Illustration of an integrated Ab-FITC/streptavidin column.

Refer to the text for a description of the process.

6.2.8 Broader applications of purification of labeled DNA

Even if the serial immunoaffinity purification does not establish strong roots for isolating synthetic genes, the incorporation of labels via PCR amplification primers for subsequent isolation of DNA may of generic benefit for a myriad of other uses. For any application that employs the hybridization of an oligo to a target sequence, the presence of a label on the hybridized oligo facilitates affinity purification of the target sequence. For instance, simple purification or detection of PCR amplicons generated with labeled primers may be performed. Another obvious application of the labeled-DNA purification is the screening

of cDNA libraries (Tagle and others 1993). In that light, individual label purification is sufficient rather than the serial arrangement with two different target labels. Even for the isolation of synthetic genes it is not an absolute truth that complete resolution of the full-length product from unwanted oligos is needed for downstream processing—the base objective is to increase the signal-to-noise ratio. Sole utilization of just one of the purification steps for synthetic genes may be adequate since all inner intermediate products that have not been assembled to completion in the chosen strand direction would be washed away, representing the vast majority of unwanted products. It is easy to see that the developed immunoaffinity workstation can be modified with little effort to a single-stage platform.

6.3 Closing thoughts

The capacity to create genes *de novo* is undoubtedly a major advancement in biotechnology. While conceptually there are no natural limitations on the exact sequences that may be constructed, many technical challenges abound. The research presented herein served to demonstrate rapid PCA and subsequent affinity purification of synthetic genes as valuable tools for researchers. A rapid thermocycler can be characterized as generally superior to the use of conventional instrumentation for PCA-based gene synthesis, particularly advantageous due to high fidelity with reduced error rates. The affinity purification of synthesized DNA to isolate the full-length gene from its reaction background has been proven a useful alternative to gel purification. Still, researchers are advised to select among available techniques for particular applications given the infinite diversity of sequences and lengths that one may desire. Especially as

the synthesis projects grow in sequence and length complexity, there will become a growing need to combine rapid PCA and affinity purification with other techniques to tailor strategies that give the highest probability of success.

6.4 References

- Allen, S.D, T.M. Luebke, and S.D. Rose. Ultramers- The longest oligonucleotides available with mass spectrometry. Taken from http://www.idtdna.com/home/Ultramers_Technical_Report_IDT.pdf (accessed January 12, 2009)
- Czar, M.J., J.C. Anderson, J.S. Bader, J. Peccoud. Gene Synthesis demystified. *Trends Biotechnol.* **27**: 63-72 (2009)
- Gibson, D.G., et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220 (2008)
- Kodumal, S.J., K.G. Patel, R. Reid, H.G. Menzella, M. Welch, and D.V. Santi. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 5573-5578 (2004)
- Linshiz, G., T.B. Yehezkel, S. Kaplan, I. Gronau, S. Ravid, R. Adar, and E. Shapiro. Recursive construction of perfect DNA molecules from imperfect oligonucleotides. *Mol. Syst. Biol.* **4**: 191. (2008)
- Louw, T.M., S.E. Whitney, J.R. TerMaat, E. Pienaar, and H.J. Viljoen. Oligonucleotide optimization for DNA synthesis. *AIChE J.* Article first published online: 29 SEP 2010. DOI: 10.1002/aic.1241

- Mamedov, T.G., N.V. Padhye, H. Viljoen, and A. Subramanian. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387 (2007)
- Margulies, M. *et al.* Genome Sequencing in open microfabricated high density picoliter reactors. *Nature.* **437**: 376-380 (2005)
- McClain, W.H., K. Foss, K.L. Mittelstadt, and J. Schneider. Variants in clones of gene-machine synthesized oligodeoxyribonucleotides. *Nucleic Acids Res.* **14**: 6770 (1986)
- Moser, A.C. and D.S. Hage. Chromatographic Immunoassays. In: Hage, D.S. Handbook of Affinity Chromatography. CRC Press. Boca Raton, FL. (2006)
- Pienaar, E., M. Theron, M. Nelson, and H.J. Viljoen, H.J., A quantitative model of error accumulation during PCR amplification. *Comput. Biol. Chem.* 30, 102-111. (2006)
- Rouillard, J.M., W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**: W176-W180 (2004).
- Smith, H.O., C.A. Hutchison, C. Pfannkock, and J.C. Venter. Generating a synthetic genome by whole genome assembly: Φ X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 100: 15440-15445 (2003)
- Stemmer, W.P.C., A. Cramer, K.D. Ha, T.M. Brennan, and H.L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene.* **164**: 49-53 (1995)
- Tagle, D.A., M. Swaroop, M. Lovett, and F.S. Collins. Magnetic bead capture of expressed sequences encoded within large genomic segments. *Nature.* **361**: 751-753 (1993).

- Takagi, M., M. Nishioka, H. Kakihara, M. Kitabayashi, H. Inoue, B. Kawakami, M. Oka, and T. Imanaka, Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl Environ Microbiol.*, **63**: 4504-10 (1997)
- Tian, J., H. Gong, N. Sheng, X. Zhou, and E. Gulari, X. Gao, and G. Church. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*. **432**: 1050-1054. (2004)
- Wittwer, C.T., and D.J. Garling. Rapid cycle DNA amplification: time and temperature optimization. *Biotechniques*. **10**: 76-83. (1991)

[illegible]

Note 2: n at position represents ambiguous nucleotide

5A.2 Oligo Hybridization Map

GAATTCCTTCTACTGCTAAAACAG-GGCGGGAGGAGTCCAGAG CCCTGCCACTGGGTGCAG-AACATGAAGACCCCTTGA
CTTAAGAAAGTGACGATTTTGTC CCGCCTCTCAGGTCTC-GGGACGGTGACCCACGTC TTGTACTTCTGGGGAAT
59.82 67.70 69.62 59.381

GTGGAAAGGGTTATACAGC-TGGACAGTGGTGGCGCAC ACCTTTAATCCCAGCACT-CGGGAGGCAGAGGCAGAC G
-CACCTTTCCCCAATATGTCG ACCTGTCACCACCGCTG-TGGAAATTAGGGTCGTGA GCCCTCCGTCTCCGTCTG-C
62.627 69.41 59.464 68.64 5

GATCTCTGAGTTCGAGG-CCAGCCTGGTCTACAGAA TGAGTTCGGGACAGCCA-GGGCCACACAGAGAACT CTTGT
CTAGAGACTCAAGCTCC GGTCGGACCAGATGTCTT-ACTCAAGGCCCTGTCCGT CCCGTGTGTCTCTTTGA-GAACA
9.81 62.017 67.19 63.310 59.20

CTCGAAAAACCAAAA-AAAAACAAAAAGGAAAGGGG TTACACAACAGAGACTCAGG-TCACAGCTACCCATCACA CA
GAGCTTTTGGTTTT TTTTGTCTTTTCTCTTCCCC-AATGTGTGTCTCTGAGTCC AGTGTGATGGGTAGTGT-GT
60.494 61.16 61.618 60

CAGGATACACATACAAAGG-TGTTACAGGCAGATGAG GAACGAGGAGAAGGGGCT-CAAGCAAGGGCCTAAAGT TTC
GTCTATGTGTATGTTTCC ACAAGTGTCCGTCTACTC-CTTGCTCCTCTTCCCCGA GTTCGTTCGGGATTTC-AAG
.90 60.61 64.86 61.27 59.

TTTTTTTTTTTTTCTTCTTTTTT-TTTTTCCCTGTGGCCTAG AGTTTCAAGAGGCTGAGG-ACCTAGGCATGAACCAAG
AAAAAAAAAAAAAGAAGAAAAA AAAAAGGGACACCGGATC-TCAAAGTTCTCCGACTCC TGGATCCGTACTTGGTTC
00 60.54 60.215 60.29

AGGGGCCAAACCACTTCA-AGAAGCAGGGGGTAGCAG CAGAATCTCACTATCAGCCT-TGAGCACAGCTGGGAAGG A
-TCCCCGGTTTGTGAAGT TCTTCGTCCCCCATCGTC-GTCTTAGAGTGATAGTCGGA ACTCGTGTGACCCCTTC-T
65.39 64.90 60.803 65.82 6

GATCCATGGAAACAACCA-AGAAAGAGCTGAAAGGGG CTGGAGAGATGGCTCAGC-AGTTAAGAGCACTGAGTGC TCT
CTAGGTACCTTTGTTGGT TCTTCTCGACTTTCCCC-GACCTCTCTACCGAGTCG TCAATTCTCGTACTCAGC-AGA
0.689 60.215 63.26 61.4 60.

TCCGAAGTCCTAGATT-CAATCCCAGCAACCATATG GTGGGTACAACCATCTG-TAATGAGATCTGATGCCCT CTT
AGGCTTCAGGATCTAAG TTTAGGGTCGTTGGTATAC-CACCCAGTGTGGTAGAC ATTACTCTAGACTACGGGA-GAA
43 59.59 61.658 59.2 62.

CTGGTGTGTCTGAAGAC-AGCTACAGTGTACTTATGTATAAT AAATAAATAAATCTTTAGAAAGGGAG-GGGGGGAGAG
GACCACACAGACTTCTG TCGATGTCACATGAATACATATTA-TTTATTTATTTAGAAATCTTCCCTC CCCCCCTCTC
44 60.27 59.74 63.27

AGAGAGAG AGAGAGAGAGAGAGAGAGAG-AGAGAGAGAGAGAGAGAGAG AGAGCTGGAAGAGGGAGA-TCTGGGAAGT
TCTCTCTC-TCTCTCTCTCTCTCTCTCTC TCTCTCTCTCTCTCTCTCTC-TCTCGACCTTCTCCCTCT AGACCTTCA
60.43 60.43 62.7 65.50

CTGCTGGC TTTATATGCTGACCATATATAGTCA-CCTGTGTTTACACACTGTG CTCATCACTTTGAAATCTCAG-TGG
GACGACCG-AAATATACGACTGGTATATATCAGT GGACACAAATGTGTGACAC-GAGTAGTGAACCTTAGAGTC ACC
60.81 60.36 59.10 60.

TTTCTTCTTTGAGCCT GTGTCTGTAAGTTCACCAG-GAGAGTGGTACATAGGCAA GAATAACAGCCAGTGGGC-ATAG
AAAGAAGAACTCGGA-CACAGACATTCAAGTGGT CTTCTACCATGTATCCGTT-CTTATTGTGCGTCACCCG TATC
9 59.4 60.21 62.13 60.9

GACACAGAGTGCATG GGCCCCAGCAAGACTGTA-GAGAGAACAGAGCTCTGG CTCCTAAGACATAGGGCC-TTCTGGG
CTGTGTCTCACGTAC-CCGGGGTCGTTCTGACAT CTCTCTTGTCTCGAGACC-GAGGATTCTGTATCCCG AAGACCC
0 65.52 60.054 59.479 60.05

AAACTCAAGCA GCCAAGCAACCTAGCCA-GCCCTTCTCTGGTGGCCC TCCTTCTGTTCCAGCAAA-GGCGAAATGG
TTTGAGTTCGT-CGGTTCGTTGGGATCGGT CGGGAAGGACCACCGGG-AGGAAGACAAGGTCGTTT CCGCCTTTACC
66.55 69.74 60.05 64.8

GAACAGG GGTGGAAGCAGAGCATTG-GCAGAGCATAGGTATGACT TAGTCTTGACTAACACAAGC-ATGGCAGTAGCC
CTTGCTCC-CCACCTTCGTCTCGTAAC CGTCTCGTATCCATACTGA-ATCAGAACTGATTGTGTTTCG TACCGTCATCGG
62.442 60.368 59.3 62.816

TGACAG TGGCCTAAATGTGGGAT-GACTGCCTTAGATGGGGA TGACTGCCTTAGATGGGC-ATGACTGCCTTAGATG
ACTGTC-ACCGGATTACACCCCTA CTGACGGAATCTACCCCT-ACTGACGGAATCTACCCG TACTGACGGAATCTAC
62.0 61.78 62.82 59.72

GG GATGACTGCCTTAGATGGG-GATGACTGCCTTAGATGGA ACAACAAACATCTATGGGC-ATGCTGTGGAACACTGG
 CC-CTACTGACGGAATCTACCC CTACTGACGGAATCTACCT-TGTTGTTTGTAGATACCCG TACGACACCTTGTGACC
 61.36 60.29 59.82 64.40

C CCACACACGGAACCTGAAG-GCACTGGCAATTTCATA GGGCAGTTAAACCTAAAAGC-ATGCTCACACTCAACAGG
 G-GGTGTGTGCCTTGACTTC CGTGACCGTTAAAGGTAT-CCCGTCAATTTGGATTTTCG TACGAGTGTGAGTTGTCC-
 61.9 59.32 61.58 60.94

CTGCCGGAATCTCATGAG-ACACCTGGAATAGACGAAT GTAGAAACAGAGCAGAGAGT-TGGTTGCCAAGGTCTGGG G
 GACGGCCTTAGAGTACTC TGTGGACCTTATCTGCTTA-CATCTTTGTCTCGTCTCTCA ACCAACGGTTCAGACCC-C
 61.24 59.822 60.94 66. 6

GCTCAGAGGACAAGCAA-GAGGCGCGGCTTTCCTTT GGGGCTGGCATGAAAGGA-AATATCGAGGTTACAGCCT GAGA
 CGAGTCTCCTGTTCGTT CTCCGCGCCGAAAGGAAA-CCCCGACCGTACTTTCCT TTATAGTCCAATGTCGGA-CTCT
 3.36 67.48 66.2 59.58 65.2

GGGCTTCCCCTGAC-ACTTCGTATTCAAAGAGGC CATGGGCACCACTGAAGA-CAAAGGAGTATGGCCTGC ACCACAG
 CCCGAAGGGGACTG TGAAGCATAAGTTCTCCG-GTACCCGTGGTCACTTCT GTTTCCTCATACCGGACG-TGGTGTCT
 6 59.22 63.39 61.820 63.641

GCTGGCNCCTGA-CAGTCAGTAAGCACACAGT CACTCTGGGTTCATCCCAT-CCCCTTCCTTGCAAGAGA AATCAAGGAA
 CGACCGNGACT GTCAGTCATTCTGTGTCA-GTGAGACCCAGTAGGGTA GGGGAAGGAACGTTCTCT-TTAGTTCCTT
 61.35 62.1 62.687 60.76

ATGTCCCCGA-GAACAATGGGGCACAGTG CCAGCAGGACATCTCTTC-CTGCCCATGGCACCCCTTT GGCACGGTATGGG
 TACAGGGCT CTGTGTACCCCGTGTAC-GGTCGTCTGTAGAGAAG GACGGGTACCGTGGGAAA-CCGTGCCATACCC
 62.6 60.77 66.94 68.78

CCCTT-CTGGGAAGGTGGCCTTCC AAATTGCTCTGCACAGGC-AGCTCCTTTTCAATGTATGC CCGACACTCTCTACA
 GGGAA GACCCTTCCACCGGAAGG-TTTAACGAGACGTGTCCG TCGAGGAAAAGTTACATACG-GGCTGTGAGAGATGT
 66.25 62.69 60.724 60.13

TGG-AGCAAGCGCCTCCACACT CTTAGAAGAATTTTAGAAAACCTCC-AGAAAAGCACCAGGAGAA GTCACCCCTCAGAT
 ACC TCGTTTCGCGGAGGTGTGA-GAATCTTCTTAAATCTTTTGAGG TCTTTTCGTGGTCTCTT-CAGTGGGAGTCTA
 68.53 59.10 59.719 60.45

GTAGC-CCGGAAGGTGAGCCTTGCTC AAAACCTCCTGTCTTGTCTT-CTATGTGACCTGTACAAATTG GAGCTCAGAA
 CATCG GGCCTGAGCTCGGAACGAG-TTTTGGAGGACAGAACAAA GATACACTGGACATGTTTAAAC-CTCAGACTT
 69.29 59.98 60.143 60.89

TTGCCTTTG-TCTGTGATGGGTTCACACC CAACCACTCAAAGTGACAC-TTGTACATTTGTCACTGATC CTATTTCT
 AACGGAAC AGACACTACCCAAGGTTGG-GTTGGTGAGTTTCACTGTG AACAGTGTAACAGTGACTAG-GATAAAGA
 63.718 60.66 61.31 61.88

TCTTTTCTGCTCCT-TCATTTTCTCCGCTTTTATA ATAAACAAGTATTACTTTTAAAGTGG-GGGAAAAATGACCAC
 AGAAAAAGACGAGGA AGTAAAAGAGGCGAAAGTAT-TATTTGTTTATAATGAAAAATTCACC CCCTTTTTTACTGGTG
 59.401 60.148 59.89

CC TTACAAAGGACTTTTAAAAATGG-CCTCCATTGTGGCCCTTG TTCCTGGCAGCCTGGGCC-TGCTCTCTCTGTGT
 GG-AATGTTTCTGAAAAATTTTACC GGAGGTAACACCGGGAAC-AAGGACCGTCGGACCCGG ACGAGAGAGACACA
 59.96 64.51 71.54 65.46

GGCC AAGAAGGAAGTGTGTAGC-CCATCTAGAGCTGTGCCA GCCTTTCCTCCACCCCA-CCCCAAAGTCTTCCTC
 CCGG-TTCTTCTTCACAAACATCG GGTAGATCTCGACACGGT-CGGAGAAGGGGTGGGT GGGGGTTTCAGAAGGAG
 59.060 62.49 70.26 63.09

C TGTGGGTCTTTAAATGCA-TCCCAGACACTCAGACAG CCATCAGTCACTTGCCCTG-ACACCGGTACCATGTGGC A
 G-ACACCCAGGAAATTTACGT AGGGTCTGTGAGTCTGTC-GGTAGTCACTGAACGGAC TGTGGCCATGGTACACCG-T
 61.31 61.60 61.73 66.17 6

GCTCACAAGCCTCTGTC-TGTTCTGTGGCCACCTGGG GAATTTCCGGCACACCAG-CTCCTCTGACTCAGTGT TCTCC
 CGAGTGTTCGGAGGACG ACAAGCACCGGTGGACCC-CTTAAAGGCCGTGTGGTC GAGGAGAACTGAGTCACA-AGAGG
 6.18 68.86 62.74 59.15 68.8

AGCAGCGAGCGTG-CCCACAGGTGCTGCGGA TCCGCAAACGTGCCAACT-CCTTCTTGAGGAGCTCC GTCACAGCA
 TCGTCGCTCGCAC GGGTGGTCCACGACGCCT-AGGCGTTTGACGGTTGA GGAAGGACCTCCTCGAGG-CAGTGTCTGT
 71.34 66.97 64.95 67.1

GCCTGGAGC-GGGAGTGCATAGAGGAGA TCTGTGACTTCGAGGAGG-CCAAGGAAATTTTCCAAAATGT GGATGACAC
 CGGACCTCG CCCTCACGTATCTCTCT-AGACACTGAAGCTCTCTC GGTTCCTTTAAAGGTTTACA-CCTACTGTG

61.12	61.67	61.086	63.392
-------	-------	--------	--------

ACTGGCCTT-CTGGTCCAAGCACGTGCA CGGTGACCAGTGCTTGGT-CTTGCCCTTGGAGCACCC GTGCGCCAGCCTG
 TGACCGGAA GACCAGGTTCTGTGCAGCT-GCCACTGGTCACGAACCA GAACGGGAACCTCGTGGG-CACGCGGTGCGAC
 66.04 66.38 67.23 71.04

TGCTG-CGGGCACGGCACGTGCAT CGACGGCATCGGCAGCTT-CAGCTGCGACTGCCGACG CGGCTGGGAGGGCCGCT
 ACGAC GCCCCGTGCCGTGCACGTA-GCTGCCGTAGCCGTGCA GTCGACGCTGACGGCGTC-GCCGACCCTCCCGGCGA
 73.05 69.8 70.68 74.18

T-CTGCCAGCGCAGGTGAG CTTCCTCAATTGCTCGCT-GGACAACGGCGGCTGCAC GCATTACTGCCTAGAGGA-GG
 A GACGGTCGCGCTCCATC-GAAGGAGTTAACGAGCGA CCTGTTGCCGCCGACGTG-CGTAATGACGGATCTCCT CC
 69.79 61.4 70.99 59.39 75

TGGGCTGGCGGCGCTG TAGCTGTGCGCCTGGCTA-CAAGCTGGGGGACGACCT CCTGCAGTGTACCCCCG-AGTGAA
 ACCCGACCGCCGCGAC-ATCGACACGCGGACCGAT GTTCGACCCCTGCTGGA-GGACGTACAGTGGGGCG TCACCT
 .83 67.64 67.95 69.83 60.46

GTTCCCTTGTGG GAGGCCCTGGAAGCGGAT-GGAGAAGAAGCGCAGTCA CCTGAAACGAGACACAGA-AGACCAAGAA
 CAGGGGAACACC-CTCCGGGACCTTCGCCTA CCTCTTCTTCGCGTCAGT-GGACTTGTCTGTGTCT TCTGGTTCTT
 68.13 63.39 60.214 59.26

GACCAAGTA GATCCGCGGCTCATTGAT-GGGAAGATGACCAGGCGG GGAGACAGCCCTGGCAG-GTGGTCTGTCTGG
 CTGGTTTCT-CTAGGCGCCGAGTAACTA CCCTTCTACTGGTCCGCC-CCTCTGTGCGGGGACCGTC CACCGGACGACC
 64.02 66.31 68.70 65.0

ACTCA AAGAAGAAGCTGGCCTGC-GGGGCAGTGCTCATCCAC CCCTCCTGGGTGCTGACA-GCGGCCCACTGCATGGA
 TGAGT-TTCTTCTTCGACCGGACG CCCCCTCACGAGTAGGTG-GGGAGGACCCACGACTGT CGCCGGGTGACGTACCT
 64.09 66.94 67.62 69.7

T GAGTCCAAGAAGCTCCTT-GTCAGGCTTGGAGAGTAT GACCTGCGGCGCTGGGAG-AAGTGGGAGCTGGACCTG GA
 A-CTCAGGTTCTTCGAGGAA CAGTCCGAACCTCTCATA-CTGGACGCCGCGACCCCTC TTCACCCCTGACCTGGAC-CT
 59.88 59.31 72.22 65.14 61

CATCAAGGAGGTCTTCG-TCCACCCCAACTACAGCA AGAGCACCACCGACAATG-ACATCGCACTGCTGCACC TGGCC
 GTAGTTCTCCAGAAGC AGGTGGGTTGATGTCGT-TCTCGTGGTGGCTGTTAC TGTAGCGTGACGACGTGG-ACCGG
 .5 64.3 63.73 67.45 77.0

CAGCCCGCCACCC-TCTCGCAGACCATAGTGC CCATCTGCCTCCCGGACA-GCGGCCTTGCGAGAGCGCG AGCTCAATC
 GTCGGGCGGTGGG AGAGCGTCTGGTATCACG-GGTAGACGGAGGGCCTGT CGCCGGAACGTCTCGCGC-TCGAGTTAG
 62.845 67.71 73.88 69.08

AGGCCGGCC-AGGAGACCCTCGTGACGG GCTGGGGCTACCACAGCA-GCCGAGAGAAGGAGGCCA AGAGAAACCGCAC
 TCGGCCGG TCCTCTGGGAGCACTGCC-CGACCCCGATGGTGTCGT CGGCTCTCTTCTCCGGT-TCTCTTGGCGTG
 66.90 68.65 67.67 63.703

CTTCG-TCCTCAACTTCATCAAGATTC CCGTGGTCCCGCACAAATG-AGTGCAGCGAGGTCATGA GCAACATGGTGTCT
 GAAGC AGGAGTTGAAGTAGTTCTAAG-GGCACGAGGCGGTGTTAC TCACGTCGCTCCAGTACT-CGTTGTACCACAGA
 60.211 67.80 65.14 60.93

GAGA-ACATGCTGTGTGCGGGCA TCCTCGGGGACCGGCAGG-ATGCCTGCGAGGGCGACA GTGGGGGGCCCATGGTCG
 CTCT TGTACGACACACGCCCGT-AGGAGCCCTGGCCGTCC TACGGACGCTCCCGCTGT-CACCCCGGGTACCAGC
 69.50 72.5 71.03 71.88

-CCTCCTTCCACGGCACCT GGTTCCTGGTGGGCCTGG-TGAGCTGGGGTGAGGGCT GTGGGCTCCTTCACAACT-ACG
 GGAGGAAGGTGCCGTGGA-CCAAGGACCACCGGACC ACTCGACCCCACTCCCGA-CACCCGAGGAAGTGTTGA TGC
 67.95 69.04 69.50 63.31 62.

GCGTTTACACCAAAG TCAGCCGCTACCTCGACT-GGATCCATGGGCACATCA GAGACAAGGAAGCCCCC-AGAAGAG
 CGCAAATGTGGTTTC-AGTCGGCGATGGAGCTGA CCTAGGTACCCGTGTAGT-CTCTGTTCTTTCGGGGG TCTTCTC
 264 66.58 63.25 65.93 64.78

CTGGGCACCTTAG
 GACCCGTGGAATC

5A.3 Sequencing Results for Gene (1+2)

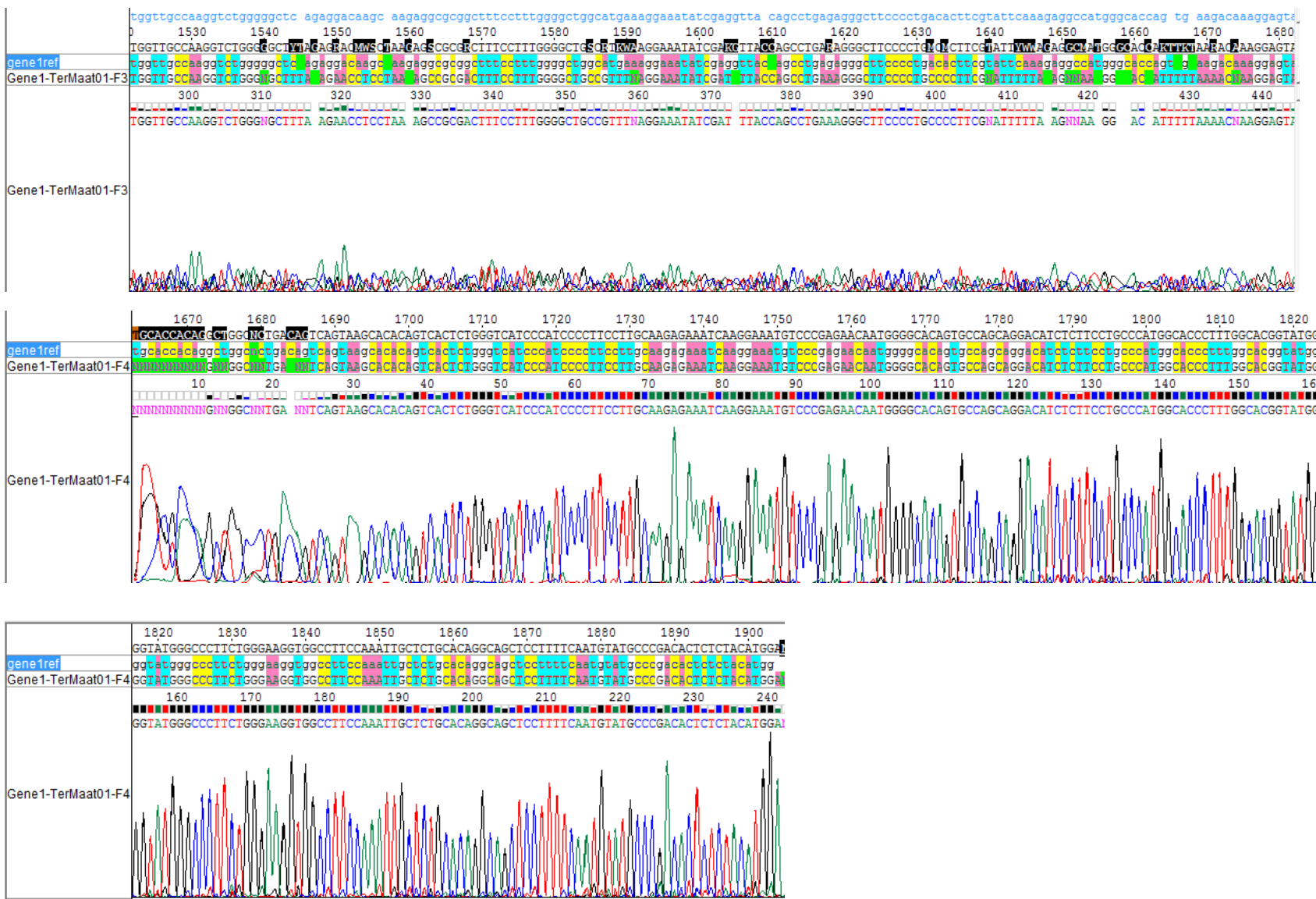
Reference (desired) sequence either at top in light blue or as “gene1ref” in each section.





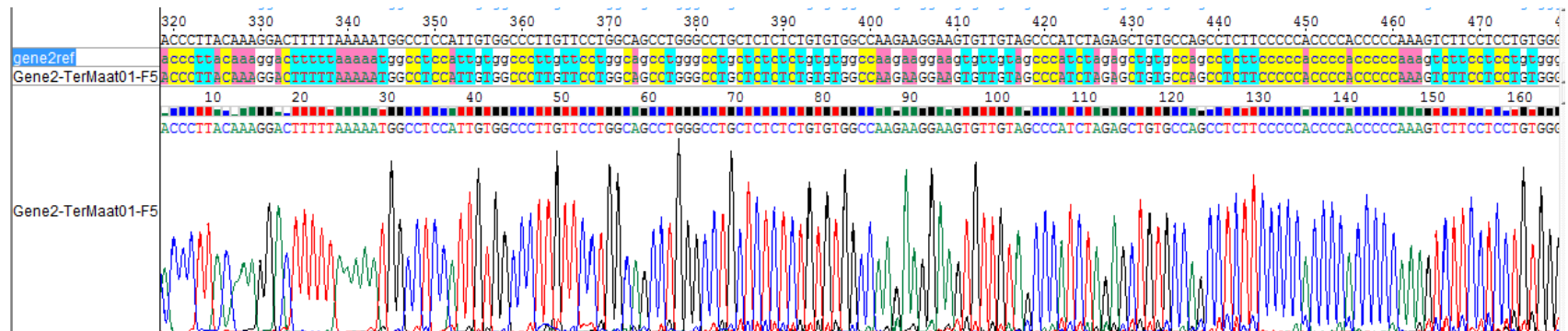
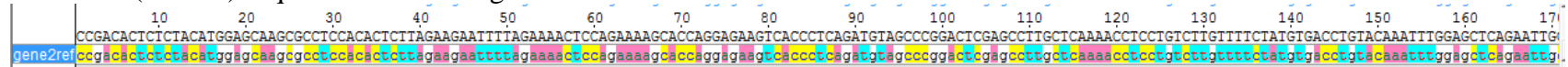


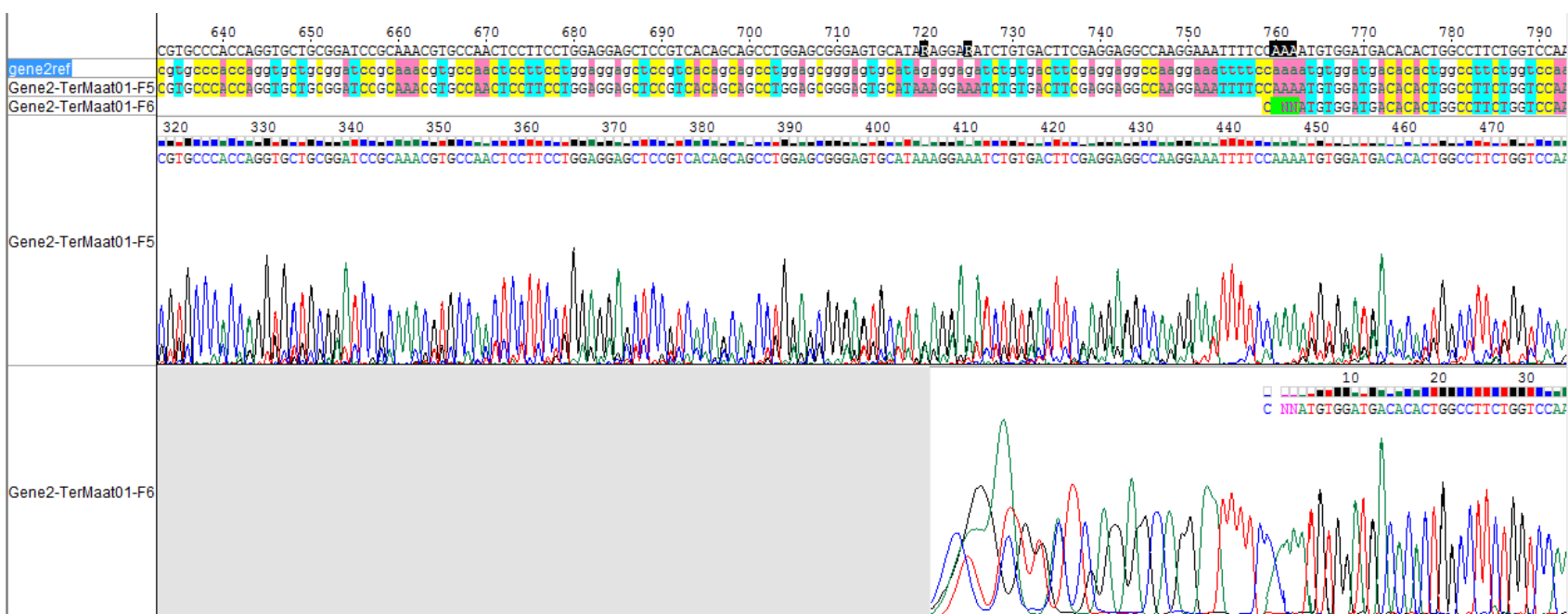
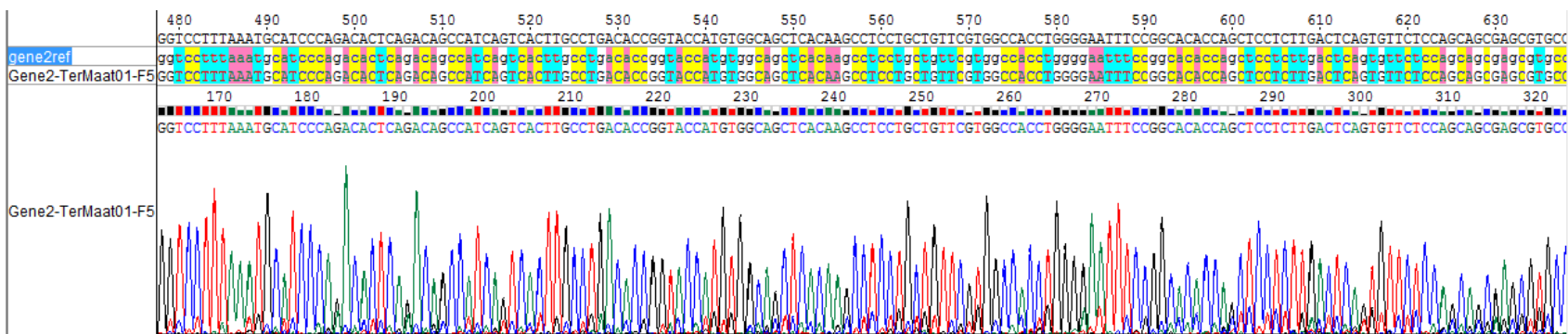


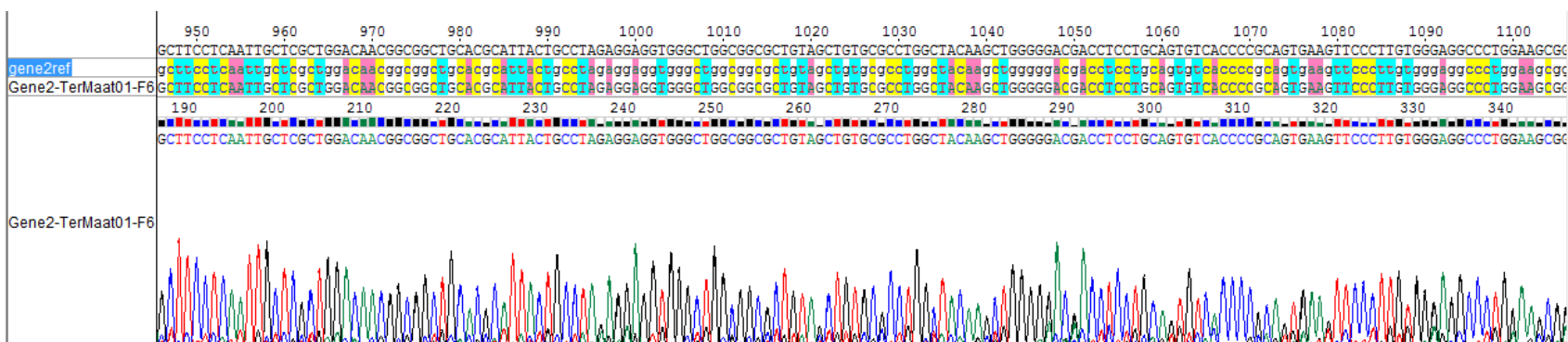
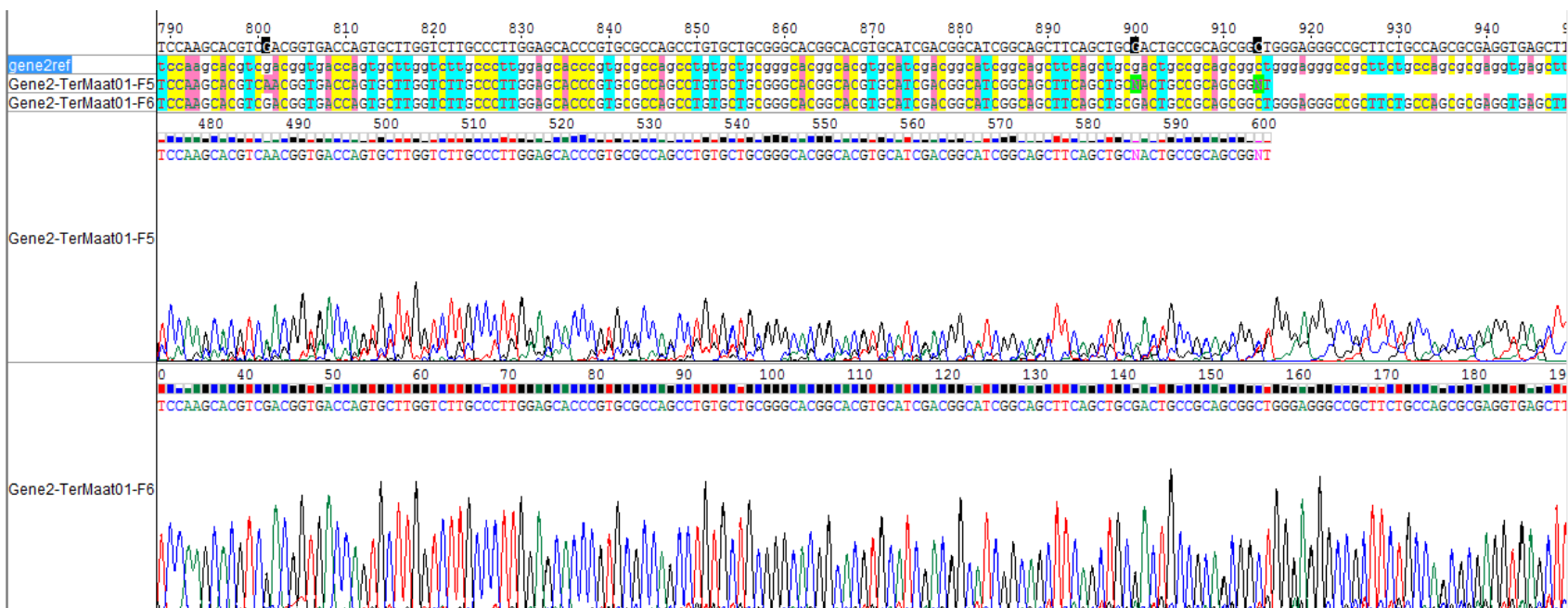


5A.4 Sequencing Results for Gene (3+4)

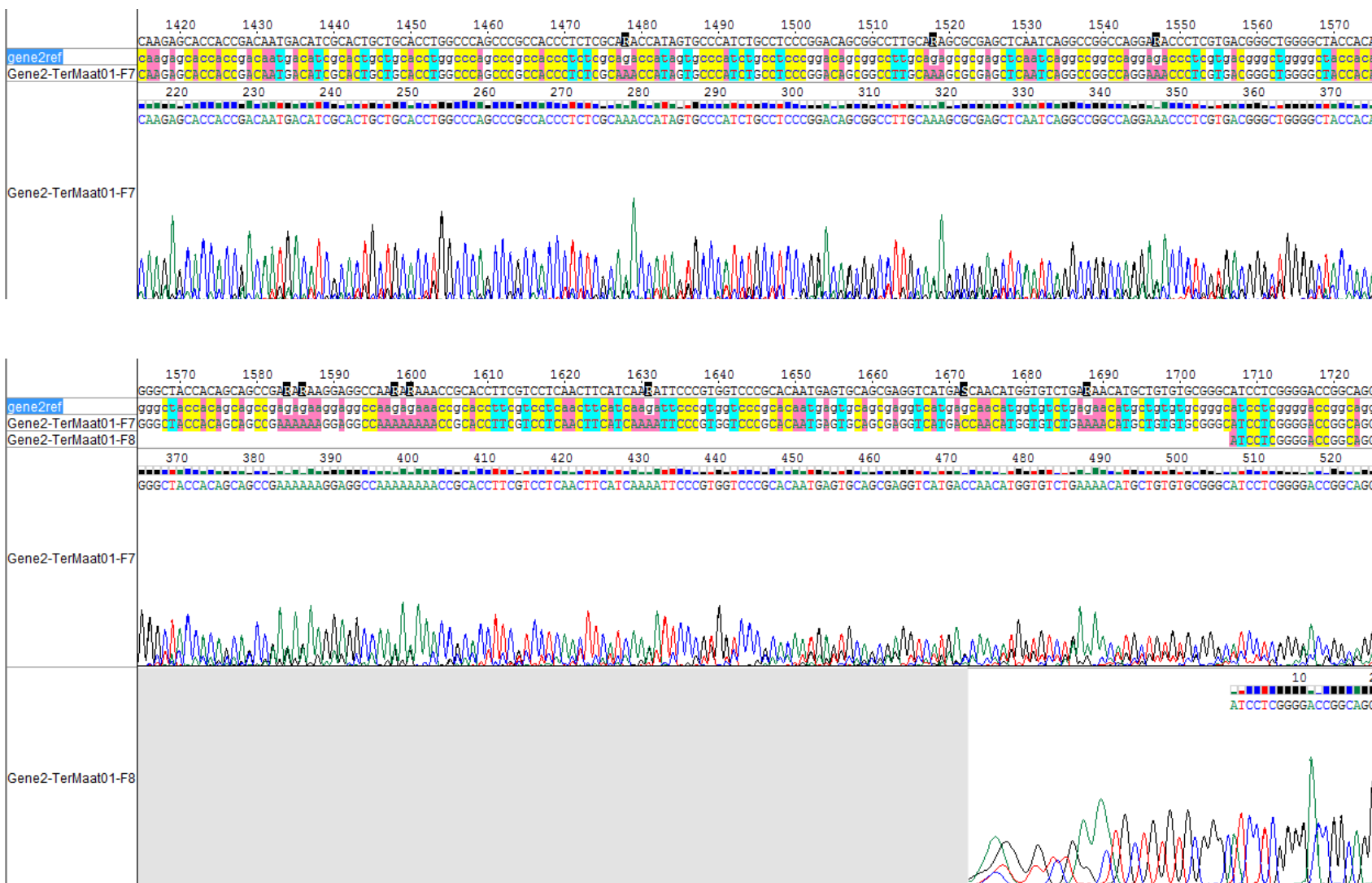
Reference (desired) sequence shown as “gene2ref” in each section.

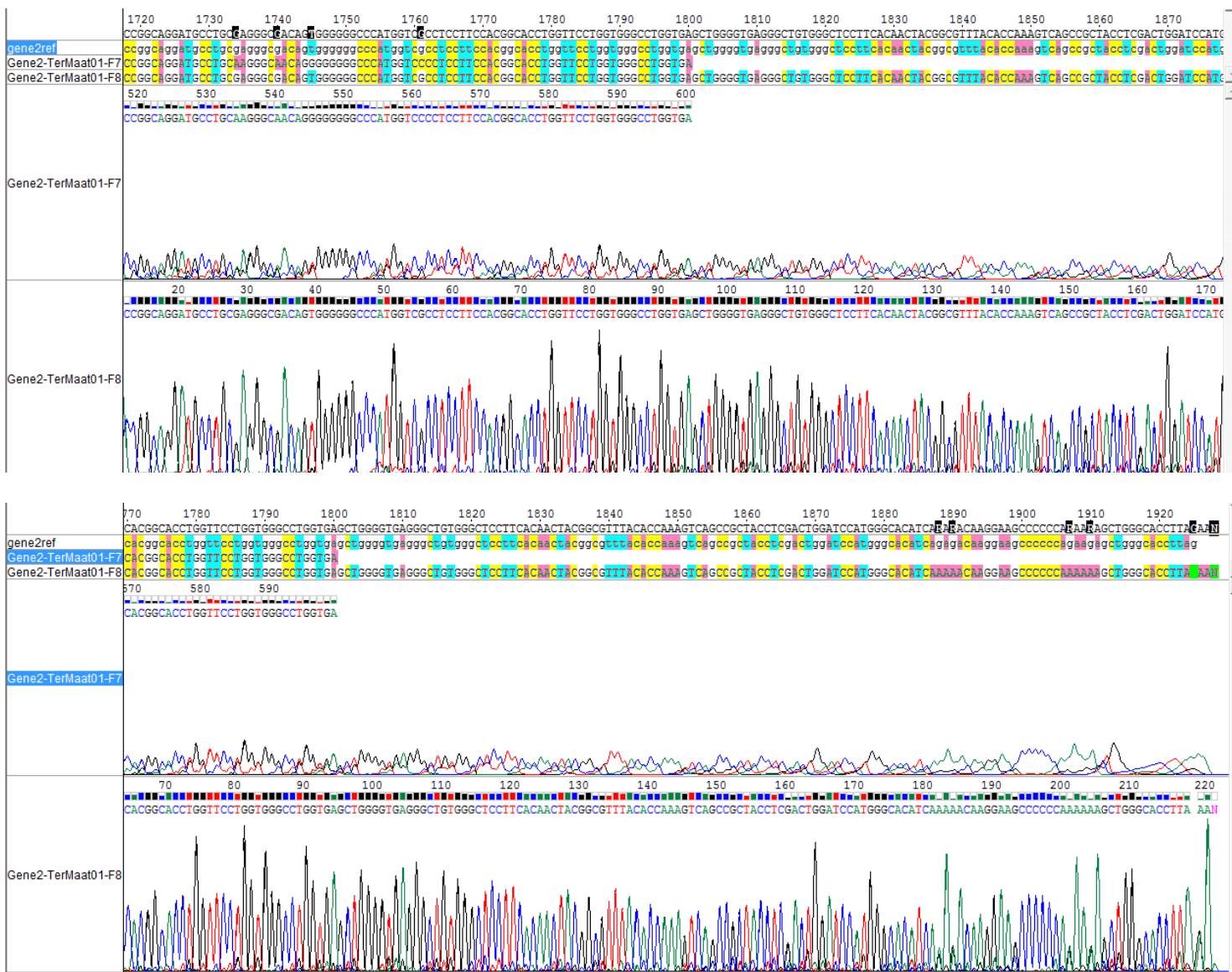












Appendix 5B- Synthesis of chimeric gene sequences

The technique of rapid PCA has been employed for the *de novo* construction of chimeric gene sequences encoding for both endothelial protein C receptor (EPCR) and thrombomodulin (TM). A mutant of TM (TM-2) was employed due to its shortened gene sequence but expected retention of high activity. Two different variants of the chimeric gene were explored, one with the amino acid sequence of TM-2 in the standard orientation and one with the amino acid sequence flipped due to the direction of the native protein state with respect to the cell wall. Two different linker epitopes, myc or flag, were incorporated into the genes. For facilitation of expression, restriction enzyme sites and His-tags were included in the design of the sequences. Both variants of the approximately 1.4 kb chimeric genes were synthesized in high-yield, with error-free clones isolated.

5B.1 Introduction

One of the most attractive advantages of PCA is the capacity to synthesize *de novo* sequences that do not naturally occur. With no inherent constraints, these gene sequences may encode for a protein that is devoid of any semblance of existing proteins. More practically, researchers are more interested in extrapolation using natural protein sequences as a basis for design. The construction of a sequence that encodes for two unique proteins to be expressed in a single cohesive stretch (i.e. chimeric gene or chimeric protein) is one such example. Even though the separate proteins are common, an integrated protein containing both is certainly a conjured entity.

Demonstration of how rapid PCA is particularly suited for synthesis of novel sequences, the construction of chimeric gene sequences was undertaken. Endothelial protein C receptor (EPCR, Genbank accession no. L35545) and the TM-2 mutant of thrombomodulin (TM, TM-1, Genbank accession no. BC035602) were selected as the two proteins of interest. Two different variants including EPCR and TM-2 were investigated—the primary difference between the two being the orientation of the TM-2 amino acid sequence. These chimeric genes serve as a reaffirmation of the rapid PCA technique's ability to construct high-fidelity *de novo* genes under extremely short thermocycling run-times. As a basic test of the convergent rapid PCA synthesis strategy, one of the gene sequences was synthesized from the stitching of two synthesized starting fragments.

EPCR and TM work synergistically in the activation of protein C and surface co-immobilization strategies have been shown to increase the catalytic conversion (Kador and Subramanian 2010). As an alternative approach, synthesis of a gene sequence that integrates both into a single protein is a perfect candidate as the resulting recombinant chimeric protein would be highly coveted if activities were retained (or even increased due to inherent proximity of the two proteins). Due to the length of the TM-1 sequence (1548bp), a shorter TM-2 sequence consisting of important domains was used that has been shown to retain high activity (Shi and others 2005). In designing a chimeric gene sequence including EPCR-1 and TM-2, there are two main considerations: 1) the orientation of the TM-2 amino acid (AA) sequence and 2) the linker region joining the two sequences. In its native orientation, the N-terminus of the TM protein

(corresponding to the 5' end of cDNA) is opposite of the cell membrane. As a secondary sequence in a chimeric gene retaining the AA order, the orientation of the AA sequence will be opposite its native with respect to the cell membrane. An alternative is to flip the AA sequence in the chimera (by reversing the cDNA), which conserves the AA orientation with respect to the cell wall. Of course, protein expression and folding may be adversely affected by this change.

The specific TM-2 for chimera #1 was extracted from the research of Shi and others (2005) using residues from Ala²²⁴ to Cys⁴⁶². Chimera #2 used residues Asp²²⁶ to Lys⁴⁶⁶. The amino acid sequences were extracted from Suzuki and others (1987). Originally, an EPCR and TM-2 chimeric sequence was designed with the TM-2 in standard sequence orientation from prior research done by Tarlan Mamedov. However, difficulties arose in purifying the expressed protein. Based on these findings, it was decided to include a His-Tag on the C-terminus of the proteins to aid in protein purification.

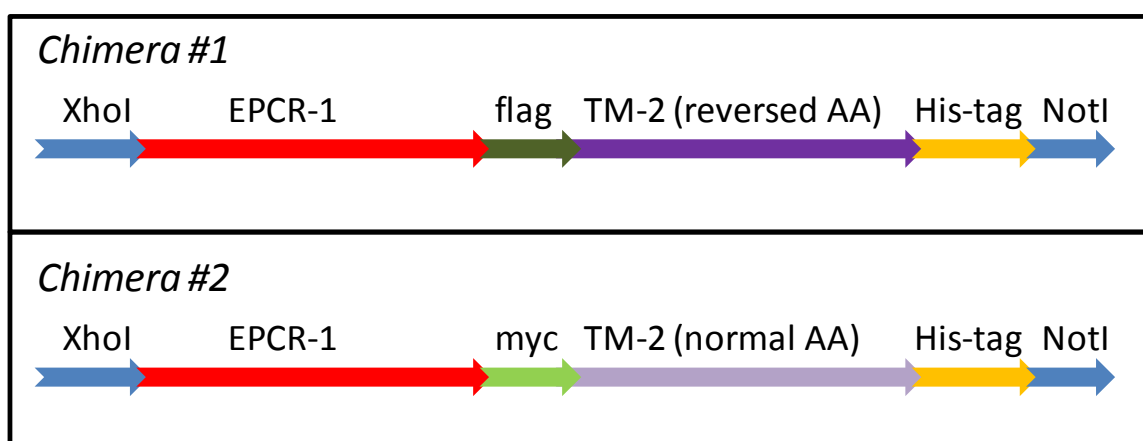


Figure 5B-1 Illustration of the EPCR and TM-2 chimeric genes.

The primary difference in the two versions of the chimera is the orientation of the TM-2 amino acids. The second aspect of the design difference is the linker between the EPCR-1 and TM-2 fragments. It is hypothesized that spacing the two proteins apart an optimal distance will result in the highest activity because of the synergy of the two proteins. Two linkers were selected as candidates, a flag epitope (DYKDDDDK) (Hopp and others 1988) for chimera #1 and a myc epitope (EQKLISEEDL) (Evan and others 1985) for chimera #2. Additionally, lysine residues flanking the flag epitope and also the end of the TM-2 were added in an attempt to increase the efficacy of the expressed chimeric #1 protein. Flanking XhoI and NotI restriction sites were designed in the gene sequence to facilitate direct insertion into pPICZ α A vector for protein expression analysis in *Pichia pastoris*.

5B.2 Chimeric genes: Design of sequence and assembly oligonucleotides

First, the hybridization oligonucleotides and AA sequences were designed (see Appendix 5C). Gene Designer from DNA 2.0 (Villalobos and others 2006) was used in the design and codon optimization of chimera #1. The codon optimization was done using a *Pichia pastoris* codon table with the usage frequency threshold set at 25%. The Gene Designer repeat minimization setting was specified to >9 bp, as this reduced sequence repeats that may be troublesome for PCA. Since XhoI and NotI sites were included at the ends to ligate the sequence into the pPICZ α A vector, avoidance of these sites internal to the sequence was also specified. The main Gene Designer window of chimera #1 is shown in Figure 5B-2 showing the pertinent segments.

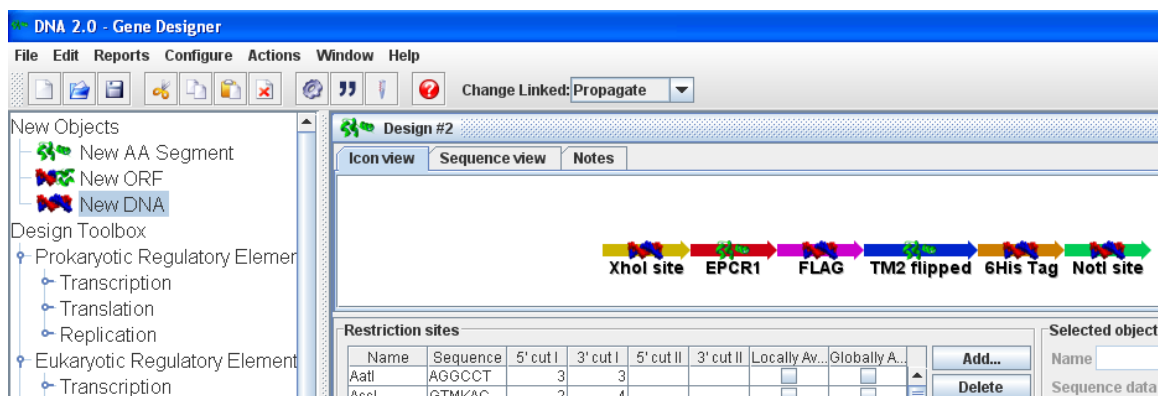


Figure 5B-2. Structure of chimera #1 as shown in Gene Designer software.

The output full nucleotide sequence for chimera #1 is shown in Appendix 5A.2. The basis for chimera #2 lacking a His tag was initially designed by Tarlan Mamedov in prior research (see Appendix 5C.6 for corresponding hybridization oligo map lacking the His-tag). Since the basis of the sequence was already designed, a slightly different approach was used. The existing sequence was first forward translated to determine the resulting AA sequence to verify the protein translation (see Appendix 5C.5). A modified oligo set was then manually generated by simply redesigning several oligos towards the 3' end of the sequence to include the His tag.

Oligo design for chimera #2 was done with the Gene2Oligo software of Rouillard and others (2004). The oligo design for chimera #1 was performed using internally developed oligo design software (Louw and others 2010). The oligo design hybridization output maps for chimera #1 and chimera #2 are provided in the appendix (Appendix 5C.3 and 5C.6, respectively). 78 oligos were needed for chimera #1 for the 1416 bp sequence. 77 oligos were needed for chimera #2 for the 1395 bp sequence. Assembly oligos and unlabelled primers were obtained from Integrated DNA technologies at 100 μ M in

nuclease-free water. The details of the chimeric gene sequences are shown in Table 5B-1.

Description	Length	# oligos	Forward Primer (5'→3')	Reverse Primer (5'→3')
Chimera #1	1416 bp	78	ATCTCGAGAAAAGATTCTGCTC	TAGCGGCCGCTTAGTG
Chimera #2	1395 bp	77	ATCTCGAGAAAAGATTTTGTAG	TAGCGGCCGCTTAGTG

Table 5B-1. Gene synthesis details for the chimeric genes.

A modified 1377 bp chimera #2 (lacking the 3' His-tag) synthesis was performed from convergent stitching of fragments. The hybridization map shown in Appendix 5C-6 shows the oligos used in this synthesis (disregard addition of oligos notation at end). The sequence was divided into an “A” fragment of 640 bp using the first 35 oligos, and a “B” fragment of 753 bp using the last 42 oligos. Note that the A and B fragments overlap each other, thus reconciling the length sum of the fragments being greater than the full-length size. Please refer to Table 4-1 in Chapter 4 for details regarding this modified chimera #2 and its fragments.

5B.3 PCA synthesis of chimeric genes

Each 25µl reaction mixture contained 200µM of each dNTP, 5 mM MgSO₄, 400 µg per ml non-acetylated BSA, and 0.5 unit KOD hot-start polymerase in 1x manufacturer's buffer obtained from Novagen (Madison, WI). 25 nM each oligonucleotide was used in the equimolar PCA reaction. PCA was conducted under the following conditions: 30 s hot start at 94°C, followed by 30 cycles of [94°C for 2 s and 56°C for 10 s and 72°C for 10 s] and a final extension at 72°C for 15 s. PCA product was subsequently PCR

amplified using 2 μ l of the PCA product as template and 0.7 μ M each primer. Note that the same reverse primer was used for both sequences due to sequence commonality on the 3' end. PCR amplification was conducted under the following conditions: 30 s hot start at 94°C, followed by 35 cycles of [94°C for 2 s and 58°C for 3 s and 72°C for 10 s] and a final extension at 72°C for 15 s.

As a very basic test of the convergent synthesis approach, the A and B fragments of the modified chimera #2 were constructed by 2-step PCA with follow-up PCR, and then directly stitched together by rapid integrated PCA-PCR. The PCA and PCR reaction compositions and thermocycling conditions were kept the same as for the full-length sequence. Note that for the particular PCR products shown in Figure 5B-4, PCA of the fragments was performed at 100 nM each oligo (done prior to Chapter 4 research). 1 μ l of each PCR product of the fragments was used as template for the stitching reaction. Stitching was performed under the same reaction conditions as PCA of chimera #2, except with 0.7 μ M of the outer primers and 1 μ l of each PCR product of the fragments used as templates.

5B.4 Cloning & Sequencing

Gel-purified DNA from both chimera #1 and chimera #1 were ligated to TOPO vector using Zero Blunt[®] TOPO[®] PCR Cloning Kit (Invitrogen, Carlsbad, CA). The ligated plasmids were transformed into chemically competent cells of *E. coli* strain TOP-10 (Invitrogen); ten positive clones of each were selected. Purified plasmids from selected clones were sequenced at the University of Nebraska-Medical Center sequencing facility

using vector and gene specific primers. Sequence analysis was performed using ChromasPro v1.5 software.

5B.5 Results & Discussion

Both of the chimeric sequences were successfully assembled and amplified by rapid 2-step PCA-PCR, as evidenced in the PCR products shown in Figure 5B-3. As can be seen from these gel results, a high yield and relative purity of the desired products were obtained using 25 nM each oligo for PCA and 2 μ l of the PCA product as PCR amplification template. The yield and purity under these conditions were dramatically better than results obtained using 100 nM / 1 μ l conditions (results not shown), in good agreement with the conclusions drawn in Chapter 4.

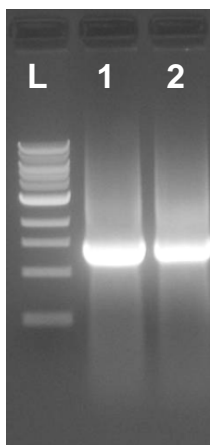


Figure 5B-3. Gel electrophoresis results of the PCR products for the 1416 bp chimera #1 and the 1395 bp chimera #2 gene synthesis.

Since labeled primers were not procured, gel purification was performed on both synthesized genes prior to cloning. Please refer to Chapter 4 that demonstrates the immunoaffinity purification of the modified chimera #2. Sequencing results verified that

the desired sequences were constructed. For chimera #1, 1 out of 9 successfully sequenced clones had no errors. For chimera #2, 3 out of 10 clones were error-free. This is indicative of the fidelity of PCA synthesis by rapid thermocycling, especially given the 1.4 kb lengths of the genes. While some of the sequences suggested that mis-annealing had occurred at the 3' end of the sequence, the majority of errors present were deletions. Deletions are primarily attributable to the quality of the phosphoramidite synthesized oligos.

The actual experimental syntheses of the two chimeric genes were straightforward using rapid 2-step PCA with follow-up PCR. The yield and purity using the optimized PCA oligo concentration of 25 nM and PCR aliquot of 2 μ l was dramatically improved over other samples in which 100 nM each oligo was used for PCA and 1 μ l each PCA product used as template (results not shown). The bulk of the research effort was concentrated in the actual design and post-sequencing analysis of the chimeric sequences. Given proper oligo design and hybridization overlaps, optimization of reaction conditions after the initial synthesis attempts was not necessary to produce error free clones for these two genes. Although, it is likely that further optimization may have reduced the number of errors by limiting mis-annealing or provide a higher discrimination of oligo deletions as evidenced by some of the sequenced clones.

Figure 5B-4 shows that the full-length modified chimera #2 lacking the His-tag was constructed by the convergent strategy. High yields of the A and B fragments were

obtained after 2-step PCA-PCR. Using these products as templates in the final reaction, the fragments successfully hybridized together coupled with amplification to construct the full-length gene. Note that a shorter fragment B band remains due to slight imbalances of the templates or from slight differences in linear amplification of the fragments with the respective primers. Not to be misleading, the final result should be expected given the ability to synthesize the full-length gene directly by 2-step PCA-PCR. However, this synthesis clearly serves as a basic example of the convergent synthesis strategy outlined in Chapter 5.

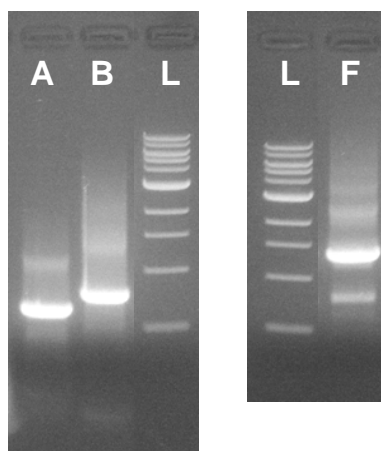


Figure 5B-4 Convergent synthesis of a modified full-length 1377 bp chimera #2

(F) from 640 bp A fragment and 753 bp B fragment. L: 1kb ladder

5B.6 Conclusions

Rapid PCA in the workstation proved highly effective for the fast and accurate synthesis of two *de novo* gene sequences. The two sequences designed consisted of chimeric couplings of EPCR and TM-2 domains. The desired codon-optimized sequences for

EPCR-myc-TM2 and EPCR-flag-TM2 (flipped) were synthesized facilely using rapid 2-step PCA and PCR. Clones screened by DNA sequencing resulted in the isolation of error-free sequences. His-tags and restriction sites were successfully included in the chimeric genes to facilitate protein purification in future work. A variant of the EPCR-myc-TM2 gene lacking the His-tag was also successfully synthesized from two shorter fragments, demonstrating the convergent strategy by rapid PCA proposed in Chapter 5.

5B.7 References

- Evan, G.I., G.K. Lewis, G. Ramsay, and J.M. Bishop. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5**: 3610-3616 (1985)
- Hopp, T.P, K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D.P. Cerretti, D.L. Urdal, and P.J. Conlon. A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Nat. Biotechnol.* **6**: 1204-1210 (1988).
- Kador K.E. and A. Subramanian. Surface Modification of Biomedical Grade Polyurethane to Enable the Ordered Co-immobilization of Two Proteins. *J Biomater Sci Polym Ed.* Appeared or available online Nov. 2nd, 2010.
- Louw, T.M., S.E. Whitney, J.R. TerMaat, E. Pienaar, and H.J. Viljoen. Oligonucleotide optimization for DNA synthesis. *AICHE J.* Article first published online: 29 SEP 2010. DOI: 10.1002/aic.1241
- Mamedov, T.G., Padhye, N.V., Viljoen, H., Subramanian, A., 2007. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *J. Biotechnol.* **131**: 379-387

- Rouillard, J.M., W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**: W176-W180 (2004).
- Shi, C.S. et al., 2005. Evidence of human thrombomodulin domain as a novel angiogenic factor. *Circulation.* **111**: 1627-1636 (2005).
- Suzuki, K. et al., 1987. Structure and expression of human thrombomodulin, a thrombin receptor on endothelium acting as a cofactor for protein C activation. *EMBO.* **6**: 1891-1897 (1987).
- Villalobos, A., J.E. Ness, C. Gustafsson, J. Minshull, and S. Govindarajan. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics.* **7**: 285 (2006).

5B.8 Contributions

- J.R. TerMaat was primarily responsible for the design and synthesis of the chimera #1, chimera #2, and convergent approach for the modified chimera #2 lacking the His-tag.
- T.G. Mamedov conducted sequence and oligo design of the modified chimera #2 gene lacking the His-tag used in this work.
- T. Louw developed the oligonucleotide hybridization software used in this work to design oligos for the chimera #1 sequence.
- A. Subramanian led the scientific direction of the research.

Appendix 5C: Details of Chimeric Sequences

5C.1 Chimera #1 Beginning Design

a t c t c g a g a a a g a **F C S Q D A S D G L Q R L H M L Q I S Y F R D P Y**
H V W Y Q G N A S L G G H L T H V L E G P D T N T T I I Q L Q P L Q E P
E S W A R T Q S G L Q S Y L L Q F H G L V R L V H Q E R T L A F P L T I
R C F L G C E L P P E G S R A H V F F E V A V N G S S F V S F R P E R A L
W Q A D T Q V T S G V V T F T L Q Q L N A Y N R T R Y E L R E F L E D
T C V Q Y V Q K H I S A E N T K G S Q T S R S Y T S K K R K K D Y K D
D D D K K R K R K C D T G I H R V L A S D P G C I C E F T G P L N H C V
G S C F G G N E C E D I D T C I F G D D L I Y G E P C E C S A Q T N P D C
D A P C A T Q N C F M Q C R H P E H P I P A F G E A C V C L Y S T Q N L
P Q C Q Y E C N A R F C P D V P E V C E G D V L D Y N P Y C H C E F G
G Q T N V C R Q P C P S P E L I C D D V D E C R H Q D A A L R Y G T E C
M C S Y S G P Q D P N P V C F H E C L D N C S Q T A S A T C S R G D A Q
L A A G A P C Q C R P A G P I A N C A H E C G G N E V S C D W A R K K
 K R **H H H H H H** Stop g c g g c c g c t a

Legend

Lower-case: nucleotides

Black- added sequence

Green- Flag epitope

Orange- His 6 tag

Upper Case: Amino Acids

Red- EPCR mature peptide

Blue- TM-2 (Ala²²⁴ to Cys⁴⁶²) flipped

Underline-Xho1 and Not1 sites

5C.2 Chimera#1 Resulting Sequence

ATCTCGAGAAAAGATTCTGCTCTCAAGATGCTTCTGATGGATTGCAAAGATTG
CATATGTTGCAAATTTCTTACTTCAGAGATCCTTACCACGTTTGGTACCAGGG
TAACGCTTCTTTGGGTGGACATTTGACTCACGTTTTTGGAGGGTCCTGATACTA
ACACTACTATCATTCAATTGCAACCTTTGCAGGAACCTGAATCTTGGGCCAGA
ACCCAATCTGGATTGCAGTCTTACTTGTTGCAGTTTCATGGTTTGGTTAGATT
GGTTCACCAGGAGAGAACTTTGGCCTTTCCTTTGACCATTAGATGTTTCTTGG
GATGTGAATTGCCTCCAGAAGGTTCTAGAGCTCATGTTTTTTTCGAAGTTGCT
GTTAACGGTTCTTCTTTTCGTTTCTTTCAGACCTGAGAGAGCCTTGTGGCAGGC
TGATACTCAGGTTACCTCTGGAGTTGTTACTTTCACCTTGCAACAGTTGAACG
CTTATAACAGAACCAGATATGAATTGAGAGAATTCTTGGAAGATACTTGTGT
TCAGTATGTTCAAAAGCACATTTCTGCCGAAAACACCAAGGGTTCTCAAACCT
CTAGATCTTACACCTCTAAGAAAAGAAAGAAGGATTACAAAGACGACGACG
ATAAGAAGAGAAAAGAGAAAATGTGACACTGGTATCCACAGAGTTTTGGCTTC
TGACCCTGGATGCATTTGTGAGTTTACTGGACCTTTGAACCACTGTGTTGGTT
CTTGTTTTTGGTGGTAACGAGTGTGAGGACATTGACACTTGCATTTTTTGGAGAC
GATTTGATTTATGGTGAGCCATGCGAATGTTCTGCCCAAACCTAACCCTGATTG
CGACGCCCCATGTGCCACTCAAACTGTTTCATGCAGTGTAGACACCCTGAA
CATCCTATTCCAGCTTTCGGTGAAGCTTGCGTTTGTGTTGATTCTACTCAGAAC
TTGCCACAGTGCCAGTACGAGTGTAACGCTAGATTTTGCCCTGACGTTCTGA
AGTTTGTGAGGGTGATGTTTTTGGACTACAACCCATACTGTCATTGCGAATTTCG
GTGGACAGACTAACGTTTGCAGACAACCTTGCCCTTCTCCAGAATTGATCTGT
GATGACGTTGATGAATGCAGACACCAGGATGCCGCCTTGAGATATGGAACCG
AATGTATGTGTTCTTATTCTGGTCCTCAAGATCCTAACCAGTTTGTGTTCCATG
AGTGCTTGGACAATTGTTCTCAGACCGCTTCTGCCACCTGTTCTAGAGGAGAT
GCCCAGTTGGCTGCTGGTGCTCCTTGCCAATGTAGACCTGCTGGTCCAATTGC
TAACTGTGCTCACGAATGTGGAGGTAATGAGGTTTCTTGTGACTGGGCTAGA
AAGAAGAAGAGACATCATCACCATCACCCTAAGCGGCCGCTA

5C.3 Chimera #1 Oligo Design Hybridization Map

```

-----
ATCTCGAGAAAAGATTCTGC TCTCAAGATGCTTCTGATG-GATTGCAAAGATTGCATATGT TGC AAAATTTCTTACTTCAGAG-ATCCTTACCACGTTTGG
TAGAGCTCTTTCTAAGACG-AGAGTTCTACGAAGACTAC CTAACGTTTCTAACGTATACA-ACGTTTAAAGATGAAGTCTC TAGGAATGGTGCAAACC
T = 57.5795 T = 56.8179 T = 57.347 T = 56.1104 T = 56.5007
-----
TACCAGGTAACGCTT-CTTTGGGTGGACATTTGA CTCACGTTTGGAGGG-TCCTGATACTAACACTACTATCA TTCAATTGCACCTTTGC-AGGAACCTGAATCTGGG
-ATGGTCCCATTTGCGAA GAAACCCACCTGTAAACT-GAGTGCAAAACCTCCC AGGACTATGATTTGTGATGATAGT-AAGTTAACGTTGSAACG TCCTTGGACTTGAACCC
T = 57.102 T = 57.5298 T = 57.2092 T = 57.3733 T = 56.9362 T = 58.9821
-----
CCAGAAACCAATCTGG-ATTGCAGTCTTACTTGTTC AGTTTCAATGGTTTGGTTAGA-TTGGTTCACAGGAGA GAACTTTGGCTTTTCT
-GGTCTTGGGTTTAGACC TAACGTCAGAATGAACAACG-TCAAGTACCAAAACAATCT AACCAAGTGGTCTCT-CTTGAACCGGAAGAGA
T = 56.6284 T = 58.8957 T = 56.8628 T = 56.8355 T = 57.0704
-----
-TTGACCATTAAGTGTTCCTGG GATGTGAATTGCTCCA-GAAGGTTCTAGAGCTCATG TTTTTCGAAGTGTCTGT-TACGGTTCCTCTTCGTT
AATCTGGTAATCTACAAAGAACC-CTACACTTAACGGAGGT CTTCCAAGATCTCAGGTAC-AAAAAAGCTTCAACGACAA TTGCAAGAAGAAAGCAA
T = 58.5048 T = 56.9309 T = 57.0395 T = 56.7238 T = 56.8126
-----
TCTTTCAGACCTGAGAGA-GCCTTGTGGCAGGCTG ATACTCAGGTTACCTCTGG-AGTGTACTTTCACCTTGC AACAGTTGAACGCTTATAACA
-AGAAAGCTGGAATCTCT CGGAACACCGTCCGAC-TATGATCCATGGAGACC TCAACAATGAAGTGGAACG-TGTCACCTTGGCAATATGT
T = 56.401 T = 64.6309 T = 57.7534 T = 58.4542 T = 57.8503
-----
-GAACCAGATATGAATTGAGAGA ATTCTTGAAGATACTTGTGT-TCAGTATGTTCAAAAGACA TTTCTGCGAAAACACC-AAGGGTCTCAAACCTCTAG
CTTGGTCTTACTTAACTCTCT-TAAGAACTTCTATGAACACA AGTCATACAAGTTTTCGTT-AAAGACGCTTTTGTGG TTCCCAAGAGTTGAAGATC
T = 57.2631 T = 56.7984 T = 57.4075 T = 58.0436 T = 56.7413
-----
ATCTTACACCTCTAAGAAAAGA-AAGAAGGATTACAAAGACA CGACGATAAGAAAGAGAAAAGA-GAAATGTGACACTGGTATC
TAGAATGTGGAGATCTTTCT TCTTCTCTAATGTTCTGCT-GCTGCTATCTCTCTCTCT CTTTACACTGTGACATAG
T = 56.0439 T = 56.3014 T = 56.5821 T = 56.2353
-----
CACAGAGTTTGGCTTCT-GACCTGGATGCATTT GTGAGTTTACTGGACCTT-TGAACACTGTGTGG TCTTGTTTTGGTGGTAAAG-AGTGTGAGACATTGACA
-GTGCTCAAACCGAAGA CTGGGACTTACGTAAA-CACTCAAATGACCTGGAA ACTTGGTGACACACC-AAGAACAAACCAACATTGC TCACACTCTGTAACTGT
T = 57.408 T = 56.4857 T = 56.0704 T = 56.3341 T = 58.7302 T = 58.1244
-----
CTTGCAATTTTGGAGACG-ATTGATTTATGTTGAGCA TGCGAATGTTCTGCC-AACTAACCTGATTGCG ACGCCCATGTGTCAC
-GAACGTAAAACCTCTGC TAAACTAAATACCACTCGGT-ACGCTTACAAGACGGG TTTGATTGGGACTAACGC-TGCGGGTACACGGTG
T = 56.8992 T = 57.5142 T = 60.409 T = 57.6783 T = 66.7021
-----
-TCAAACTGTTTTCATGCG TGTAGACACCTGAACA-TCTATTCCAGCTTTCGG TGAAGCTTGGGTTTGT-TTGTATTCTACTCAGAACTTGC
AGTTTGTGACAAAGTACGTC-ACATCTGTGGGACTGT AGGATAAGGTCGAAAGCC-ACTTCGAACGCAACA AACATAAGATGAGTCTTGAACG
T = 56.3193 T = 56.9429 T = 59.0842 T = 56.942 T = 57.8555
-----
CACAGTCCAGTACGA-GTGAACGCTAGATTTTGC CCTGACGTTCTGAAG-TTGTGAGGGTGATGTT TGGACTACAACCATACT-GTCATTGCGAAATTCGGT
-GTGTCACGGTCATGCT CACATTGCGATCTAAACG-GGACTGCAAGGACTTC AAACACTCCACTACAAA-ACCTGATGTTGGGTATGA CAGTAACGCTTAAGCCA
T = 58.6769 T = 56.7828 T = 56.4518 T = 56.5278 T = 56.9998 T = 58.0003
-----
GGACAGACTAACGTTTGC-AGACAACCTTGCCCTT CTCAGAAATTGATCTGTGA-TGACGTTGATGAATGCA GACACAGGATGCCGC
-CCTGTCTGATTGCAACG TCTGTTGGAAACGGAA-GAGGCTTAACTAGACACT ACTGCAACTACTTACGT-CTGTGCTCTACGGCG
T = 58.3361 T = 58.1022 T = 56.4545 T = 56.5597 T = 63.9534
-----
-CTTGAGATATGGAACCGAA TGTATGTGTTCTTATTCTGGT-CTCAAGATCTAACCA GTTGTGTTCCATGAGTGC-TTGGAACAATTGTTCTCAGA
GAACCTCTATACCTTGGCTT-ACATACACAAGAATAAGACA GGAGTTCTAGGATTGGT-CAAACAAGGTACTCACG AACCTGTTAAACAAGATCT
T = 56.5937 T = 56.5149 T = 58.0162 T = 56.7897 T = 56.3285
-----
CCGCTTCTGCCACCTG-TTCTAGAGAGATGCC AGTTGGCTGCTGGTGC-TCTTGCCAATGTAGAC CTGCTGGTCCAAATGTC-TAACTGTGCTCACGAATG
-GGCGAAGACGGTGGAC AAGATCTCTCTACGG-TCAACCGACGACCAAG AGGAACGGTTACATCTG-GACGACAGGTTAACG ATTGACACGAGTGCTTAC
T = 63.8609 T = 57.2886 T = 63.5895 T = 56.1306 T = 58.506 T = 57.2194
-----
TGAGGTAATGAGGTTTCT-TGTGACTGGCTAGAA AGAAGAAGACATCATCAC-CATCAACCTAAGCGG CCGCTA
-ACCTCATTAACCTCAAGA ACACGACCGATCTT-TCTTCTCTCTGTAGTAGTG GTAGTGGTATTCGCC-GGCGAT
T = 57.7677 T = 56.5621 T = 56.554 T = 57.0889

```

5C.4 Chimera #2 Starting original nucleotide sequence

ATCTCGAGAAAAGATTTTGTAGTCAAGATGCTTCTGATGGTCTACAAAGATTG
CATATGTTGCAAATCTCTTACTTTCGTGATCCTTATCATGTTTGGTACCAAGGT
AATGCTTCTTTGGGTGGACATTTGACTCATGTTTTGGAAGGTCCAGATACTAA
TACTACTATCATTCAATTGCAACCTTTGCAAGAACCTGAAAGTTGGGCTCGTA
CTCAAAGTGGTTTGCAATCTTACTTGTTACAATTCATGGTCTAGTTCGTTTGG
TTCATCAAGAACGTACTTTGGCTTTTCCTTTGACTATCCGTTGTTTTTTGGGTT
GTGAATTGCCTCCTGAAGGTTCTAGAGCTCATGTTTTTTTTGAAGTTGCTGTTA
ATGGTAGTTCTTTTTGTTAGTTTTTCGTCCAGAAAGAGCTTTGTGGCAAGCTGAT
ACTCAAGTCACCTCTGGAGTTGTCACCTTTACCTTGCAACAATTGAATGCTTA
CAATCGTACTCGTTATGAATTACGTGAATTTCTTGAAGATACTTGTGTTCAAT
ATGTTCAAAAACATATTTCTGCTGAAAATACTAAAGGTAGTCAAAGTAGTCG
TTCTTACACTTCTGAACAAAACTCATCTCAGAAGAGGATCTGAATAGCGCC
GTCGACTGTTCTGTTGAAAATGGTGGTTGTGAACATGCTTGTAATGCTATTCC
AGGTGCTCCAAGATGTCAATGTCCAGCTGGTGCTGCTTTGCAAGCTGATGGTA
GATCTTGTACTGCTTCTGCTACTCAATCTTGTAATGATTTGTGTGAACATTTCT
GTGTTCCAAATCCAGATCAACCAGGTTCTTATTCTTGTATGTGTGAAACTGGT
TATAGATTGGCTGCTGATCAACATAGATGTGAAGATGTTGATGATTGTATTTT
GGAACCATCTCCATGTCCACAAAGATGTGTTAATACTCAAGGTGGTTTTCGAAT
GTCATTGTTATCCAAATTATGATTTGGTTGATGGTGAATGTGTTGAACCAGTT
GATCCATGTTTCAGAGCTAATTGTGAATATCAATGTCAACCATTGAATCAAAC
TTCTTATTTGTGTGTTTGTGCTGAAGGTTTCGCTCCAATTCCACATGAACCACA
TAGATGTCAAATGTTCTGTAATCAAAGTCTTGTCAGCTGATTGTGATCCAA
ATACTCAAGCTTCTTGTGAATGTCCAGAAGGTTATATTTTGGATGATGGTTTC
ATTTGTACTGATATTGATGAATGTGAAAATGGTGGTTTCTGTTCTGGTGTTTG
TCATAATTTGCCAGGTACTTTCGAATGTATTTGTGGTCCAGATTCTGCTTTGGT
TAGACATATTGGTACTGATTGTGATTCTGGTAAGCATCATCACCATCACCCT
AAGCGGCCGCTA

5C.5 Resulting Forward Translation of Chimera #2

a t c t c g a g a a a g a **FCSQDASDGLQRLHMLQISYFRDPYHV**
WYQGNASLGGHLTHVLEGPDTNTTIIQLQPLQEPESWART
QSGLQSYLLQFHGLVRLVHQERTLAFPLTIRCFLGCELPP
EGSRAHVFFEVAVNGSSFVSFRPERALWQADTQVTSGVV
TFTLQQLNAYNRTRYELREFLEDTCVQYVQKHISAENTK
GSQTSRSYTSEQKLISEEDLNSAVDCSVENGGCEHACNAI
PGAPRCQCPAGAALQADGRSCTASATQSCNDLCEHFCVP
NPDQPGSYSCMetCETGYRLAADQHRCEDVDDCILEPSPCP
QRCVNTQGGFECHCYPNYDLVDGECVEPVDPCFRANCEY
QCQPLNQTSYLCVCAEGFAPIPHRCQMetFCNQTTACPA
DCDPNTQASCECPEGYILDGFICTDIDECENGGFCSGVC
HNLPGTFCICGPDSALVRHIGTDCDSGKHHHHHHHStop
g c g g c c g c t a

Legend

Lower-case: nucleotides

Black- added sequence

Green- Flag epitope

Orange- His 6 tag

Upper Case: Amino Acids

Red- EPCR mature peptide

Blue- TM-2 (Asp²²⁶ to Lys⁴⁶⁶) normal

Underline-Xho1 and Not1 sites

5C.6 Chimera #2 Oligo Design Hybridization Map

```

tatagaacagatctatCTCGAG-AAAAGATTTTGTAGTCAAGAT GCTTCTGATGGTCTACA-AAGATTGCATATGTTGC
atatcttgtctagaTAGAGCTC TTTTCTAAAACATCAGTTCTA-CGAAGACTACCAGATGT TTCTAACGTATACAACG
62.732          62.35          64.03          63.68

A AATCTCTTACTTTTCGTGAT-CCITTATCATGTTTGGTACC AAGGTAATGCTTCTTTTGG-GTGGACATTTGACTCATG
T-TTAGAGAATGAAAGCACTA GGAATAGTACAAACCATGG-TTCCATTACGAAGAAACC CACCTGTAAACTGAGTAC-
62.32          64.10          63.60          64.33

TTTTGGAAGGTCCAGAT-ACTAATACTACTATCATTTCAATTG CAACCTTTGCAAGAACC-TGAAAGTTGGGCTCGTA C
AAAACCTTCCAGGTCTA TGATTATGATGATAGTAAAGTTAAC-GTTGGAACGTTCTTGG ACTTTCAACCCGAGCAT-G
63.895          63.14          64.89          66.87          6

TCAAAGTGGTTTGCA-ATCTTACTTGTACAAATTTTCAT GGTCTAGTTCGTTTGGT-TCATCAAGAACGTACTTTG GCT
AGTTTCACCAAACGT TAGAATGAACAATGTTAAAGTA-CCAGATCAAGCAAACCA AGTAGTTCTTGCATGAAAC-CGA
2.93          62.812          64.61          63.892          64.

TTTCCTTTGACTATCC-GTTGTTTTTTGGGTTTG G AATTGCCTCCTGAAGG-TTCTAGAGCTCATGTTTTT TTTGAAG
AAAGGAAACTGATAGG CAACAAAAAACCCACAC-TTAACGAGGACTTCC AAGATCTCGAGTACAAAAA-AAACTTC
69          63.98          64.42          62.89          62.824

TTGCTGTTAATG-GTAGTTCTTTTGTAGTTTTCG TCCAGAAAGAGCTTTGT-GGCAAGCTGATACTCA AGTCACCTC
AACGACAATTAC CATCAAGAAAACAATCAAAAGC-AGGTCTTTCTCGAAACA CCGTTCGACTATGAGT-TCAGTGGAG
64.68          64.1          64.07          64.68

TGGAGTT-GTCACCTTTACCTTGC AACAAATTGAATGCTTACAA-TCGTACTCGTTATGAATTAC GTGAATTTCTTGAA
ACCTCAA CAGTGGAAATGGAACG-TTGTTAACTTACGAATGTT AGCATGAGCAATACTTAAAG-CACTTAAAGAACTT
62.73          62.157          63.47          62.7

GATACTT-GTGTTC AATATGTTCAAAAAC ATATTTCTGCTGAAAATACTAA-AGGTAGTCAAAGTATGCTG TTCTTACA
CTATGAA CACAAGTTATACAAGTTTTTG-TATAAAGACGACTTTTATGATT TCCATCAGTTTGTATCAGC-AAGAATGT
62.96          62.330          63.97          62.72

CTTCTGAACAA-AAACTCATCTCAGAAGAGG ATCTGAATAGCGCCGT-CGACTGTTCTGTTGAAA ATGGTGGTTGTGA
GAAGACTTGT TTTGAGTAGAGTCTTCTCC-TAGACTTATCGCGGCA GCTGACAAGACAACCTT-TACCACCAACACT
65.02          66.54          63.00          64.15

ACA-TGCTTGTAATGCTATTCC AGGTGCTCCAAGATGT-CAATGTCCAGCTGGTG CTGCTTTGCAAGCTGA-TGGTAG
TGT ACGAACATTACGATAAGG-TCCACGAGGTCTTACA GTTACAGGTCGACCAC-GACGAAACGTTGACT ACCATC
63.40          65.99          65.9          66.19          64.41

ATCTTGTAAGTGC TTCTGCTACTCAATCTTG-TAATGATTTGTGTGAACATT TCTGTGTTCCAAATCCA-GATCAACCA
TAGAACATGACG-AAGACGATGAGTTAGAAC ATTACTAAACACACTTGTAA-AGACACAAGGTTTAGGT CTAGTTGGT
62.576          62.29          64.20          62.27

GGTTCTTAT TCTTGATGTGTGAAACTG-GTTATAGATTGGCTGCTG ATCAACATAGATGTGAAGAT-GTTGATGATT
CCAAGAATA-AGAACATACACACTTTGAC CAATATCTAACCGACGAC-TAGTTGTATCTACACTTCTA CAACTACTAA
63.84          63.85          62.854          62.72

GTATTTTGGG ACCATCTCCATGTCCA-CAAAGATGTGTTAATACTCAA GGTGGTTTCGAATGTC-ATTGTTATCCAAA
CATAAAACCT-TGGTAGAGGTACAGGT GTTCTACACAATTATGAGTT-CCACCAAAGCTTACAG TAACAATAGGTTT
65.28          62.74          63.15          62.

```

```

TTATGATTG GTTGATGGTGAATGTGT-TGAACCAGTTGATCCA TGTTTCAGAGCTAATTGT-GAATATCAATGTCAA
AATACTAAAC-CAACTACCACTTACACA ACTTGGTCAACTAGGT-ACAAAGTCTCGATTAAACA CTTATAGTTACAGTT
        63.14                63.03                62.44                62.510

CCATT GAATCAAACCTTCTTATTTGTGT-GTTTGTGCTGAAGGTT TCGCTCCAATTCCACA-TGAACCACATAGATGTC
GGTAA-CTTAGTTTGAAGAATAAACACA CAAACACGACTTCCAA-AGCGAGGTTAAGGTGT ACTTGGTGTATCTACAG
        64.18                62.94                66.72                63.9

A AATGTTCTGTAATCAAACCTG-CTTGTCAGCTGATTG TGATCCAAATACTCAAGC-TTCTTGTGAATGTCCAG AAG
T-TTACAAGACATTAGTTTGAC GAACAGGTCGACTAAC-ACTAGGTTTATGAGTTCG AAGAACACTTACAGGTC-TTC
        62.64                62.81                63.00                62.36                62.

GTTATATTTTGGATGATG-GTTTCATTGTACTGATATTGA TGAATGTGAAAATGGTGG-TTCTGTCTGGTGTGTTG
CAATATAAACCTACTAC CAAAGTAAACATGACTATAACT-ACTTACACTTTTACCACC AAAGACAAGACCACAAAC-
64                63.44                64.24                64.05

TCATAATTTGCCAGGTAC-TTTCGAATGTATTTGTGGT CCAGATTCTGCTTTGG-TTAGACATATTGGTACTGATT GT
AGTATTAAACGGTCCATG AAAGCTTACATAAACACCA-GGTCTAAGACGAAACC AATCTGTATAACCATGACTAA-CA
63.01                63.88                62.84                62.88                64

GATTCTGGTAAGTAAGC-GGCCGCTAgtatagtc
CTAAGACCAATTCATTG CCGGCGATcatatcag
.05                63.545

```

Note that the above does not contain the His tag. Manual addition resulted in the following three oligos substituted at the 3' end.

R1343his 5' TGATGCTTACCAGAATCACAATCAGTACCAATATGTCTAA 3'

F1364his 5' GTGATTCTGGTAAGCATCATCACCATCACCCTAAGC 3'

R1383his 5' TAGCGGCCGCTTAGTGGTGATGGTGA 3'