Microfluidic Gene Synthesis

David Sun Kong

B.S. Chemical Engineering, Massachusetts Institute of Technology, 2001
S.M. Media Arts and Sciences, Massachusetts Institute of Technology, 2004

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences
at the Massachusetts Institute of Technology

June 2008

Signature of Author.....
Program in Media Arts and Sciences
May 23, 2008

Certified by.....
Joseph M. Jacobson
Associate Professor
Program in Media Arts and Sciences

Accepted by.....
Deb Roy
Chairman, Department Committee on Graduate Students
Program Media Arts and Sciences

Microfluidic Gene Synthesis

by

David Sun Kong

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

The ability to synthesize custom *de novo* DNA constructs rapidly, accurately, and inexpensively is highly desired by researchers, as synthetic genes and longer DNA constructs are enabling to numerous powerful applications in both traditional molecular biology and the emerging field of synthetic biology, from the synthesis of large sets of novel proteins to the complete re-writing of bacterial genomes. However, the current cost of *de novo* synthesis—driven largely by reagent and handling costs—is a significant barrier to the widespread availability of such technology. The use of microfluidic technology greatly reduces reaction volumes and corresponding reagent and handling costs. Additionally, microfluidic technology enables large numbers of complex reactions to be performed in parallel, while facilitating the automation and integration of multiple processes in a single device. While microfluidic devices have been used to miniaturize a variety of chemical and biological processes, the benefits of such devices have yet to be realized in the area of *de novo* DNA synthesis.

This thesis reports the first demonstration of gene synthesis in a microfluidic environment. A variety of DNA constructs with sizes as large as 1 kb were fabricated in parallel in a multi-chamber microfluidic device at volumes one to two orders of magnitude lower than those utilized in conventional bench top techniques. This thesis also reports on progress toward the direct synthesis of genes from hybrid microfluidic-DNA microarray devices, the integration of microfluidic gene synthesis with on-chip protein synthesis, and the microfluidic hierarchical synthesis of long DNA molecules.

Thesis supervisor:
Joseph M. Jacobson
Associate Professor of Media Arts and Sciences, MIT

Microfluidic Gene Synthesis

David Sun Kong

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning in partial fulfillment of the requirements of the Degree of Doctor of Philosophy.

Massachusetts Institute of Technology.

June 2008.

Joseph M. Jacobson — **Thesis Advisor**
Associate Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Peter A. Carr — **Thesis Reader**
Research Scientist, Media Laboratory
Massachusetts Institute of Technology

Scott R. Manalis — **Thesis Reader**
Associate Professor of Biological Engineering
Massachusetts Institute of Technology

Franco Cerrina — **Thesis Reader**
Lynn H. Matthias Professor in Engineering
Department of Electrical and Computer Engineering
University of Wisconsin-Madison

Table of Contents

Certification 1

Abstract 3

Table of Contents 4

Acknowledgements 8

1 Introduction 9

2 Gene synthesis 14

2.1 10^0 to 10^2 bases: short oligonucleotides by chemical synthesis 14

2.2 10^2 to 10^3 bases: single genes by *in vitro* enzymatic synthesis 16

2.2.1 Factors and tradeoffs for synthesis of short genes 19

2.2.2 Parsing Software 20

2.3 10^3 to 10^4 bases: Multiple genes and viral genomes by *in vitro* synthesis and cloning 22

2.3.3 Multiple genes 23

2.3.2 Viral Genomes and vaccines 24

2.3.3 Construction Methods 25

2.4 10^4 to 10^6 bases: gene clusters, bacterial genomes by cloning 26

2.5 Error Correction 27

3 Microfluidics and biotechnology 28

3.1 Why microfluidics 28

3.2 Technology for fluid manipulation 28

3.2.1 Elastomeric valves 29

3.2.2 Hard/soft hybrid valving 29

3.2.3 Passive systems 30

3.3 Fouling and material choice 31

3.4 This work 32

4 Microfluidic gene synthesis: integrated road map 35

4.1 Power of integration in microfluidics 36

5 Parallel gene synthesis in a microfluidic device 38

5.1 Introduction 39

5.2 Materials and Methods 40

5.2.1 Master mold fabrication 41

5.2.2 Microfluidic device fabrication. . . . 42
5.2.3 Passive systems. . . . 42
5.2.4 PCA reaction mixtures. . . . 43
5.2.5 PDMS microchannel preparation. . . . 43
5.2.6 Sample evaporation. . . . 44
5.2.7 Device design and operation. . . . 45
5.2.8 Control experiments. . . . 46
5.2.9 Polyacrylamide gel electrophoresis. . . . 46
5.2.10 DNA sequencing. . . . 47
5.3 Results. . . . 48
5.4 Discussion. . . . 53
6 Direct microarray gene synthesis. . . . 57
6.1 Introduction. . . . 57
6.1.1 Hybrid microfluidic-microarray device fabrication: challenges. . . . 58
6.2 DNA Microarray platforms 59
6.3 Microfluidic-microarray integration 60
6.3.1 Bonding. . . . 61
6.3.2 Device Design. . . . 61
6.3.3 Alignment. . . . 62
6.4 Oligonucleotide cleavage. . . . 64
6.4.1 Type II restriction enzymes. . . . 66
6.4.2 USER mix, UDG and endo IV. . . . 70
6.5 Single pot cleavage and synthesis *in vitro* and *in fluidic*. . . . 73
6.6 Microfluidic-microarray direct cleavage of oligos 75
7 Integrated microfluidic gene and protein synthesis. . . . 80
7.1 Introduction 82
7.2 *In vitro* gene and protein synthesis. . . . 85
7.2.1 Fluorescent Protein Selection. . . . 86
7.2.2 Cell-free protein synthesis with circular and linear templates. . . . 88
7.2.3 Design of linear FP gene expression constructs. . . . 90
7.2.4 Analysis of *in vitro* gene and protein synthesis by PAGE and fluorometer. . . . 93
7.3 Gene and protein synthesis in microfluidic devices. . . . 95

7.3.1 Design and fabrication of microfluidic devices for gene and protein synthesis. 99
7.3.2 Gene and protein synthesis in microfluidic devices. 100
7.4 Integrated gene and protein synthesis in a microfluidic device. 103
8 Hierarchical gene synthesis in a microfluidic device. 107
9 Concluding remarks. 108
A Appendix. 111

Acknowledgements

JP Urbanski;

Kurt Broderick;

Scott Manalis;

Kim Hamad-Schifferli;

Shuguang Zhang;

Molecular Machines group;

Linda Peterson;

UROPS: Lu Chen, Eddie Liu, Kelly Chang

Friends: Giles, Eric Rosenbaum, Chet, Diya, Victor, Hsindy,

2 Gene Synthesis: applications and methods

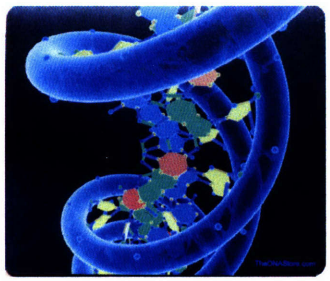


Figure 2.1: Double-stranded B-DNA. From TheDNAStore.com

This chapter of the thesis overviews two major topics: (1) current and envisioned applications for synthetic DNA over orders of magnitude of construct size, and (2) an evaluation of the various technologies that have been used for the assembly of such constructs to date. Put simply, what are the uses of synthetic DNA, and how do you build it?

Additionally, I will briefly discuss several complementary technologies in sections 2.2.3 and 2.4, namely gene parsing (i.e. the software-based design of short, chemically synthesized oligonucleotides used in gene synthesis), and DNA error correction. I will also introduce the role of high-density DNA microarrays in section 2.1.1. These foundational technologies have played, and will continue to play, significant roles in increasing the availability of cheap, accurate, and long synthetic DNA.

2.1 10^0 to 10^2 bases: short oligonucleotides by chemical synthesis

Short, chemically synthesized oligonucleotides are indispensable tools in the tool kit of modern biology. As related to the work of this thesis, short oligos serve primarily as a set of building materials for the construction of larger synthetic DNA constructs, gene-length and longer. They are the bricks of DNA manufacturing.

Modern synthesis of DNA is based on phosphoramidite chemistry originating from the work of Caruthers in the 1980s[1], and is now a widespread, commercially available service. Short

[1] Beaucage, S.L. & Caruthers, M.H. in Bioorganic Chemistry: Nucleic Acids. (ed. S.M. Hecht) 36-74 (Oxford University Press, Oxford; 1996).

oligonucleotides of approximately 100 bases in length can be designed and ordered on a computer, and physical molecules can arrive by mail overnight. Achieving the ease and relative low cost of this process is, in the short term, a modest and, by some metrics, an already realized goal for both academic and commercial entities working on the synthesis of constructs gene-length and longer. At the writing of this thesis, synthetic genes ordered from Codon Devices, a commercial synthetic DNA vendor, costs ____.

Synthesizing short oligonucleotides is accomplished via traditional DNA phosphoramidite chemistry[1], whereby chemically protected nucleotides are added one at a time in a linear polymerization reaction to a solid support (e.g. controlled-pore glass (CPG)). After each nucleotide base addition, acid is utilized to cleave the protecting group, leaving a reactive hydroxyl group to which subsequent protected bases can couple. This process is shown schematically in Figure 2.2.

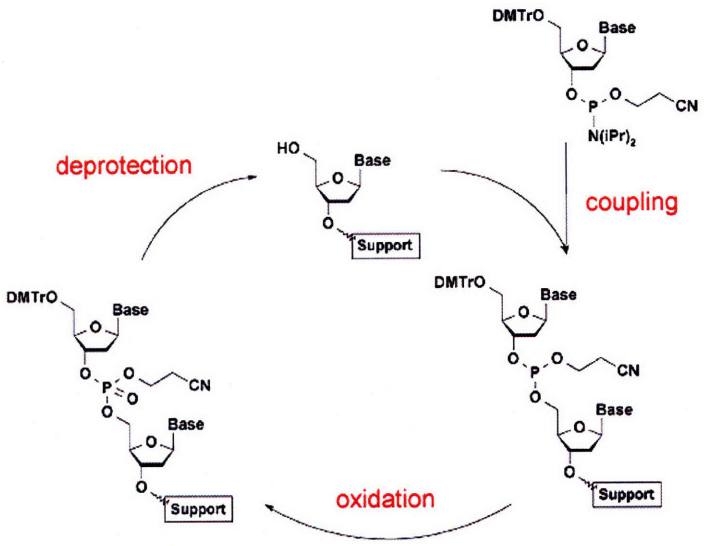


Figure 2.2: Phosphoramidite synthesis cycle.

The phosphoramidites themselves are derived from sugar cane[2] and are commercially available. Four bottles containing phosphoramidites, as sold by Glen Research, are shown in Figure 2.3a. It is remarkable to consider that this purified chemistry provides the set of synthetic small molecules from which entire synthetic bacterial genomes can be constructed. The entire process

[2] Y. Sanghvi (2007). A Roadmap to the Assembly of Synthetic DNA from Raw Materials, http://hdl.handle.net/1721.1/39657

of chemical synthesis is automated utilizing commercially available DNA synthesis machines, such as the one shown in Figure 2.3b.

(A) (B)

Figure 2.3: (a) Phosphoramidites from Glen Research. **(b)** An ABI 394 synthesizer from Applied Biosystems.

As building materials that, when properly assembled, yield constructs that encode biological programs, minimizing errors in the synthesis of short oligonucleotides is of the utmost importance. As shall be seen throughout this thesis, the error rate of each individual DNA fabrication process contributes overall to the level of effort and difficulty of assembling perfect constructs of the desired size; thus, reduction of error in any of the synthesis steps described, particularly high-error-rate processes, is always welcome (although, as we shall see later in section **2.2**, high error rates in chemical synthesis or subsequent steps may prove acceptable if error *correction* is sufficiently robust).

Currently, errors in final synthetic DNA products arise largely from phosphoramidite synthesis, rendering it the most fallible step in the manufacturing process. The most common source of error from chemical synthesis arises when the protecting group of a growing DNA strand is not removed, resulting in either a truncated product (when no further addition occurs) or an internal deletion (when in subsequent rounds addition occurs)[3]. Other types of errors include depurination and various types of DNA damage resulting from chemical treatment. The efficiency of the protecting group removal step is ~99%; thus, the final crude yield is $\sim 0.99^{N} * 100\%$, where N is the number of bases in the final strand. Figure 2.4, from Stewart et al.[4], nicely summarizes the impact of even small alterations in step-wise yield upon the final population of full-length desired

[3] For example, see Table 5.1 for a comparison of the frequency of various error types for gene synthesis in microfluidic and macroscopic environments. Deletions account for 83% and 70% of the total types of errors observed, respectively.

[4] Stewart, L. and Burgin, A.B. (2005) Whole Gene Synthesis: A Gene-O-Matic Future. *Frontiers in Drug Design & Discovery*, **1**, 297-341.

species relative to species with deletions (i.e. N-minus 1 mer). As we can see, even slight variations in the efficiency of chemical reactions (here comparing 99.0% to 99.5% removal efficiency of the protecting group) can have a significant impact upon both the quality and quantity of the synthesized oligonucleotides. The cumulative effect of these minor inefficiencies renders the process of constructing a high percentage of full-length oligonucleotides larger than even 100 bases extremely difficult. These deficiencies in the synthesis process ultimately limit the size of the oligonucleotides that can be utilized for gene synthesis, as even modest quantities of low-quality oligos (present in larger proportion for longer oligos) can have considerable impact upon the probability of fabricating a perfect DNA construct. As this probability decreases, the subsequent resources (e.g. cloning, sequencing) that must be devoted to producing this error free gene similarly increases. As we shall see in section 2.2.2, the enzymatic methods utilized to assemble these short oligos into gene-length constructs necessitates minimum oligo sizes as well, thus high-lighting the importance of stringency and accuracy in the oligo synthesis process itself.

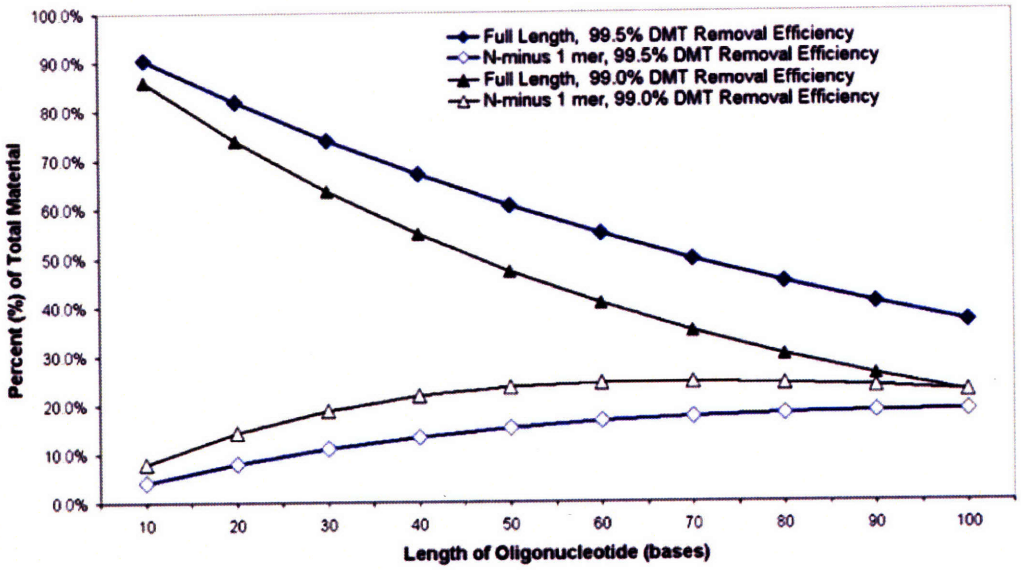


Figure 2.4: Effects of DNA strand length and chemical reaction efficiency in oligonucleotide synthesis. From Stewart et al.[4]

Finally, when built massively in parallel on a solid support (e.g. microscope slide) via DNA microarray technology, short, chemically synthesized oligonucleotides provide a wealth of cheap construction material for gene synthesis. Just as the advent of microarray technology made possible the dramatic leap from low to high throughput for gene expression assays, thus revealing new vistas of research and analysis, it is hoped that the use of microarrays for gene synthesis will have a similar groundbreaking effect, enabling levels of synthesis throughput heretofore

unattainable. Efforts in employing DNA microarrays in concert with microfluidic devices will be touched upon in Chapter 5 and explored more deeply in Chapter 6.

2.2 10^2 to 10^3 bases: single genes by *in vitro* enzymatic synthesis

Given the capacity to easily obtain single genes (which covers a vast range of construct size; constructs from hundreds to even tens of thousands of base pairs can be considered "single genes." The focus of this work is on single genes approximately one thousand base pairs long or shorter), myriad applications are possible, from the synthesis of libraries of genes for large-scale combinatorial studies of synthetic protein designs[5], the reengineering of proteins for improved crystallization[6], to enabling novel projects such as the complete synthesis of a library of canine olfactory receptors (OR) genes[7]. Currently, these projects are considerably difficult to pursue given the extensive requirements in time and resources devoted simply to manipulating and manufacturing large sets of desired DNA constructs. Readily available cheap, on-demand, synthetic DNA would significantly enable such research efforts.

Synthesis of genes on the order of hundreds of base pairs is readily accomplished utilizing enzyme-mediated methods, which can broadly be divided into those directed by DNA ligase or DNA polymerase for construct assembly from oligos. In both cases, a heterogeneous mixture of products is generated, so polymerase is typically used to subsequently amplify synthesized genes to desired quantities via polymerase chain reaction (PCR). Prominent examples of gene synthesis using both construction methods have been reported with steady advances in technique since the work of Khorana et al.[8], who famously synthesized the yeast alanine tRNA gene by ligating short 6 to 8 base oligos.

Gene synthesis by ligase or polymerase mediated assembly is based fundamentally on the same paradigm, whereby the ability of oligonucleotides to form hydrogen-bonded Watson-Crick double-stranded DNA is exploited to initially hybridize overlapping short oligos. The process of designing the oligonucleotides given a starting DNA sequence-taking into account, for example,

---

5 Hecht;

6 Dyda et al.; Crystal structure of the catalytic domain of HIV-1 integrase; similarity to other polynucleotidyl transferases. *Science* **266** 1981-1986.

7 Shuguang Zhang, DARPA RealNose project

8 Khorana, H.G. (1968) Nucleic acid synthesis in the study of the genetic code, in Nobel Lectures: Physiology or Medicine (1963-1970). Elsevier Science Ltd, Amsterdam, pp. 341-369.

oligo length and positioning relative to other oligos-is referred to as "parsing," and will be discussed in greater depth in sections 2.2.2 and 2.2.3. Once annealed, ligase can be employed to covalently link the ends of each oligo, or alternatively, polymerase can be utilized to extend overlapping pairs of oligos, yielding the desired full-length construct after multiple rounds of extension via thermocycling (as in PCR). Two examples of gene synthesis mediated by ligase and polymerase, respectively, are shown schematically in Figures 2.5 (a) and (b). The method of assembly shown in 2.5(a), from Bang and Church[9], relies initially upon a stringent, high-temperature annealing condition (70°C) and thermostable ligase to covalently join a pool of 5'-phosphorylated oligos. The use of thermostable ligase builds upon the work of Au et al.[10], who similarly employed high temperature annealing to avoid the formation of secondary structure and mismatches that occurs at lower temperatures (e.g. 37°C for T4 DNA ligase), and were prominent during early gene synthesis work prior to the advent of PCR (e.g. Khorana et al.[8]). Two selection steps utilizing exonucleases and endonucleases are employed to purify the desired species before PCR is used to amplify the final product.

The method from Stemmer et al.[11], shown in figure 2.5 (b), utilizes polymerase for the assembly of gene-length constructs from short oligos. This protocol similarly readily yields full-length genes, beginning again with the initial annealing of oligos based upon designed Watson-Crick base-pairing, followed by multiple rounds of thermocycling with polymerase. There are, however, two important distinctions between the ligase and polymerase mediated methods: firstly, the oligos used with polymerase do *not* require 5'-phosphorylation; thus, the additional resources and effort to phosphorylate the 5'-terminus of each oligo, either during chemical synthesis or prior to assembly via polynucleotide kinase, is not required. Eliminating this additional step removes the associated resources required for phosphorylation, which can be expensive when performed during chemical synthesis, and also removes another unnecessary source of error or inefficiency during biochemical addition prior to assembly. Secondly, polymerase is able to fill in gaps *between* oligos, meaning they are not required to be designed such that they meet end to end as required by ligase. This design flexibility provides additional degrees of freedom for any software tool utilized in parsing the synthetic gene and designing

---

[9] Bang, D. and Church, G. (2007) Gene synthesis by circular assembly amplification. *Nature Methods*, **5**, 37-39.

[10] Au, L.C., Yang, F.Y., Yang, W.J., Lo, S.H., and Kao, C.F. (1998) Gene Synthesis by a LCR-Based Approach: High-level Production of Leptin-L54 Using Synthetic Gene in *Escherichia coli*. *Biomedical and Biophysical Research Communications* **248** 200-203.

[11] Stemmer, W.P., Crameri, A., Ha,K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligonucleotides. *Gene*, **164**, 49-53.

oligos. Thus, these two factors have led our research group to favor the use of polymerase over ligase as the primary engine to drive gene synthesis.

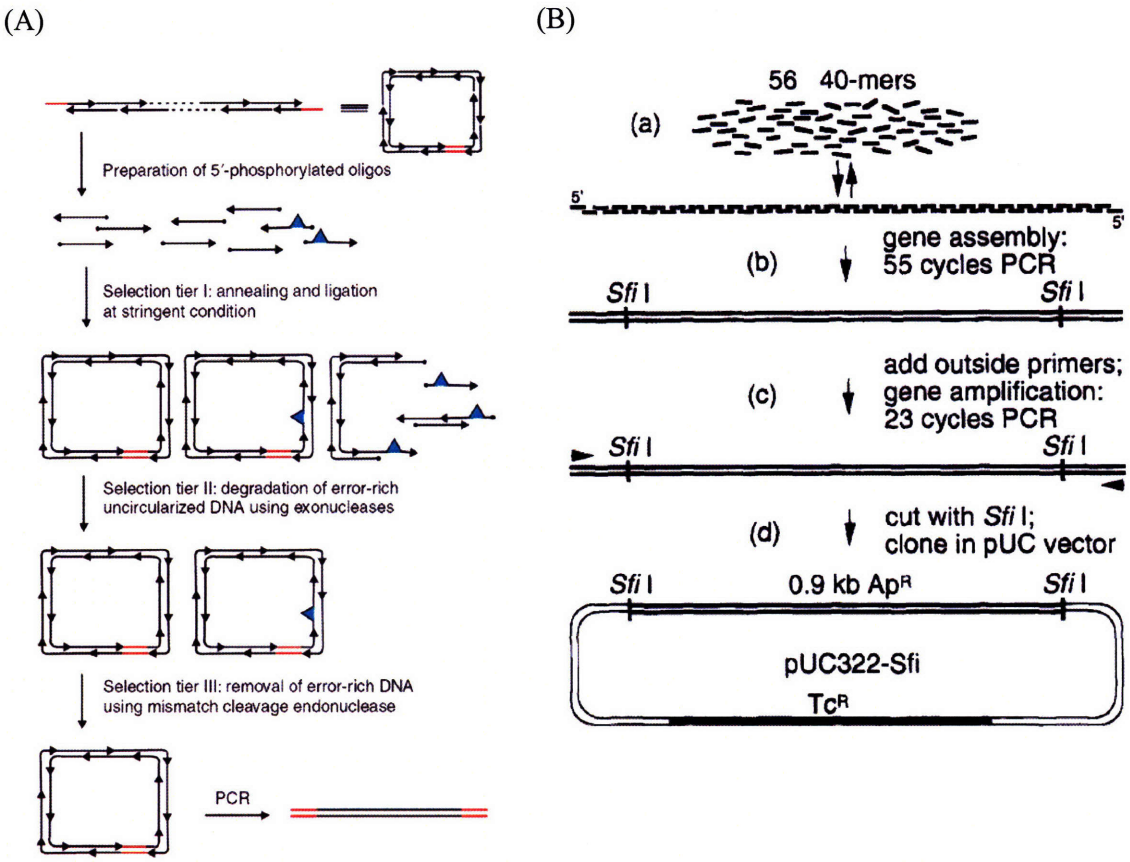


Figure 2.5: Schematics for gene synthesis mediated by **(a)** ligase, from Bang and Church[9] and by **(b)** polymerase, from Stemmer et al.[11].

As we shall see, these techniques for construct assembly on the order of hundreds of base pairs will prove to be foundational for the fabrication of longer constructs, up to complete bacterial genomes. The techniques employed to assemble any long construct is ultimately built from intermediate pieces on this size-scale, and as I will discuss in sections 2.3 and 2.4, a hierarchical approach is generally employed for synthesis of long DNA. The miniaturization into microfluidic format of this hierarchical synthesis approach for long DNA molecules is the subject of Chapter 8.

As a final point regarding nomenclature, multiple terms have been coined to describe polymerase-mediated synthesis of genes. Our research group has settled on the term "Polymerase

Construction and Amplification" (PCA), as first named by Mullis et al[12], and I will use this throughout this thesis when describing polymerase-mediated synthesis of genes, which is the primary technique utilized in this work.

2.2.1 Factors and tradeoffs for synthesis of short genes

In designing the oligos that will utilized for gene synthesis, a number of factors must be carefully considered in order to best optimize the assembly reaction. As mentioned in section 2.1, oligonucleotide length is a crucial factor: if oligos are too long, errors from chemical synthesis can reach unacceptable levels, (see figure 2.4), leading to poor synthesis yield, or in some cases incomplete assembly. If oligos are too short, the overlapping regions may not be sufficiently long to ensure stable priming. Typical overlap length is approximately the length used for PCR primers, e.g. 15 to 20 bases, depending upon the composition of the oligo and thus annealing temperature. Longer overlaps generally promote hybridization specificity, and are thus desirable. As already discussed in reference to the described ligase-based assembly techniques, the overall annealing temperature of each oligo pair should be kept as high as possible to again maximize hybridization specificity and reduce undesired oligo interactions. Annealing temperatures in the range of 60 to 70°C are preferred. Similarly, annealing temperatures should be kept homogeneous throughout the designed pairs of oligo interactions; non-homogeneous annealing will ultimately decrease the synthesis efficiency.

Another crucial factor is the number of oligos per assembly reaction. As the complexity of the oligo pool increases, the likelihood of mispriming and other undesirable interactions increases, thus inhibiting assembly or ultimately preventing assembly entirely. Additionally, for ligase-based synthesis, large numbers of oligos lead to greater numbers of ligations per assembly reaction. As the number of oligos is reduced, however, of course only smaller final constructs can be manufactured (depending upon the length of the oligos). As we shall see in sections 2.3 and 2.4, constructs of approximately 500 bp or smaller are widely reported to be of an appropriate size for subsequent manipulations when building large structures, while in the work reported in this thesis, constructs of approximately 1 kb in size can be readily assembled in a single reaction. Increasing the size of constructs that can be built per reaction can potentially reduce the total

---

[12] Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, **51**, 263-273.

number of reactions necessary when building large constructs. The factors discussed are summarized in Table 2.1 below.

TABLE 2.1: Summary of several factors that impact gene synthesis

Factor	Commentary
Oligo length	Too short, overlaps not long enough for stable priming; too long, will have errors from chemical synthesis
Annealing (melting) temperature (T_m)	High as possible to maximize hybridization specificity, homogeneous as possible to maximize PCR efficiency
Number of oligos per assembly reaction	Too many, likelihood of mispriming increases, preventing assembly; smaller pools lead to smaller final products

Numerous other factors ultimately affect the quality of gene synthesis. The impact of additional factors, such as oligo vendor and polymerase choice, are characterized and reported in Carr et al.[13]

2.2.2 Parsing Software

Given the variety of factors that ultimately impact the quality of gene synthesis, a diverse set of useful software tools have been developed to optimize these parameters in assisting researchers in oligonucleotide design. Parsing genes, or, taking the DNA sequence of a desired gene and partitioning the sequence into overlapping oligos (for both strands), can be readily accomplished using a number of these tools. The genes synthesized in this thesis were parsed utilizing DNA Works, developed by Hoover et al.[14] The parsing software available for gene synthesis is summarized and mapped according to publication date and citation in Wu et al.[15]. The relevant figure is reproduced in figure 2.6, and the reader should investigate Wu et al. for the complete list of references. Genes can also be parsed without the aid of such tools (i.e. "naively"), by simply setting an oligo length (e.g. 40 bases) and overlap length (e.g. 20 bases), thus yielding the desired oligos "by hand," without optimizing any of the discussed parameters. This method has been

---

[13] Carr et al. (2008) Practical Gene Synthesis. Manuscript in preparation.

[14] Hoover, D.M. and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30,** e43.

[15] Wu, G., Dress, L., and Freeland, S.J. (2007) Optimal encoding rules for synthetic genes: the need for a community effort. *Mol Sys Bio* **3**:134

2.3.1 Multiple genes

Easy accessibility to DNA constructs on the order of thousands of base pairs will ultimately enable some of the most exciting applications of synthetic biology. Effective microbial engineering for drug[17,18] energy[19,20], and material[21] production ultimately requires a more complete understanding of the complex processes of cells, and given constructs on this size scale systematic experimentation and analysis of gene and multiple gene constructs could be conducted, allowing researchers to unravel and optimize their behavior. A fine example of this type of experimentation is the work done by Elowitz et al.,[22] where they constructed a synthetic oscillating network which they termed a "repressilator" via 'rational network design.' The motivation of this type of work is again two-fold: firstly, in the process of building synthetic systems, can we learn something about natural occurring systems? And secondly, can we get good enough at building synthetic systems so we can engineer new cellular behaviors? The DNA machinery driving repressilator function is shown in figure 2.7, and as in the common refrain, required significant resources and effort simply to manufacture, highlighting yet again the potential enabling power of DNA synthesis on the size-scale of thousands of bases.

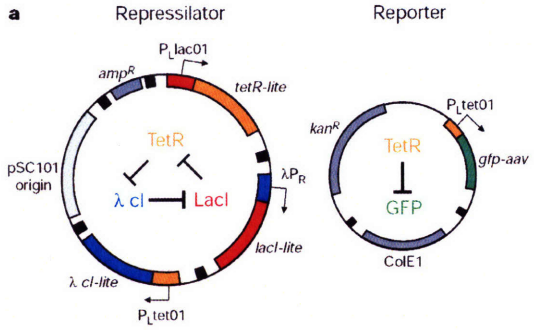


Figure 2.7: (a) A schematic of a repressilator. From Elowitz and Leibler.[22]

---

17 R, Keasling JD. 2006. Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast. *Nature 440*: 940-943.

18 Martin, V.J., Pietera, D.J., Withers, S.T., Newman, J.D. and Keasling, J.D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, **21**, 796-802.

19 Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated Bioprocessing of Cellulosic Biomass: An Update. *Current Opinion in Biotechnology 16*: 577-583

20 United States Department of Energy, Biological and Environmental Research Advisory Committee. 2004. *Synthetic Genomes: Technologies and Impact.* http://www.sc.doe.gov/ober/berac/SynBio.pdf

21 Aldor IS, Keasling JD. 2003. Process Design for Microbial Plastic Factories: Metabolic Engineering of Polyhydroxyalkanoates. *Current Opinion in Biotechnology 14*: 475-483.

22 Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335-338.

This process of design (rational or naive), synthesis, and characterization is a concept that will be applied for the motivation of the work in Chapter 7 regarding the microfluidic synthesis of DNA constructs followed by protein synthesis.

2.3.2 Viral Genomes and vaccines

As a special class of constructs on the size-scale of thousands of base pairs, viral genomes have also recently been synthesized, with some fanfare[23, 24]. Significantly, these demonstrations, while illustrative of some of the dangerous applications of gene synthesis, also point toward the great promise that synthesis on this size-scale has for vaccine development[20]. Given the outbreak of a new virus, once sequenced, any laboratory in the world with gene synthesis capabilities could produce vaccines even without the original infectious organism. Again, large variant libraries could also be produced to create vaccines with broad immune responses against similarly diverse viruses like human immunodeficiency virus (HIV) and hepatitis C.

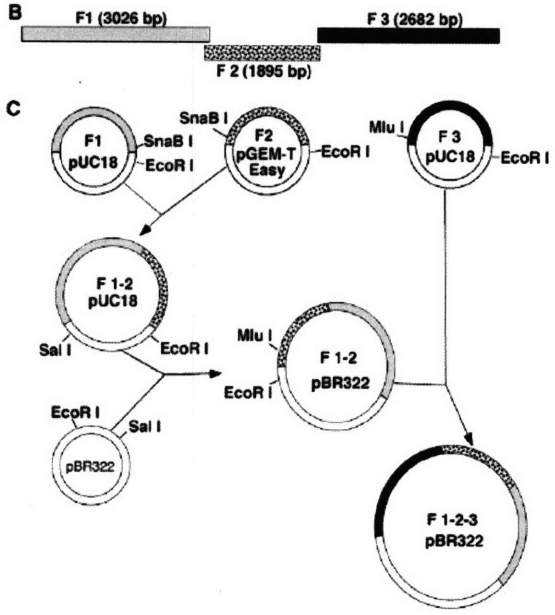


Figure 2.7: Schematics for gene synthesis mediated by **(a)** ligase, from Bang and Church**[ref]** and by **(b)** polymerase, from Stemmer et al.[REF].

---

[23] Cello, J., Paul, A.V., and Wimmer, E. (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, **297**, 1016-1018.

[24] Smith, H.O., Hutchison, C.A., 3rd, Pfannkoch, C., and Venter, J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci U S A*, **100**, 15440-15445.

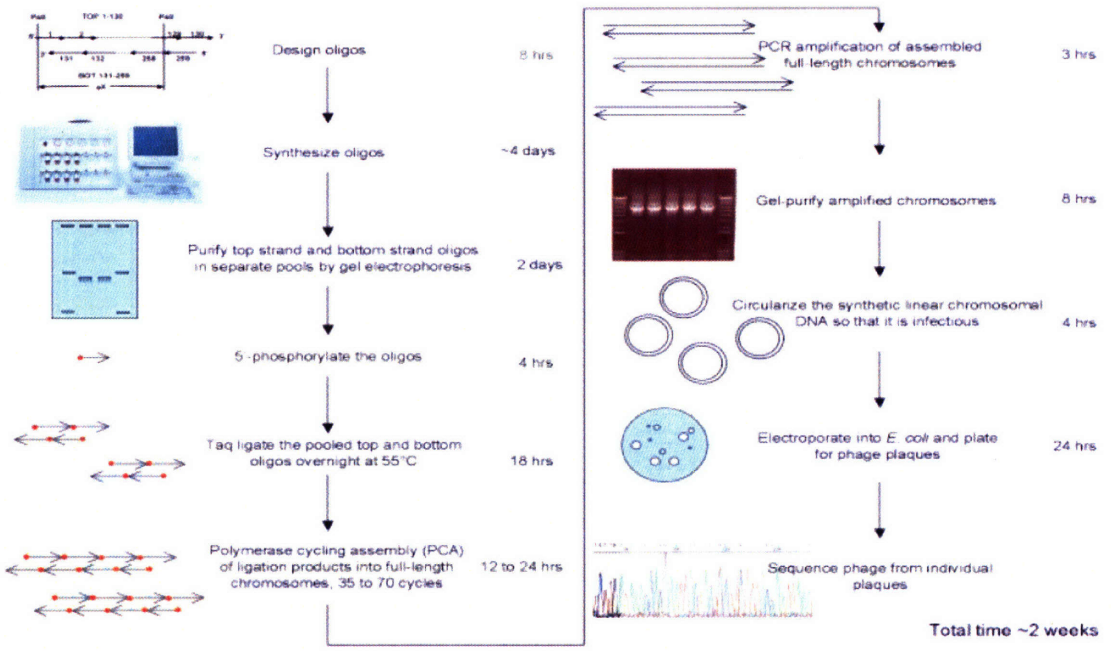


Figure 2.8: Effects of DNA strand length and chemical reaction efficiency in oligonucleotide synthesis. From Stewart et al.

5 Parallel gene synthesis in a microfluidic device

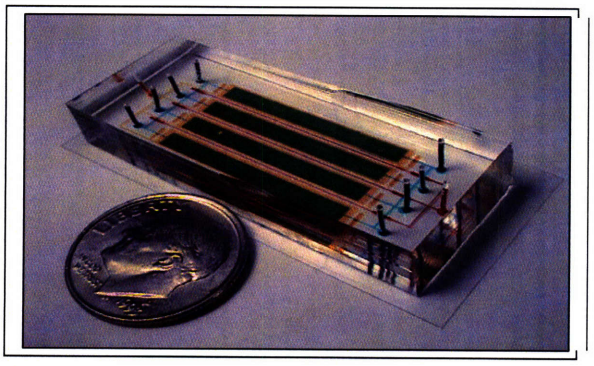


Figure 5.1: Optical image of a multi-layer PDMS device used for microfluidic gene synthesis

The work in this chapter was published in "D.S. Kong, P.A. Carr, L. Chen, S. Zhang, J.M. Jacobson, 'Parallel gene synthesis in a microfluidic device,' *Nucleic Acids Research*, vol. 35, no. 8, e61, Apr, 2007."

5.1 Introduction

It has long been recognized that the capacity to design and synthesize genes and longer DNA constructs can be enabling to a broad cross section of applications within molecular biology[1] including the study of large sets of single genes[2], the design of genetic circuitry[3], the engineering of entire metabolic pathways for target molecule manufacture[4], and even the construction and re-engineering of viral and bacterial genomes[5,6,7].

---

1 Khorana, H.G. (1968) Nucleic acid synthesis in the study of the genetic code, in Nobel Lectures: Physiology or Medicine (1963-1970). Elsevier Science Ltd, Amsterdam, pp. 341-369.

2 The MGC Project Team (2004) The status, quality, and expansion of the NIH full-length cDNA project: the mammalian gene collection (MGC). *Genome Res.*, **14**, 2121-2127.

3 Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335-338.

4 Martin, V.J., Pietera, D.J., Withers, S.T., Newman, J.D. and Keasling, J.D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, **21**, 796-802.

5 Cello, J., Paul, A.V., and Wimmer, E. (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, **297**, 1016-1018.

6 Smith, H.O., Hutchison, C.A., 3rd, Pfannkoch, C., and Venter, J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci U S A*, **100**, 15440-15445.

7 Hutchison, C. A., III, Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. and Venter, J. C. (1999) Global Transposon Mutagenesis and a Minimal Mycoplasma Genome. *Science* **286,** 2165–2169.

The core technology for custom DNA synthesis centers on the assembly of pools of oligonucleotides (oligos), typically less than 50 nucleotides in length, into increasingly larger DNA molecules. These oligos, hereafter referred to as "construction oligos," are synthesized by variations of phosphoramidite chemistry[8], and are the building blocks for the different gene synthesis techniques developed thus far. The most widely reported methods for building long DNA molecules involve variations of the polymerase-mediated assembly technique shown in Figure 5.2, collectively termed Polymerase Construction and Amplification (PCA)[9,10]. Here, much like in the more conventional Polymerase Chain Reaction (PCR), three temperature steps are employed to denature, anneal, and elongate the various overlapping construction oligos until, after multiple rounds of thermocycling, the desired full length DNA construct is obtained. Furthermore, assembly and amplification can be performed in a single reaction with the introduction of amplifying primers[11]. Thus, once a minute quantity of full length product is assembled, this product is amplified as per PCR. Using such polymerase-mediated techniques, researchers have successfully synthesized DNA constructs as large as 12 kb[12] and 15 kb. A PCA process was also employed as the first step in generating a 32 kb DNA construct by Santi and co-workers[13]. In addition, significant progress has been made in correcting synthesis errors, which originate primarily from the phosphoramidite synthesis of initial oligonucleotide building blocks. The use of protein-mediated error correction has been effective in increasing the accuracy of synthetic DNA[14,15,16], with error rates as low as 1 per 10,000 base pairs reported.

---

[8] Caruthers, M.H. (1985) Gene synthesis machines: DNA chemistry and its uses. *Science*, **230**, 281-285.

[9] Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, **51**, 263-273.

[10] Stemmer, W.P., Crameri, A., Ha,K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligonucleotides. *Gene*, **164**, 49-53.

[11] Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432**, 1050-1054.

[12] Xiong, A.S., Yao, Q.H., Peng, R.H., Duan, H., Li, X., Fan, H.Q., Cheng, Z.M., and Li, Y. (2006) PCR-based accurate synthesis of long DNA sequences. *Nature Protocols*, **1**, 791-797.

[13] Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M., and Santi, D.V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA*, **101**, 15573-15578.

[14] Carr, P.A., Park, J.S., Lee, Y.J., Yu, T., Zhang, S. and Jacobson, J.M. (2004) Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res.*, **32**, e162.

[15] Binkowski, B.F., Richmond, K.E., Kaysen, J., Sussman, M.R., and Belshaw, P.J. (2005) Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res.*, **33**, e55.

[16] Fuhrmann, M., Oertel, W., Berthold, P., and Hegemann, P. (2005) Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage. *Nucleic Acids Res.*, **33**, e58.

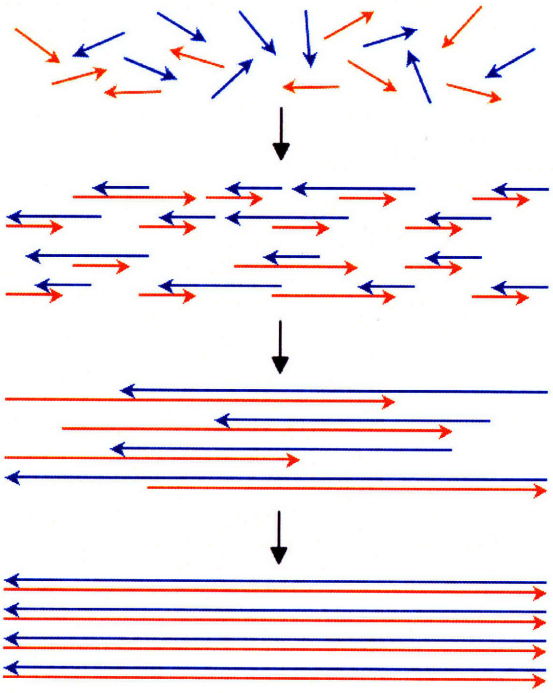


Figure 5.2: Schematic for gene synthesis by Polymerase Construction and Amplification (PCA). Multiple rounds of oligo annealing and extension by DNA polymerase generate successively longer DNA assemblies from a starting pool of construction oligos, typically < 50 nt, until the full-length gene is produced. The pool of heterogeneous DNA products is enriched for the full-length species by amplification in a separate subsequent reaction, or in the same reaction by including amplifying primers in the reaction mixture.

Despite these promising results significant challenges remain, most significantly the cost and time of synthesizing long constructs. Currently, while conventionally synthesized oligos are available at a cost on the order of $0.1 dollars per nucleotide, the cost for custom gene synthesis services is significantly higher, on the order of $1.00-$1.60 dollars per base pair, with the major expenditure components for such long syntheses being attributable to reagent and sample handling. Microfluidic technology provides an elegant means to overcome these limitations. By scaling reactions down to volumes of less than a microliter, reagent costs can be substantially reduced[17]. Furthermore, microfluidic technology enables highly parallelized synthesis along with the potential for automated sample handling and process integration.

[17] Liu, J., Hansen, C., and Quake, S.R. (2003) Solving the "world-to-chip" interface problem with a microfluidic matrix. *Anal. Chem.*, **75**, 4718-4723.

In this paper we report what is to our knowledge the first gene synthesis conducted in a microfluidic environment. We have successfully conducted synthesis and amplification in a single reaction for a variety of genes and gene segments, including GFP, OR128-1, DsRed, ble (bleomycin resistance), a Holliday junction cleavase (hjc) gene from the bacteriophage *SIRV-1*, and a variant alba gene from *S. solfataricus*. The identity of all synthetic genes was verified by sequencing, and extensive sequencing of DsRed enabled the determination of an error rate for genes synthesized in a microfluidic environment, along with a comparison of error rates for genes synthesized in standard PCR tubes. In other reports construction oligos were synthesized on the microscale, cleaved from the surface and subsequently assembled in macroscopic ($\geq$ 5 µl) reactions[11,18,19]. In contrast, we have synthesized these DNA constructs in parallel within four 500 nanoliter reactors of a microfluidic device. Furthermore, the minute construction oligo concentrations utilized (10-25 nM each oligo) are significantly lower than concentrations attainable (without amplification) from high density oligonucleotide microarrays. Thus, such a microfluidic approach should be compatible with DNA microarray-derived oligonucleotides[11], further reducing the cost of this crucial reagent.

5.2 Materials and Methods

5.2.1 Master mold fabrication

Devices utilized in this work employed "push-down" valve geometries for fluidic valve actuation[20]. Two master molds were fabricated, one from which the fluidic "flow layer" could be cast, the other from which the fluidic "control layer" could be cast. The flow layer master was fabricated by first rinsing a 4" silicon wafer (WaferNet) in acetone and isopropyl alcohol, followed by wafer dehydration at 200°C on a hot plate. Next, hexamethyldisilizane (HMDS, Sigma) was spun on the wafer at 4000 rpm to promote adhesion of the photoresist. A layer of

---

[18] Richmond, K.E., Li, M.H., Rodesch, M.J., Patel, M., Lowe, A.M., Kim, C., Chu, L.L., Venkataramaian, N., Flickinger, S.F., Kaysen, J., Belshaw, P.J., Sussman, M.R., and Cerrina, F. (2004) Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.*, **32**, 5011-5018.

[19] Zhou, X., Cai, S., Hong, A., You, Q., Yu, P., Sheng, N., Srivannavit, O., Muranjan, S., Rouillard, J.M., Xia, Y., Zhang, X., Xiang, Q., Ganesh, R., Zhu, Q., Matejko, A., Gulari, E., and Gao, X. (2004) Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.*, **32**, 5409-5417.

[20] Unger, M.A., Chou, H.P., Thorsen, T., Scherer, A., and Quake, S.R. (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. Science. **288**, 113-6.

AZP 4620 positive photoresist (AZ Electronic Materials) was then coated at 1500 rpm for 40s followed by a one hour soft-bake at 90°C. Upon completion of the soft-bake, the wafer was then exposed for 20 seconds at 50% intensity using a UV floodlight (Uvitron, Int.), followed by development. Next, the resist was placed on a hotplate at 150°C for 1 minute to reflow the resist and achieve rounded fluid channels, thus enhancing sealing during valve actuation.

The control layer master was fabricated by again employing a solvent wash followed by wafer dehydration. A layer of SU-8 50 negative photoresist (MicroChem) was then coated at 1000 rpm, followed by pre-exposure bake steps of 65°C for 10 minutes and 95°C for 30 minutes. The resist was then exposed for 40 seconds at 50% intensity and post-exposure baked at 65°C for 1 minute and 95°C for 10 minutes before being developed.

Finally, both flow layer and control layer masters were briefly exposed to Chlorotrimethylsilane (Sigma) vapors for several minutes to promote release of the elastomer from the master molds. All transparency masks used for the various exposure steps were designed in Adobe Illustrator and printed by PageWorks (Cambridge, MA).

5.2.2 Microfluidic device fabrication

Approximately 30 grams of liquid PDMS pre-polymer (GE, RTV 615) at a component A to B ratio of 5:1 was poured onto the control layer master to a thickness of approximately ~1 cm, followed by partial curing in a convection oven at 80°C for 45 minutes. Liquid PDMS pre-polymer at a component A to B ratio of 20:1 was coated onto the flow layer master at 2000 rpm for 60 seconds and also partially cured at 80°C for 45 minutes. The PDMS control layer was then peeled from its master and individual devices were cut out with a razor blade. Holes for control line inlet ports were cored with an 18 G needle whose tip had been beveled and sanded down for clean coring. Next, control layer devices (typically 6 per wafer) were aligned and bonded to the PDMS-coated flow layer master, followed by additional curing for 45 minutes at 80°C. These two-layer devices were then cut and peeled off the flow-layer molds, cored, and bonded overnight at 80°C to 1 mm thick glass cover slips coated with a thin layer of partially cured PDMS (typically spun on at 2000 rpm for 40s, with a 20:1 polymer to curing agent ratio and cured at 80°C for 45 minutes).

An example of a three-layer PDMS device capable of parallel gene synthesis is shown in figure 2. Colored food dyes are used to emphasize various features of the device, with red indicating actuation lines in the PDMS control layer, blue (and green) indicating the four gene synthesis reactors, and yellow indicating a mesh of fluid lines in the control layer, hereafter referred to as a 'water jacket', placed above the reactors to minimize sample evaporation during thermocycling.

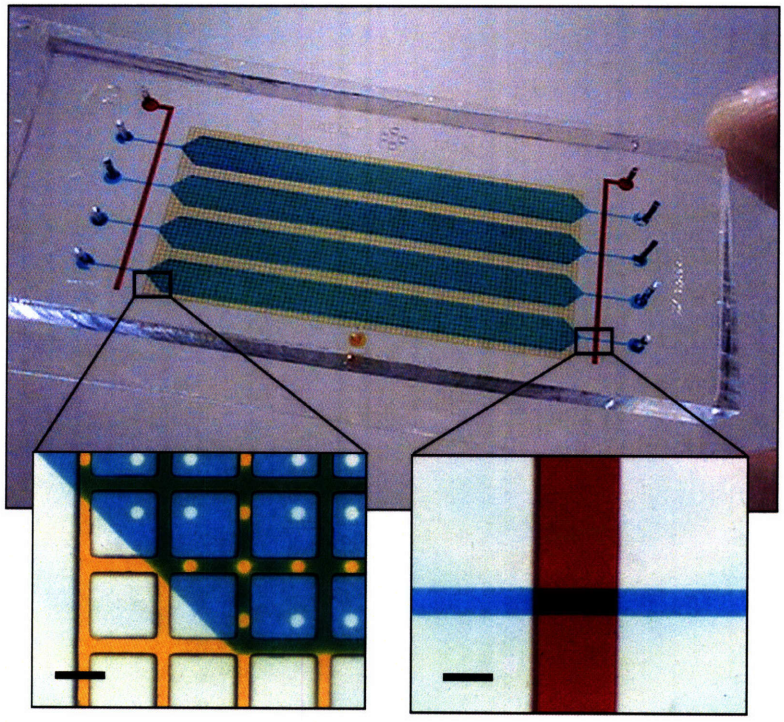


Figure 5.3: Optical images of a microfluidic device capable of conducting four parallel 500 nL reactions with various features emphasized with food coloring. Left inset: gene synthesis chamber (blue and green) and water jacket (yellow) layers. Right inset: fluid inlet channel (blue) overlaid with red valve channel (red). Scale bars correspond to 200 μm.

5.2.3 Parsing of genes

Several genes and gene segments were selected for synthesis and parsed utilizing the program DNAWorks[21] to generate the desired oligonucleotides sequences for assembly and amplification. The genes selected for synthesis were: (1) a randomized amino acid sequence of the alba gene

[21] Hoover, D.M. and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30,** e43.

from *S. solfataricus* (total length 327 bp, 16 oligos); (2) a Holliday junction cleavase (hjc) gene from the bacteriophage *SIRV-1* (total length 390 bp, 16 oligos); (3) ble (bleomycin resistance, total length 461 bp, 16 oligos); (4) DsRed (total length 733 bp, 26 oligos); (5) OR128-1 (total length 942 bp, 32 oligos); and (6) a GFP construct including a promoter and regulatory elements (total length 993 bp, 42 oligos), using the same sequence reported in Carr et al[14]. All genes were parsed in protein-mode utilizing codon optimization with the exception of GFP, which was parsed in DNA-only mode. Relevant parameters for the parses selected from DNAWorks for all synthesized gene and gene segments are summarized in Table 5.1. Complete DNAWorks output files can be found in the Appendix.

TABLE 5.1 Key parameters for the selected parses for each gene and gene segment synthesized in this work as output by DNAWorks.

Gene	Total size (nt)	Number of oligos	Anneal Temp (°C)	Construction Oligo size (nt)	Amplifying Primer sizes (nt)	[Oligonucleo -tide] (nM)	$[Na^+]$ (mM)	$[Mg^{2+}]$ (mM)
alba	327	16	59	38	35, 32	25	5	2
hjc	390	16	60	48	25, 25	20	50	2
dsRed	733	26	60	50	25, 20	25	50	2
GFP	993	42	59	42	29, 29	20	50	2

5.2.4 PCA reaction mixtures

PCA reaction mixtures for each desired gene or gene segment were prepared for utilization with the microfluidic device. Each reaction mixture contained the following concentration of reagents: 1 mM dNTPs (250 μM each), 0.15 U/μL of Pfu Hotstart Turbo Polymerase (Stratagene), 1X cloned Pfu Buffer (Stratagene), 0.1% *n*-Dodecyl-β-D-maltoside (Sigma), 10 or 25 nM of each construction oligo depending on the construct, and 500 nM of each amplifying outside primer. The addition of amplifying outside primers enabled the synthesis and amplification of the desired DNA construct in a single reaction. For synthesis of the full GFP construct and dsRed, 10 nM of each construction oligo was utilized, while for all other gene and gene segments 25 nM of each construction oligo was used.

Two segments of the GFP gene were also synthesized; for these experiments, the first pool consisted of oligonucleotides 1-22, with 1 and 22 used as the primers to amplify segment 1, which was 531 bp in length. The second pool consisted of oligonucleotides 21-42, with 21 and 42

used as the primers to amplify segment 2, which was 529 bp in length. Oligonucleotides were purchased from Integrated DNA Technologies and Operon Biotechnologies without additional purification.

5.2.5 PDMS microchannel preparation

While PDMS has a number of superb characteristics that make it, in many cases, an ideal material from which automated biological platforms can be built, its hydrophobicity has inhibited certain biological processes due to a strong tendency for non-specific protein adsorption. PCR in μL and nL volumes generally suffer from such surface effects for a variety of materials because of the high surface area to volume ratio of reactors[22], thus mandating some type of surface passivation. To address this problem in PDMS, we have successfully employed a nonionic surfactant, *n*-Dodecyl-β-D-maltoside (DDM), as a passivating agent[23]. DDM adsorbs strongly to hydrophobic surfaces, and when included in reaction mixtures is capable of successfully eliminating the majority of protein adsorption. Reaction mixtures that did not include DDM or any other passivating reagent failed to generate desired synthesis products.

Additionally, we found that devices exhibited the most robust, reliable performance after having been extensively thermocycled prior to conducting gene synthesis reactions. While the mechanism for this is not yet clear, experiments have shown a substantial increase in product yields when devices were first thermocycled with reactors containing a mixture of 0.1% DDM, 1X Pfu Buffer and water for 100 cycles utilizing the following program: 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds (data not shown). An Eppendorf Mastercycler Gradient thermocycler with an *in situ* adapter that facilitated thermal contact between the heating block and the glass slide was utilized for all thermocycling of microfluidic devices in this work.

5.2.6 Sample evaporation

Because of the high porosity of PDMS, during the course of thermocycling significant sample evaporation can occur, thus altering reactant concentrations and subsequently reducing reaction

---

[22] Shoffner, M.A., Cheng, J., Hvichia, G.E., Kricka, L.J., and Wilding, P. (1996) Chip PCR. I. Surface passivation of microfabricated silicon-glass chips for PCR. *Nucleic Acids Res,* **24,** 375-379.

[23] Huang, B., Wu, H., Kim, S., and Zare, R.N. (2005) Coating of poly(dimethylsiloxane) with n-dodecyl-beta-D-maltoside to minimize nonspecific protein adsorption. *Lab Chip,* **5**, 1005-1007.

efficiency, and in some cases completely inhibiting synthesis. It has been found that the addition of fluid reservoirs in the vicinity of reaction chambers can reduce sample evaporation[17]; thus, a water jacket composed of a mesh of fluid lines 50 μm wide with 300 μm spacing was designed in the control layer above the four reactors. When filled with water and actuated during thermocycling, the water jacket substantially decreased reactor evaporation as observed qualitatively.

5.2.7 Device design and operation

The microfluidic device was designed with individual reactor volumes of 500 nL to facilitate analysis of reaction products by polyacrylamide gel electrophoresis (PAGE). The overall device architecture is quite simple, with only three control lines necessary: a single valve to address all reactor inputs, a single valve to address all reactor outputs, and a control line for water jacket actuation. An array of 50 μm diameter posts present in each reactor prevented chamber ceiling collapse. Reactor input and output channels were 100 μm wide while control lines were 300 μm wide, thus ensuring a strong seal to prevent sample evaporation from the reactor inlets and outlets during thermocycling. Without such valving evaporation occurs almost instantaneously upon reaching the denaturation temperature.

All control lines were dead-end loaded with water by backing with pressurized air to force any air initially within the control lines out through the porous bulk PDMS. PCA mixes were introduced into the device by first actuating the reactor output valve at 15 psi and then dead-end loading the four reaction mixes at 5-10 psi into the reactor. Once all air bubbles were pushed out of the device, the inlet valve was closed to seal the reaction mix for thermocycling. All control valves, including the water jacket, were actuated and maintained at 15 psi for the duration of the synthesis reaction. Fresh devices that had been extensively thermocycled as described were used for each experiment.

Upon completion of sample loading, the device was placed on the *in situ* adapter of the Eppendorf Mastercycler Gradient and adhered with a small volume of mineral oil. Thermocycling commenced by heating first at 94°C for two minutes to activate the polymerase, followed by either 35 or 45 cycles of the subsequent program: 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds. For synthesizing GFP and dsRed, 35 cycles were utilized, while for synthesizing all other gene and gene segments 45 cycles were used. Upon completion

of cycling, a final two minute extension at 72°C was conducted. Samples were collected by flushing with 5 μL of deionized water in preparation for analysis by PAGE.
It should be noted that while steel pins (New England Small Tube Corp) were utilized to interface polymer control line tubing to the fluidic device, for all reaction mix introduction and collection steps, only polymer pins were utilized to interface to device inlets and outlets, as it has been reported that prolonged contact between reaction mixes and steel can inhibit PCR[24].

All fluid manipulations, including valving and pressure-driven flows, were controlled by individually actuated solenoid valves (The Lee Co) connected through a custom printed circuit board to a National Instruments DAQ card. A LabVIEW software interface allowed control over individual valves and fluid lines, while air flow for pressure-driven fluid manipulation was controlled by standard pressurized air regulators (McMaster).

5.2.8 Control experiments

Several sets of control experiments were conducted. For each PCA reaction mix, synthesis reactions were performed both within the fluidic and also *in vitro* in standard 0.2 ml PCR tubes to compare the performance of fluidic versus *in vitro* synthesis. Additionally, negative controls where conducted where construction oligos for synthesis reactions were omitted from the mixes. These 'primers-only' negative controls were run side-by-side in the microfluidic device with synthesis mixtures containing construction oligos. *In vitro* positive control experiments were conducted in an MWG Primus 2500 thermocycler utilizing the same thermocycle programs described above. All 'in fluidic' control experiments were similarly conducted with the 100-thermocycle microchannel treatment discussed previously.

5.2.9 Polyacrylamide gel electrophoresis

Reaction mixtures collected from the four 500 nL reactors for all 'in fluidic' syntheses and negative control experiments along with 0.5 μL of each positive *in vitro* control were analyzed by PAGE (4%-12% gradient TBE gel, Invitrogen) and visualized by SYBR Gold staining (Molecular Probes). Band intensities for synthesized gene and gene segments were approximated utilizing AlphaEaseFC software from Alpha Innotech Corporation.

---

[24] Panaro, N.J., Lou, X.J., Fortina, P., Kricka, L.J., and Wilding, P. (2004) Surface effects on PCR reactions in multichip microfluidic platforms. *Biomedical Microdevices*, **6**, 75-80.

5.2.10 DNA Sequencing

Gene synthesis products were sequenced to confirm the identities of the six target genes. Upon completion of microfluidic gene synthesis and visualization by PAGE, reaction mixtures that demonstrated successful synthesis along with successful 'in fluidic' negative controls were further PCR amplified for 25 or 30 cycles to produce larger quantities of DNA for sequencing. 'Primers-only' negative controls were again conducted alongside this amplification step to verify that only gene products from the original microfluidic synthesis reaction and not a contaminant species were amplified. Upon completion of PCR, the resultant reaction mixtures were visualized by PAGE to verify successful amplification and the absence of product in the negative controls. PCR products were purified using QIAquick PCR Purification Kits (QIAGEN) prior to sequencing. It was subsequently demonstrated (with the alba and DsRed genes) that gene assembly products taken directly from the microfluidic devices provided sufficient material for DNA sequencing, after first using ethanol precipitation to remove salts and enzymes.

The GFP gene product was sequenced using internal sequencing primers as in Carr et al[14]. All other gene products were sequenced (top and bottom strands) using the amplifying primers as sequencing primers, by the MIT Biopolymers Laboratory.

To quantify the errors present in these synthetic genes, one gene was chosen for further analysis. DsRed gene synthesis products were cloned (without purification or secondary amplification) into vector pCR4Blunt-TOPO (Invitrogen) and transformed into chemically competent DH5α cells. Individual colonies were picked and grown in Luria-Bertani broth. Glycerol stocks of these cultures were sent to Cogenics for plasmid extraction and sequencing. One 96-well plate of samples was sequenced (48 from cloned microfluidic-synthesized DsRed genes, 48 from the positive control synthesis reactions performed in standard 0.2 ml PCR tubes). All sequence reads were analyzed using sequence-alignment tool ClustalX, and each error was verified by direct visual confirmation of electropherograms using Chromas (Technylesium)

5.3 Results

Parallel gene syntheses were successfully conducted in a PDMS-based microfluidic device, as visualized in the gel shown in Figure 3 and ultimately verified by DNA sequencing. Here,

parallel synthesis of four gene and gene segments, namely GFP segment 1 (531 bp), GFP segment 2 (529 bp), the hjc gene from bacteriophage *SIRV-1* (390 bp) and the randomized alba gene from *S. solfataricus* (327 bp), is demonstrated. Successful assembly was also achieved for the four positive *in vitro* controls, while successful 'primers-only' negative controls were conducted both 'in fluidic' and *in vitro* to confirm that the presence of desired-length product was not a consequence of amplification of contaminant species (not shown). Strong, dominant bands are evident for the desired products of all four fluidic syntheses, with product yields greater than 50% relative to the positive *in vitro* controls (i.e. in PCR tubes). The ladder of lower molecular weight species below the product bands indicates normal levels of assembly intermediates for a single reaction PCA.

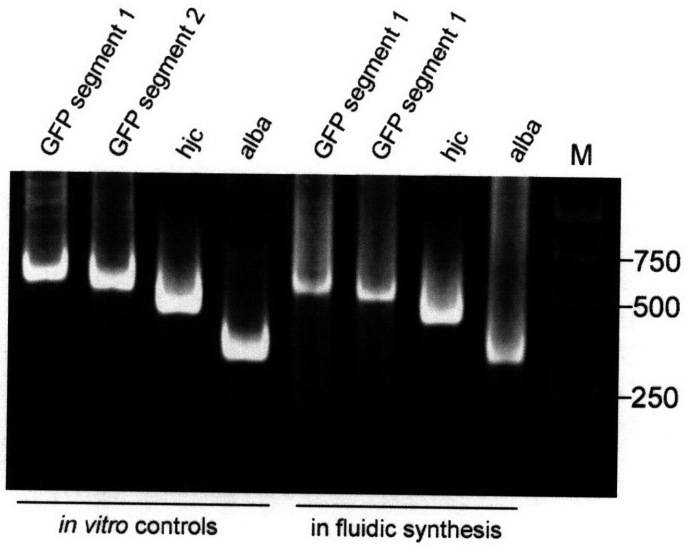


Figure 5.4: Polyacrylamide gel electrophoresis (PAGE) showing successful parallel synthesis of four gene and gene segments: GFP segment 1 (531 bp), GFP segment 2 (529 bp), the hjc gene from *SIRV-1* (390 bp), and a variant alba gene from *S. solfataricus* (327 bp). Positive *in vitro* controls are shown side-by-side. Molecular weight markers are shown (M) with 250, 500, and 750 bp positions indicated.

Additionally, the synthesis of four additional constructs, the full length GFP construct (993 bp), OR128-1 (942 bp), DsRed (733 bp), and ble (461 bp) was also accomplished, thus demonstrating the generality and robustness of microfluidic gene synthesis. Lower oligonucleotide concentrations (10 nM) were required for the longer genes (GFP, OR128-1 and DsRed), as it is hypothesized that at higher construction oligo concentrations all dNTPs are consumed generating intermediate products. The results of the parallel syntheses of these four genes along with their

respective negative controls are shown in Figure 4. Again, strong, dominant product bands are observed for all four assemblies, while the negative controls exhibit no discernable product bands. To obtain successful negative controls as shown in Figure 4, significant care must be taken to eliminate all contamination, as the presence of even minute quantities of template molecules can lead to undesired amplification—and thus erroneous results—in both PCA and PCR. These negative controls have yet to fail when appropriate care is taken to avoid contamination (fresh reagents, thorough cleanliness of all lab surfaces and equipment—pipettors and tips, PCR tubes, fluidic tubing, etc.).

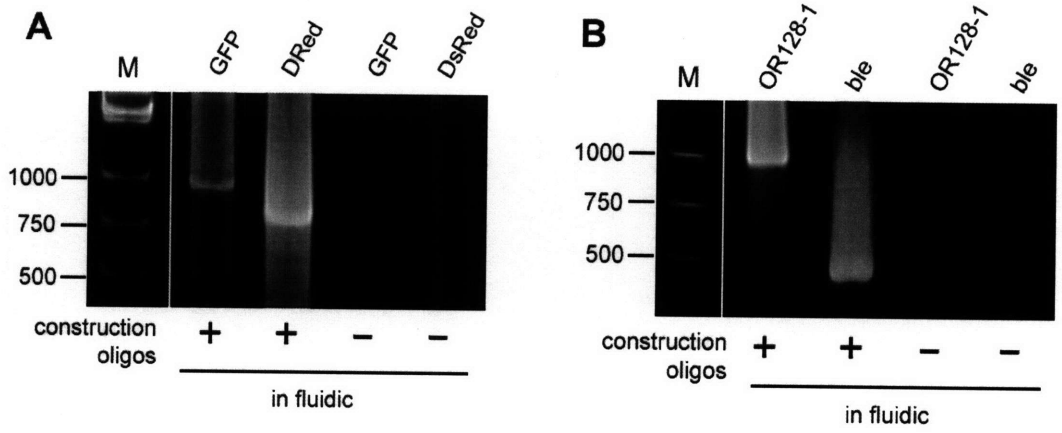


Figure 5.5: Polyacrylamide gel electrophoresis (PAGE) showing successful parallel synthesis of genes along with negative controls. In the presence of construction oligos, DNA constructs GFP and dsRed (993 and 733 bp respectively, figure 4A) and OR128-1 and ble (942 and 461 bp respectively, figure 4B) are synthesized and amplified. Without construction oligos, no product bands are generated. Molecular weight markers are shown (M) with 500, 750, and 1000 bp positions indicated.

In all cases, direct sequencing of microfluidic gene synthesis products unambiguously confirmed the identity of each target gene. However, such sequencing does not effectively report on the rate of error in the product material, as errors in individual molecules are effectively averaged out in the ensemble of products. Thus, one gene product (DsRed, 733 bp) was cloned, and the resultant clones sequenced to quantify error rates. For DsRed sequencing, purification (by length or secondary amplification) was deliberately omitted to prevent the addition or masking of errors in such processing. For the same reason, clones were not screened prior to sequencing other than blue/white screening to confirm successful insertion into the cloning vector. Thus, gene synthesis products (which include the desired full-length species along with other incomplete, intermediary products, as seen in Figure 4) were cloned directly from the microfluidic device or PCR tube

upon verification of synthesis by PAGE. The results of this sequencing are shown in Table 5.3. 48 clones for both 'in fluidic' and *in vitro* DsRed synthesis yielded 16,250 and 13,389 bases of sequence information, respectively. A total of 29 and 30 errors were identified for the 'in fluidic' and *in vitro* DsRed syntheses, thus generating error rates, per base, of 0.0018 and 0.0022, respectively, with an overall per-base error rate for all sequence reads of 0.0020. These values correspond well with the 0.0018 per-base error rate for the un-purified synthesis products reported by Carr et al.[14] and Hoover et al.[21], along with the 0.0027 per-base error rate reported by Kodumal et al.[13] Given the 0.0018 per-base error rate for 'in fluidic synthesis,' as calculated in Carr et al.[14], approximately 9 DsRed clones are required for sequencing to have a high probability (95%) of at least one which is error free. Ultimately, 12.5% of full-length clones were error-free, in agreement with theoretical expectations. For detailed tabulation of sequencing results see Appendix A.5.

Table 5.3: Summary of errors for synthesis of DsRed in the microfluidic device as compared to in a standard PCR-tube.

Error Type		**Microfluidic Device**	**PCR tube**
Deletion	Single-base	19	16
Deletion	Multiple-base	5	5
Transition	G/C to A/T	3	6
Transition	A/T to G/C	0	2
Transversion	G/C to C/G	0	0
Transversion	G/C to T/A	1	1
Transversion	A/T to C/G	1	0
Transversion	A/T to T/A	0	0
	Total Errors:	29	30
	Bases Sequenced:	16,250	13,389
	Error Rate (per base):	0.0018	0.0022

5.4 Discussion

Currently, the cost and time required to generate long, high-fidelity DNA molecules prevents such synthesis technology from being an extensively utilized resource. For example, at current oligo costs of approximately $1x10^{-1}$ dollars per base, applications such as the *de novo* synthesis of bacterial genomes 10^6 bp in size become prohibitively costly, requiring on the order of $100,000 in oligos alone. Similarly, the ability to generate sets of hundreds or thousands of single genes is restricted. The costs of expensive reagents such as polymerase and oligonucleotides can be significantly reduced by utilizing microfluidic technology to minimize reaction volumes to a fraction of a microliter as compared to tens of microliters required in conventional syntheses.

Further reductions in oligonucleotide costs by several orders of magnitude can be achieved by utilizing the oligos synthesized from DNA microarrays[11, 18-19]. In such arrays large numbers of distinct oligos are synthesized massively in parallel (10^4-10^5 or more for a single high density array[25,26]) but in minute quantities (femtomoles or less). Thus each oligo in a microarray can cost as little as $1x10^{-5}$ to $1x10^{-3}$ dollars per base, depending upon the array, which typically cost between a few hundred to a few thousand dollars (e.g. $420 for a 22,000 spot Agilent microarray). These costs per base are orders of magnitude less than for conventional oligo synthesis. Thus, the current significant contribution of oligo costs to the overall price of gene synthesis could be reduced to an almost trivial amount if the wealth of raw building material provided by microarrays could be successfully utilized. If maximally employed, oligo costs for building a 10^6 bp genome could potentially be reduced to tens of dollars. To achieve this goal, two difficulties must be addressed: 1) conducting synthesis from the low yields of each oligo in a microarray; and 2) problems that arise from manipulating highly complex pools of oligonucleotides (>10^4 distinct sequences). In this work, successful gene synthesis from minute oligo quantities (femtomoles) utilizing a microfluidic device architecture has been demonstrated, while such an architecture employed in conjunction with a microarray has the potential to overcome the limitations associated with complex pool manipulation.

---

[25] Cleary, M.A., Kilian, K., Wang, Y., Bradshaw, J., Cavet, G., Ge, W., Kulkarni, A., Paddison, P.J., Chang, K., Sheth, N., Leproust, E., Coffey, E.M., Burchard, J., McCombie, R.W., Linsley, P., and Hannon, G.J. (2004) Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis. *Nature Methods*, **1**, 241-248.

[26] Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J. et al. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.*, **12**, 1749-1755.

In prior applications, oligos synthesized in microarray format have been cleaved from the arrays and collected in "large" volumes (e.g. 5 μl or more) [11, 18-19]. The resulting low concentrations of oligo have been below the minimum needed to perform gene synthesis. Thus, additional process steps such as DNA concentration and/or amplification by PCR were required in order to assemble genes from this raw material. Direct gene synthesis of microarray oligos in microfluidic reactors such as the ones presented here can circumvent these requirements by confining synthesis reactions to individual chambers, thus maintaining oligo concentrations at levels sufficient for synthesis. Table 5.4 indicates the concentrations of construction oligos which are expected for two different microarrays[25, 26] (Agilent, Nimblegen) for a reactor enclosing 16 oligo spots, sufficient to build a 400 bp gene. Using a reasonable estimate for oligo yields as function of spot area (0.1 picomoles/mm^2, as in Richmond et al.[18]; as much as 4 picomoles/mm^2 have been estimated[27]. See also Pirrung[28] for further discussion of oligo density), the spot size and spacing for the two microarrays, and assuming a chamber with the same height as the reactors used in this work (~10 μm), we estimate that construction oligos can be confined to yield concentrations in excess of 200 nM each. This is substantially larger than the 10-25 nM per oligo utilized for microfluidic synthesis reported here. Thus, ample room for error is provided to account for low oligo synthesis and/or cleavage yields, as well as chambers enclosing more oligos spots to synthesize larger genes. Employing such direct synthesis without concentration or an initial amplification step not only reduces the time and cost of the overall synthesis protocol, but also eliminates the possibility that additional errors will be generated during the amplification procedure. The oligonucleotide building blocks themselves are currently the greatest source of error in synthesized products, so reducing the likelihood of further inaccuracies is crucial for obtaining high quality synthetic DNA.

---

[27] Elder, J.K., Johnson, M., Milner, N., Mir, K.U., Sohail, M., and Southern, E.M. (1999) DNA Microarrays. A Practical Approach (Ed.: M. Schena), Oxford Press, New York, 77-99.

[28] Pirrung, M.C. (2002) How to make a DNA chip. *Angew. Chem. Int. Ed.*, **41**, 1276-1289.

TABLE 5.4 Calculations for expected oligonucleotide yields from a typical DNA microarray for 16 oligonucleotides, sufficient to build a 400 bp gene. Values for spot area and spot spacing are for commercially available Agilent and Nimblegen DNA microarrays.

	Agilent	**Nimblegen**
Area of spot	1.4×10^4 μm^2	2.56×10^2 μm^2
Oligo density	0.1 picomole/mm^2	0.1 picomole/mm^2
Maximum expected yield per spot	1.4 femtomole	0.0256 femtomole
Dimensions of spot spacing	212 μm by 188 μm	25 μm by 25 μm
Minimal footprint of 16 oligo spots	6.4×10^5 μm^2	1×10^4 μm^2
Minimal chamber volume (10 μm height)	6.4 nanoliters	100 picoliters
Estimated concentration of each oligo	**220 nM**	**256 nM**

Resolving hundreds of thousands of oligos into reactions generating thousands of genes is a non-trivial challenge. For example, while multiplex gene synthesis utilizing bulk-sample handling has been impressively demonstrated from an oligo pool containing ~600 distinct oligonucleotides[11], we expect such amplifications to become unfeasible for pools of higher complexity. Just as multiplex PCR suffers from inconsistencies such that each template may not be equally amplified[29], similarly the simultaneous amplification and subsequent assembly of 10^5 or more sequences is unlikely to proceed evenly. For gene synthesis, this is expected to be limiting; if the pool becomes dominated by a few DNA species, the required pool diversity would be lost, rendering assembly impossible. The absence of a single oligo prevents the assembly of its corresponding gene, so that losses even as low as 0.1% could interfere with the production of dozens or hundreds of genes. Correspondingly, other reagent concentrations become impacted by the complexity of oligo pools. For example, if only 1 nM of each construction oligo were required for synthesis (a low estimate), for a pool of 10^5 oligos the starting material would be 0.1 mM, meaning that virtually all the required deoxynucleotide (dNTP) precursors used by DNA polymerase would be consumed in the first cycle of a PCA reaction, terminating the reaction before generating the desired product.. Use of a microfluidic device architecture such as the one presented in this work to enclose sets of oligo spots for gene synthesis would maintain reagent concentrations at desired levels while eliminating unwanted interference between sets of oligonucleotides in a complex pool. In the case of parallel synthesis of genes with related

[29] Edwards, M.C. and Gibbs, R.A. (1994) Multiplex PCR: advantages, development, and applications. *PCR Methods Appl.*, **3**, S65-S75.

sequences (e.g. many variants of the same gene), avoiding undesired oligo annealing events during assembly will be crucial.

In this work we report, to our knowledge, the first gene synthesis in a microfluidic environment. Genes and gene segments with sizes as large as a kilobase were assembled in four parallel reactors in a single device. Reactions were conducted in 500 nL chambers, which are reaction volumes one to two orders of magnitude smaller than those used in conventional gene synthesis, thus achieving substantial reductions in reagent costs. This work also demonstrates the feasibility of utilizing such device architecture in conjunction with high-density oligonucleotide microarrays to potentially further reduce costs by several orders of magnitude. Microfluidic syntheses were successfully conducted at low construction oligonucleotide concentrations of 10-25 nM, values substantially lower than the anticipated concentration attainable from microarrays. By enclosing microarray oligos in microfluidic chambers, the currently required complex pool handling would be eliminated while enabling researchers, in principle, to maximally harness the high density of oligonucleotides present on a microarray. The effective use of such architecture in combination with high density oligo microarrays would constitute a major step toward realizing the goal of low-cost *de novo* gene synthesis.

While this work utilized four parallel 500 nL chambers to facilitate analysis of reaction products via PAGE, both the number and volume of reactors can be scaled substantially. Previous work has demonstrated PCR in volumes as small as 86 pL[30], and a 100 pL chamber with dimensions of 100 μm x 100 μm x 10 μm capable of enclosing groups of 16 oligonucleotides (described in the calculations in table 5.4) can be fabricated with ease using existing techniques. Extension of current work to microfluidic devices containing a high-density of reactors for massively parallel gene synthesis is being currently investigated.

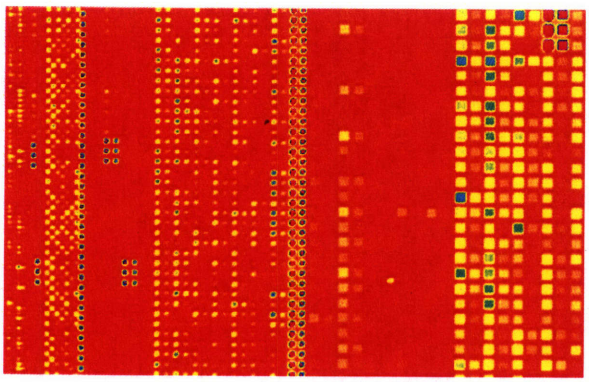
While fusion of microfluidic handling with oligo microarrays will provide the first step in making gene synthesis more available to researchers, integration with further microfluidic functions will allow this technology to mature. These advances will include: (1) incorporation of existing DNA error correction techniques[11, 14-16] on-chip to improve the quality of the synthesis products. This will help minimize the need for another substantial contribution to the cost and time of gene

---

[30] Nagai, H., Murakami, Y., Morita, Y., Yokoyama, K., and Tamiya, E. (2001) Development of a microchamber array for picoliter PCR. *Anal. Chem.* **73**, 1043-1047.

synthesis: quality control (i.e. typically cloning and sequencing). While the device described in this work does not integrate on-chip error correction, it can be used readily with existing DNA error correction techniques both before and after synthesis. For example, construction oligos can be first gel-purified, as demonstrated by Hutchinson et al.[6], prior to conducting gene synthesis in the microfluidic device, or alternatively the MutS error-filter described by Carr et al.[14] could be performed on reaction mixtures collected from the device upon completion of synthesis. Thus, the microfluidic device can complement these bench-top error correction methods while providing the associated benefits of reduced reagent costs during synthesis. For certain *in vitro* applications, cloning will not necessarily be required. Thus a second application will be (2) the integration of *in vitro* protein expression using high-quality synthetic DNA as a template. (3) Finally, assembly of constructs larger than single genes can be achieved with microfluidic devices, employing the same types of hierarchical *in vitro* assembly reactions used to create 12 kb and larger segments[11-13].

6 Direct Microarray Gene Synthesis



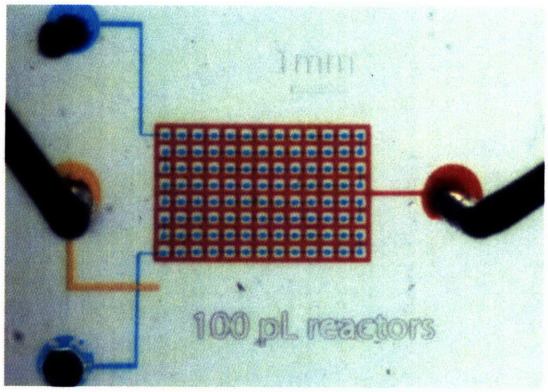


Figure 6.1 (left): A fluorescence image of a DNA microarray fabricated via Maskless Array Synthesizer (MAS) technology
Figure 6.2 (right): An optical image of a microfluidic device featuring an array of 96 reactors, each 100 picoliters in volume.

6.1 Introduction

The fluorescence micrograph shown in Figure 6.1 from Nuwaysir et al.[1] displays hybridization of biotin labeled bacterial cRNAs to a custom DNA microarray. Nearly 200,000 distinct features, each a potentially unique probe, are present with characteristic dimensions as small as 14 μm (seen on the left edge of the image). Currently, DNA microarray vendors boast chip capacities of nearly one million distinct oligo sequences, which translates into a staggering wealth of raw sequence information per array-upwards of tens of megabases. While such information density has already been demonstrated to be an extraordinarily powerful technology for biological assays, the potential for such arrays to be utilized as a feedstock for gene synthesis has only begun to be realized. As discussed at length in Section 5.4, if maximally utilized, an array of even "modest" density (100,000 features) could yield enough material for the assembly of a complete synthetic bacterial genome at a total oligo cost in the tens of dollars-a remarkable and tantalizing possibility.

In this chapter, I will describe progress toward realizing this vision by employing microfluidic technology in combination with high density oligonucleotide arrays to yield hybrid, integrated devices for the synthesis of gene-length constructs in microfluidic volumes directly at the array surface. The strategic advantage of such an approach, as detailed in Section 5.4, lies in avoiding the significant technical challenges associated with manipulating the highly complex

[1] Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J. et al. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.*, **12**, 1749-1755.

oligonucleotide pools produced when cleaving oligos off the array into macroscopic volumes. Performing gene synthesis in controlled microfluidic environments at the surface of the array circumvents this difficulty by enclosing only the oligos of interest for synthesis, keeping reactions simple while crucially preserving the total information density present in each array.

Techniques published to date utilizing microarray oligos as building blocks for gene synthesis, while impressive, are still modest in terms of the number of distinct oligonucleotides used-approximately 600 at best[2]-highlighting the difficulties of complex pool manipulation, and also the considerable amount of material wasted given the massive information density of microarrays. Clearly, significant work remains to realize the awesome potential of microarray oligos as a feedstock for gene and genome synthesis. The technology presented in this chapter will hopefully lead to reducing that gap.

6.1.2 Hybrid microfluidic-microarray device fabrication: challenges

The type of hybrid microarray-microfluidic device envisioned would combine a DNA microarray of complexity comparable to the one shown in Figure 6.1 with a microfluidic device of architecture similar to the one shown in Figure 6.2. Here, 100 picoliter (pL) reactors are arranged in a canonical 96 well-plate format; the device was fabricated via multi-layer soft lithography techniques described in sections 5.2.1 and 5.2.2. Blue food coloring indicates fluidic lines and reactors where gene synthesis mixture would be loaded, red food coloring indicates control-lines In an ideal embodiment, such an array of reactors would be directly aligned and bonded to a high-density microarray, with each reactor enclosing the oligonucleotides necessary to build a small (several hundred bp) gene. Given chamber dimensions of 100 μm x 100 μm x 10 μm, such a 100-picoliter reactor, as described in Table 5.2, would be sufficient to enclose 16 distinct oligonucleotides-enough material to build a 400-bp gene. Successful gene synthesis conducted at such reactor volumes and microarray feature sizes could yield compelling technology of the type shown in Figure 6.3, where in the area currently devoted to a single well of a canonical, macroscopic 96-well plate, a microarray-microfluidic 96-well "gene plate" could be synthesized. In such a configuration, it would be possible to synthesize some ten thousand genes in a single plate!

---

[2] Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432**, 1050-1054.

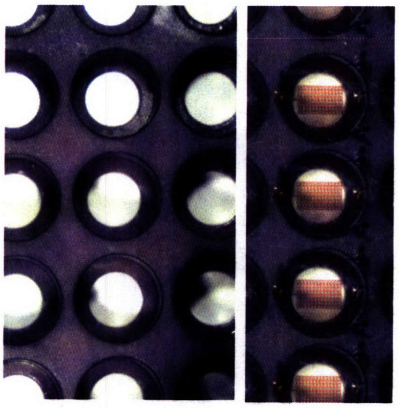


Figure 6.3: An optical image of a canonical 96-well plate, juxtaposed with a vision for the future: in each well, a 96-reactor "gene plate," yielding a high-density plate featuring nearly ten thousand genes.

Let us consider an example from structural biology to again highlight the potential power of microarray oligos. Integral membrane proteins (IMPs) such as the small multi-drug resistance (SMR) transporters expressed by Klammt et al.[3], are notoriously difficult to crystallize, thus resulting in a startling lack of existing structures despite their significant biological relevance; despite the fact that 20 to 25% of all cellular proteins are IMPs, they account for only ~0.3% (~80 of >26,000) structures in the Protein Data Bank[4]. One way to increase the likelihood of obtaining an IMP crystal is to follow the strategy of Dyda et al.[5], where hydrophobic residues in the core domain of HIV-1 integrase were systematically replaced until it was discovered, empirically, that the single amino acid substitution of Lys for Phe^{185} resulted in considerably improved protein solubility while maintaining its biological activity. This type of systematic study could be performed with significantly greater ease and elegance by employing gene synthesis to generate a library of variants. Typically composed of 110 amino acids, these SMR transporters consist of four trans-membrane segments forming a tightly packed four-helix bundle. The associated gene for these membrane proteins would be 330 bp, a size readily amenable for single-step gene polymerase-mediated gene assembly. Using oligo design tools such as DNAWorks, specific amino acids of the four-helix bundle of the SMR could be systematically altered, generating a variety of mutants. Assuming the use of 50mers, a total of 14 oligos would be required to synthesize the full 330 bp gene. Given a Nimblegen gene chip composed of ~800,000 spots, ~57,000 different mutants could be synthesized from a single microarray!

---

[3] Klammt C et al.; High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry* **271** 586-580, 2004.

[4] http://www.wwpdb.org/

[5] Dyda et al.; Crystal structure of the catalytic domain of HIV-1 integrase; similarity to other polynucleotidyl transferases. *Science* **266** 1981-1986.

While an appealing vision, numerous technical challenges need to be overcome. On the device side, how do we align and bond the fluidic to the microarray surface? Given the variety of microarrays available, which would be most suitable for the proposed synthesis scheme? Once attached, how are the oligonucleotides to be liberated from the surface? Once cleaved from the surface, what additional, if any, modifications to the microfluidic gene synthesis protocols described in Chapter 5 are required? What are the limitations in terms of reactor volume, and perhaps more importantly reactor surface area to volume ratio?

The remaining sections of this chapter will be devoted to addressing the above challenges. In section 6.2, a brief overview of the various existing microarray platforms will be presented, emphasizing spot feature size, known substrate material, and details on the microarray technology selected for this work; strategies for aligning and bonding will be discussed in section 6.3; methodologies for cleaving oligos from the surface of the array will be outlined in section 6.4; and finally, sections 6.5 to 6.7 will describe experimental results to date for integrated microarray-microfluidic devices, which will hopefully convince the reader that the vision articulated in this section of the thesis is realizable.

6.2 DNA Microarray platforms

DNA microarrays can be fabricated a variety of different ways, as reflected in the numerous platforms commercially available and in development[6, 7]. Figure 6.4, re-printed from Gao et al., nicely summarize a variety of processes utilized to fabricate high-density oligonucleotide arrays.

---

[6] Pirrung, M.C. (2002) How to make a DNA chip. *Angew. Chem. Int. Ed.*, **41**, 1276-1289.

[7] Gao, X., Gulari, E., and Zhou, X. (2004) In Situ Synthesis of Oligonucleotide Microarrays. *Biopolymers*, **73**, 579-596

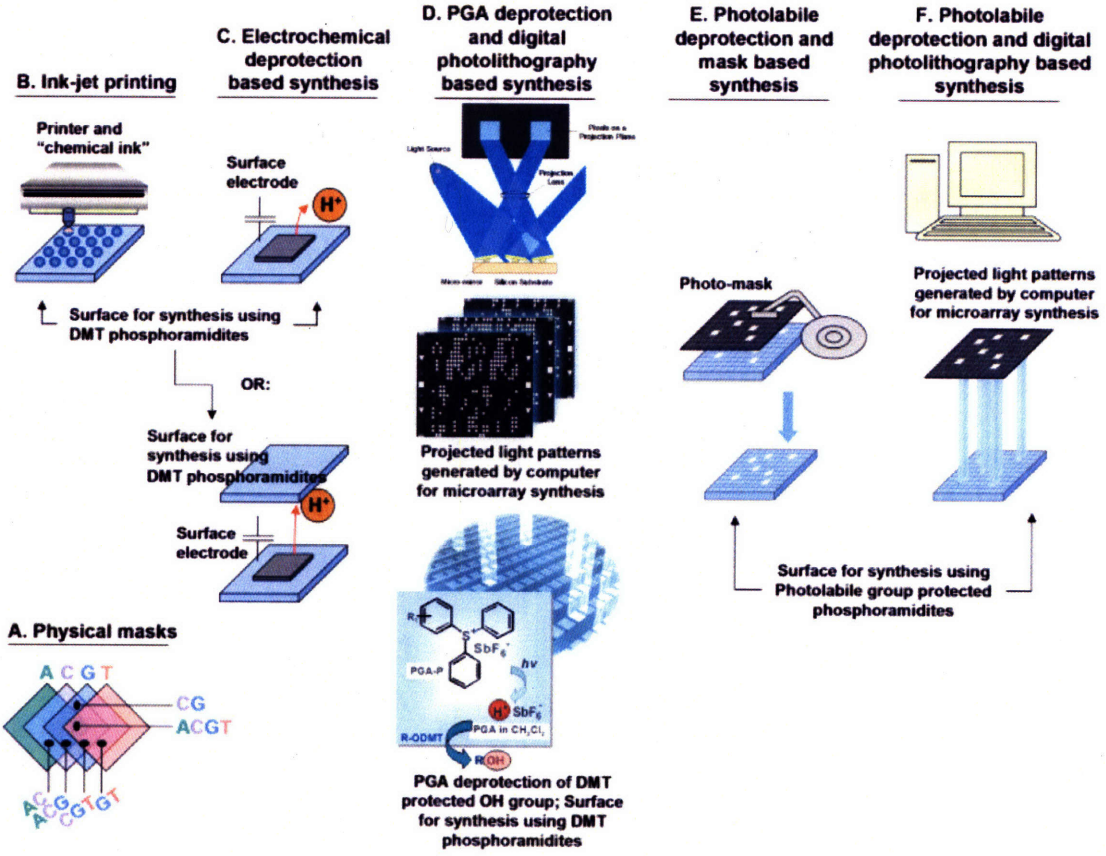


Figure 6.4: A summary of various techniques utilized for the fabrication of DNA microarrays. From Gao et al.[7]

While each of these techniques has associated strengths and weaknesses, the characteristics of greatest interest for the work of this thesis are: (1) frequency of errors in the synthesized oligos (can genes even be built from these oligos?); (2) spot feature size and density (are the oligos present in high enough concentration?); and (3) compatibility with the proposed scheme of direct microarray gene synthesis (how easily can hybrid fluidic-array devices be constructed?).

Overall, reports of genes synthesized from DNA microarray oligos have come from three different microarray platforms: photo-generated acid deprotection in a closed fluidic (Atactic/Xeotron)[8,9]; photolabile deprotection by Maskless Array Synthesizer (MAS) technology (Nimblegen)[10]; and physically addressable ink-jet based deprotection (Agilent)[11]. Crucially, the

[8] Zhou, X., et al. (2004) Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.*, **32**, 5409-5417.

[9] Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432**, 1050-1054.

[10] Richmond, K.E., et al. (2004) Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.*, **32**, 5011-5018.

array technologies most suitable for direct synthesis requires an 'open-face', or some access to bonding to a fluidic, which eliminates closed systems such as the Atactic/Xeotron platform seen in figure 6.5, unless it could be dramatically re-configured. Engineering a complete system that allowed for both oligonucleotide synthesis and gene synthesis in a single, integrated device leveraging the same architecture, plumbing, and fluid manipulation scheme would certainly be the most challenging, but perhaps in the long term the most desirable. More immediately, the goal of direct gene synthesis proposed here should be facilitated by simply taking a two-layer PDMS-fluidic device and aligning and bonding it to a microarray with 'exposed' oligos. Taking into account this design requirement, several microarray platforms were considered for fabricating integrated fluidic-microarray devices: Agilent[12] (physical deprotection by ink-jet), Nimblegen[1,13] (UV photocleavage using DMD), Affmyetrix[14] (UV photocleavage using physical masks), Combimatrix[15] (electrochemical deprotection utilizing a CMOS chip) and a photoelectrochemical deprotection method developed by Brian Y. Chow[16]. Ultimately, Agilent and Nimblegen (i.e. MAS) arrays were first tested, due to the apparent low-error rate in oligo synthesis relative to other microarray platforms in the case of the former[11], and because of the

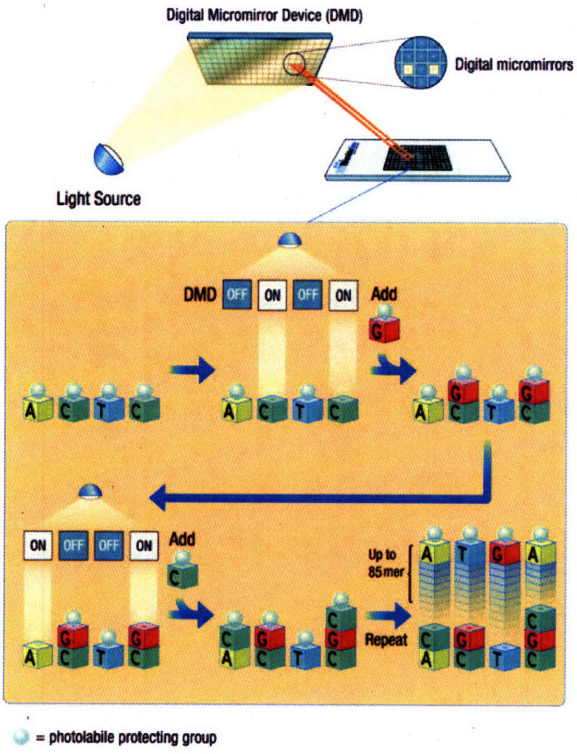


Figure 6.5: Nimblegen (MAS) technology for generating microarrays with accessible oligos. From www.nimblegen.com

11 Personal communication, P.A. Carr and G.M. Church

12 Hughes, T.R. et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342-347.

13 Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnology*, **17**, 974-978.

14 Fodor, S.P., et al. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767-773.

15 Maurer, K. et al. (2006) Electrochemically Generated Acid and Its Containment to 100 Micron Reaction Areas for the Production of DNA Microarrays. *Plos one* **1**, e34.

16 Chow, B.Y. (2008) Photoelectrochemical Synthesis of Low-Cost DNA Microarrays. Ph.D. Thesis.

similarly demonstrated gene synthesis and importantly high density of oligo spots in the case of the latter. The process flow for Nimblegen (MAS) synthesis is shown in figure 6.5.

6.3 Microfluidic-microarray integration

Experiments to directly synthesize genes on the surface of microarrays in a microfluidic environment followed a trajectory of first overcoming the microfluidic design challenges of device integration (device design, bonding and alignment), followed by an investigation of the molecular biology necessary to cleave oligos from the surface and assemble them into genes. In this section, the various microfluidic-microarray integration challenges will be discussed.

6.3.1 Bonding

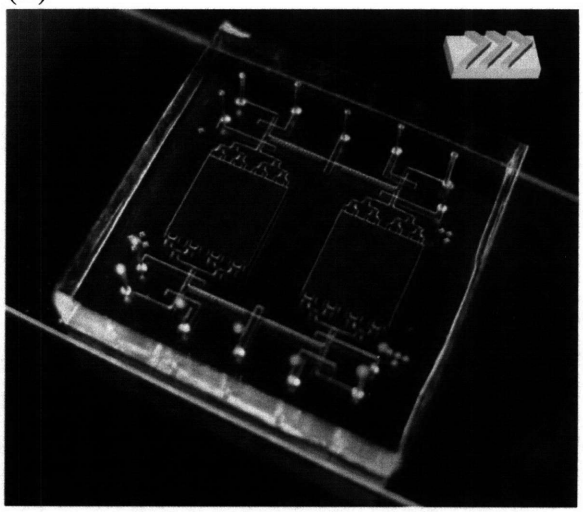
The major design constraint for any bonding methodology between a fluidic and microarray is the fact that the exposed oligos on the microarray surface must survive the process unscathed. Thus, many conventional methods for bonding PDMS and glass substrates such as plasma bonding or corona discharge were not possible due to the inevitable DNA damage. Similarly, high-temperature processes such as anodic bonding could not be pursued. Ultimately, several methods that would not cause DNA damage during the bonding process were evaluated. As reported by Liu et al.[17], for example, multi-layer fluidic devices can be successfully sealed to open-face microarrays, such as the spotted arrays described, simply by bringing together the clean, native glass surface of the array with a clean PDMS surface, as seen in Figure 6.5a. In this case, sealing was performed without any surface chemistry or external pressure, and is reversible. Another alternative involves utilizing an external pressure source of some kind to forcefully bond the microarray with a fluidic device. In Wei et al.[18], for example, a steel clamp is used to bring into contact a spotted microarray with a fluidic device with reactor features etched in glass in a sandwich-like structure, as seen in Figure 6.5b.

---

[17] Liu J, Williams BA, Gwirtz RM, Wolf BJ, Quake SR. "Enhanced Signals and Fast Nucleic Acid Hybridization By Microfluidic Chaotic Mixing " *Angew. Chem.* 2006 45:3618-3623

[18] Wei, C.W., Cheng, J.Y., Huang, C.T., Yen, M.H., and Young, T.H. (2005) Using a microfluidic device for 1 μL DNA microarray hybridization in 500s. *Nucleic Acids Research* **33**, e78.

(A)



(B)

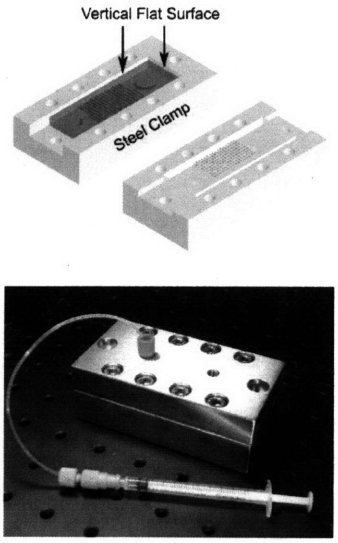


Figure 6.5(a) Multi-layer PDMS device reversibly sealed to a spotted microarray without surface chemistry or external pressure. From Liu et al.[16] **(b)** Sealing a glass-based fluidic structure with a spotted microarray utilizing a steel clamp. From Wei et al.[17]

Ultimately, for the work in this chapter, an adhesive bonding methodology was utilized to seal a multi-layer PDMS device with a high-density *in situ* (oligos synthesized directly on the surface; as opposed to an *ex situ*, or spotted array) DNA microarray. The reasons were several-fold. Firstly, while the adhesive-free sealing from Liu et al. can in fact sustain internal pressures sufficient for mixing by peristaltic pumping, ultimately such reversible sealing proves incapable of surviving the internal pressures generated during PCR thermocycling (e.g. 94°C), and leads to delamination. While an external pressure source of some kind was an intriguing possibility using either clamps or, as had been previously demonstrated in our group, an air bladder[19], ultimately this path was abandoned due to the inability to achieve uniform pressure across the entire surface of the device area. Given inhomogeneous sealing, valve performance was either poor or non-functional. It should be noted that an applied force could in fact be quite a robust sealing method in the case of single-layer fluidics using soft or hard materials (as seen by Wei et al.[17]); however, multi-layer structures prove more difficult to work with due to the more complex internal pressures involved.

[19] Personal communication, Eric Wilhelm

The bonding method utilized is shown in figure 6.6, and is a modified version of the adhesion-transfer techniques described by Satyanarayana et al.[20] and Wu et al.[21] Here, a transfer substrate (typically a glass microscope slide) is first prepared by spinning a pre-polymer solution of PDMS (10:1, monomer:curing agent, in toluene at a mass ratio of 1:7) at 500 rpm for 3s, followed by 1500 rpm for 60s. Next, the fluidic device, with raised features indicating reactors, is gently inked to the PDMS prepolymer-coated transfer substrate. Contact is maintained for 60s without any additional pressure applied from above. The fluidic is then removed from the transfer surface, leaving a thin layer of prepolymer adhesive on only the raised features of the fluidic device. This layer has previously been measured to be several hundred nanometers thick[20], which is negligible relative to the width (>100 μm) and height (10 μm) of the channels, thus ensuring that none of the prepolymer adhesive will spread into the channels, thus avoiding interaction with any of the microarray oligos. Finally, the device is aligned (described in the subsequent section) and bonded to the DNA microarray and cured at low temperature (60°C) overnight.

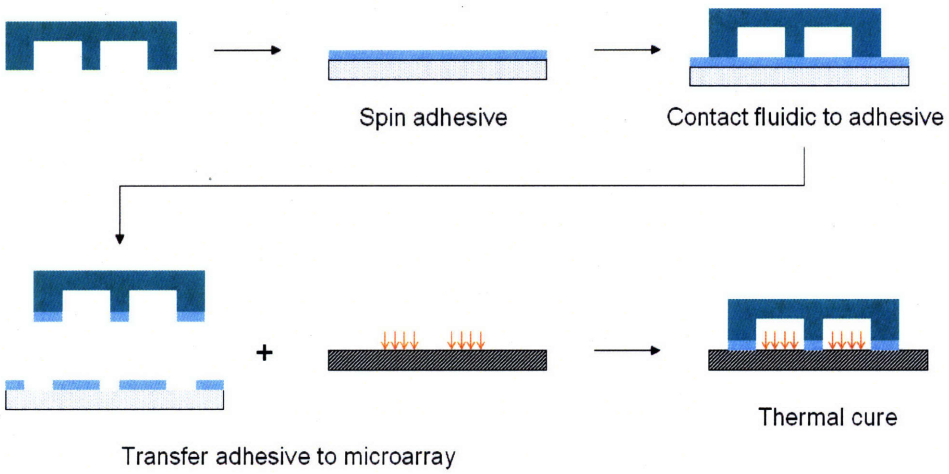


Figure 6.6: Methodology for bonding fluidic devices to DNA microarrays.

Using this type of bonding process, irreversibly sealed fluidic-microarray devices could be fabricated where the DNA was unperturbed during the bonding itself. Internal pressures in the 10-15 psi range were tolerated and thermocycling was possible without causing device delamination, thus enabling all of the necessary device functions.

---

[20] Satyanarayana S., Karnik R.N., and Majumdar, A. (2005) Stamp-and-Stick Room-Temperature Bonding Technique for Microdevices. *Journal of Microelectromechanical Systems* **14**, 392-399.

[21] Wu, H., Huang, B., and Zare, R.N. (2005) Construction of microfluidic chips using polydimethylsiloxane for adhesive bonding. *Lab on a Chip*, **5**, 1393-1398.

This adhesive bonding technique worked best for bare glass substrates as would be the case with *ex situ*, or spotted microarrays. In the case of *in situ* DNA microarrays, however, typically large-scale, homogeneous surface functionalization is utilized (e.g. capping), thus rendering the surface chemically different from native glass. This points to an inherent disadvantage of adhesive bonding; it is not general and must be compatible with the surface chemistry of the exposed microarray surface. To that point, Agilent chips were tested for bonding utilizing adhesive transfer and delaminated easily, due to incompatibility with the proprietary surface coating. DNA microarrays fabricated via Nimblegen technology from Professor Franco Cerrina were ultimately utilized for the various tests described in this chapter. Bonding to these arrays was successful given a regiment of three washes in Tween 20 prior to adhesive transfer.

6.3.2 Device Design

For this first generation set of devices and experiments, the ultimate goal was to demonstrate the synthesis of a gene-length construct, from microarray oligos, in an integrated fluidic-microarray chamber. Assuming success, even in a large volume relative to ultimate desired volumes (e.g. 100 pL), scaling-up should proceed readily. Thus, initial device designs were borrowed to the extent possible from the previously successful designs shown in Chapter 5. The microfluidic device designed for integration with DNA microarrays is shown in figure 6.7, with food coloring used to illustrate the various components: two 500 nL reactors are featured, shown in blue; each reactor has an inlet and outlet with bifurcation channels to evenly distribute flows entering and exiting the reactors. In the control layer, there are two valves, one addressing all inlets and the other addressing all outlets, shown in red. Finally, there is also a water jacket, shown in yellow, which sits atop both reactors to help minimize sample evaporation during thermocycling. "Push-down" valving geometry was also employed (and required).

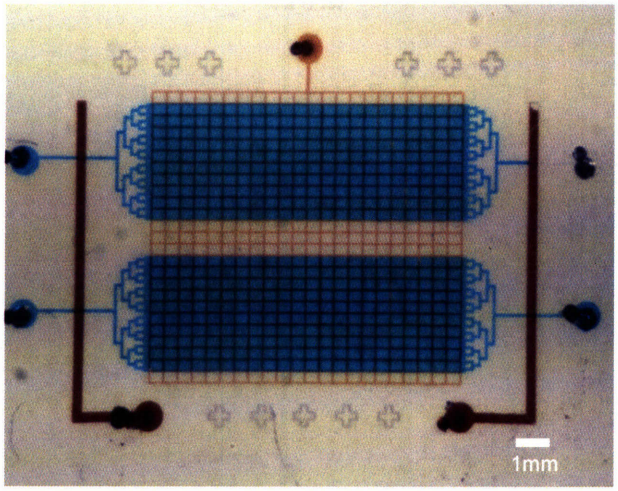


Figure 6.7: Optical image of the two-layer microfluidic device to be aligned and bonded to the DNA microarray. Food coloring is utilized to highlight various components.

The device is fabricated using the same protocol described in section 5.2.2, with the important difference being that instead of bonding to a third layer of PDMS that would ultimately serve as the 'floor' of the device, the two-layer device was bonded to the DNA microarray according to the protocol described in 6.3.1. In order to achieve the most reproducible, strongest bonds, the device needed to undergo the adhesive-transfer protocol while still in a partially-cured state, i.e. immediately after trimming and coring the bonded control and flow layers after their 45 minute back at 80°C.

6.3.3 Alignment

For these first generation devices, alignment was accomplished by hand, aided by optical microscopy, using the edge of the microarray itself as the alignment marker. From array to array, the location of the oligo spots themselves is nearly identical-at the center of each microscope slide, covering a total area of 10.5 mm x 14.0 mm. Three alignment pins set the position of the microscope slide upon which the array is fabricated, and such positioning is sufficient for array placement accuracy in the tens of microns. As can be seen in figure 6.8, two layers of photoresist were exposed 'blindly' using the DMD system, the first layer leaving an outer cross, the second layer leaving an inner cross. Given that the only alignment used between resist processing steps was the pin positioning, the accuracy of such placement resulted in a misalignment error of +/- 10 to 20 μm. Thus, the predominant source of error in terms of array placement comes from differences in the dimensions of each slide (25 +/- 0.2mm x 76 +/- 0.3mm). These several

hundred micron differences in slide dimension ultimately determine the alignment tolerance when using the microscope slide edge as the registration mark.

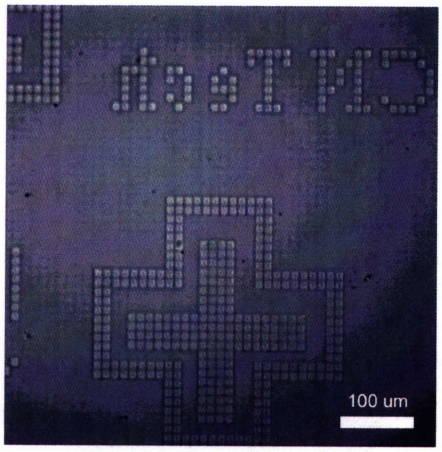


Figure 6.8 Image showing alignment error from two resist layers exposed with the DMD system utilized for array DNA microarray fabrication. Misalignment between inner and outer crosses is approximately ten to twenty microns. Image courtesy of Mike Bassetti.

The initial device was designed such that the two 500 nL reactors would sit directly above the 10.0 mm x 14.0 mm array area, with the total footprint of the two reactors being sufficiently small enough that they would be surrounded by more than 1 mm of DNA on all sides, as seen in Figure 6.9a. Here, the dashed line indicates the DNA array area. Given this geometry, the several hundred micron alignment error present from the variation in microscope slide dimension would be more than offset, ensuring that the reaction chambers would be enclosing the desired oligo spots.

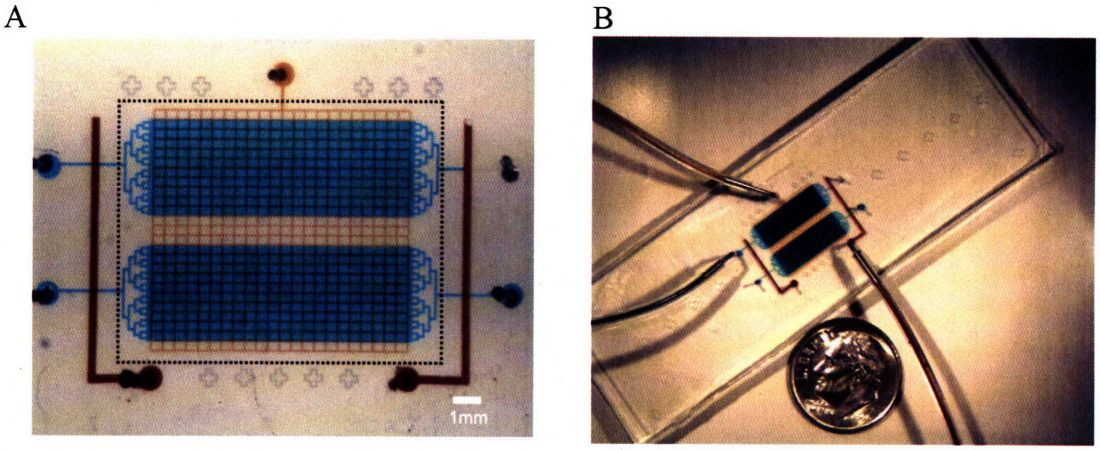


Figure 6.9 (a) Optical image of the designed microfluidic device overlaying the DNA array, with the DNA array area indicated by a dashed line. **(b)** Optical image of the two-layer PDMS microfluidic aligned and bonded to a DNA microarray.

For future device designs, registration markers on both the microfluidic device and the array itself will be utilized to ensure alignment accuracy below 10 microns.

6.4 Oligonucleotide cleavage

The next major technical challenge involved deciding upon the methodology for releasing the anchored oligonucleotides, which are bound covalently to the microscope slide surface, into solution in the microfluidic chamber, whereupon they could be assembled into genes via PCA. Three techniques were initially considered: chemical cleavage[8-10], enzymatic cleavage by type II restriction enzymes, enzymatic cleavage by Uracil DNA glycosylase (UDG) and an endonuclease, or finally enzymatic cleavage by Ribonuclease A (RNase A)[9]. In evaluating these options, as has been the theme, simplicity in adaptation to microfluidics was forefront in mind. Cleavage alternatives requiring purification or separation steps were eliminated due to the required additional on-chip functionality. An ideal cleavage method would have a high cleavage yield and be compatible with the molecular biology necessary to perform PCA, thus allowing for single-pot syntheses and simplified microfluidic operation. Even moderate yield should be sufficient given the high expected concentrations of oligos (section 5.4). Chemical cleavage by ammonium hydroxide (NH_4OH), despite its high cleavage yield, was eliminated for these reasons of necessary on-chip purification. Ultimately enzymatic cleavage by type II restriction enzymes and a blend of UDG and endonuclease was explored.

6.4.1 Type II restriction enzymes

Type II restriction enzymes, which have previously been demonstrated to be successful at cleaving micro oligos (e.g. Bulyk et al.[22]), were first examined as a cleavage mechanism. The mechanism for such cleavage is shown schematically in Figure 6.10. Because Type II restriction enzymes require a double-stranded DNA substrate, for this type of cleavage, in addition to the enzyme itself a separate single-stranded DNA molecule is required to hybridize with the anchored microarray oligo for the appropriate recognition site to be formed. This additional oligo, or 'helper oligo,' is shown as an orange arrow, which hybridizes to a complementary region of DNA on the anchored microarray oligo, as seen in stage 2. All microarray oligos were designed such that every oligo shared a common distance from the array surface (in terms of nucleotide spacing and strand length), and also a common recognition sequence to which the helper oligo, and subsequently restriction enzyme, can bind. After annealing, the restriction enzyme then binds to the double-stranded recognition site (stage 3), cutting and liberating the microarray oligos, now ready for construction by PCA, into solution (stage 4).

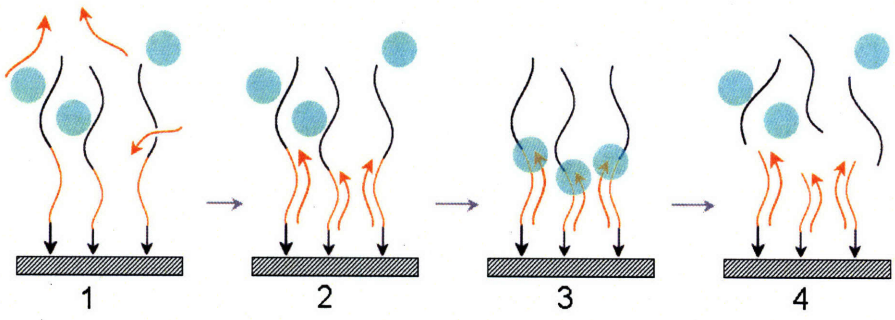


Figure 6.10. A schematic depicting the enzymatic cleavage method to release construction oligos from a microarray surface using type II restriction enzyme digestion. Initially (stage 1) the oligos are anchored by their 3' ends (arrowheads) to the array surface. Enzymes (blue circles) and helper oligos (short orange arrows) are also in solution. Helper oligos hybridize to complementary regions, indicated in orange, on the anchored oligos (stage 2). The restriction enzyme binds to the recognition site (stage 3) and cleaves the oligo (stage 4).

In order to assess whether type II restriction enzymes were a viable option for microarray oligo cleavage, first, *in vitro* tests were performed, whereby this cleavage methodology would be tested with un-anchored oligos in solution. Mly1, an 'outside cutter,' was the type II restriction enzyme selected for initial testing, and recognizes the 10-base sequence shown below:

[22] Bulyk, M.L., Gentalen, E., Lockhard, D.J., and Church, G.M. (1999) Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nature Biotechnology*, **17**, 573-577

```
5' AGGTGGACTC 3'
3' TCCACCTGAG 5'
```

A randomized amino acid sequence of the alba gene from *S. solfataricus* (total length 327 bp) was parsed using DNAWorks to yield a design consisting of 16 oligos, of which the 14 construction oligos were 38 bases long each. A randomized protein sequence was used for *in vitro* tests while the wild type protein sequence was utilized for the design of oligos to be synthesized on the microarray itself, thus keeping with an appropriate level of stringency to ensure that any contamination from *in vitro* tests would not impact tests with actual microarrays. Parse and full sequence information for both alba variants is available in Appendix A6. As we have seen throughout our work in gene synthesis, amplification on contaminant DNA can be a non-trivial source of error during experimentation. The full-length design for each oligo, as it would appear on the microarray surface, consisted of (5' to 3'): the 38 base 'payload' to be utilized in PCA, a universal (across all 14 oligos) 10-base sequence for helper oligo annealing and enzyme recognition, and finally a poly-T segment 12 bases long to yield a final oligo 60 bases in length. 60-mer oligos were ordered from Integrated DNA Technologies (IDT).

In vitro cleavage tests were performed by incubating, at 37°C, a mixture of the 60 base oligos (14 oligos, 25 nM each), the 15 base helper oligo (500 nM) complementary to the Mly1 recognition site, and finally the restriction enzyme itself. 500 nM of the helper oligo was chosen to ensure saturation of the 350 nM of 60-mers. Digestion products were visualized by denaturing PAGE, as is seen in Figure 6.11, and are as expected: bands for the full-length 60-mer and 15-mer helper oligo are evident for undigested product (lane 2) but, after digestion with Mly1, yield the 38-mer construction payload and 22-mer 'waste' products.

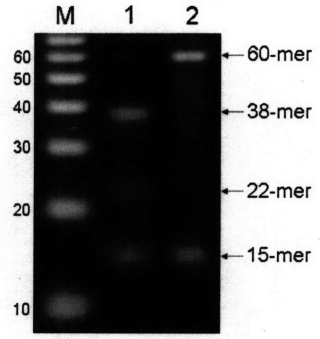


Figure 6.11 Denaturing PAGE showing the products of the *in vitro* digestion of sixteen 60-mers, facilitated by a 15-mer helper oligo. The sixteen released 38-mers are the construction oligos to be used in DNA assembly. Lane 1 shows the digested product; lane 2 the original undigested mix.

A more rigorous characterization of oligo cleavage efficiency as a function of Mly1 concentration (units/μL) and helper oligo relative to 60-mer (reactant) concentration is shown in Figure 6.12. As might be anticipated, when the concentration of helper oligo is only a fraction of the reactant concentration, cleavage does not proceed fully; however, once those ratios are equivalent or greater, all reactions with an Mly1 concentration greater than 0.5 U/μL proceed to completion (i.e., all 60-mer reactants are consumed). Reactant/product ratios were calculated by measuring gel band intensities utilizing AlphaEaseFC software from Alpha Innotech Corporation.

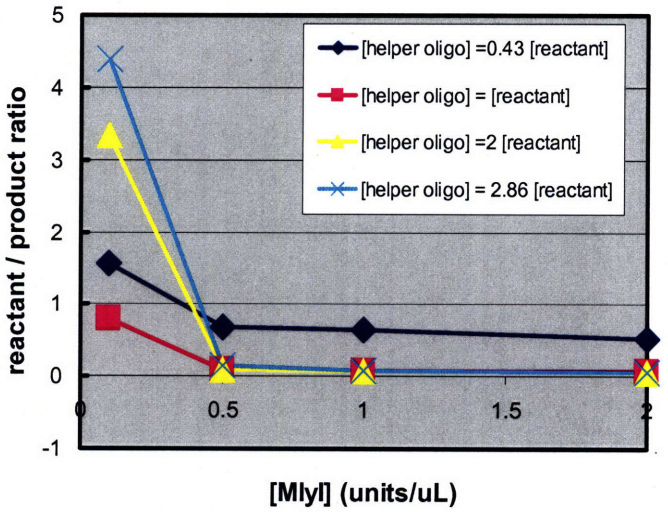


Figure 6.12 Graph showing digestion of 60-mer reactants as a function of Mly1 concentration and also the concentration of helper oligo relative to reactant concentration.

6.4.2 USER mix, UDG and endo IV

The second type of enzymatic cleavage tested utilized a blend of UDG and endonuclease, either endonuclease IV or endonuclease VIII. All enzymes were purchased from New England Biolabs (NEB), with the UDG and endonuclease VIII coming in a commercially available mix, sold as USER (Uracil-Specific Excision Reagent) by NEB. The cleavage mechanism is shown schematically in figure 6.13. Here, each microarray oligo is synthesized with at least a single Uracil base present. The UDG works by excising Uracil, leaving an abasic site which the endonuclease recognizes and cleaves, thus releasing the construction payload into solution. Endonuclease IV and VIII operate slightly differently, however, in that after cleaving the abasic site, endo IV leaves a hydroxyl group at the 3' terminus and a deoxyribose 5'-phosphate at the 5' terminus, while endo VIII leaves phosphate groups at both 3' and 5' termini. It was thought that the 3' hydroxyl group would be necessary for gene synthesis, but as we shall see in the subsequent section, USER still enables gene synthesis. One general advantage of using this enzyme blend is that no 'helper oligo' is required and all protein-DNA interaction is with single-stranded DNA, thus potentially eliminating some of the steric issues involved when using Mly1.

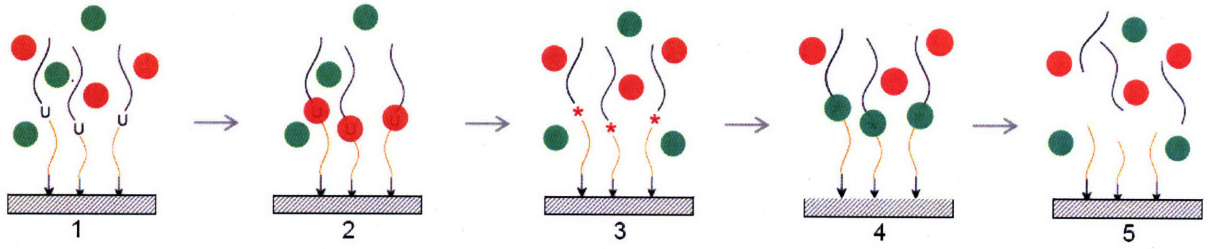


Figure 6.13. A schematic depicting the enzymatic cleavage method to release construction oligos from a microarray surface utilizing UDG and endonuclease. Initially (stage 1) the oligos are anchored by their 3' ends (arrowheads) to the array surface. Each oligo features a Uracil base. UDG (red circles) and endonuclease (green circles) are also in solution. UDG binds to Uracil bases in the single stranded DNA (stage 2), excising the base and leaving an apyrimidinic site (asterisk, stage 3). Endonuclease then binds (stage 4), cleaving at the abasic site and liberating the construction payload (black lines) into solution (stage 5).

As with Mly1, cleavage by UDG/endonuclease was first characterized *in vitro* with un-anchored 60-mer oligos. An antifreeze gene (300 bp) was parsed utilizing DNAWorks into 12 oligos, and the 10 construction oligos were synthesized with a single Uracil base present in each. The design for each oligo, as it would appear on the microarray surface, consisted of (5' to 3'): a 44 base construction payload, a single Uracil base, and a poly-T segment 15 bases long serving as a spacer. An example of such a sequence is shown below for oligo "b2," with the full parse and sequence information for all oligos found in Appendix A6:

Anti-b2 GCAGCACAAACAGCAGGCCGGTCAGTATTACGCTTTTCATATGC/ideoxyU/TTTTTTTTTTTTTTT

In vitro cleavage characterization was accomplished for both sets of enzymes by incubating the 10 construction oligos (each at 25 nM) at 37°C for 1 hour, viewing the digested products with denaturing gel electrophoresis, and tabulating the resultant band intensities. As can be seen in figure 6.14, in the case of UDG with endonuclease IV, the concentration of endonuclease has the most significant impact; for three separate UDG concentrations, endonuclease IV concentration greater than 0.2 U/μL leads to almost complete digestion of the 60-mer reactants. Utilizing the USER mix, for concentrations greater than 0.1 U/μL, similarly all products are digested.

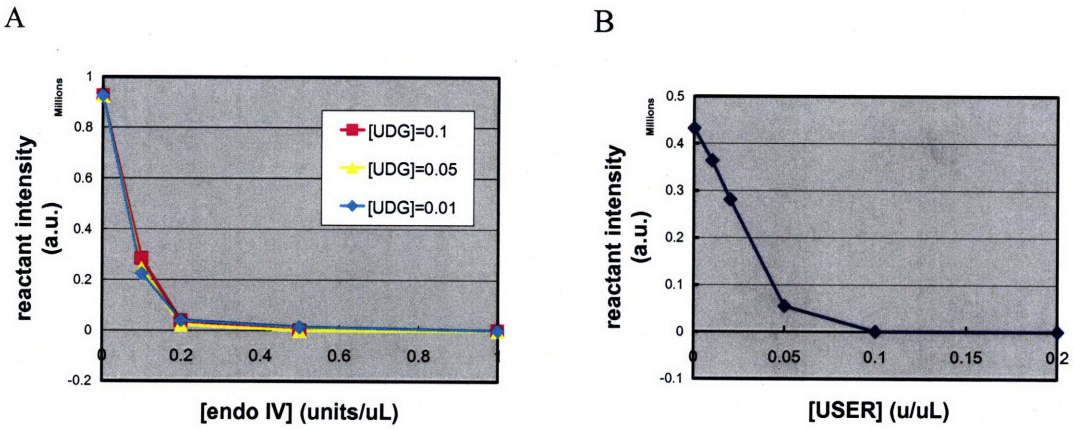


Figure 6.14 Graphs showing digestion of reactant antifreeze oligos as a function of **(a)** UDG and endonuclease IV and **(b)** USER mix.

6.5 Single pot cleavage and synthesis *in vitro* and *in fluidic*

Because microarrays are an expensive commodity, to the extent possible, experiments that would proof the essential concepts without using arrays were conducted. So, upon completion of characterizing the cleavage reactions *in vitro*, the next step involved performing single pot cleavage followed by gene synthesis, both *in vitro* and also in a microfluidic environment, again using un-anchored oligos (no microarray). Reaction mixes testing the Mly1 cleavage system were prepared containing 1 mM dNTPs (250 μM each), 0.15 U/μL of Pfu Hotstart Turbo Polymerase (Stratagene), 1X cloned Pfu Buffer (Stratagene), 0.1% *n*-Dodecyl-β-D-maltoside (Sigma), 25 nM of each construction oligo of the 'shuffled' alba gene, 500 nM of amplifying primers, 350 nM of the helper oligo, and 1 U/μL of Mly1. This mixture was kept on ice until placed either in a PCR tube or loaded into a microfluidic device such as the one described in

section 5.2.2. The reaction proceeded by first incubating at 37°C, where cleavage of the 60-mers into 22-mers and the desired 38-mer construction oligo occurred, followed by thermocycling whereby the 327 bp gene are subsequently assembled and amplified. This one-pot cleavage, assembly, and amplification is successfully verified in the gel shown in Figure 6.15 for the *in vitro* case.

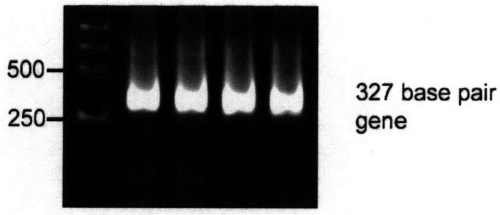


Figure 6.15 PAGE showing the single-pot, *in vitro* cleavage and synthesis of the 327 bp 'shuffled' alba gene.

For the *in fluidic* case, the 4-reactor microfluidic device (500 nL each reactor) was loaded with three separate reaction mixes; in reactor 1 only oligos are present; in reactor 2 MlyI was introduced in addition, facilitating cleavage; in lane 3 polymerase is present, thus enabling PCA. These reactions were carried out in parallel, again initiated with a 1 hour 37°C incubation followed by thermocycling.

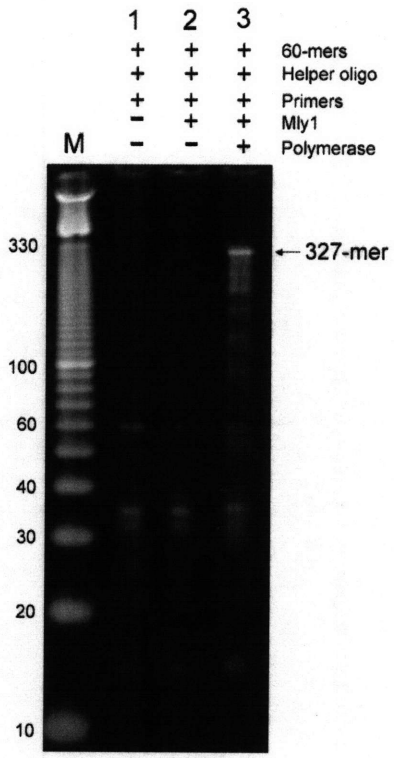


Figure 6.16 Denaturing PAGE demonstrating the *in situ* cleavage of 60-mers yielding construction oligos which are subsequently assembled to the desired 327-mer gene and amplified in a single-step PCA, all in one microfluidic device.

For the UDG and endonuclease blends, similar experiments were carried out. Each combined cleavage and synthesis reaction contained the same reaction mix as indicated previously, with Mly1 and the helper oligo exchanged for either UDG (0.1 U/μL) and endo IV (0.2 U/μL) or USER (0.1 U/μL). As seen in figure 6.15, again, both conditions do in fact yield the desired full-length 300 bp antifreeze gene, for both *in vitro* and also *in fluidic* cases. It is of interest to note here that, despite the presence of a 3'-phosphate for oligos cleaved via USER, synthesis still proceeds; this could be explained by the inherent exonuclease activity of the polymerase. It is also possible that the endonuclease VIII generates 3' hydroxyls instead of 3' phosphates at some low frequency. While such a process yields successful gene synthesis for the

small number of construction oligos seen here (10 oligos), for larger oligo pools such inefficient generation of 3' hydroxyls could adversely affect the concentration of even a single oligo, thus disrupting the assembly process and causing the synthesis to fail.

Figure 6.17 PAGE showing the single-pot, cleavage and synthesis of the 300 bp antifreeze gene both *in vitro* and *in fluidic*.

6.6 Microfluidic-microarray direct cleavage of oligos

Once these various *in vitro* and *in fluidic* tests were completed, the next step was to verify whether the enzymatic cleavage methods would work for microarrays. DNA microarrays were fabricated via a custom Maskless Array Synthesizer (MAS)[1] from the lab of Professor Franco Cerrina. For the Mly1-based array, 60-mers were synthesized for the alba gene (not randomized) with sequences shown in Appendix A6.6, while for the arrays tested for UDG and endonuclease cleavage, 60-mers for alba featuring an individual Uracil base were synthesized. These sequences can also be found in A6.6. The oligo layout is shown in figure 6.18, where the fourteen construction oligos (top strand oligos with prefix 't' and bottom strand oligos with prefix 'b') are tiled throughout the 10.0 mm x 14.0 mm area. Each oligo spot has a 13 μm feature size. Oligos were synthesized with a "1 in 4" format, where only one out of every four spot areas is used for oligo synthesis with the remaining areas capped; such a strategy helps to reduce errors that occur at the edge of each oligo spot exposure[23].

[23] Kim, C., Kaysen, J., Richmond, K., Rodesch, M., Binkowski, B., Chu, L., Li, M., Heinrich, K., Blair, S., Belshaw, P., Sussman, M., and Cerrina, F. (2006) Progress in gene assembly from a MAS-drive DNA microarray. *Microelectronic Engineering* **83**, 1613-1616.

b02	t03	b04	t05	b06	t07	b08	t09	b10	t11	b12	t13	b14	t15	b02
t15	b02	t03	b04	t05	b06	t07	b08	t09	b10	t11	b12	t13	b14	t15
b14	t15	b02	t03	b04	t05	b06	t07	b08	t09	b10	t11	b12	t13	b14
t13	b14	t15	b02	t03	b04	t05	b06	t07	b08	t09	b10	t11	b12	t13

Figure 6.18 Layout for alba construction oligos on the array surface. The dimension of each spot is 13 μm.

Integrated microfluidic-microarray devices were fabricated as described in section 6.3. To test for the various cleavage schemes with microarrays, the following reaction mixes were prepared: 1X cloned Pfu Buffer (Stratagene), 0.1% *n*-Dodecyl-β-D-maltoside (Sigma), and 0.4% Polyvinylpyrrolidone (PVP), combined with either: (a) helper oligo (2 μM) and Mly1 (2 U/μL); (b) UDG (0.3 U/μL) and endo IV (2 U/μL); or (c) USER mix (0.5 U/μL). The PVP was added as it has previously been shown to be effective as a dynamic passivating agent in the case of glass-PDMS systems[24]. Each of these mixes was loaded into the two 500 nL reactors of 3 different microarrays (one for Mly1 digestion and two featuring Uracil bases) and incubated at 37°C for 1 hour. The sample was collected from the device reactors with 4.5 μL of dH_2O (into a 5 μL total volume), and was further diluted 20-fold before three PCR reactions were carried out to assay whether certain individual oligos were successfully liberated from the microarray surface within each reactor. PCR reactions that amplify top strand oligos t03, t07, and t011 (see Appendix A6.6 for primer sequences) were prepared with standard concentrations of polymerase, buffer, and dNTPs and thermocycled for 30 cycles according to the following protocol: 94°C for 2 minutes; 30 cycles of 94°C for 30s, 55°C for 30s, 72°C for 30s; 72°C for 1 minute; and finally down to 4°C. The presence of PCR bands (along with successful negative controls) would indicate that that particular oligo was successfully cleaved from the device. Denaturing PAGE as seen in 6.19a reveals that for Mly1, only oligo t3 yielded a PCR band, while oligos t7 and t11 were not cleaved in sufficient quantity to be assayable by PCR. The "double band" present is a common artifact of denaturing PAGE[25]. For UDG/endoIV and USER cleavage, however, PCR bands of varying intensity were observed for all three sets of assayed oligos (figure 6.19b and c). The band intensities for each experiment are inhomogeneous, however, which is of course not

[24] Xia, Y-M., Hua, Z-S., Srivannavit, O., Ozel, A.B., and Gulari, E. (2007) Minimizing the surface effect of PDMS-glass microchip on polymerase chain reaction by dynamic polymer passivation. *Journal of Chemical Technology & Biotechnology*, **1**, 33-38.
[25] Peter Carr, personal communication

desirable given the fickle nature of oligo concentration as it relates to gene synthesis; however, these initial tests verify that, at a minimum, the enzymatic cleavage methods, without any optimization, do in fact yield cleaved products. The next set of experiments will involve the ultimate step of *in fluidic* cleavage and gene synthesis from the microarray surface.

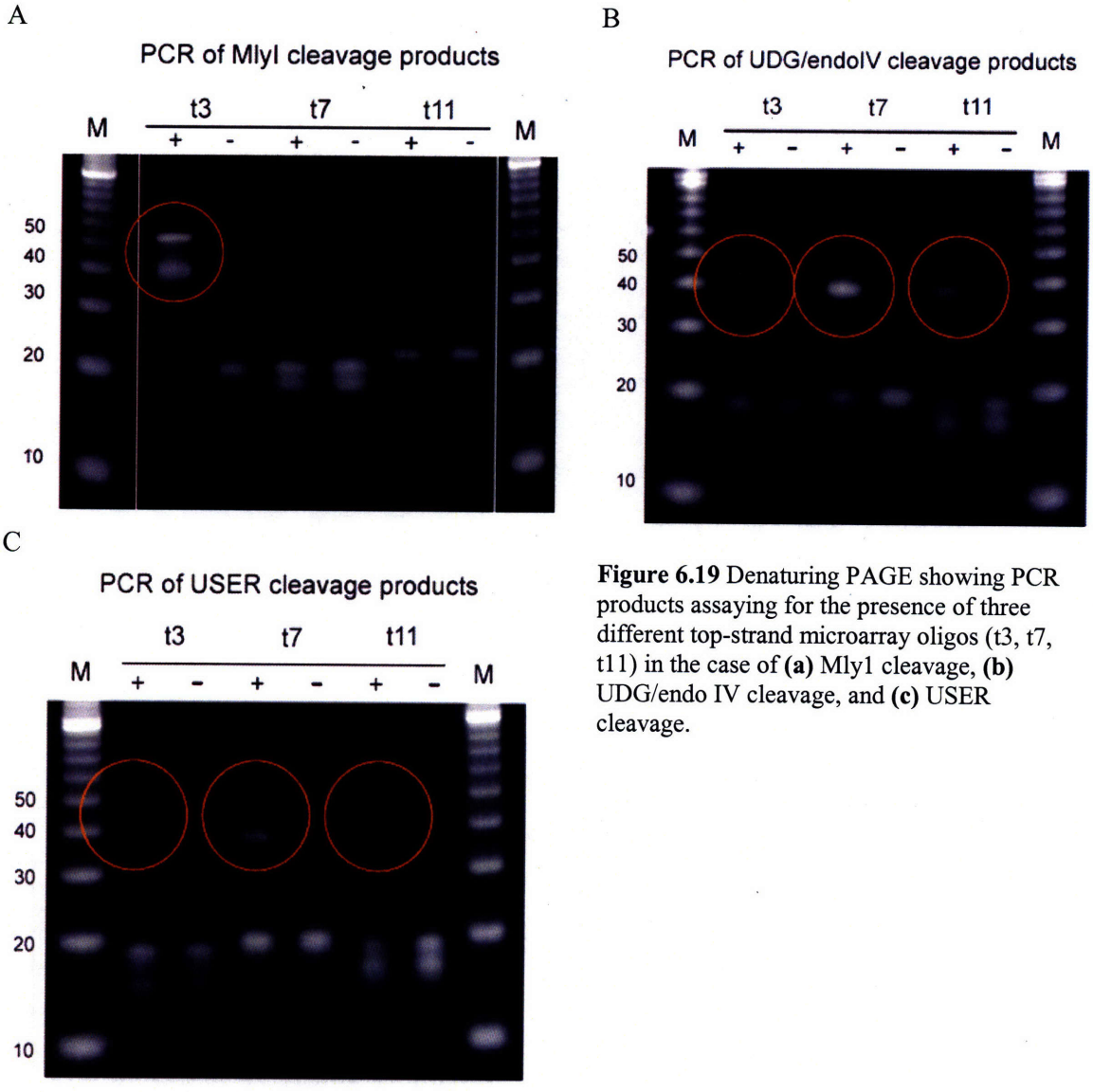


Figure 6.19 Denaturing PAGE showing PCR products assaying for the presence of three different top-strand microarray oligos (t3, t7, t11) in the case of **(a)** Mly1 cleavage, **(b)** UDG/endo IV cleavage, and **(c)** USER cleavage.

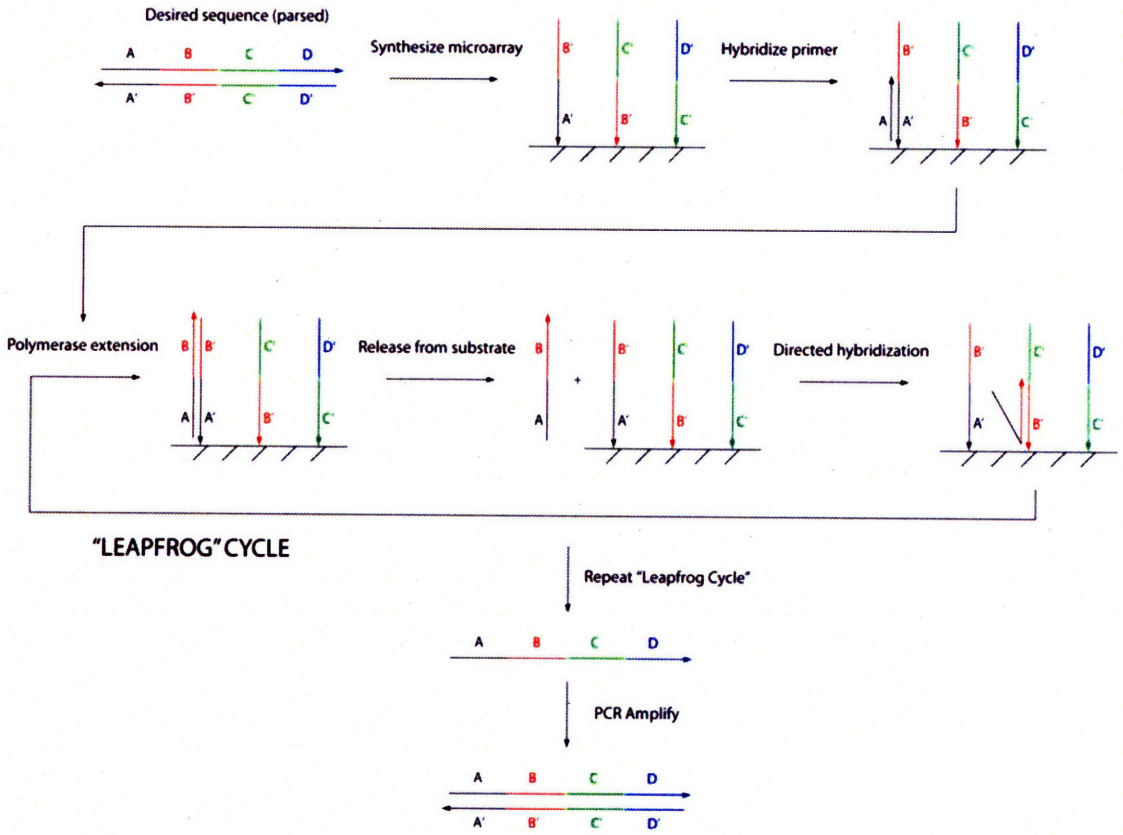


Figure 6.1 "Leapfrog" gene synthesis from microarray oligos. Image courtesy of Brian Chow.

7 Integrated microfluidic gene and protein synthesis

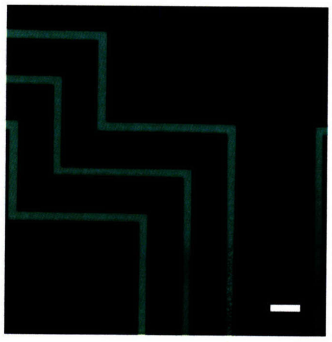


Figure 7.1 (left): Structure of green fluorescent protein (GFP). From Tsien[1].

Figure 7.2 (right): Confocal image of GFP synthesized in a microfluidic device. Scale bar indicates 300 μm. .

7.1 Introduction

Large-scale, high throughput integrated gene and protein synthesis would be an invaluable prototyping tool. Prototyping on the scale of individual proteins would be useful to researchers synthesizing novel proteins, while the potential exists even for those working on constructing genetic circuits *in vitro*. *Many* of the *in vivo* tests of genetic circuits could be replaced with *in fluidic* examinations[2] and directed evolution[3].

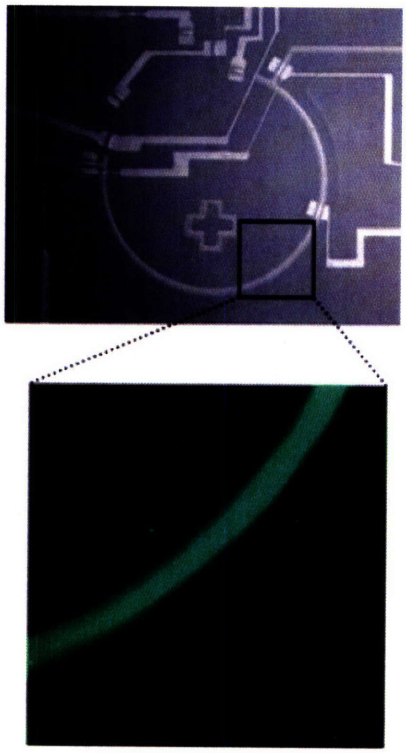


Figure 7.2

[1] http://www.tsienlab.ucsd.edu/Images.htm
[2] Sprinzak D. and Elowitz, M.B. (2005) Reconstruction of genetic circuits. *Nature*, **438**, 443-448.
[3] Griffiths, A.D., and Tawfik, D.S. (2000) Man-made enzymes – from design to *in vitro* compartmentalization. *Current Opinion in Biotechnology*, **11**, 338-353.

7.2 *In vitro* gene and protein synthesis

In order to realize this vision of an integrated gene and protein synthesis device, as has been the guiding philosophy throughout much of this thesis work, steps were first taken to proof the essential concepts *in vitro* before moving to the more challenging microfluidic environment. The first steps in this regard were selection of target fluorescent proteins, then the design of gene expression constructs, followed by the synthesis and expression of these genes *in vitro*. Functioning protein could then be assayed by fluorometer.

7.2.1 Fluorescent Protein Selection

In selecting fluorescent proteins (FPs), several factors were examined: compatibility with the laser lines of the available confocal scope (Zeiss LSM Pascal confocal microscope); intrinsic brightness (i.e. product of extinction coefficient and quantum yield); and lastly compatibility with cell-free transcription/translation systems (e.g. will the protein fold properly?). Tsien et al.[4] was a valuable reference in evaluating the various available fluorescent proteins. The proteins initially selected for gene and protein synthesis are shown in Table 7.1. eGFP, eYFP, and mOrange[5] were well-suited because of their brightness and compatibility with the laser lines of the available confocal scope (488 nm Argon, 514 nm Argon, 548 nm Helium Neon). DsRed was initially tested as well because of its already demonstrated microfluidic gene synthesis from previous work (chapter 5), and Cerulean[6] was later synthesized to add a relatively bright blue fluorescent protein that was also an *Aequorea Victoria* GFP mutant, as those proteins expressed well in initial tests.

Table 7.1: The various fluorescent proteins selected for gene and protein synthesis along with associated characteristics of note.

Fluorescent Protein	Organism	Excitation (nm)	Emission (nm)	Brightness (% of fluorescein)

[4] Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005) A guide to choosing fluorescent proteins. *Nature Methods*, **2**, 905-909.
[5] Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnolol.*, **22**, 1567-1572.
[6] Rizzo, M.A., Springer, G.H., Granada, B. and Piston, D.W. (2004) An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnol.*, **22**, 445-449.

Cerulean	*Aequorea victoria*	433	475	39%
eGFP	*Aequorea victoria*	484	511	49%
eYFP	*Aequorea victoria*	514	527	74%
mOrange	*Discosoma sp.*	548	562	71%
DsRed	*Discosoma sp.*	556	586	5.1%

As an aside, the methods of directed evolution employed to generate these various mutant fluorescent proteins are ones of great interest for miniaturization and automation, and could be a strong application for the technology described in this thesis given sufficient technological maturation and advancement. For example, for the development of both mOrange**[ref]** and Cerulean **[ref]**, a combined rational and combinatorial design was employed, where select amino acid residues were targeted for mutation and mutants were then randomly generated (e.g., two hydrophobic residues in the case of Cerulean). Such an approach would be well-suited for microfluidic gene synthesis (generation of mutant library) and subsequent protein synthesis (prototyping mutants, selecting winners) given the relatively small combinatorial parameter space.

7.2.2 Cell-free protein synthesis with circular and linear templates

Initially, in addition to protein selection, testing of cell-free protein synthesis was also accomplished. While a variety of commercially available cell-free expression systems exist, ultimately the Rapid Translation System (RTS) 100 *E. coli* HY Kit from Roche was chosen for initial tests due to its widespread use in the literature and also its robustness in the hands of other lab members. Initial experiments were performed to assess firstly, whether linear constructs of the type generated during gene synthesis could be successfully expressed, and secondly, whether unpurified products post-gene synthesis could yield functional protein. To assess both of these factors, firstly T7 regulatory elements that would enable gene expression were added to the 993 bp eGFP construct built in Chapter 5 by the "RTS *E. coli* Linear Template Generation Set, His-tag" Kit from Roche. This entails two PCR reactions: one for the additions of overlap regions using custom primers, and a second for the addition of the T7 regulatory elements (downstream promoter, upstream terminator). Addition of T7 regulatory elements to generate linear constructs was also performed for a circular eGFP plasmid (pEGFP, BD Biosciences). Negative controls were also run for these PCR reactions without any template and only primers present.

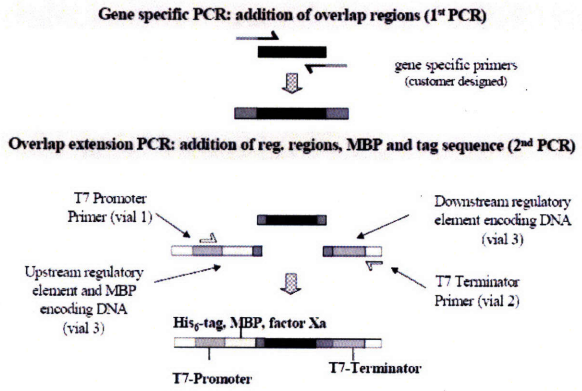


Figure 7.3 Protocol for adding T7 regulatory elements for *in vitro* expression. From the Roche manual for "Rapid Translation System RTS *E. coli* Linear Template Generation Set, His-tag."

Upon completion of regulatory element addition, linearized samples were purified utilizing a QIAquick PCR purification kit. Next, protein synthesis was performed utilizing the RTS 100 *E. coli* HY Kit from Roche. 20 μL reactions containing kit master mix (*E. coli* lysate, Reaction Mix, Amino Acids, Methionine, and Reconstitution Buffer) along with 4 μL of PCR product (linear template from plasmid or PCA, clean and unclean, primer-only negative control, 20% of the reaction volume) was incubated at 30°C for 6 hours in an Eppendorf Mastercycler Gradient thermocycler and maintained at 4°C (either on the thermocycler or in a 4°C refrigerator) prior to being imaged by a Spex Fluoromax fluorometer. Fluorometer measurements were taken in a 120 μL total volume (20 μL protein synthesis + 100 μL dH_2O) with 1-nm steps, a 1 second integration time, entrance/exit slits 2.0 nm, $\lambda_{excitation}$ = 488 nm. Signals were averaged 3 times per sample. The fluorometer scans are shown in figure 7.4

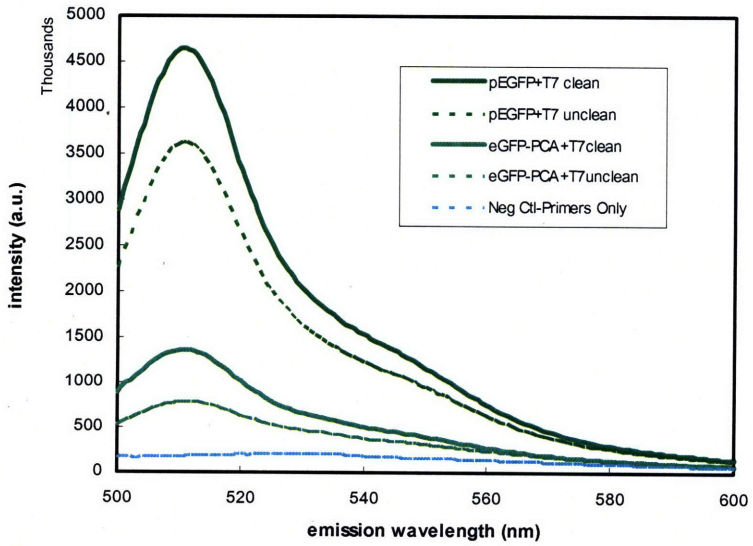


Figure 7.4 Graph showing fluorometer scans with 488 nm excitation.

As can be seen from the fluorometer scans, all eGFP samples yielded functional, fluorescent protein; however, purified samples gave 1.3 and 1.7 times greater fluorescence for linear pEGFP and linear PCA eGFP samples, respectively. The presence of fluorescent protein without purification was a promising result in that, for microfluidic integration, such an on-chip purification step would not be necessary. However, clearly purification has a substantial benefit in terms of increasing signal intensity; thus, future device designs could include some on-chip refinement (e.g. via microfluidic affinity columns[7]).

7.2.3 Design of linear FP gene expression constructs

Given the success of linear constructs in yielding functional FPs, the next step was to design and parse gene expression constructs into oligos for synthesis by PCA. Because the T7 regulatory elements from the Roche linear template generation kit worked well, those exact sequences were added to the coding regions of Cerulean, eGFP, eYFP, and mOrange; the DsRed construct already featured previously designed regulatory elements. DNAWorks was used to generate oligos with these regulatory DNA sequences fixed while codon optimization was utilized for the protein sequences. Complete parses for these constructs can be found in Appendix A.7.2. A schematic depicting the approximate break-up of the oligo design relative to the position of regulatory elements is shown in figure 7.5. Because initial gene synthesis did not fare well for the full length, 30 oligo constructs (Cerulean, eGFP, eYFP, and mOrange were all parsed into 30 oligos as shown in figure 7.5), constructs shorter than the full length were generated. A set of constructs was generated that would preserve functionality of essential regulatory elements for protein synthesis (promoter, RBS). Because oligos t1 and b30 contained potentially unnecessary sequence information (again, these sequences originated from the Roche kit), the following sets of additional constructs were pooled and synthesized for Cerulean and eYFP: t1 to b28 (all regulatory elements), t1 to b26 (no terminator), t3 to b30 (all regulatory elements), t3 to b28 (all regulatory elements), t3 to b26 (no terminator). For eGFP and mOrange, the following sets of additional constructs were pooled and synthesized: t1 to b28 (all regulatory elements), t1 to b24 (no terminator), t3 to b30 (all regulatory elements), t3 to b28 (all regulatory elements), t3 to b24 (no terminator).

---

[7] Hong JW, Studer V, Hang G, Anderson WF, Quake SR. 2004. A nanoliter-scale nucleic acid processor with parallel architecture. *Nat. Biotechnol.* 22(4):435–39

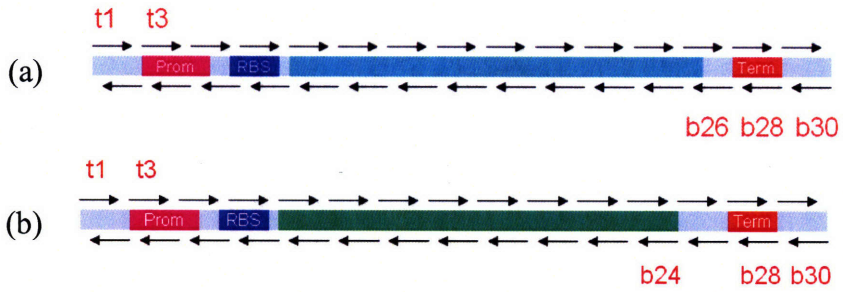


Figure 7.5 Schematic of a gene expression construct for **(a)** Cerulean and eYFP and **(b)** mOrange and eGFP, with oligo positions relative to regulatory elements indicated. Promoter, RBS, and Terminator are labeled, with the blue and green regions indicating the open reading frame for Cerulean and eGFP, respectively.

7.2.4 Analysis of i*n vitro* gene and protein synthesis by PAGE and fluorometer

Once some problematic reagents were eliminated (a bad stock of dNTPs), all variants were eventually generated by PCA, as seen in figures 7.6 and 7.7. PCA reaction mixtures contained: 1 mM dNTPs (250 μM each), 0.05 U/μL of Pfu Hotstart Turbo Polymerase (Stratagene), 1X cloned Pfu Buffer (Stratagene), 10 nM of each construction oligo depending on the construct, and 500 nM of each amplifying outside primer. Negative controls were also conducted where construction oligos were excluded (primers only). Thermocycling was performed on an Eppendorf Mastercycler Gradient thermocycler according to the following protocol: 94°C for two minutes to activate the polymerase; 45 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds; and finally, a final extension period of no less than 2 minutes.

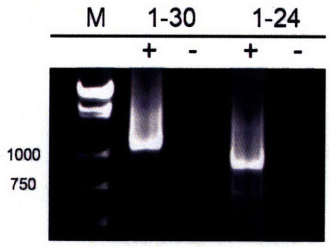


Figure 7.6 PAGE showing synthesis of mOrange variants (1-30) and (1-24) along with primers-only negative controls.

Upon completion of gene synthesis by PCA, cell-free protein synthesis was attempted for all constructs. Protein synthesis and fluorometer scans were conducted as described in section 7.2.2. While all gene expression constructs were in fact synthesized via PCA, ultimately only the *Aequorea Victoria* mutants of GFP yielded functional protein, as seen in figure 7.7. The *Discosoma sp.*FPs (DsRed and mOrange) did not yield fluorescence above background levels; it

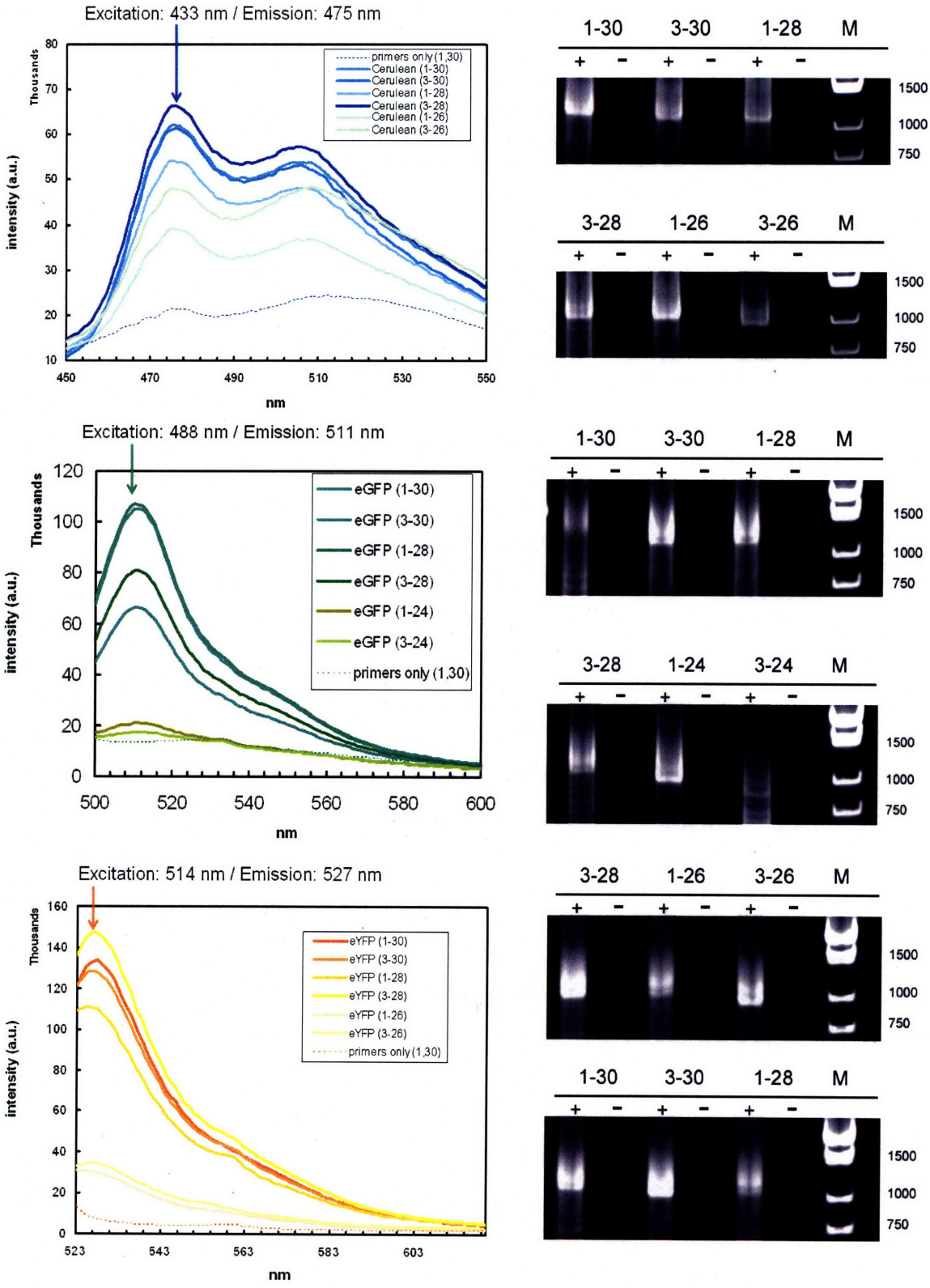


Figure 7.7 Fluorometer and PAGE results for Cerulean, eGFP, and eYFP. Excitation frequencies are indicated for fluorometer scans, while PAGE images for all constructs, including primers-only negative controls, are also shown.

is unknown during what stage of the protein synthesis that failure occurred. Poor synthesis for *Discosoma sp.*FPs could be affected by a variety of factors, including reaction temperature and duration, or the lack of chaperones, for example. The strongest sets of potential concerns, however, deal with a lack of generality in regards to the RTS protein synthesis system. For the proposed integrated gene and protein synthesis tool to be maximally effective, a robust, widely applicable cell-free expression system is highly desirable.

Of particular note for the results shown in figure 7.7 is that significant fluorescence signal was measured for FPs expressed from all constructs despite the lack of a purification step in between gene and protein synthesis. This result boded well for the prospect of miniaturization in microfluidic devices as, again, incorporation of on-chip purification would have been a non-trivial addition. This also highlights the power and ease of gene synthesis; it is possible within a 24 to 48 hour period to go from protein design (virtual) to protein on-hand (physical) via mail order oligos and two simple reactions, thus eliminating the need for any of the laborious steps associated with traditional DNA manipulation (e.g. cloning).

When examining the fluorescence data for this set of experiments, one notable feature in terms of correlating content of the DNA construct to fluorescence intensity (and thus efficiency of expression) was the substantial decrease in fluorescence for all constructs that did not feature terminators. The utility of a highly parallelized gene to protein synthesis device would depend greatly upon its ability to generate similar data sets which could lead to conclusions linking genotype and phenotype. Again, such a tool could be invaluable for large scale comparative structural studies (e.g. biobrick prototyping: do I get better expression with terminator A or terminator B?).

7.3 Gene and protein synthesis in microfluidic devices

Given successful *in vitro* gene and protein synthesis results, the next step involved miniaturizing each of these processes in a microfluidic environment, first separately and finally in an integrated system. Given the already successful microfluidic gene synthesis results of chapters 5 and 6, the next major step was the demonstration of microfluidic protein synthesis. Several other research

groups have recently demonstrated synthesis of proteins in microfabricated volumes[8, 9,10,11,12], notably all in a well-based format (no on-chip fluid manipulations; all automation must come from external agents, e.g. fluid dispensing robots). Key characteristics of each of these works are shown in table 7.2; device images from Yamomoto, Mei, and Kinpara can be seen in figure 7.8. The most impressive demonstration was the work of Kinpara et al., where GFP synthesis from 10 DNA molecules was measured in 1 pL wells, a volume only ~100 larger than an *E. coli* cell (tens of femtoliters). The use of a glass and PDMS material set for that work also bodes well in terms of gauging the potential ultimate limits of minimum reactor volume and maximum reactor density for the work presented in this thesis. Both microfluidic gene and protein synthesis reaction volumes should scale down well from the tens to hundreds of nLs used in this work. As an aside, it would also be interesting to test via a system such as Kinpara's reactors how reactor volume (and ultimately surface effects) affects the resolution of gene expression. At what point would I be unable to compare the performance between two terminators with known performance differences? At what point do we start to sacrifice resolution for throughput?

Table 7.2: A comparison of various *in vitro* protein syntheses performed in miniaturized systems.

Reference	Year	Expression System	Microchip material	Reactor Volume	Protein
Tabuchi **[REF]**	2002	RTS	PMMA	10 μL	Adipose-type fatty acid binding protein (A-FABP)
Yamamoto	2002	RTS	Glass, ITO, PDMS	125 nL	GFP, Blue Fluorescent Protein (BFP)
Angenendt	2004	RTS	glass	1.5 μL to 100 nL	GFP, □-galactosidase
Kinpara	2004	RTS	PDMS, glass	150, 5, and 1 pL	GFP
Mei	2005	RTS	Acrylic	13 μL	GFP, CAT, luciferase

[8] Tabuchi M, Hino M, Shinohara Y, Baba Y (2002) Cell-free protein synthesis on a microchip. *Proteomics,* **2**, 430–435

[9] Yamamoto T, Nojima T, Fujii T (2002) PDMS-glass hybrid microreactor array with embedded temperature control device. Application to cell-free protein synthesis. *Lab on a Chip*, **2,** 197–202

[10] Angenendt P, Nyarsik L, SzaflarskiW, Glokler J, Nierhaus KH, Lehrach H, Cahill DJ, Lueking A (2004) Cell-free protein expression and functional assay in nanowell chip format. *Anal Chem* **76,** 1844–1849.

[11] Kinpara T, Mizuno R, Murakami Y, KobayashiM, Yamaura S, Hasan Q, Morita Y, Nakano H, Yamane T, Tamiya E (2004) A picoliter chamber array for cell-free protein synthesis. *J Biochem* **136**, 149–154

[12] Mei Q, Fredrickson CK, Jin SG, Fan ZH (2005) Toxin detection by a miniaturized in vitro protein expression array. *Anal Chem* **77**, 5494–5500.

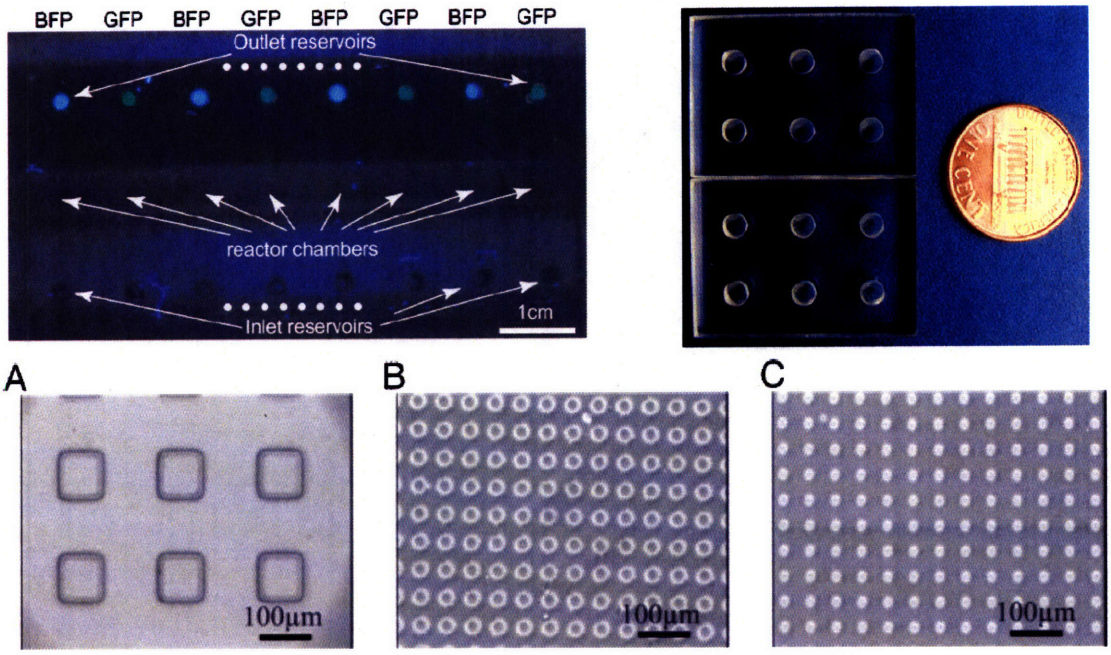


Figure 7.8 Images from of various microchips utilized for *in vitro* protein synthesis, including: 125 nL reactors integrated with ITO heaters from Yamamoto **(upper left)**; 13 μL wells in acrylic from Mei **(upper right)**; and 150 pL **(a)**, 5 pL **(b)**, and 1 pL **(c)** PDMS-on-glass well arrays from Kinpara.

7.3.1 Design and fabrication of microfluidic devices for gene and protein synthesis

Microfluidic devices for gene and protein synthesis were designed and fabricated for conducting initial experiments testing out the individual modules. The microfluidic device presented in section 5.2.2 was modified by adding two additional 500 nL reactors (7 total) and also bifurcation channels to evenly distribute flow at the inlets and outlets of the device, as seen in figure 7.9. Other slight modifications on the wafer-level in terms of device spacing and placement were made to facilitate easier fabrication.

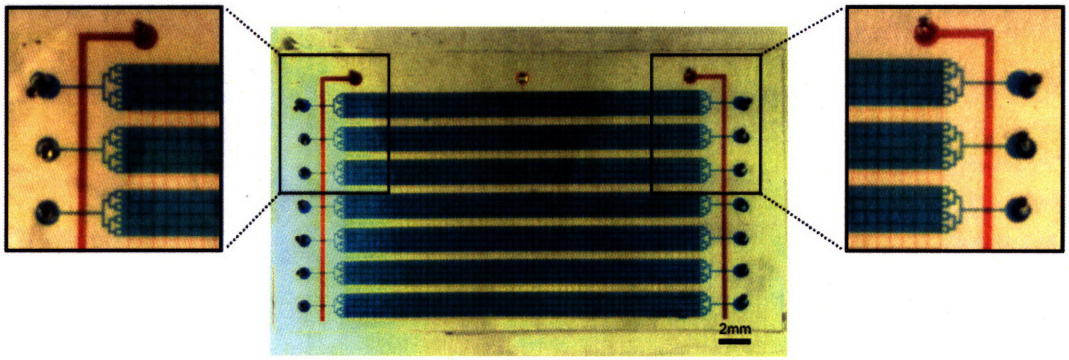
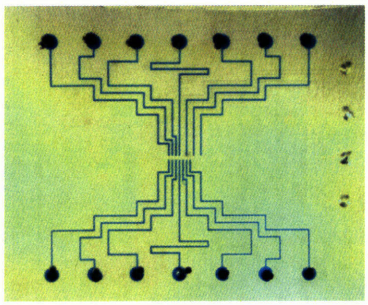


Figure 7.9 Optical image of an updated device design for microfluidic gene synthesis featuring 7 parallel 500 nL reactors. Insets show bifurcation channels at inlets and exits.

Microfluidic devices for protein synthesis were designed and fabricated by multi-layer soft lithography (section 5.2.2) as shown in figure 7.10. The devices feature 14 parallel reactors, each 15 nL in volume (excluding the dead volume from the device inlet), with channel widths of 100 μm. The devices were designed such that all fourteen reactors could be viewed within a single field of view of the confocal scope (Zeiss LSM Pascal confocal microscope) at the center of the chip. The reactors also have no outlets; they are designed to be dead-end pressure loaded and imaged, so protein is never extracted from the device.

The device shown in figure 7.10a is a single-layer device, while the one in figure 7.10b features two-layers. The 'control' layer here does not feature valves, but instead channels from which air or oxygen can be supplied to nearby reaction channels. The presence or absence of molecular oxygen has a significant impact upon the maturation of FPs, as O_2 is required to dehydrogenate amino acids during chromophore formation[13]. Because fluorescence formation is prevented by anoxic conditions, it was thought that, despite the porosity of PDMS, maturation efficiency could be enhanced by an increased oxygen flux. In an ideal case, such oxygen channels could potentially increase the signal from each reactor in a tunable fashion (as a function of oxygen flux), which would of course be a welcome outcome. Such experiments will be the subject of future work.

A



B

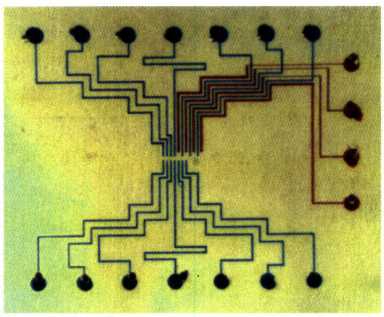


Figure 7.10 Optical image of two microfluidic devices designed for protein synthesis, with food coloring utilized to highlight features: **(a)** a single-layer device featuring 14 parallel reactors (blue), each 15 nL in volume and **(b)** a two-layer device with channels to supply oxygen to potentially improve protein maturation.

[13] Hansen, M.C., Palmer, R.J., Jr, Udsen, C., White, D.C. & Molin, S. (2001) Assessment of GFP fluorescence in cells of *Streptococcus gordonii* under conditions of low pH and low oxygen concentrations. *Microbiology*, **147**, 1383-1391.

7.3.2 Gene and protein synthesis in microfluidic devices

Upon completion of device fabrication, the next step was to demonstrate successful microfluidic gene synthesis for the linear FP gene expression constructs detailed in section 7.2.3. Reaction mixes were prepared as described in section 7.2.4, and microfluidic gene synthesis was performed as described in sections 5.2.5 and 5.2.7. The results of PAGE (section 5.2.9) are shown in figure 7.11 for Cerulean, eYFP, and eGFP constructs, indicating that microfluidic gene synthesis yielded FP gene expression constructs of the appropriate length.

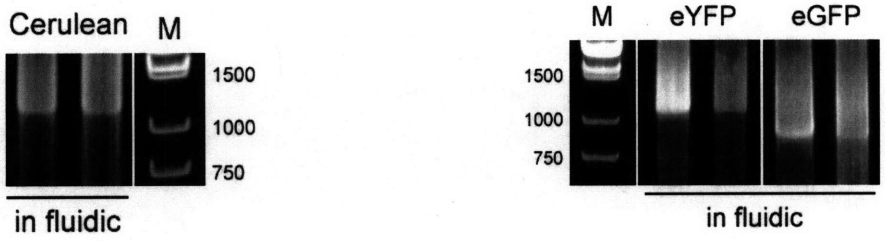


Figure 7.11 PAGE results for microfluidic synthesis of Cerulean, eYFP, and eGFP.

Upon completion of microfluidic gene synthesis, microfluidic protein synthesis was then tested. Initial experiments were performed with the original microfluidic device design of section 5.2.5. Protein synthesis mixtures were prepared as in section 7.2.2; the DNA template utilized was the pIVEX GFP expression vector included with the Roche kit, which was diluted to a concentration of 0.01 μg/μL in the final reaction mix. This mixture was then loaded into the 500 nL reactors of the microfluidic device and incubated for 6 hours at 30°C, followed by storage at 4°C prior to imaging via confocal microscopy. Scanning was accomplished utilizing the 488 nm Argon line with a 505 nm long pass (LP) filter. The results of the imaging are shown in figure 7.12, with green false coloring. As can be seen, the microfluidic protein synthesis yielded strong fluorescence relative to the background bulk PDMS. Given this successful result, the next step was to assess whether the linear FP gene expression constructs assembled via PCA could similarly yield functional protein in a microfluidic environment.

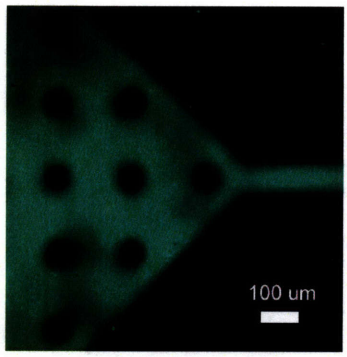


Figure 7.12 Confocal microscopy image of GFP synthesized in a 500 nL microfluidic reactor. Excitation was performed at 488 nm while imaging was accomplished a 505 nm long-pass filter.

Microfluidic protein synthesis experiments were first conducted with linear gene synthesis constructs assembled *in vitro*. eGFP construct (1-24, no terminator) was synthesized via PCA (section 7.2.4) and then prepared as the DNA template in a standard protein synthesis mixture (section 7.2.2). Negative controls with protein synthesis mixture and primers-only (negative control from PCA) were also prepared. Mixes were then dead-end loaded into the 15 nL reaction channels of the protein synthesis device at 10 psi for 5 minutes to ensure complete reactor filling (i.e. no air bubbles). Again, given the device design, channels with protein synthesis experiments and negative controls could be compared side-by-side in a single field of view of the confocal scope. Incubation was performed at 30°C for 6 hours on the Eppendorf Mastercycler Gradient thermocycler with *in situ* adapter suitable for four devices; following incubation, chips were maintained at 4°C either until imaging via confocal microscopy. Scanning was accomplished utilizing the 488 nm Argon line with a 505 nm long pass (LP) filter; images are shown in figure 7.13, with the autofluorescence negative control labeled. All other channels giving fluorescent signal contained protein synthesis mixture with the eGFP (1-24) construct, indicating successful synthesis of functional protein from PCA assembled linear gene expression constructs.

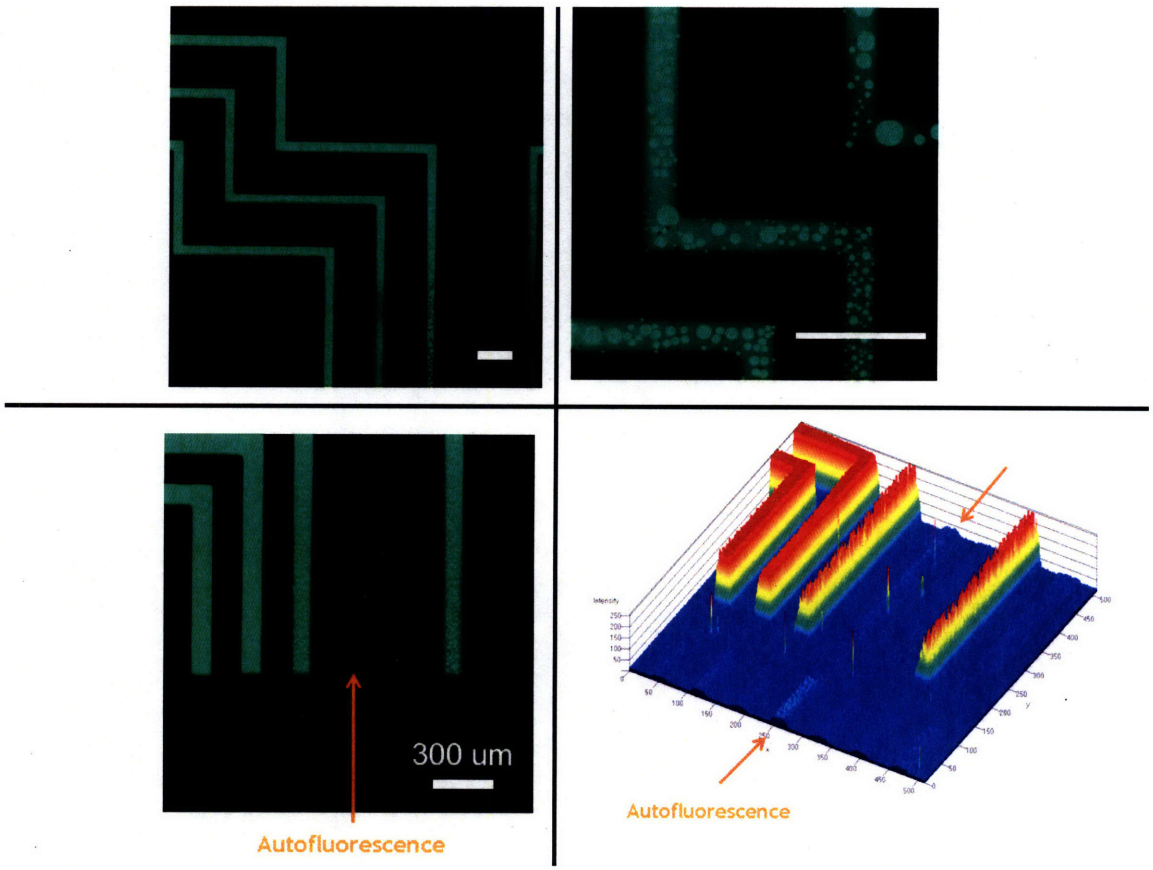


Figure 7.13 Confocal microscope images showing strong fluorescent signal relative to autofluorescence indicating successful synthesis of eGFP from linear templates assembled by PCA. Excitation was performed at 488 nm while imaging was accomplished a 505 nm long-pass filter. Scalebars represent 300 μm.

While strong fluorescence is observable relative to autofluorescence indicating the successful synthesis of functional protein, some notable and potentially undesirable features include the presence of what are believed to be concentrated agglomerations of protein, as seen in figure 7.13. While not characterized as a function of time for this work, over long periods (many hours) protein progressively agglomerates until fairly regularly spaced islands are resolved, as seen for example in Figure 7.14. Here, eYFP (constructed by PCA) was synthesized in the microfluidic device and imaged after approximately 30 hours of storage at 4°C post-synthesis. While such aggregation may not be an issue for assay and prototyping applications (e.g. comparing effects of terminator function on expression), it may prove problematic for applications requiring long-term protein stability, storage or on-chip utilization (e.g. olfactory sensing on a chip). Protein aggregation could potentially be reduced via improved surface passivation; as shall be seen in subsequent sections, several passivating agents were tested to improve device bio-compatibility.

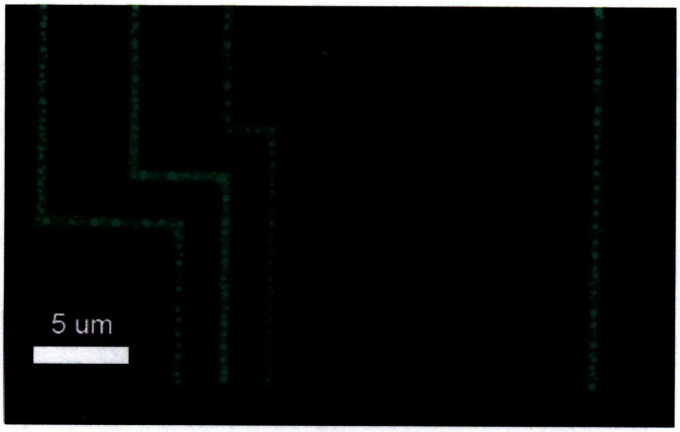


Figure 7.14 Confocal microscopy image of microfluidic synthesis of eYFP. Imaging was performed at approximately 30 hours of storage at 4°C, showing regularly spaced islands.

Upon completion of microfluidic protein synthesis from linear FP gene expression constructs assembled by *in vitro* PCA, the next step toward proofing the full integration paradigm involved first synthesizing the constructs in a microfluidic environment followed by microfluidic protein synthesis. For this procedure gene synthesis devices shown in section 7.3.1 were first utilized to construct eYFP and eGFP constructs (section 7.3.2). Upon completion of synthesis, protein synthesis mixture (2 μL) was utilized to collect the 500 nL gene synthesis volumes, thus yielding a reaction mix of the appropriate concentration for protein synthesis, and this mix was then loaded into the protein synthesis channels. The reaction mixture containing protein synthesis mix and eluted microfluidic gene synthesis mix (~2.5 μL) was sufficient to load several protein synthesis channels. Once loaded, the devices were then incubated at 30°C for 6 hours followed by storage at 4°C prior to imaging. The results of such an experiment are shown in figure 7.15, where eYFP and eGFP were synthesized in parallel (following parallel microfluidic gene synthesis). Three channels were utilized for the synthesis of both fluorescent proteins, as indicated. Scans were taken utilizing 488 nm and 514 nm excitation for eGFP and eYFP excitation along with 505 nm and 514 nm long pass filters, respectively. As can be seen from figure 7.15, strong fluorescent signal relative to autofluorescence was detected, indicating the successful synthesis of fluorescent protein. There is, however, again the presence of regularly-spaced aggregates, including in the autofluorescence for this particular sample, indicating that components of the cell-free expression kit can also agglomerate over time and may also be mediating the aggregation effects observed during synthesis of fluorescent protein.

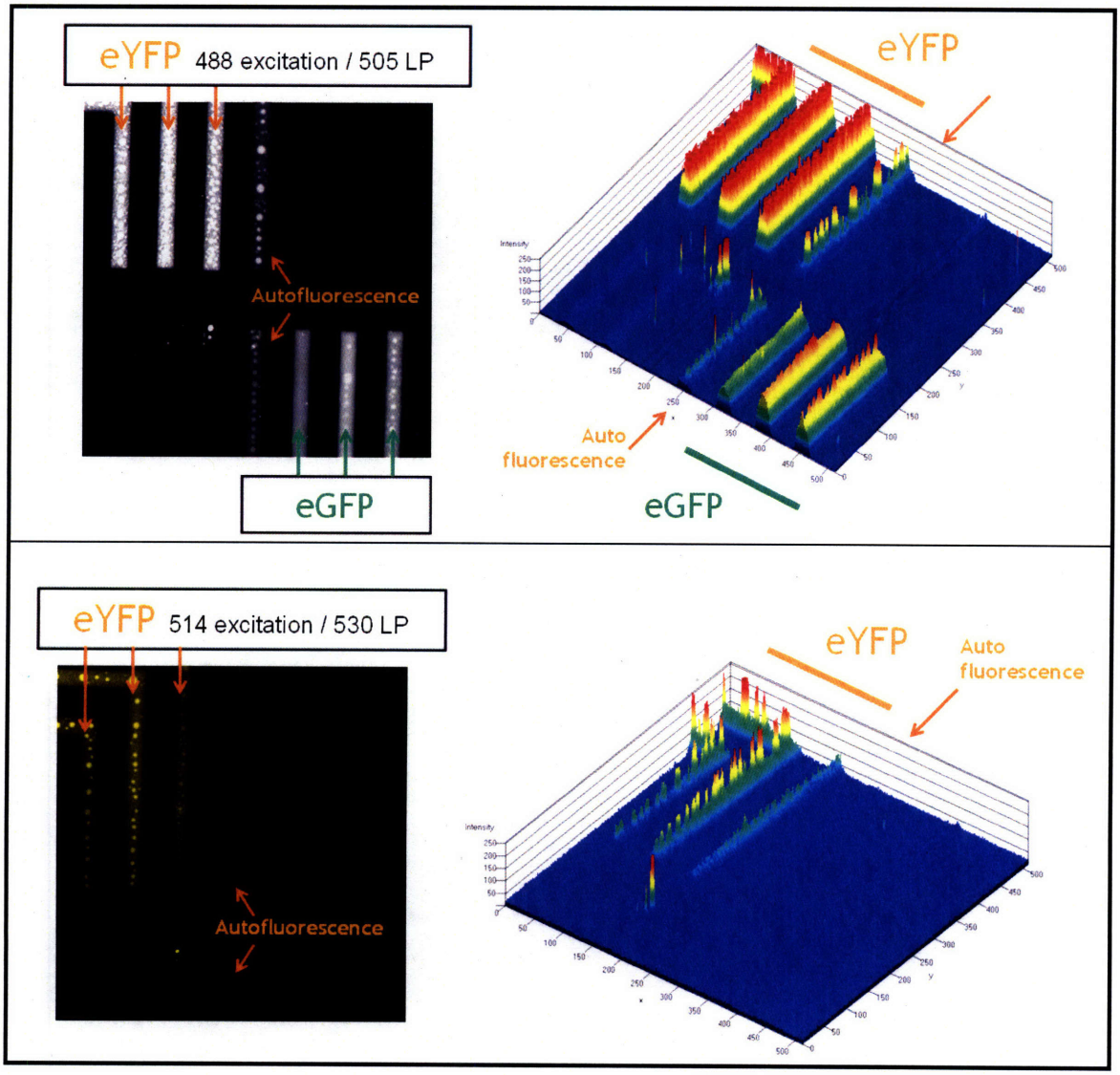


Figure 7.15 Confocal microscope images of eYFP and eGFP synthesized in a microfluidic environment following parallel microfluidic synthesis of gene expression constructs.

One major challenge for future work will be the optimization of microfluidic protein synthesis such that robust, reliable data can be consistently generated. Consistency is important not only from reactor to reactor, but also for within a single microchannel. As can be seen from figure 7.16, for even the same reaction mix within a single microfluidic channel significant variation is evident. Here, an eYFP gene expression construct was first synthesized in a microfluidic environment followed by microfluidic protein synthesis in channels 2 and 5, from left to right. Channel 4 is an autofluorescence negative control. Such inhomogeneity could have arisen from a variety of sources, from non-uniformly mixed reaction mixtures to variability in the composition of the PDMS microchannels. Particularly because PDMS is an elastomer, its surface properties

can be dynamic; as such, surface functionalization, coatings, and highly standardized device manufacture should be explored to increase the homogeneity of surface properties.

While for the previous microfluidic protein synthesis results in this chapter utilized no surface coatings, the results shown in figure 7.16 relied upon a 100 cycle pre-treatment with 0.1% *n*-Dodecyl-β-D-maltoside (Sigma) as described in section 5.2.5. More extensive studies are required to establish the potential benefit of such treatment, but anecdotally protein aggregation appeared to be diminished, though artifacts such as inhomogeneous expression in microchannels is of course still evident.

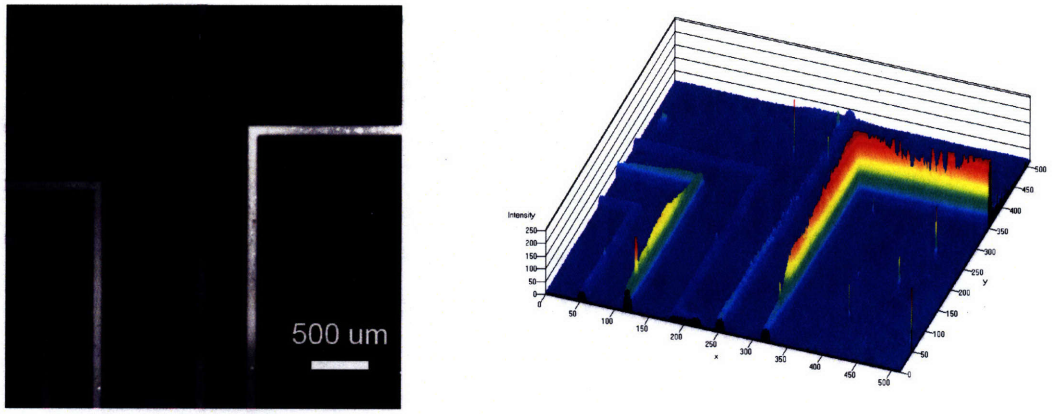


Figure 7.16 Inhomogeneous expression in a microfluidic channel for eYFP synthesis (channels 2 and 5 from left to right). Autofluorescence (channel 4 from left to right) is also shown.

The successful demonstration of microfluidic protein synthesis utilizing gene expression constructs similarly synthesized in a microfluidic environment strongly implies that the proposed integrated system should work. These experiments in and of themselves, however, are already a non-trivial demonstration of the power of miniaturization. The Roche cell-free expression system utilized provides enough reaction mix for conducting twenty-four 50 μL reactions, or, a total of 1.2 mL worth of mix, at a cost of $475 (~$24/reaction); in comparison, given the 15 nL reaction volumes utilized here, approximately 80,000 reactions could be performed given the same amount of starting material (0.6 cents/reaction)! While the simple design and substantial dead-volume at the reaction inlet (~0.5 μL) for each reaction channel of the protein synthesis device utilized here prevents maximal use of 80,000 reactions, given a more sophisticated design with better amortization of reagents (e.g. a single reaction inlet addressing multiple reactors, such as in

the designed integrated system shown in the next section. Liu et al.[14] provides a good example of efficient reagent amortization in microfluidics), such efficient use of reaction mixture should be realizable. Other simple advancements in dead volume reduction have also been demonstrated for fluidic inlets utilizing inert filler materials[15]. Even with the substantial dead-volume requirement for channel loading in the design shown in figure 7.10, volumes are still small enough to yield approximately 2,400 reactions for the Roche kit ($0.20/reaction). Combined with the similarly minimal reagent costs associated with *de novo* microfluidic gene synthesis, use of these two sets of microfluidic processes already represents a low-cost protocol for allowing researchers to rapidly transition from biological design to physical molecule. For assay and prototyping applications, such high throughput devices should prove to be a useful enabling technology given the minimal reagent input per information output.

7.4 Integrated gene and protein synthesis in a microfluidic device

Having proofed the core technologies for both gene and protein synthesis, I then designed the integrated device shown schematically in figure 7.16. The device features several core elements, notably three parallel serpentine gene synthesis reactors (45 nL each) and also three protein synthesis reactors (12 nL each). Device operation is described in figure 7.17.

[14] Liu, J., Hansen, C., and Quake, S.R. (2003) Solving the "world-to-chip" interface problem with a microfluidic matrix. *Anal. Chem.*, **75**, 4718-4723.
[15] Will Grover, personal communication.

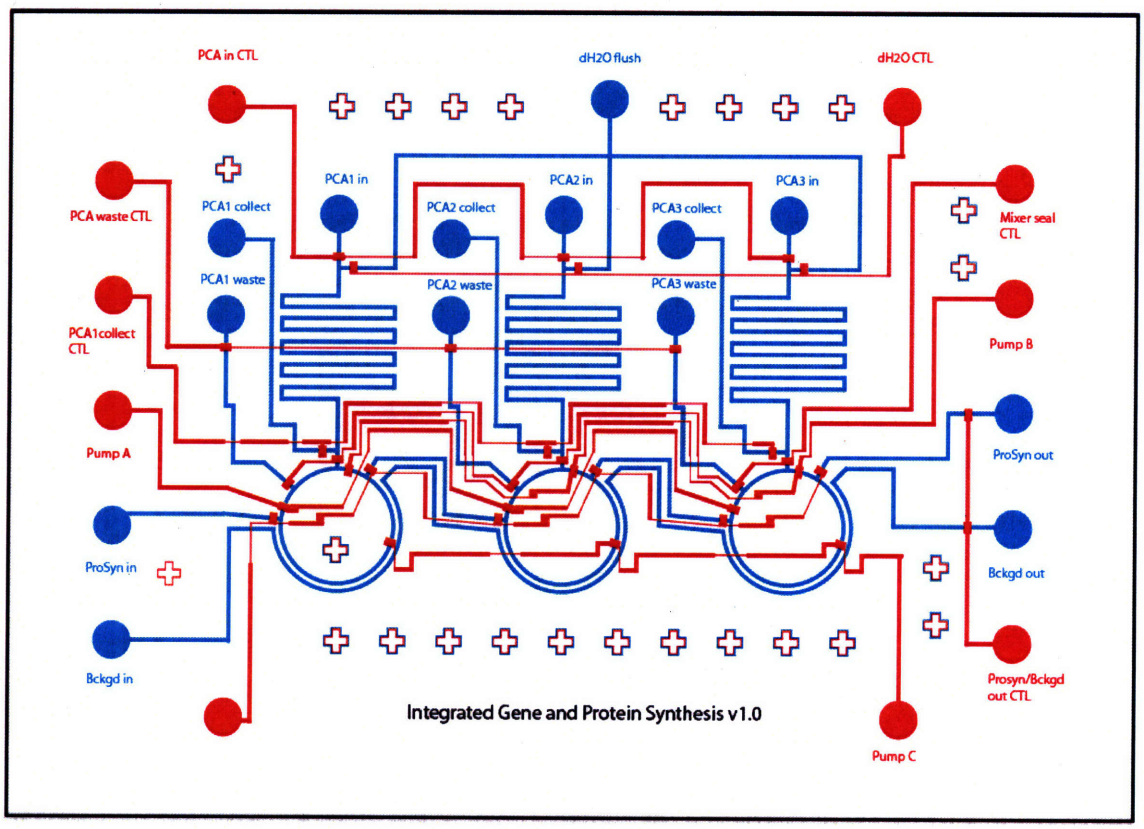


Figure 7.17 Denaturing PAGE showing the products of the *in vitro* digestion of sixteen 60-mers, facilitated by a 15-mer helper oligo. The sixteen released 38-mers are the construction oligos to be used in DNA assembly. Lane 1 shows the digested product; lane 2 the original undigested mix.

Device Fabrication

Devices were fabricated utilizing multi-layer soft lithography with "push-up" geometry. In such a configuration, the control valves are positioned beneath the flow channels instead of above, thus reducing the required sealing pressure by as much as a factor of three [**ref**]. Utilizing "push-up" geometry proved to be particularly crucial as the high sealing pressures (~15 psi) required for "push-down" valving, along with the long sealing times and high temperatures required during device pre-treatment and thermocycling, sometimes resulted in permanent valve sealing, thus rendering devices useless.

Pump characterization

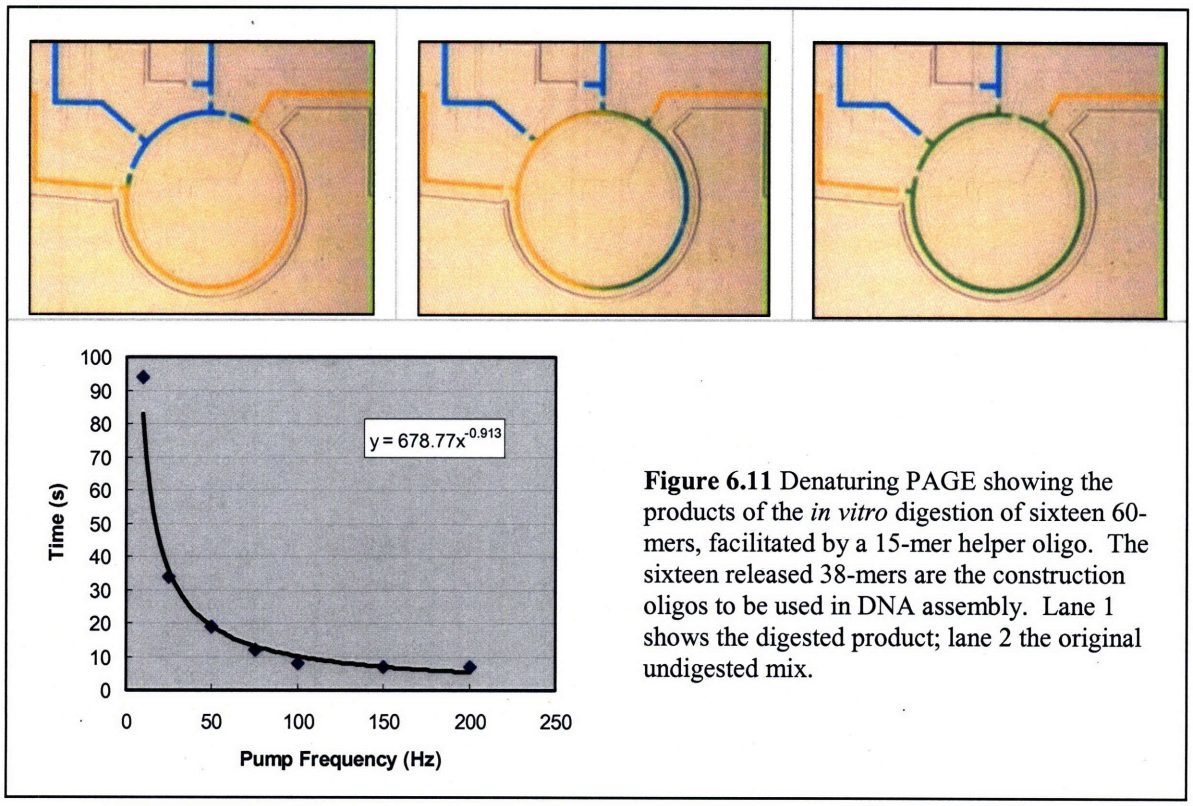


Figure 6.11 Denaturing PAGE showing the products of the *in vitro* digestion of sixteen 60-mers, facilitated by a 15-mer helper oligo. The sixteen released 38-mers are the construction oligos to be used in DNA assembly. Lane 1 shows the digested product; lane 2 the original undigested mix.

*protein could be pulled off the chip.

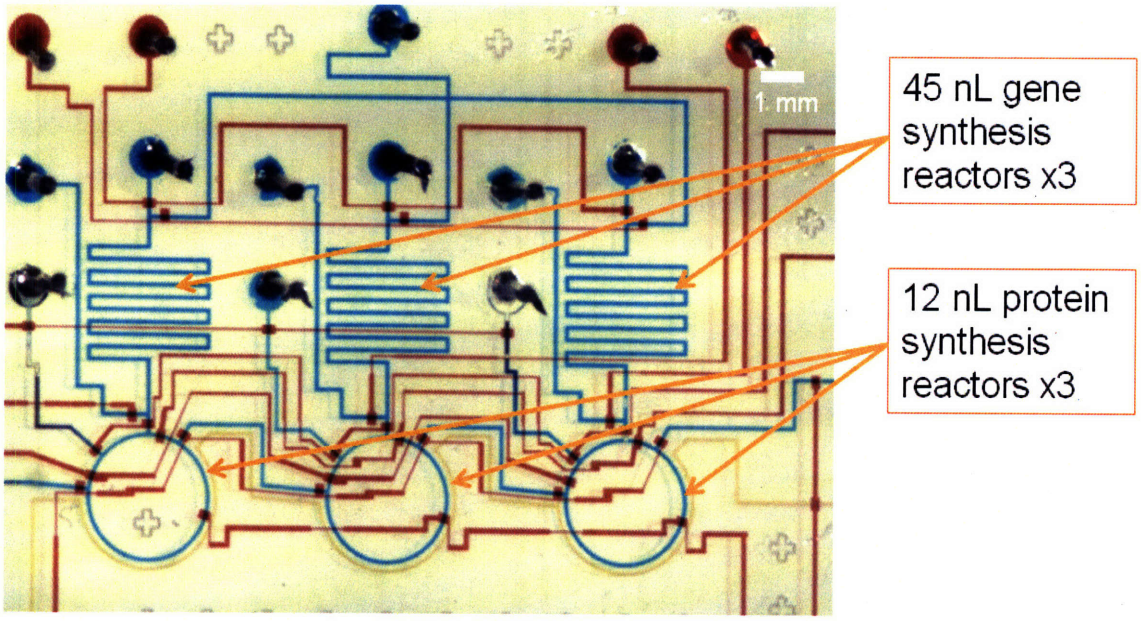


Figure 6.11 Denaturing PAGE showing the products of the *in vitro* digestion of sixteen 60-mers, facilitated by a 15-mer helper oligo. The sixteen released 38-mers are the construction oligos to be used in DNA assembly. Lane 1 shows the digested product; lane 2 the original undigested mix.

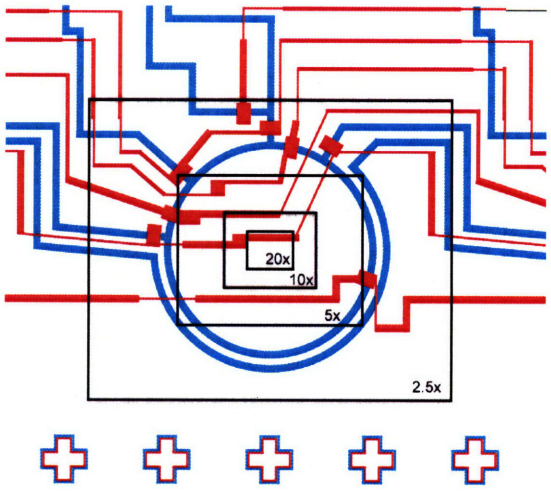


Figure 6.11 Denaturing PAGE showing the products of the *in vitro* digestion of sixteen 60-mers, facilitated by a 15-mer helper oligo. The sixteen released 38-mers are the construction oligos to be used in DNA assembly. Lane 1 shows the digested product; lane 2 the original undigested mix.

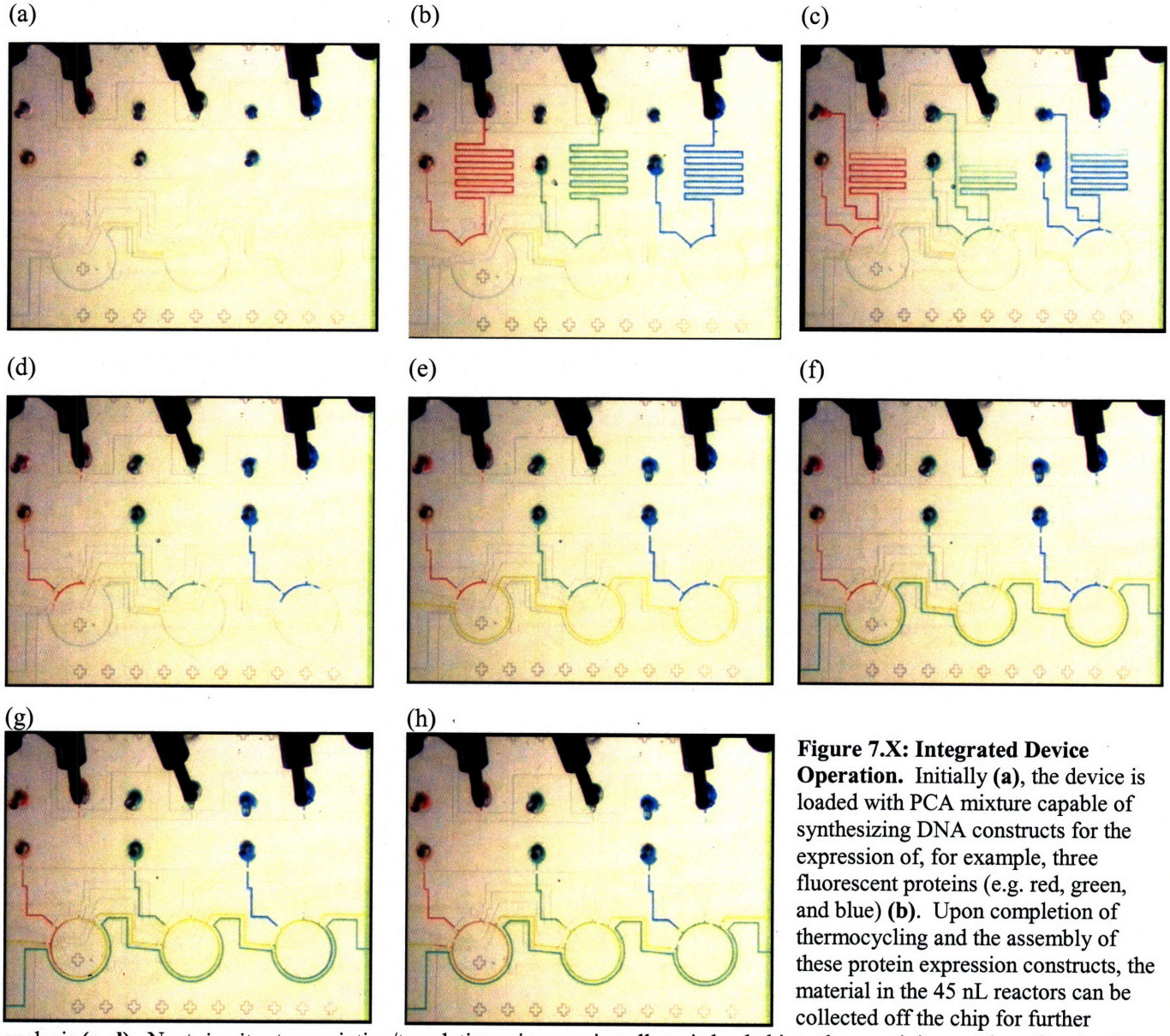
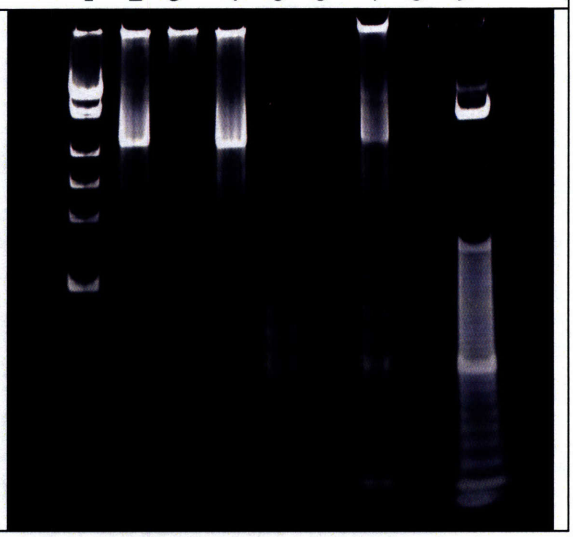
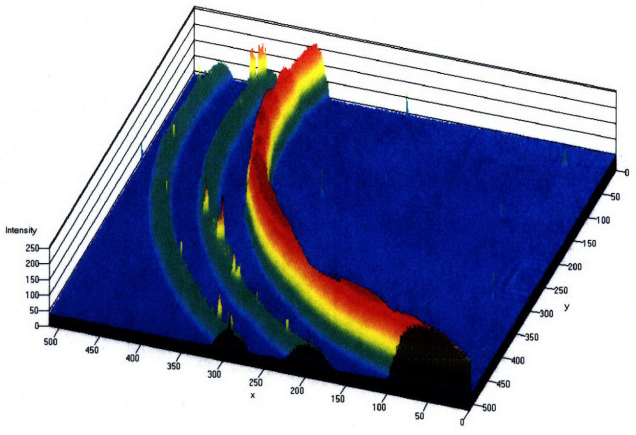


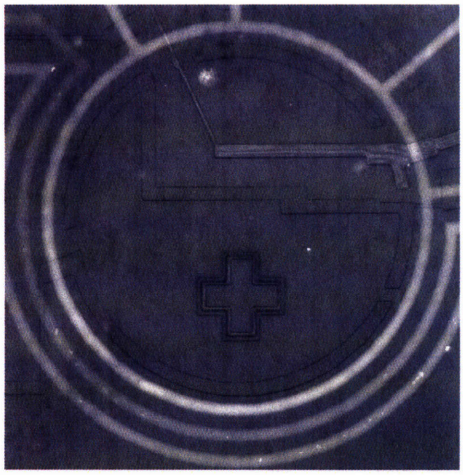
Figure 7.X: Integrated Device Operation. Initially **(a)**, the device is loaded with PCA mixture capable of synthesizing DNA constructs for the expression of, for example, three fluorescent proteins (e.g. red, green, and blue) **(b)**. Upon completion of thermocycling and the assembly of these protein expression constructs, the material in the 45 nL reactors can be collected off the chip for further analysis **(c, d)**. Next, *in vitro* transcription/translation mix, seen in yellow, is loaded into the remaining portion of each of the three mixers **(e)**, followed by the introduction of an auto-fluorescence negative control **(f)**. Finally, each of the three mixer rings is operated **(g)**, mixing the *in vitro* transcription/translation mix with each synthesized DNA construct. Upon completion of mixing **(h)**, the device can be incubated at 30°C to initiate the synthesis of fluorescent protein.

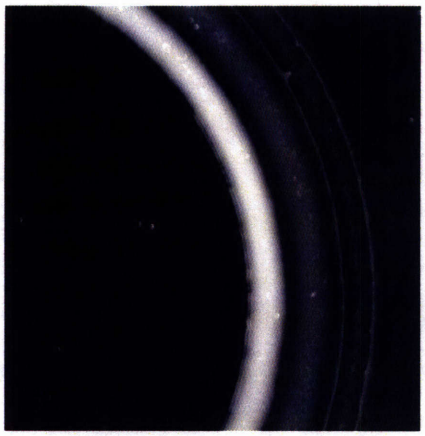
Below, showing 45 nL synthesis of eGFP in reactor 2.

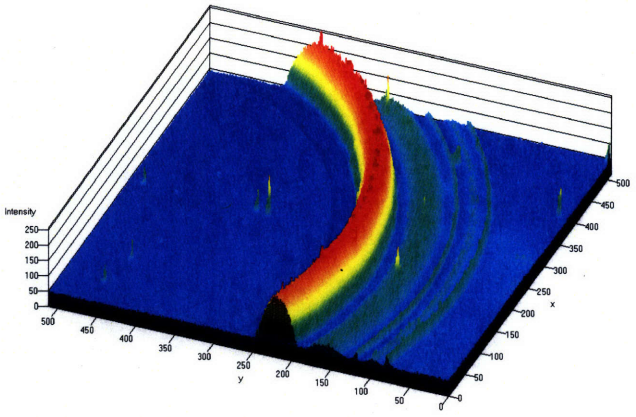
4-12% GPS gel	1 2 3 4 5 6 7 8 9
1. kb ladder 2. eYFP (+) 3. eYFP (-) 4. eGFP (+) 5. eGFP (-) 6. GPS8/13/07C-eYFP reactor 1 7. GPS8/13/07C- eGFP reactor 2 8. GPS8/13/07C- eYFP reactor 3 9. 10 bp ladder	

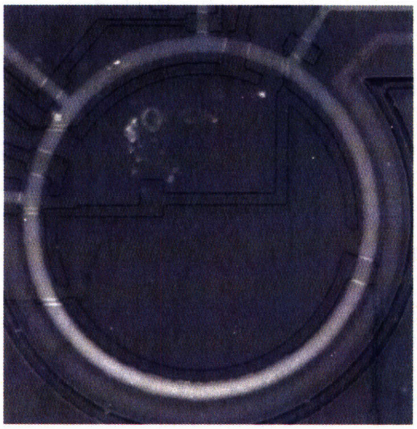
Surface Treatment of the Chip. At the beginning of each experiment, Pluronic F-127 (Sigma-Aldrich Corp., St. Louis, MO; 0.2% w/w in PBS, filter-sterilized) was incubated for 1 h inside the entire network of flow channels except the culture chambers. This passivated the PDMS surfaces and therefore prevented adsorption of proteins and adherence of the cells to the channels.

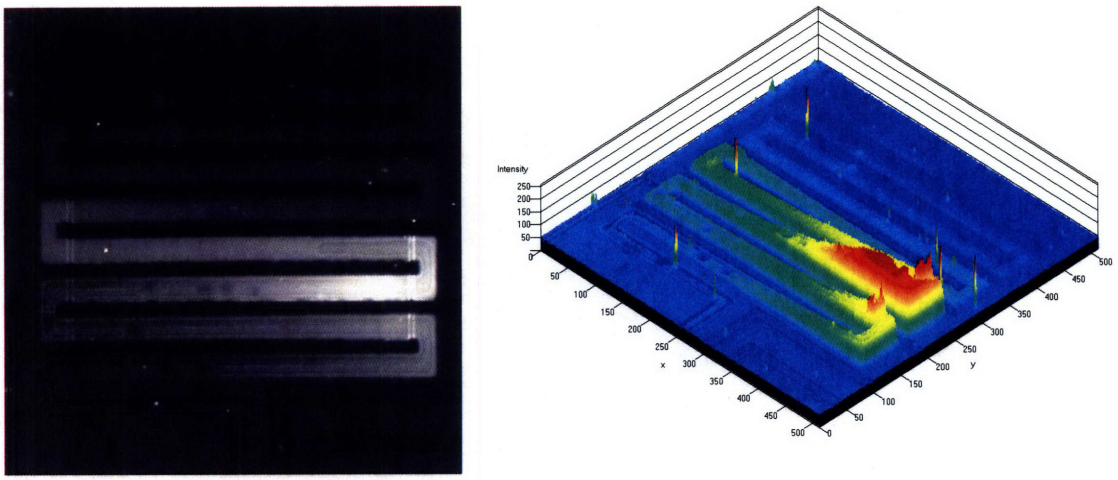




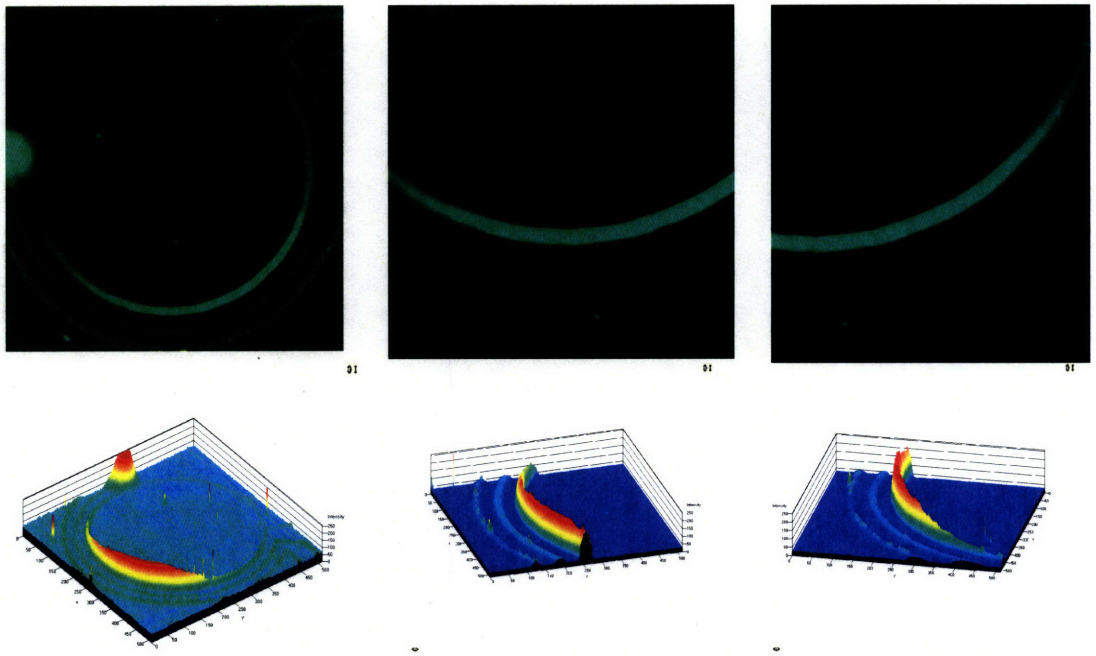



Intensity
250
200
150
100
50
0
500
450
400
350
300
250
y
200
150
100
50
0
x
500
450
400
350
300
250
200
150
100
50
0




Intensity
250
200
150
100
50
0
50
100
150
200
250
x
300
350
400
450
500
0
50
100
150
200
250
y
300
350
400
450
500

Discussion



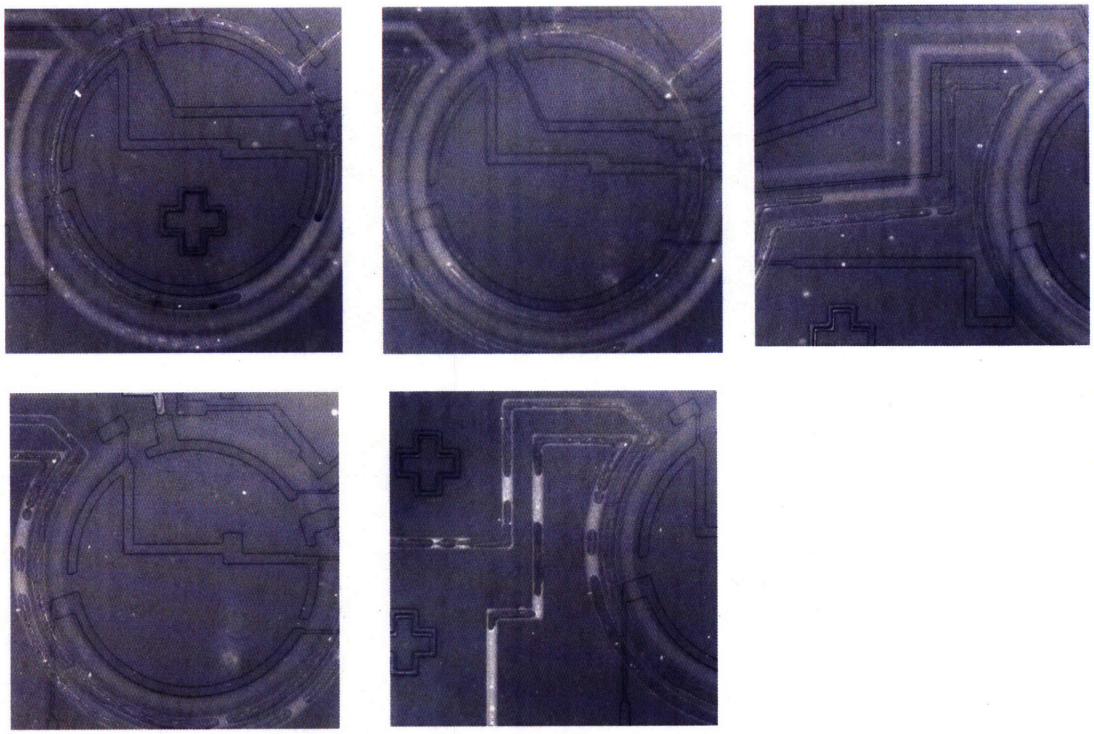
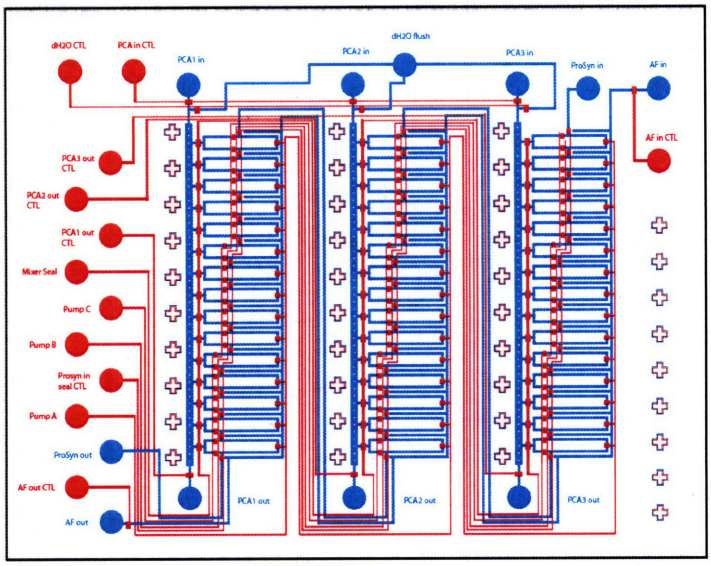
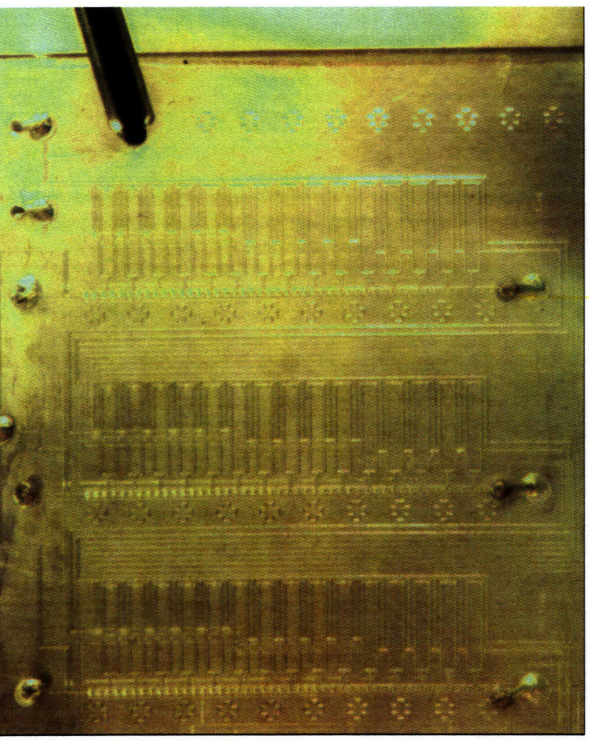


Figure 7.8 Images from of various microchips utilized for *in vitro* protein synthesis, including: 125 nL reactors integrated with ITO heaters from Yamamoto **(upper left)**; 13 µL wells in acrylic from Mei **(upper right)**; and 150 pL **(a)**, 5 pL **(b)**, and 1 pL **(c)** PDMS-on-glass well arrays from Kinpara.

Scaling up





Appendix

A.5 Chapter 5 appendix

DNAWorks output file for the alba gene from *S. solfataricus*. The amino acid sequence was randomized prior to conducting the parse.

```
-------------------------------------------------------------
                  PARAMETERS FOR TRIAL 9
 -------------------------------------------------------------
              Total Size Of Gene ..... 327 nt
              Protein Residues ....... 101
              Mutatable Residues ..... 98
              Fixed Nucleotides ...... 33 nt
              Oligo Size ............. 38 nt
              Annealing Temp ......... 59 +/- 1*C
              Oligo Concentration .... 2.50E-8 M
              Sodium Concentration ... 5.00E-2 M
              Mg2+ Concentration ..... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ....... 8 nt
              Mispriming Threshold ... 8/18 (6 exact) nt

-------------------------------------------------------------------------------

 The DNA sequence #   9 is:
 -------------------------------------------------------------
   1 CGCAGGCTTGCTGATGGTGGTGTACATCAAAGTGCGCCGTGCTGACATTATGCAAGAATC
  61 TCAGCCGCGTCGTCGCGGTGTGGAAAAGAAAACGATCGGTTTCAAGGACCAGCTCGAAAT
 121 CGGCAAAGTCAGCAACTCCAAGGGCATCATCATCTCCGTTGGCATGAAGGAGGCGAACAC
 181 CATCTCTGCCGTCAGCCTGACGAGCGTGCAGATTAACACCCTGGTCAAGCTGTCTAACAT
 241 CGTTATCACCCGCGGCAAACGTCTGGAGGTGCTGGCGGCAAGCGTTACTCCGCCTCGTGT
 301 TGACGAACCGGCGTAAGCAAGAACTAC
 -------------------------------------------------------------

 The oligonucleotide assembly is:
 -------------------------------------------------------------
     1       10        20        30        40        50        60
     |       |         |         |         |         |         |

     1 --->                              3 --->
   1 CGCAGGCTTGCTGATGGTGGTGTACATCAAAGTGCgccgtgctgacattatgcaagaatc
                   CCACCACATGTAGTTTCACGCGGCACGACTGTAATACG    tag
                                                    <---  2
                 M  V  V  Y  I  K  V  R  R  A  D  I  M  Q  E  S

     |       |         |         |         |         |         |

                   5 --->                                   7 --->
  61 tcagccgcgtcgt  CGGTGTGGAAAAGAAAACGATCGGTTTCAAGGACCAGC cgaaat
     agtcggcgcagcagcgccacacctttttctttgctAGCCAAAGTTCCTGGTCGAGCTTTA
                              <---  4
       Q  P  R  R  R  G  V  E  K  K  T  I  G  F  K  D  Q  L  E  I

     |       |         |         |         |         |         |

                                     9 --->
 121 cggcaaagtcagcaactccaagggcatcatca CTCCGTTGGCATGAAGGAGGCGAACAC
     GCCGTTTCAGTCGttgaggttcccgtagtagtagaggcaaccgtacttcct  GCTTGTG
           <---  6                              <---  8
```

```
      G  K  V  S  N  S  K  G  I  I  I  S  V  G  M  K  E  A  N  T

    |        |        |        |        |        |        |

                   11 --->                                  13 --->
181 CATCTCTGCCG    cctgacgagcgtgcagattaacaccctggtcaagctgtCTAACAT
    GTAGAGACGGCAGTCGGACTGCTCGCACGTC    tgtgggaccagttcgacagattgta
                              <--- 10
      I  S  A  V  S  L  T  S  V  Q  I  N  T  L  V  K  L  S  N  I

    |        |        |        |        |        |        |

                                      15 --->
241 CGTTATCACCCGCGGCAAACGTCTGGAGGTG   gcggcaagcgttactccgcctcgtgt
            ********                                                < GC rich
    gcaatagtgggcgCCGTTTGCAGACCTCCACGACCGCCGTTCGCAATGAGG    gcaca
          <--- 12                                  <--- 14
      V  I  T  R  G  K  R  L  E  V  L  A  A  S  V  T  P  P  R  V

    |        |        |        |        |        |        |


301 tgacgaaccggc
    actgcttggccgcattcgttcttgatg
                        <--- 16
      D  E  P  A  X
-----------------------------------------------------------------

The total codon usage score ..........     0.000
The total length score ...............     0.000
The total hairpin score ..............     0.000
The total melting temperature score ...    0.000
The total repeat score ...............     0.000
The total pattern score ..............     0.000
The total mispriming score ...........     0.000
The total AT content score ...........     0.000
The total GC content score ...........     0.489
               The OVERALL score ......    0.489


                  DETAILED CODON FREQUENCY REPORT
 [ Codon, AA, Frequency, # of times used in coding sequence ]
-----------------------------------------------------------------

TTT F 0.29   0  TCT S 0.32   3  TAT Y 0.35   0  TGT C 0.39   0
TTC F 0.71   1  TCC S 0.27   2  TAC Y 0.65   1  TGC C 0.61   0
TTA L 0.03   0  TCA S 0.05   0  TAA X 0.63   1  TGA X 0.35   0
TTG L 0.06   0  TCG S 0.07   0  TAG X 0.08   0  TGG W 1.00   0

CTT L 0.06   0  CCT P 0.11   1  CAT H 0.30   0  CGT R 0.64   5
CTC L 0.08   1  CCC P 0.02   0  CAC H 0.70   0  CGC R 0.33   3
CTA L 0.01   0  CCA P 0.15   0  CAA Q 0.19   1  CGA R 0.01   0
CTG L 0.77   5  CCG P 0.72   3  CAG Q 0.81   3  CGG R 0.01   0

ATT I 0.34   2  ACT T 0.29   1  AAT N 0.17   0  AGT S 0.05   0
ATC I 0.66   9  ACC T 0.54   3  AAC N 0.83   4  AGC S 0.24   4
```

ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	4	AGA	R	0.01	0
ATG	M	1.00	3	ACG	T	0.13	2	AAG	K	0.22	5	AGG	R	0.00	0
GTT	V	0.40	4	GCT	A	0.28	1	GAT	D	0.46	0	GGT	G	0.51	2
GTC	V	0.14	3	GCC	A	0.16	1	GAC	D	0.54	3	GGC	G	0.43	4
GTA	V	0.20	0	GCA	A	0.24	1	GAA	E	0.75	4	GGA	G	0.02	0
GTG	V	0.27	6	GCG	A	0.32	3	GAG	E	0.25	2	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	1
10% - 14%	6
15% - 19%	2
20% - 24%	12
25% - 29%	10
30% - 34%	11
35% - 39%	4
40% - 44%	4
45% - 49%	0
>= 50%	51

Total Codons Used = 101

Tm Range	# of Overlaps
<47^C	0
47-49^C	0
50-52^C	0
53-55^C	0
56-58^C	8
59-61^C	7
62-64^C	0
65-67^C	0
68-70^C	0
71-73^C	0
>=74^C	0

Tm Range = 1.9

Ovrlap Len Range	# of Oligos
<10	0
10-11	0
12-13	0
14-15	0
16-17	4
18-19	8
20-21	3
22-23	0
24-25	0
26-27	0
28-29	0
>=30	0

Lowest Overlap = 16

Length Range	# of Oligos
<20	0
20-24	0
25-29	0
30-34	0
35-39	14
40-44	0
45-49	0
50-54	0
55-59	0
>=60	0

Longest = 38

Sequence Patterns Screened (As Supplied By User)

None found

16 oligonucleotides need to be synthesized

1	CGCAGGCTTGCTGATGGTGGTGTACATCAAAGTGC	35
2	GCATAATGTCAGCACGGCGCACTTTGATGTACACCACC	38
3	GCCGTGCTGACATTATGCAAGAATCTCAGCCGCGTCGT	38
4	TCGTTTTCTTTTCCACACCGCGACGACGCGGCTGAGAT	38
5	CGGTGTGGAAAAGAAAACGATCGGTTTCAAGGACCAGC	38
6	GCTGACTTTGCCGATTTCGAGCTGGTCCTTGAAACCGA	38
7	CGAAATCGGCAAAGTCAGCAACTCCAAGGGCATCATCA	38
8	TCCTTCATGCCAACGGAGATGATGATGCCCTTGGAGTT	38
9	CTCCGTTGGCATGAAGGAGGCGAACACCATCTCTGCCG	38
10	CTGCACGCTCGTCAGGCTGACGGCAGAGATGGTGTTCG	38
11	CCTGACGAGCGTGCAGATTAACACCCTGGTCAAGCTGT	38
12	GCGGGTGATAACGATGTTAGACAGCTTGACCAGGGTGT	38
13	CTAACATCGTTATCACCCGCGGCAAACGTCTGGAGGTG	38
14	GGAGTAACGCTTGCCGCCAGCACCTCCAGACGTTTGCC	38
15	GCGGCAAGCGTTACTCCGCCTCGTGTTGACGAACCGGC	38
16	GTAGTTCTTGCTTACGCCGGTTCGTCAACACG	32

TABLE IIb

DNAWorks output file for the Holliday junction cleavase (hjc) gene from bacteriophage *SIRV-1*.

```
-----------------------------------------------------------------
                  PARAMETERS FOR TRIAL 14
 ---------------------------------------------------------------
              Total Size Of Gene ..... 390 nt
              Protein Residues ..... 122
              Mutatable Residues ..... 121
              Fixed Nucleotides ..... 27 nt
              Oligo Size ..... 48 nt
              Annealing Temp ..... 60 +/- 1*C
              Oligo Concentration ..... 2.00E-8 M
              Sodium Concentration ..... 5.00E-2 M
              Mg2+ Concentration ..... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ..... 8 nt
              Mispriming Threshold ..... 8/18 (6 exact) nt

-------------------------------------------------------------------------------

 The DNA sequence #  14 is:
 ---------------------------------------------------------------
   1 CAGGTAATTCCATATGAACATCCGTCAGTCTGGTAAATACTACGAGTACAAAACTCTGGA
  61 GATCCTGGAAAAGAATGGTTTCAAAGCGCTGCGTATCCCGGTTTCTGGTACCGGCAAACA
 121 GGCGCTGCCGGACCTGATCGCGACCAAAAACAACACCATCTACCCTATTGAAGTTAAATC
 181 TACCTCTAAAGACGTTGTTACCGTTCGTAATTTCCAGATCGAAAAACTGTTCAAATTCTG
 241 CGAAATCTTCAACTTCTGTGAATGCCACCCGCTGGTAACCGTTTACTACAAGAAATACAA
 301 AATCGTTATCGTTTATGAACTGTCTCAGGACGTTCGCACCAAAGAAAAAATCAAGTTCAA
 361 GTACGGCATCAACTCCTAACTCGAGCGGAC
 ---------------------------------------------------------------

 The oligonucleotide assembly is:
 ---------------------------------------------------------------
     1       10        20        30        40        50        60
     |        |         |         |         |         |         |

     1 --->                  3 --->
   1 CAGGTAATTCCATATGAACATCCGTcagtctggtaaatactacgagtacaaaactctgga
       CCATTAAGGTATACTTGTAGGCAGTCAGACCATTTATGATGCTCATGT ttgagacct
                                                <---  2
                  M  N  I  R  Q  S  G  K  Y  Y  E  Y  K  T  L  E

     |        |         |         |         |         |         |

                          5 --->
  61 gatcctggaaaag         CAAAGCGCTGCGTATCCCGGTTTCTGGTACCGGCAAACA
     ctaggacctttttcttaccaaagtttcgcgacgcatagggg              CGTTTGT
                                     <---  4
       I  L  E  K  N  G  F  K  A  L  R  I  P  V  S  G  T  G  K  Q

     |        |         |         |         |         |         |

                        7 --->
```

```
121 GGCGCTGCC           cgaccaaaaacaacaccatctaccctattgaagttaaatc
    CCGCGACGGCCTGGACTAGCGCTGGTTTTTGTTGTGGTAGA gggataacttcaatttag
                                        <---  6
      A  L  P  D  L  I  A  T  K  N  N  T  I  Y  P  I  E  V  K  S

    |        |        |        |        |        |        |

             9 --->                                            1
181 tacctcta AGACGTTGTTACCGTTCGTAATTTCCAGATCGAAAAACTGTTCAAATT  g
    atggagatttctgcaacaatggcaagcatt   GGTCTAGCTTTTTGACAAGTTTAAGAC
                          <---  8
      T  S  K  D  V  V  T  V  R  N  F  Q  I  E  K  L  F  K  F  C

    |        |        |        |        |        |        |

    1 --->                                          13 --->
241 cgaaatcttcaacttctgtgaatgccacccgctggtaaccgtttact CAAGAAATACAA
    GCTTTAGAAGTTGAAGACACT       ggcgaccattggcaaatgatgttctttatgtt
                  <--- 10
      E  I  F  N  F  C  E  C  H  P  L  V  T  V  Y  Y  K  K  Y  K

    |        |        |        |        |        |        |

                                          15 --->
301 AATCGTTATCGTTTATGAACTGTCTCAGGACGTTCG  ccaaagaaaaaatcaagttcaa
    ttagcaatagcaaataCTTGACAGAGTCCTGCAAGCGTGGTTTCTTTTTTAGTTCAAGTT
            <--- 12                                          <--
      I  V  I  V  Y  E  L  S  Q  D  V  R  T  K  E  K  I  K  F  K

    |        |        |        |        |        |        |


361 gtacggcatcaactcctaactcgagc
    CATG cgtagttgaggattgagctcgcctg
    - 14                  <--- 16
      Y  G  I  N  S  X
-----------------------------------------------------------------

The total codon usage score ..........     0.000
The total length score ...............     0.000
The total hairpin score ..............     0.000
The total melting temperature score ...    0.000
The total repeat score ...............     0.000
The total pattern score ..............     0.718
The total mispriming score ...........     0.000
The total AT content score ...........     0.000
The total GC content score ...........     0.000
              The OVERALL score ......     0.718


                DETAILED CODON FREQUENCY REPORT
 [ Codon, AA, Frequency, # of times used in coding sequence ]
-----------------------------------------------------------------

TTT F 0.29   0  TCT S 0.32   5  TAT Y 0.35   1  TGT C 0.39   1
TTC F 0.71   7  TCC S 0.27   1  TAC Y 0.65   8  TGC C 0.61   2
```

TTA	L	0.03	0	TCA	S	0.05	0	TAA	X	0.63	1	TGA	X	0.35	0
TTG	L	0.06	0	TCG	S	0.07	0	TAG	X	0.08	0	TGG	W	1.00	0
CTT	L	0.06	0	CCT	P	0.11	1	CAT	H	0.30	0	CGT	R	0.64	3
CTC	L	0.08	0	CCC	P	0.02	0	CAC	H	0.70	1	CGC	R	0.33	1
CTA	L	0.01	0	CCA	P	0.15	0	CAA	Q	0.19	0	CGA	R	0.01	0
CTG	L	0.77	8	CCG	P	0.72	3	CAG	Q	0.81	4	CGG	R	0.01	0
ATT	I	0.34	1	ACT	T	0.29	1	AAT	N	0.17	2	AGT	S	0.05	0
ATC	I	0.66	11	ACC	T	0.54	7	AAC	N	0.83	5	AGC	S	0.24	0
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	13	AGA	R	0.01	0
ATG	M	1.00	1	ACG	T	0.13	0	AAG	K	0.22	4	AGG	R	0.00	0
GTT	V	0.40	9	GCT	A	0.28	0	GAT	D	0.46	0	GGT	G	0.51	3
GTC	V	0.14	0	GCC	A	0.16	0	GAC	D	0.54	3	GGC	G	0.43	2
GTA	V	0.20	1	GCA	A	0.24	0	GAA	E	0.75	7	GGA	G	0.02	0
GTG	V	0.27	0	GCG	A	0.32	3	GAG	E	0.25	2	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	0
10% - 14%	1
15% - 19%	2
20% - 24%	7
25% - 29%	2
30% - 34%	10
35% - 39%	11
40% - 44%	2
45% - 49%	0
>= 50%	87

Total Codons Used = 122

Tm Range	# of Overlaps
<47^C	0
47-49^C	0
50-52^C	0
53-55^C	0
56-58^C	0
59-61^C	15
62-64^C	0
65-67^C	0
68-70^C	0
71-73^C	0
>=74^C	0

Tm Range = 1.9

Ovrlap Len Range	# of Oligos
<10	0
10-11	0
12-13	0

14-15	0
16-17	1
18-19	2
20-21	4
22-23	3
24-25	2
26-27	2
28-29	1
>=30	0

Lowest Overlap = 16

Length Range	# of Oligos
<20	0
20-24	0
25-29	1
30-34	0
35-39	0
40-44	0
45-49	13
50-54	0
55-59	0
>=60	0

Longest = 48

Sequence Patterns Screened (As Supplied By User)

Name	Seq	Pos	Notes
NdeI	CATATG	11	forward
XhoI	CTCGAG	380	forward

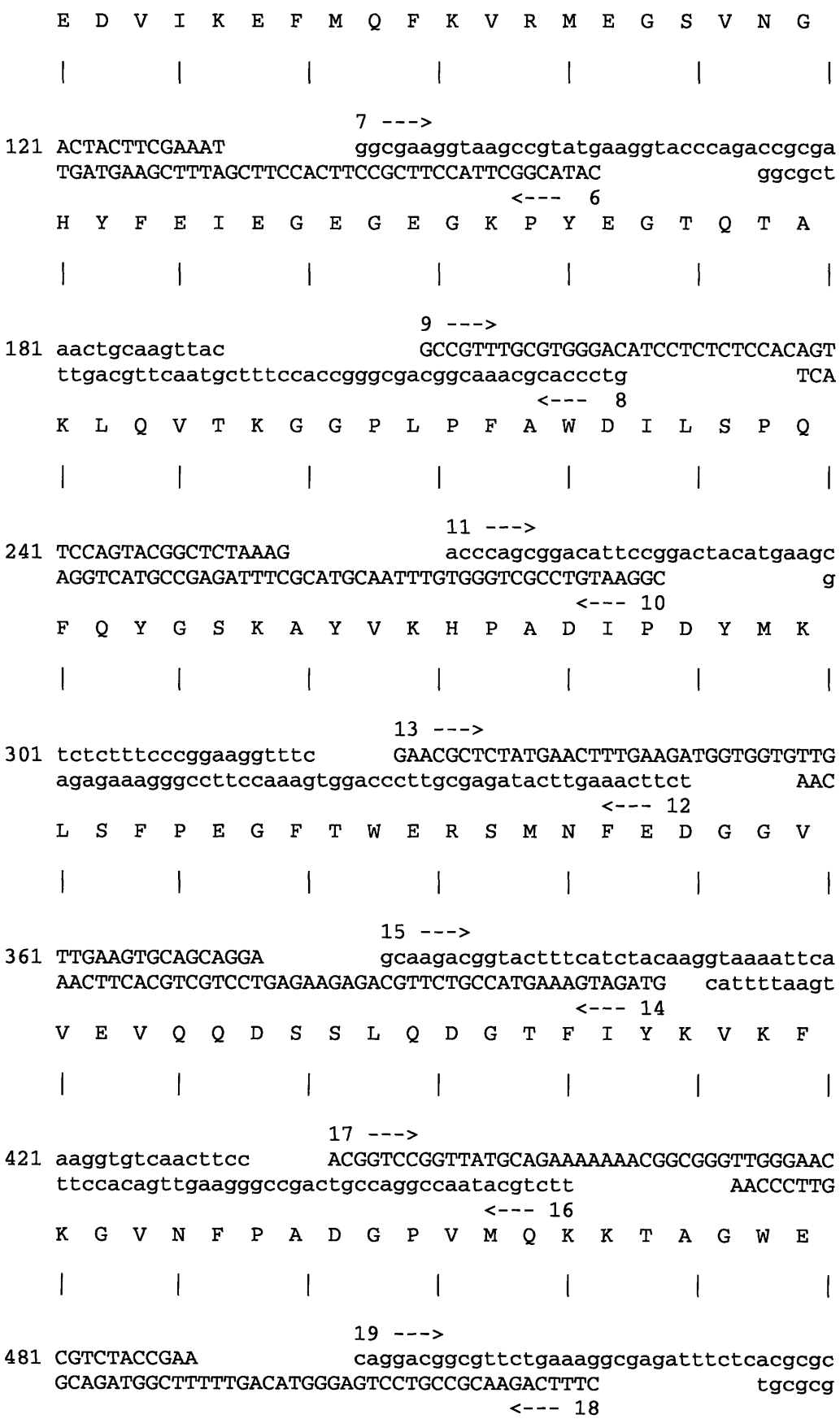
16 oligonucleotides need to be synthesized

1	CAGGTAATTCCATATGAACATCCGT	25
2	TGTACTCGTAGTATTTACCAGACTGACGGATGTTCATATGGAATTACC	48
3	CAGTCTGGTAAATACTACGAGTACAAAACTCTGGAGATCCTGGAAAAG	48
4	GGGATACGCAGCGCTTTGAAACCATTCTTTTCCAGGATCTCCAGAGTT	48
5	CAAAGCGCTGCGTATCCCGGTTTCTGGTACCGGCAAACAGGCGCTGCC	48
6	AGATGGTGTTGTTTTTGGTCGCGATCAGGTCCGGCAGCGCCTGTTTGC	48
7	CGACCAAAAACAACACCATCTACCCTATTGAAGTTAAATCTACCTCTA	48
8	TTACGAACGGTAACAACGTCTTTAGAGGTAGATTTAACTTCAATAGGG	48
9	AGACGTTGTTACCGTTCGTAATTTCCAGATCGAAAAACTGTTCAAATT	48
10	TCACAGAAGTTGAAGATTTCGCAGAATTTGAACAGTTTTTCGATCTGG	48
11	GCGAAATCTTCAACTTCTGTGAATGCCACCCGCTGGTAACCGTTTACT	48
12	ATAAACGATAACGATTTTGTATTTCTTGTAGTAAACGGTTACCAGCGG	48
13	CAAGAAATACAAAATCGTTATCGTTTATGAACTGTCTCAGGACGTTCG	48
14	GTACTTGAACTTGATTTTTTCTTTGGTGCGAACGTCCTGAGACAGTTC	48
15	CCAAAGAAAAAATCAAGTTCAAGTACGGCATCAACTCCTAACTCGAGC	48
16	GTCCGCTCGAGTTAGGAGTTGATGC	25

TABLE IIc

DNAWorks output file for DsRed.

```
-----------------------------------------------------------------
                    PARAMETERS FOR TRIAL 33
 ---------------------------------------------------------------
              Total Size Of Gene ......... 733 nt
              Protein Residues ........... 225
              Mutatable Residues ......... 216
              Fixed Nucleotides .......... 85 nt
              Oligo Size ................. 50 nt
              Annealing Temp ............. 60 +/- 1*C
              Oligo Concentration ........ 2.50E-8 M
              Sodium Concentration ....... 5.00E-2 M
              Mg2+ Concentration ......... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ........... 8 nt
              Mispriming Threshold ....... 8/18 (6 exact) nt

-------------------------------------------------------------------------------

 The DNA sequence #  33 is:
 ---------------------------------------------------------------
   1 CTAATACGACTCACTATAGGGAGACACAACGGTAAGGAGATATACATATGGATAACACGG
  61 AAGACGTTATCAAAGAATTCATGCAGTTCAAGGTTCGTATGGAAGGCTCTGTTAACGGCC
 121 ACTACTTCGAAATCGAAGGTGAAGGCGAAGGTAAGCCGTATGAAGGTACCCAGACCGCGA
 181 AACTGCAAGTTACGAAAGGTGGCCCGCTGCCGTTTGCGTGGGACATCCTCTCTCCACAGT
 241 TCCAGTACGGCTCTAAAGCGTACGTTAAACACCCAGCGGACATTCCGGACTACATGAAGC
 301 TCTCTTTCCCGGAAGGTTTCACCTGGGAACGCTCTATGAACTTTGAAGATGGTGGTGTTG
 361 TTGAAGTGCAGCAGGACTCTTCTCTGCAAGACGGTACTTTCATCTACAAGGTAAAATTCA
 421 AAGGTGTCAACTTCCCGGCTGACGGTCCGGTTATGCAGAAAAAAACGGCGGGTTGGGAAC
 481 CGTCTACCGAAAAACTGTACCCTCAGGACGGCGTTCTGAAAGGCGAGATTTCTCACGCGC
 541 TGAAACTGAAAGACGGCGGTCACTACACCTGCGACTTTAAAACCGTTTACAAAGCTAAAA
 601 AGCCGGTTCAGCTGCCGGGTAACCACTACGTTGACTCTAAACTGGACATCACCAACCACA
 661 ATGAAGACTATACCGTAGTTGAGCAGTATGAACACGCGGAAGCGCGTCACTCTGGTTCTC
 721 AACTCGAGCGGAC
 ---------------------------------------------------------------

 The oligonucleotide assembly is:
 ---------------------------------------------------------------
     1       10        20        30        40        50        60
     |        |         |         |         |         |         |

     1 --->                     3 --->
   1 CTAATACGACTCACTATAGGGAGAC  aacggtaaggagatatacatatggataacacgg
       TTATGCTGAGTGATATCCCTCTGTGTTGCCATTCCTCTATATGTATACCT   gtgcc
                                                 <---  2
                                                    M  D  N  T

     |        |         |         |         |         |         |

                           5 --->
  61 aagacgttatcaaagaa      CAGTTCAAGGTTCGTATGGAAGGCTCTGTTAACGGCC
     ttctgcaatagtttcttaagtacgtcaagttccaagcatacctttc       ATTGCCGG
                                          <---  4
```


E D V I K E F M Q F K V R M E G S V N G
7 --->
121 ACTACTTCGAAAT ggcgaaggtaagccgtatgaaggtacccagaccgcga
TGATGAAGCTTTAGCTTCCACTTCCGCTTCCATTCGGCATAC ggcgct
<--- 6
H Y F E I E G E G E G K P Y E G T Q T A
9 --->
181 aactgcaagttac GCCGTTTGCGTGGGACATCCTCTCTCCACAGT
ttgacgttcaatgctttccaccgggcgacggcaaacgcaccctg TCA
<--- 8
K L Q V T K G G P L P F A W D I L S P Q
11 --->
241 TCCAGTACGGCTCTAAAG acccagcggacattccggactacatgaagc
AGGTCATGCCGAGATTTCGCATGCAATTTGTGGGTCGCCTGTAAGGC g
<--- 10
F Q Y G S K A Y V K H P A D I P D Y M K
13 --->
301 tctctttcccggaaggtttc GAACGCTCTATGAACTTTGAAGATGGTGGTGTTG
agagaaagggccttccaaagtggaccttgcgagatacttgaaacttct AAC
<--- 12
L S F P E G F T W E R S M N F E D G G V
15 --->
361 TTGAAGTGCAGCAGGA gcaagacggtactttcatctacaaggtaaaattca
AACTTCACGTCGTCCTGAGAAGAGACGTTCTGCCATGAAAGTAGATG cattttaagt
<--- 14
V E V Q Q D S S L Q D G T F I Y K V K F
17 --->
421 aaggtgtcaacttcc ACGGTCCGGTTATGCAGAAAAAAACGGCGGGTTGGGAAC
ttccacagttgaagggccgactgccaggccaatacgtctt AACCCTTG
<--- 16
K G V N F P A D G P V M Q K K T A G W E
19 --->
481 CGTCTACCGAA caggacggcgttctgaaaggcgagatttctcacgcgc
GCAGATGGCTTTTTGACATGGGAGTCCTGCCGCAAGACTTTC tgcgcg
<--- 18

```
    P  S  T  E  K  L  Y  P  Q  D  G  V  L  K  G  E  I  S  H  A

    |        |        |        |        |        |        |

                          21 --->
541 tgaaactgaaaga          CTACACCTGCGACTTTAAAACCGTTTACAAAGCTAAAA
    actttgactttctgccgccagtgatgtggacgctgaaattttgg          CGATTTT
                                          <--- 20
    L  K  L  K  D  G  G  H  Y  T  C  D  F  K  T  V  Y  K  A  K

    |        |        |        |        |        |        |

                          23 --->
601 AGCCGGTTCAGC           accactacgttgactctaaactggacatcaccaaccaca
    TCGGCCAAGTCGACGGCCCATTGGTGATGCAACTGAGATTTGA       gtggttggtgt
                                        <--- 22
    K  P  V  Q  L  P  G  N  H  Y  V  D  S  K  L  D  I  T  N  H

    |        |        |        |        |        |        |

                          25 --->
661 atgaagactat             AGCAGTATGAACACGCGGAAGCGCGTCACTCTGGTTCTC
    tacttctgatatggcatcaactcgtcatacttgtgcgcc               CCAAGAG
                                     <--- 24
    N  E  D  Y  T  V  V  E  Q  Y  E  H  A  E  A  R  H  S  G  S

    |        |        |        |        |        |        |


721 AACTCGAGCGG
    TTGAGCTCGCCTG
          <--- 26
    Q
```

```
The total codon usage score ...........     0.000
The total length score ................     0.000
The total melting temperature score ...     0.000
The total repeat score ................     0.000
The total pattern score ...............     0.382
The total mispriming score ............     0.000
The total AT content score ............     0.000
The total GC content score ............     0.000
               The OVERALL score ......     0.382
```

DETAILED CODON FREQUENCY REPORT

[Codon, AA, Frequency, # of times used in coding sequence]

```
TTT F 0.29   3  TCT S 0.32  12  TAT Y 0.35   3  TGT C 0.39   0
TTC F 0.71   9  TCC S 0.27   0  TAC Y 0.65   9  TGC C 0.61   1
TTA L 0.03   0  TCA S 0.05   0  TAA X 0.63   0  TGA X 0.35   0
TTG L 0.06   0  TCG S 0.07   0  TAG X 0.08   0  TGG W 1.00   3

CTT L 0.06   0  CCT P 0.11   1  CAT H 0.30   0  CGT R 0.64   2
```

CTC	L	0.08	2	CCC	P	0.02	0	CAC	H	0.70	8	CGC	R	0.33	1
CTA	L	0.01	0	CCA	P	0.15	2	CAA	Q	0.19	3	CGA	R	0.01	0
CTG	L	0.77	9	CCG	P	0.72	10	CAG	Q	0.81	10	CGG	R	0.01	0
ATT	I	0.34	2	ACT	T	0.29	1	AAT	N	0.17	1	AGT	S	0.05	0
ATC	I	0.66	5	ACC	T	0.54	8	AAC	N	0.83	6	AGC	S	0.24	0
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	17	AGA	R	0.01	0
ATG	M	1.00	6	ACG	T	0.13	3	AAG	K	0.22	5	AGG	R	0.00	0
GTT	V	0.40	13	GCT	A	0.28	2	GAT	D	0.46	2	GGT	G	0.51	14
GTC	V	0.14	1	GCC	A	0.16	0	GAC	D	0.54	13	GGC	G	0.43	8
GTA	V	0.20	2	GCA	A	0.24	0	GAA	E	0.75	17	GGA	G	0.02	0
GTG	V	0.27	1	GCG	A	0.32	8	GAG	E	0.25	2	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	2
10% - 14%	5
15% - 19%	6
20% - 24%	9
25% - 29%	7
30% - 34%	23
35% - 39%	16
40% - 44%	8
45% - 49%	2
>= 50%	147

Total Codons Used = 225

Tm Range	# of Overlaps
<57	0
57	0
58	0
59	16
60	9
61	0
62	0
63	0
64	0
>=65	0

Tm Range = 1.8

Ovrlap Len Range	# of Oligos
<17	1
17	1
18	2
19	8
20	0
21	3
22	6
23	2

24	0
25	2
26	0
>=27	0

Lowest Overlap = 16

Length Range	# of Oligos
<39	2
39-40	0
41-42	0
43-44	0
45-46	0
47-48	0
49-50	24
51-52	0
53-54	0
55-56	0
57-58	0
>=59	0

Longest = 50

Sequence Patterns Screened (As Supplied By User)

Name	Seq	Pos	Notes
NdeI	CATATG	45	forward
XhoI	CTCGAG	723	forward

26 oligonucleotides need to be synthesized

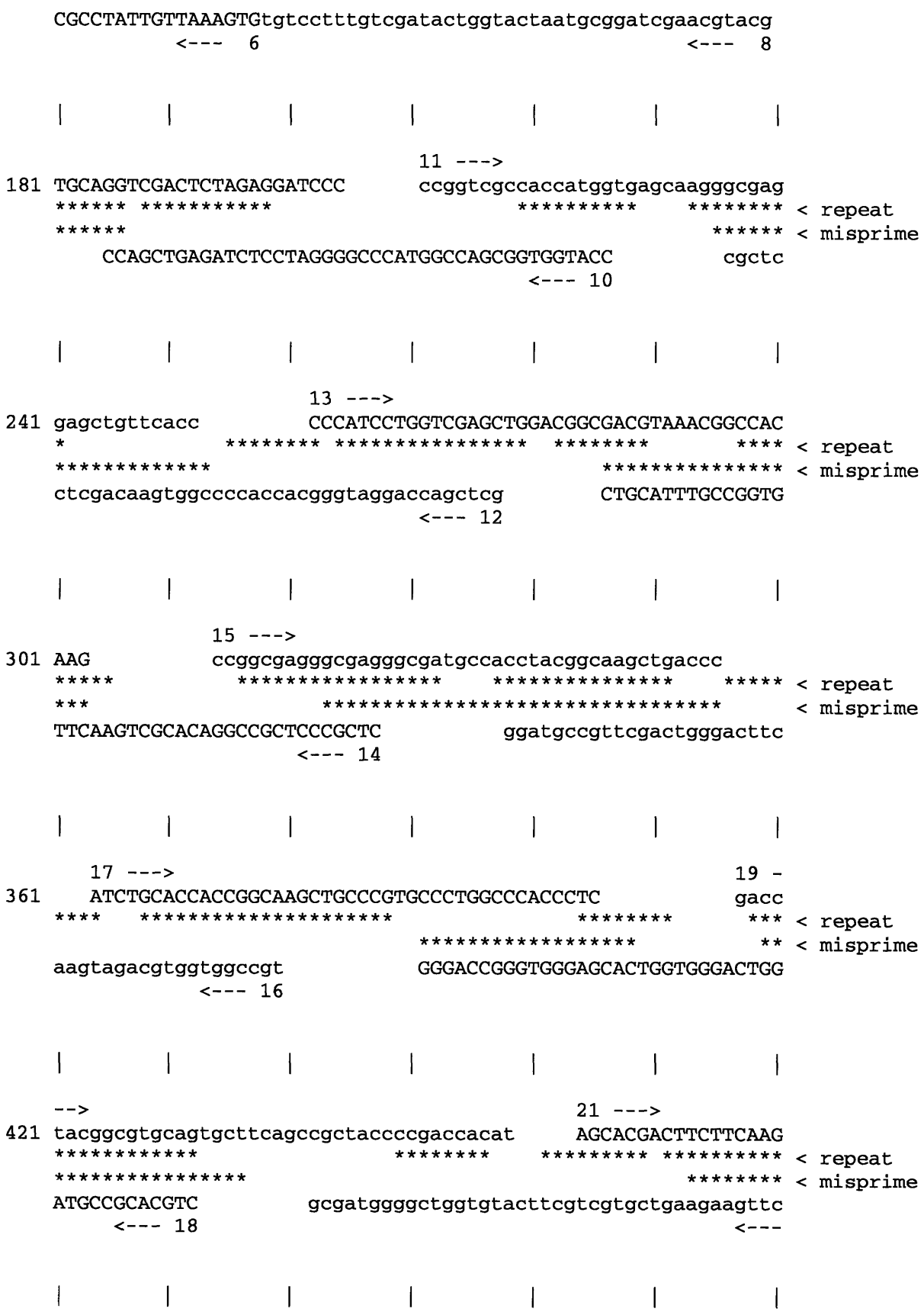
1	CTAATACGACTCACTATAGGGAGAC	25
2	TCCATATGTATATCTCCTTACCGTTGTGTCTCCCTATAGTGAGTCGTATT	50
3	AACGGTAAGGAGATATACATATGGATAACACGGAAGACGTTATCAAAGAA	50
4	CTTCCATACGAACCTTGAACTGCATGAATTCTTTGATAACGTCTTCCGTG	50
5	CAGTTCAAGGTTCGTATGGAAGGCTCTGTTAACGGCCACTACTTCGAAAT	50
6	CATACGGCTTACCTTCGCCTTCACCTTCGATTTCGAAGTAGTGGCCGTTA	50
7	GGCGAAGGTAAGCCGTATGAAGGTACCCAGACCGCGAAACTGCAAGTTAC	50
8	GTCCCACGCAAACGGCAGCGGGCCACCTTTCGTAACTTGCAGTTTCGCGG	50
9	GCCGTTTGCGTGGGACATCCTCTCTCCACAGTTCCAGTACGGCTCTAAAG	50
10	CGGAATGTCCGCTGGGTGTTTAACGTACGCTTTAGAGCCGTACTGGAACT	50
11	ACCCAGCGGACATTCCGGACTACATGAAGCTCTCTTTCCCGGAAGGTTTC	50
12	TCTTCAAAGTTCATAGAGCGTTCCCAGGTGAAACCTTCCGGGAAAGAGAG	50
13	GAACGCTCTATGAACTTTGAAGATGGTGGTGTTGTTGAAGTGCAGCAGGA	50
14	GTAGATGAAAGTACCGTCTTGCAGAGAAGAGTCCTGCTGCACTTCAACAA	50
15	GCAAGACGGTACTTTCATCTACAAGGTAAAATTCAAAGGTGTCAACTTCC	50
16	TTCTGCATAACCGGACCGTCAGCCGGGAAGTTGACACCTTTGAATTTTAC	50
17	ACGGTCCGGTTATGCAGAAAAAAACGGCGGGTTGGGAACCGTCTACCGAA	50
18	CTTTCAGAACGCCGTCCTGAGGGTACAGTTTTTCGGTAGACGGTTCCCAA	50
19	CAGGACGGCGTTCTGAAAGGCGAGATTTCTCACGCGCTGAAACTGAAAGA	50
20	GGTTTTAAAGTCGCAGGTGTAGTGACCGCCGTCTTTCAGTTTCAGCGCGT	50
21	CTACACCTGCGACTTTAAAACCGTTTACAAAGCTAAAAAGCCGGTTCAGC	50

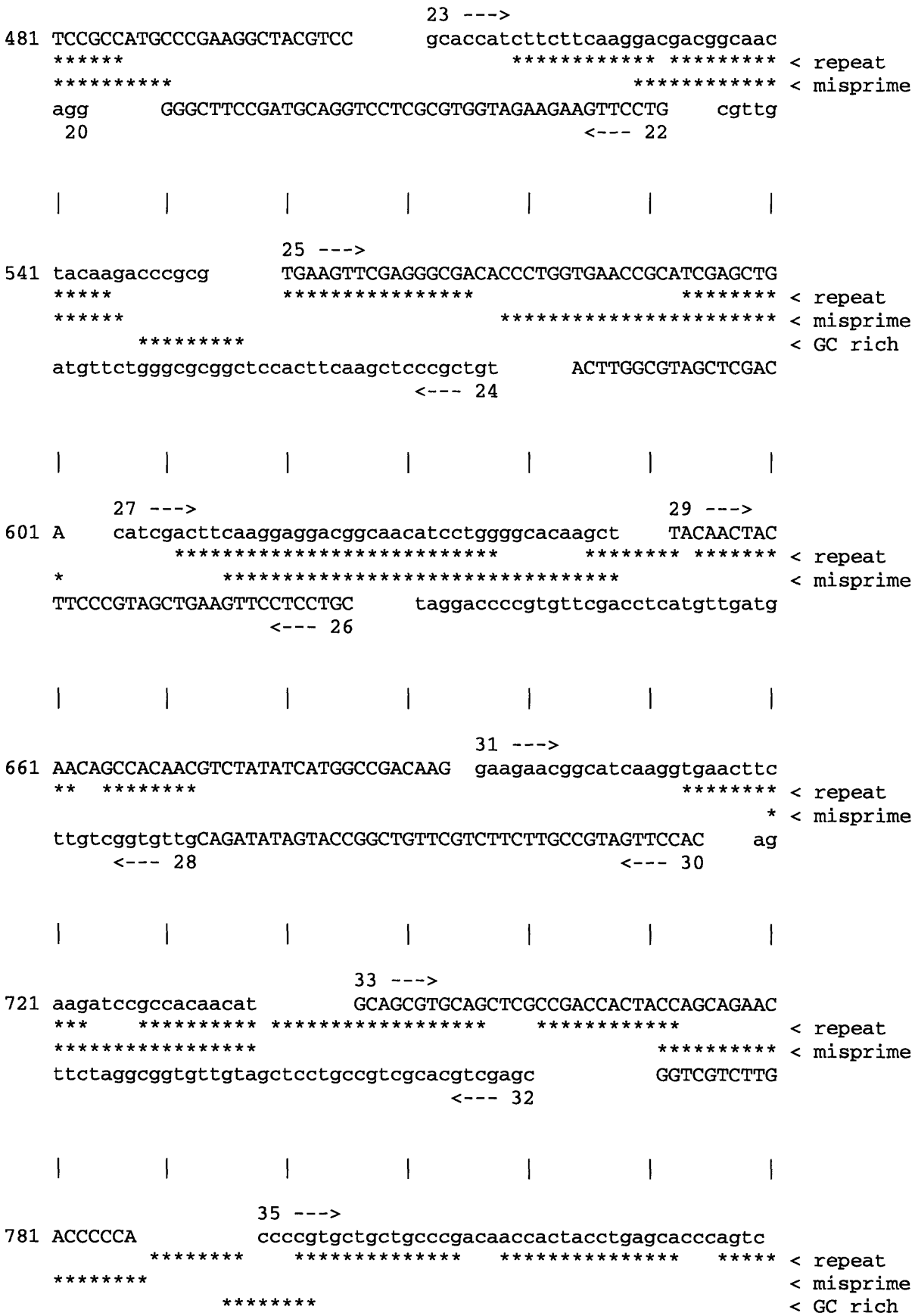
22 AGTTTAGAGTCAACGTAGTGGTTACCCGGCAGCTGAACCGGCTTTTTAGC 50
23 ACCACTACGTTGACTCTAAACTGGACATCACCAACCACAATGAAGACTAT 50
24 CCGCGTGTTCATACTGCTCAACTACGGTATAGTCTTCATTGTGGTTGGTG 50
25 AGCAGTATGAACACGCGGAAGCGCGTCACTCTGGTTCTCAACTCGAGCGG 50
26 GTCCGCTCGAGTTGAGAACC 20

TABLE IId

DNAWorks output file for an eGFP construct with a promoter and regulatory elements.

```
 The DNA sequence #   5 is:
 ------------------------------------------------------------
   1 GGGGACCACTTTGTACAAGAAAGCTGGGTCGCAACGCAATTAATGTGAGTTAGCTCACTC
  61 ATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGA
 121 GCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCTAGCTTGCATGCC
 181 TGCAGGTCGACTCTAGAGGATCCCCGGGTACCGGTCGCCACCATGGTGAGCAAGGGCGAG
 241 GAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC
 301 AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 361 TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACC
 421 TACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
 481 TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC
 541 TACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
 601 AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC
 661 AACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTC
 721 AAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
 781 ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCC
 841 GCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
 901 GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGA
 961 ATTCAGCCTGCTTTTTTGTACAAACTTGTGGGG
 ------------------------------------------------------------

 The oligonucleotide assembly is:
 ------------------------------------------------------------
     1       10        20        30        40        50        60
     |        |         |         |         |         |         |

     1 --->                       3 --->
   1 GGGGACCACTTTGTACAAGAAAGCTGGGT gcaacgcaattaatgtgagttagctcactc
             **********                                          < repeat
      ******************                                         < misprime
             GAAACATGTTCTTTCGACCCAGCGTTGCGTTAATTACACTCA   gagtgag
                                                <---  2


     |        |         |         |         |         |         |

                  5 --->                                    7 ---
  61 attaggcacccc GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATtgtga
                                               ********           < repeat
                                               ******************  < misprime
     taatccgtggggtccgaaatgtgaaatacgaaggcCGAGCATACAACACACCTTAACACT
                                  <---  4


     |        |         |         |         |         |         |

     >                                        9 --->
 121 gcggataacaatttcacacaggaaacagctatgacca     ACGCCTAGCTTGCATGCC
                                                          ** < repeat
                                                ********** < misprime
```


CGCCTATTGTTAAAGTGtgtcctttgtcgatactggtactaatgcggatcgaacgtacg
<--- 6
<--- 8
11 --->
181 TGCAGGTCGACTCTAGAGGATCCC ccggtcgccaccatggtgagcaagggcgag
< repeat
< misprime
CCAGCTGAGATCTCCTAGGGGCCCATGGCCAGCGGTGGTACC cgctc
<--- 10
13 --->
241 gagctgttcacc CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC
< repeat
< misprime
ctcgacaagtggccccaccacgggtaggaccagctcg CTGCATTTGCCGGTG
<--- 12
15 --->
301 AAG ccggcgagggcgagggcgatgccacctacggcaagctgaccc
< repeat
< misprime
TTCAAGTCGCACAGGCCGCTCCCGCTC ggatgccgttcgactgggacttc
<--- 14
17 --->
19 -
361 ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC gacc
< repeat
< misprime
aagtagacgtggtggccgt GGGACCGGGTGGGAGCACTGGTGGGACTGG
<--- 16
-->
21 --->
421 tacggcgtgcagtgcttcagccgctaccccgaccacat AGCACGACTTCTTCAAG
< repeat
< misprime
ATGCCGCACGTC gcgatggggctggtgtacttcgtcgtgctgaagaagttc
<--- 18
<---


23 --->
481 TCCGCCATGCCCGAAGGCTACGTCC gcaccatcttcttcaaggacgacggcaac
****** ********** ********* < repeat
********** ********** < misprime
agg GGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAAGTTCCTG cgttg
20 <--- 22
25 --->
541 tacaagacccgcg TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
***** ********** ******** < repeat
****** ********** < misprime
********* < GC rich
atgttctgggcgcggctccacttcaagctcccgctgt ACTTGGCGTAGCTCGAC
<--- 24
27 ---> 29 --->
601 A catcgacttcaaggaggacggcaacatcctggggcacaagct TACAACTAC
********** ******** ******* < repeat
* ********** < misprime
TTCCCGTAGCTGAAGTTCCTCCTGC taggaccccgtgttcgacctcatgttgatg
<--- 26
31 --->
661 AACAGCCACAACGTCTATATCATGGCCGACAAG gaagaacggcatcaaggtgaacttc
** ******** ******** < repeat
* < misprime
ttgtcggtgttgCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTTCCAC ag
<--- 28 <--- 30
33 --->
721 aagatccgccacaacat GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
*** ********** ********** ********** < repeat
********** ********** < misprime
ttctaggcggtgttgtagctcctgccgtcgcacgtcgagc GGTCGTCTTG
<--- 32
35 --->
781 ACCCCCA ccccgtgctgctgcccgacaaccactacctgagcacccagtc
******** ********** ********** ***** < repeat
******** < misprime
******** < GC rich

```
    TGGGGGTAGCCGCTGCCGGGGCACGACGACGG       tgatggactcgtgggtcagg
                               <--- 34


    |        |         |         |         |         |         |

        37 --->                                                39
841     TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGC             acc
    **********                                     **************** < repeat
                                      ******************            < misprime
                                                                 ** < GC rich
    cgggactcgtttctggggttgc      CGCTAGTGTACCAGGACGACCTCAAGCACTGG
                      <--- 36


    |        |         |         |         |         |         |

    --->                                         41 --->
901 gccgccgggatcactctcggcatggacgagctgtacaag      CGGCCGCGACTCTAGA
                                    ********    ***************** < repeat
                        ********************                      < misprime
    *********                                   *********         < GC rich
    CGGCGGCCCT          cgtacctgctcgacatgttcatttcgccggcgctgagatct
       <--- 38                                              <--- 4


    |        |         |         |         |         |         |


961 ATTCAGCCTGCTTTTTTGTACAAACT
                  *****************   < repeat
    t   TCGGACGAAAAAACATGTTTGAACACCCC
    0                           <--- 42
------------------------------------------------------------------

The total codon usage score ...........     0.000
The total length score ................     0.000
The total hairpin score ...............     0.000
The total melting temperature score ...     0.038
The total repeat score ................    14.340
The total pattern score ...............     0.000
The total mispriming score ............     9.426
The total AT content score ............     0.000
The total GC content score ............     1.450
               The OVERALL score ......    25.255


             Tm Range   # of Overlaps
            ------------------------------------
               <47^C          0
              47-49^C         0
              50-52^C         0
              53-55^C         0
              56-58^C        23
              59-61^C        18
```

```
  62-64^C          0
  65-67^C          0
  68-70^C          0
  71-73^C          0
   >=74^C          0
-------------------------------------
 Tm Range =  5.1

 Ovrlap Len Range  # of Oligos
-------------------------------------
      <10           0
     10-11          0
     12-13          1
     14-15          4
     16-17         11
     18-19         11
     20-21         11
     22-23          3
     24-25          0
     26-27          0
     28-29          0
      >=30          0
-------------------------------------
 Lowest Overlap = 13

 Length Range   # of Oligos
-------------------------------------
      <20            0
     20-24           0
     25-29           1
     30-34           0
     35-39           0
     40-44          39
     45-49           0
     50-54           0
     55-59           0
      >=60           0
-------------------------------------
 Longest =  42

 There are  13 potential misprimings with <=  8 non-identical nts:
-----------------------------------------------------------------
 Oligo Type 5'-start     Sequence     3'-start  Identical
-----------------------------------------------------------------
   39   DS    922  ATGGACGAGCTGTACAAG   939      10/18
                   |  |    ||||||||
               2  ggGacCacttTGTACAAG    19

   31   IS    720  CAAGATCCGCCACAACAT   737      13/18
                  ||  || |  ||||||||
              119  CAcaATtCcaCACAACAT   102

   27   IS    630  CATCCTGGGGCACAAGCT   647      11/18
                  | | | ||    ||||||
              186  CcTgCaGGcatgCAAGCT   169
```

```
11   IA    235  GGCGAGGAGCTGTTCACC   252      12/18
                ||  |  || |||||||
           599  aGCtcGatGCgGTTCACC   582

12   DA    253  CGGTGAACAGCTCCTCGC   236      10/18
                ||||       ||||||
           340  aGGTGgcatcgcCCTCGC   323

13   DS    286  GACGTAAACGGCCACAAG   303      11/18
                || |    || ||||||
           628  aACaTcctgGGgCACAAG   645

25   DS    584  TGAACCGCATCGAGCTGA   601      10/18
                 | |  |  | ||||||
           335  ccAcCtaCggCaAGCTGA   352

16   DA    355  GGGTCAGCTTGCCGTAGG   338      12/18
                 |  | ||  ||||||||
           436  aGcaCtGCacGCCGTAGG   419

18   DA    408  CACGAGGGTGGGCCAGGG   391      10/18
                |     |||   ||||||
           595  CgatgcGGTtcaCCAGGG   578

40   IA    937  TGTACAGCTCGTCCATGC   920      12/18
                | | ||  ||  ||||||
           473  TcTtCAagTCcgCCATGC   490

39   DS    922  ATGGACGAGCTGTACAAG   939      10/18
                   ||||     ||||||
           529  gacGACGgcaacTACAAG   546

31   DS    720  CAAGATCCGCCACAACAT   737      11/18
                ||||     |  ||||||
           615  CAAGgaggaCggCAACAT   632

34   IS    788  ATGGGGGTGTTCTGCTGG   771      11/18
                ||      || |||||||
           872  ATcacatgGTcCTGCTGG   889
```

There are 41 repeats greater than 8 nt:

```
DR Pos1 =   10   Pos2 =  975   Size =  9   Seq1 = TTTGTACAA
                                           Seq2 = TTTGTACAA
IR Pos1 =   10   Pos2 =  976   Size =  9   Seq1 = TTTGTACAA
                                           Seq2 = TTGTACAAA
PR Pos1 =   11   Pos2 =   11   Size =  8   Seq1 = TTGTACAA
                                           Seq2 = TTGTACAA
DR Pos1 =   12   Pos2 =  932   Size =  8   Seq1 = TGTACAAG
                                           Seq2 = TGTACAAG
IR Pos1 =  102   Pos2 =  730   Size =  8   Seq1 = ATGTTGTG
                                           Seq2 = CACAACAT
PR Pos1 =  179   Pos2 =  179   Size =  8   Seq1 = CCTGCAGG
                                           Seq2 = CCTGCAGG
```

```
DR Pos1 =  188   Pos2 =  951   Size = 10   Seq1 = CGACTCTAGA
                                           Seq2 = CGACTCTAGA
PR Pos1 =  191   Pos2 =  191   Size =  8   Seq1 = CTCTAGAG
                                           Seq2 = CTCTAGAG
PR Pos1 =  219   Pos2 =  219   Size = 10   Seq1 = CACCATGGTG
                                           Seq2 = CACCATGGTG
DR Pos1 =  233   Pos2 =  320   Size =  9   Seq1 = AGGGCGAGG
                                           Seq2 = AGGGCGAGG
IR Pos1 =  255   Pos2 =  368   Size =  8   Seq1 = GGTGGTGC
                                           Seq2 = GCACCACC
DR Pos1 =  264   Pos2 =  630   Size =  8   Seq1 = CATCCTGG
                                           Seq2 = CATCCTGG
DR Pos1 =  272   Pos2 =  593   Size =  8   Seq1 = TCGAGCTG
                                           Seq2 = TCGAGCTG
DR Pos1 =  282   Pos2 =  789   Size =  8   Seq1 = CGGCGACG
                                           Seq2 = CGGCGACG
IR Pos1 =  297   Pos2 =  983   Size =  9   Seq1 = CCACAAGTT
                                           Seq2 = AACTTGTGG
DR Pos1 =  316   Pos2 =  322   Size = 11   Seq1 = GGCGAGGGCGA
                                           Seq2 = GGCGAGGGCGA
DR Pos1 =  318   Pos2 =  567   Size =  9   Seq1 = CGAGGGCGA
                                           Seq2 = CGAGGGCGA
DR Pos1 =  324   Pos2 =  567   Size =  9   Seq1 = CGAGGGCGA
                                           Seq2 = CGAGGGCGA
DR Pos1 =  337   Pos2 =  418   Size =  9   Seq1 = ACCTACGGC
                                           Seq2 = ACCTACGGC
DR Pos1 =  342   Pos2 =  375   Size = 10   Seq1 = CGGCAAGCTG
                                           Seq2 = CGGCAAGCTG
DR Pos1 =  356   Pos2 =  560   Size =  8   Seq1 = TGAAGTTC
                                           Seq2 = TGAAGTTC
IR Pos1 =  356   Pos2 =  713   Size =  9   Seq1 = TGAAGTTCA
                                           Seq2 = TGAACTTCA
DR Pos1 =  381   Pos2 =  807   Size =  8   Seq1 = GCTGCCCG
                                           Seq2 = GCTGCCCG
DR Pos1 =  404   Pos2 =  893   Size =  8   Seq1 = TCGTGACC
                                           Seq2 = TCGTGACC
DR Pos1 =  425   Pos2 =  749   Size =  8   Seq1 = GCGTGCAG
                                           Seq2 = GCGTGCAG
DR Pos1 =  449   Pos2 =  761   Size =  8   Seq1 = CCGACCAC
                                           Seq2 = CCGACCAC
IR Pos1 =  461   Pos2 =  801   Size =  9   Seq1 = AGCAGCACG
                                           Seq2 = CGTGCTGCT
DR Pos1 =  471   Pos2 =  519   Size = 10   Seq1 = CTTCTTCAAG
                                           Seq2 = CTTCTTCAAG
DR Pos1 =  479   Pos2 =  836   Size =  8   Seq1 = AGTCCGCC
                                           Seq2 = AGTCCGCC
DR Pos1 =  522   Pos2 =  612   Size =  9   Seq1 = CTTCAAGGA
                                           Seq2 = CTTCAAGGA
DR Pos1 =  532   Pos2 =  622   Size =  9   Seq1 = GACGGCAAC
                                           Seq2 = GACGGCAAC
DR Pos1 =  537   Pos2 =  654   Size =  9   Seq1 = CAACTACAA
                                           Seq2 = CAACTACAA
IR Pos1 =  560   Pos2 =  714   Size =  8   Seq1 = TGAAGTTC
                                           Seq2 = GAACTTCA
DR Pos1 =  611   Pos2 =  716   Size =  8   Seq1 = ACTTCAAG
                                           Seq2 = ACTTCAAG
```

```
DR Pos1 =  619   Pos2 =  739   Size = 10   Seq1 = GAGGACGGCA
                                           Seq2 = GAGGACGGCA
DR Pos1 =  645   Pos2 =  885   Size =  8   Seq1 = GCTGGAGT
                                           Seq2 = GCTGGAGT
DR Pos1 =  665   Pos2 =  728   Size =  8   Seq1 = GCCACAAC
                                           Seq2 = GCCACAAC
DR Pos1 =  764   Pos2 =  818   Size =  9   Seq1 = ACCACTACC
                                           Seq2 = ACCACTACC
DR Pos1 =  825   Pos2 =  843   Size =  8   Seq1 = CCTGAGCA
                                           Seq2 = CCTGAGCA
PR Pos1 =  944   Pos2 =  944   Size =  8   Seq1 = GCGGCCGC
                                           Seq2 = GCGGCCGC
PR Pos1 =  975   Pos2 =  975   Size = 10   Seq1 = TTTGTACAAA
                                           Seq2 = TTTGTACAAA
-------------------------------------------------------------------


         Sequence Patterns Screened (As Supplied By User)
-------------------------------------------------------------------
                       None found
-------------------------------------------------------------------

       42 oligonucleotides need to be synthesized
-------------------------------------------------------------------
 1 GGGGACCACTTTGTACAAGAAAGCTGGGT  29
 2 ACTCACATTAATTGCGTTGCGACCCAGCTTTCTTGTACAAAG  42
 3 GCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCC  42
 4 CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAG  42
 5 GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAAT  42
 6 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC  42
 7 TGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCA  42
 8 GCATGCAAGCTAGGCGTAATCATGGTCATAGCTGTTTCCTGT  42
 9 ACGCCTAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCC  42
10 CCATGGTGGCGACCGGTACCCGGGGATCCTCTAGAGTCGACC  42
11 CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACC  42
12 GCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGC  42
13 CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG  42
14 CTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTC  42
15 CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC  42
16 TGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGG  42
17 ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC  42
18 CTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGG  42
19 GACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT  42
20 GGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCG  42
21 AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC  42
22 GTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGG  42
23 GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG  42
24 TGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGC  42
25 TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA  42
26 CGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCA  42
27 CATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCT  42
28 GTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGAT  42
29 TACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAG  42
30 CACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATAGAC  42
31 GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACAT  42
32 CGAGCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGA  42
```

33 GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA 42
34 GGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGG 42
35 CCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTC 42
36 CGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGT 42
37 TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGC 42
38 TCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGC 42
39 ACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG 42
40 TTCTAGAGTCGCGGCCGCTTTACTTGTACAGCTCGTCCATGC 42
41 CGGCCGCGACTCTAGAATTCAGCCTGCTTTTTTGTACAAACT 42
42 CCCCACAAGTTTGTACAAAAAAGCAGGCT 29

A.6 Chapter 6 appendix

Helper oligos:

MlyI-help15 AAAAAGAGTCCACCT

MlyI-help16 AAAAAAGAGTCCACCT

MlyI-help18 AAAAAAAAGAGTCCACCT

MlyI-help20 AAAAAAAAAAGAGTCCACCT

MlyI-help22 AAAAAAAAAAAAGAGTCCACCT

A.7 Chapter 7 Appendix

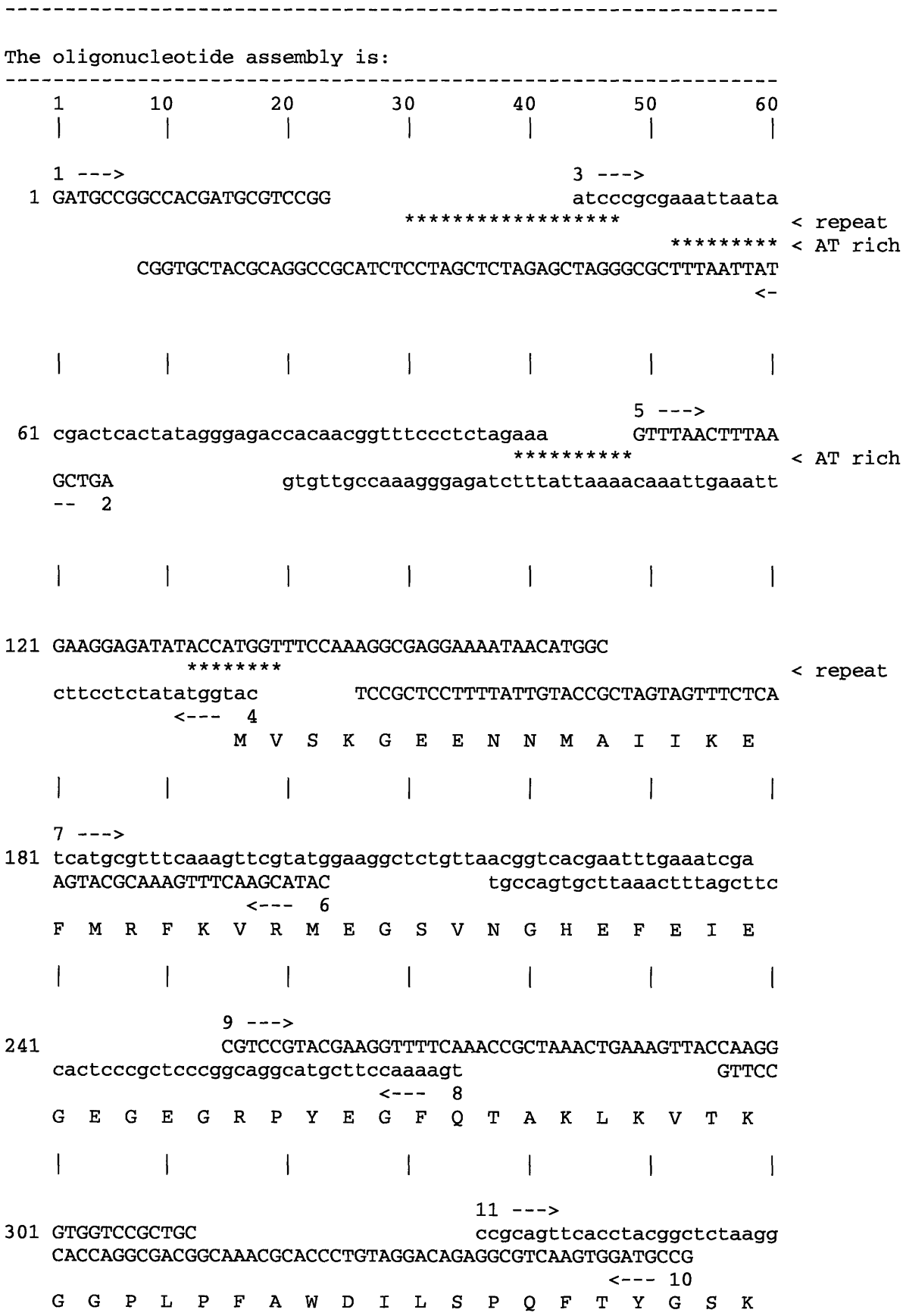
A.7.1 custom primers for T7 regulatory element addition

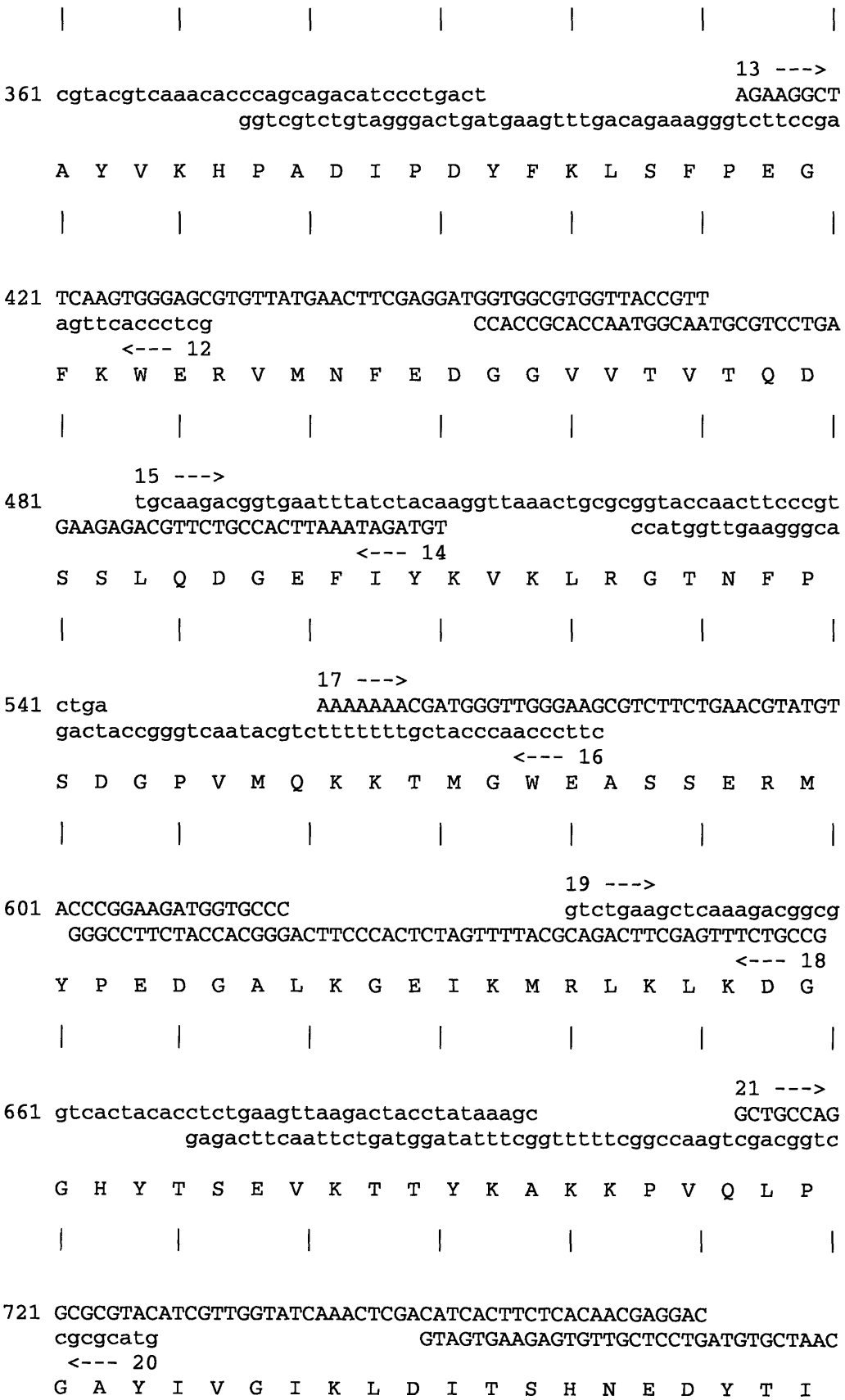
-custom primers for T7 addition to...pEGFP/eGFP?

A.7.2. DNAworks output files for fluorescent protein gene expression constructs

mOrange

```
 ---------------------------------------------------------------
                    PARAMETERS FOR TRIAL 42
 ---------------------------------------------------------------
              Total Size Of Gene ......... 1079 nt
              Protein Residues ........... 237
              Mutatable Residues ......... 224
              Fixed Nucleotides .......... 407 nt
              Oligo Size ................. 58 nt
              Annealing Temp ............. 61 +/- 1*C
              Oligo Concentration ........ 2.00E-8 M
              Sodium Concentration ....... 5.00E-2 M
              Mg2+ Concentration ......... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ........... 8 nt
              Mispriming Threshold ....... 8/18 (6 exact) nt

--------------------------------------------------------------------------------

 The DNA sequence #  42 is:
 ---------------------------------------------------------------
   1 GATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATA
  61 CGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAA
 121 GAAGGAGATATACCATGGTTTCCAAAGGCGAGGAAAATAACATGGCGATCATCAAAGAGT
 181 TCATGCGTTTCAAAGTTCGTATGGAAGGCTCTGTTAACGGTCACGAATTTGAAATCGAAG
 241 GTGAGGGCGAGGGCCGTCCGTACGAAGGTTTTCAAACCGCTAAACTGAAAGTTACCAAGG
 301 GTGGTCCGCTGCCGTTTGCGTGGGACATCCTGTCTCCGCAGTTCACCTACGGCTCTAAGG
 361 CGTACGTCAAACACCCAGCAGACATCCCTGACTACTTCAAACTGTCTTTCCCAGAAGGCT
 421 TCAAGTGGGAGCGTGTTATGAACTTCGAGGATGGTGGCGTGGTTACCGTTACGCAGGACT
 481 CTTCTCTGCAAGACGGTGAATTTATCTACAAGGTTAAACTGCGCGGTACCAACTTCCCGT
 541 CTGATGGCCCAGTTATGCAGAAAAAAACGATGGGTTGGGAAGCGTCTTCTGAACGTATGT
 601 ACCCGGAAGATGGTGCCCTGAAGGGTGAGATCAAAATGCGTCTGAAGCTCAAAGACGGCG
 661 GTCACTACACCTCTGAAGTTAAGACTACCTATAAAGCCAAAAAGCCGGTTCAGCTGCCAG
 721 GCGCGTACATCGTTGGTATCAAACTCGACATCACTTCTCACAACGAGGACTACACGATTG
 781 TTGAACAGTACGAACGTGCCGAAGGTCGCCACTCTACCGGTGGTATGGATGAACTGTACA
 841 AATGAGGGGGGGGTTCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACA
 901 CTGGCGGCCGTTACTAGTGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
 961 TGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAG
1021 GGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCC
```


The oligonucleotide assembly is:
1 10 20 30 40 50 60
1 ---> 3 --->
1 GATGCCGGCCACGATGCGTCCGG atcccgcgaaattaata
< repeat
< AT rich
CGGTGCTACGCAGGCCGCATCTCCTAGCTCTAGAGCTAGGGCGCTTTAATTAT
<-
5 --->
61 cgactcactatagggagaccacaacggtttccctctagaaa GTTTAACTTTAA
< AT rich
GCTGA gtgttgccaaagggagatctttattaaaacaaattgaaatt
-- 2
121 GAAGGAGATATACCATGGTTTCCAAAGGCGAGGAAAATAACATGGC
< repeat
cttcctctatatggtac TCCGCTCCTTTTATTGTACCGCTAGTAGTTTCTCA
<--- 4
M V S K G E E N N M A I I K E
7 --->
181 tcatgcgtttcaaagttcgtatggaaggctctgttaacggtcacgaatttgaaatcga
AGTACGCAAAGTTTCAAGCATAC tgccagtgcttaaactttagcttc
<--- 6
F M R F K V R M E G S V N G H E F E I E
9 --->
241 CGTCCGTACGAAGGTTTTCAAACCGCTAAACTGAAAGTTACCAAGG
cactcccgctcccggcaggcatgcttccaaaagt GTTCC
<--- 8
G E G E G R P Y E G F Q T A K L K V T K
11 --->
301 GTGGTCCGCTGC ccgcagttcacctacggctctaagg
CACCAGGCGACGGCAAACGCACCCTGTAGGACAGAGGCGTCAAGTGGATGCCG
<--- 10
G G P L P F A W D I L S P Q F T Y G S K


13 --->
361 cgtacgtcaaacacccagcagacatccctgact AGAAGGCT
ggtcgtctgtagggactgatgaagtttgacagaaagggtcttccga
A Y V K H P A D I P D Y F K L S F P E G
421 TCAAGTGGGAGCGTGTTATGAACTTCGAGGATGGTGGCGTGGTTACCGTT
agttcaccctcg CCACCGCACCAATGGCAATGCGTCCTGA
<--- 12
F K W E R V M N F E D G G V V T V T Q D
15 --->
481 tgcaagacggtgaatttatctacaaggttaaactgcgcggtaccaacttcccgt
GAAGAGACGTTCTGCCACTTAAATAGATGT ccatggttgaagggca
<--- 14
S S L Q D G E F I Y K V K L R G T N F P
17 --->
541 ctga AAAAAAACGATGGGTTGGGAAGCGTCTTCTGAACGTATGT
gactaccgggtcaatacgtctttttttgctacccaacccttc
<--- 16
S D G P V M Q K K T M G W E A S S E R M
19 --->
601 ACCCGGAAGATGGTGCCC gtctgaagctcaaagacggcg
GGGCCTTCTACCACGGGACTTCCCACTCTAGTTTTACGCAGACTTCGAGTTTCTGCCG
<--- 18
Y P E D G A L K G E I K M R L K L K D G
21 --->
661 gtcactacacctctgaagttaagactacctataaagc GCTGCCAG
gagacttcaattctgatggatatttcggtttttcggccaagtcgacggtc
G H Y T S E V K T T Y K A K K P V Q L P
721 GCGCGTACATCGTTGGTATCAAACTCGACATCACTTCTCACAACGAGGAC
cgcgcatg GTAGTGAAGAGTGTTGCTCCTGATGTGCTAAC
<--- 20
G A Y I V G I K L D I T S H N E D Y T I

```
    |       |       |       |       |       |       |

          23 --->
781       gtacgaacgtgccgaaggtcgccactctaccggtggtatggatgaactgtaca
    AACTTGTCATGCTTGCACGGCTTCCA             accatacctacttgacatgt
                      <--- 22
    V  E  Q  Y  E  R  A  E  G  R  H  S  T  G  G  M  D  E  L  Y

    |       |       |       |       |       |       |

              25 --->
841 aatga     GGTTCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACA
                   ******************                          < repeat
        ********                                             < GC rich
                                    **********               < AT rich
    ttactccccccccaagagtagtagtagtagtaatt              GTCGTGT
                                  <--- 24
    K  X

    |       |       |       |       |       |       |

                                  27 --->
901 CTGGCGGCC                     taacaaagcccgaaaggaagctgagttggc
      ********                                               < GC rich
    GACCGCCGGCAATGATCACCTAGGCCGACGATTGTTTCGGGCTTTCCTTCG
                                                 <--- 26


    |       |       |       |       |       |       |

                                                  29 --->
961 tgctgccaccgctgagcaataactagca                  TAAACGGGTCTTGAG
            ggcgactcgttattgatcgtattggggaaccccggagatttgcccagaactc
                                                                <

    |       |       |       |       |       |       |


1021 GGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAG
                               **************                 < repeat
    cccaaa              CCTTGATATAGGCCTATAGGTGTCCTGCCCACACCAGCGG
    --- 28                                              <--- 30

 ------------------------------------------------------------

 The total codon usage score ...........     0.000
 The total length score ................     0.000
 The total melting temperature score ...     0.002
 The total repeat score ................     2.298
 The total pattern score ...............     0.000
 The total mispriming score ............     0.000
 The total AT content score ............     1.186
 The total GC content score ............     0.297
               The OVERALL score ......     3.783
```

DETAILED CODON FREQUENCY REPORT

[Codon, AA, Frequency, # of times used in coding sequence]

TTT	F	0.29	4	TCT	S	0.32	12	TAT	Y	0.35	1	TGT	C	0.39	0
TTC	F	0.71	8	TCC	S	0.27	1	TAC	Y	0.65	11	TGC	C	0.61	0
TTA	L	0.03	0	TCA	S	0.05	0	TAA	X	0.63	0	TGA	X	0.35	1
TTG	L	0.06	0	TCG	S	0.07	0	TAG	X	0.08	0	TGG	W	1.00	3
CTT	L	0.06	0	CCT	P	0.11	1	CAT	H	0.30	0	CGT	R	0.64	7
CTC	L	0.08	2	CCC	P	0.02	0	CAC	H	0.70	5	CGC	R	0.33	2
CTA	L	0.01	0	CCA	P	0.15	4	CAA	Q	0.19	2	CGA	R	0.01	0
CTG	L	0.77	10	CCG	P	0.72	7	CAG	Q	0.81	5	CGG	R	0.01	0
ATT	I	0.34	1	ACT	T	0.29	2	AAT	N	0.17	1	AGT	S	0.05	0
ATC	I	0.66	10	ACC	T	0.54	8	AAC	N	0.83	5	AGC	S	0.24	0
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	16	AGA	R	0.01	0
ATG	M	1.00	10	ACG	T	0.13	3	AAG	K	0.22	8	AGG	R	0.00	0
GTT	V	0.40	13	GCT	A	0.28	1	GAT	D	0.46	4	GGT	G	0.51	16
GTC	V	0.14	1	GCC	A	0.16	3	GAC	D	0.54	8	GGC	G	0.43	10
GTA	V	0.20	0	GCA	A	0.24	1	GAA	E	0.75	16	GGA	G	0.02	0
GTG	V	0.27	1	GCG	A	0.32	5	GAG	E	0.25	8	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	2
10% - 14%	5
15% - 19%	10
20% - 24%	17
25% - 29%	9
30% - 34%	20
35% - 39%	15
40% - 44%	10
45% - 49%	4
>= 50%	145

Total Codons Used = 237

Tm Range	# of Overlaps
<59	0
59	1
60	11
61	17
62	0
63	0
64	0
65	0
66	0
>=67	0

Tm Range = 2.0

Ovrlap Len Range	# of Oligos
<17	3
17	2
18	2
19	2
20	5
21	3
22	5
23	1
24	2
25	1
26	0
>=27	2

Lowest Overlap = 16

Length Range	# of Oligos
<48	2
48-49	0
50-51	0
52-53	0
54-55	0
56-57	0
58-59	28
60-61	0
62-63	0
64-65	0
66-67	0
>=68	0

Longest = 58

There are 5 repeats greater than 8 nt:

```
PR Pos1 =   30   Pos2 =   30   Size = 18   Seq1 = GGATCGAGATCTCGATCC
                                           Seq2 = GGATCGAGATCTCGATCC
PR Pos1 =  132   Pos2 =  132   Size =  8   Seq1 = ACCATGGT
                                           Seq2 = ACCATGGT
DR Pos1 =  857   Pos2 =  860   Size = 16   Seq1 = TCATCATCATCATCAT
                                           Seq2 = TCATCATCATCATCAT
PR Pos1 = 1046   Pos2 = 1046   Size = 12   Seq1 = ATATCCGGATAT
                                           Seq2 = ATATCCGGATAT
PR Pos1 = 1052   Pos2 = 1052   Size =  8   Seq1 = GGATATCC
                                           Seq2 = GGATATCC
```

Sequence Patterns Screened (As Supplied By User)

None found

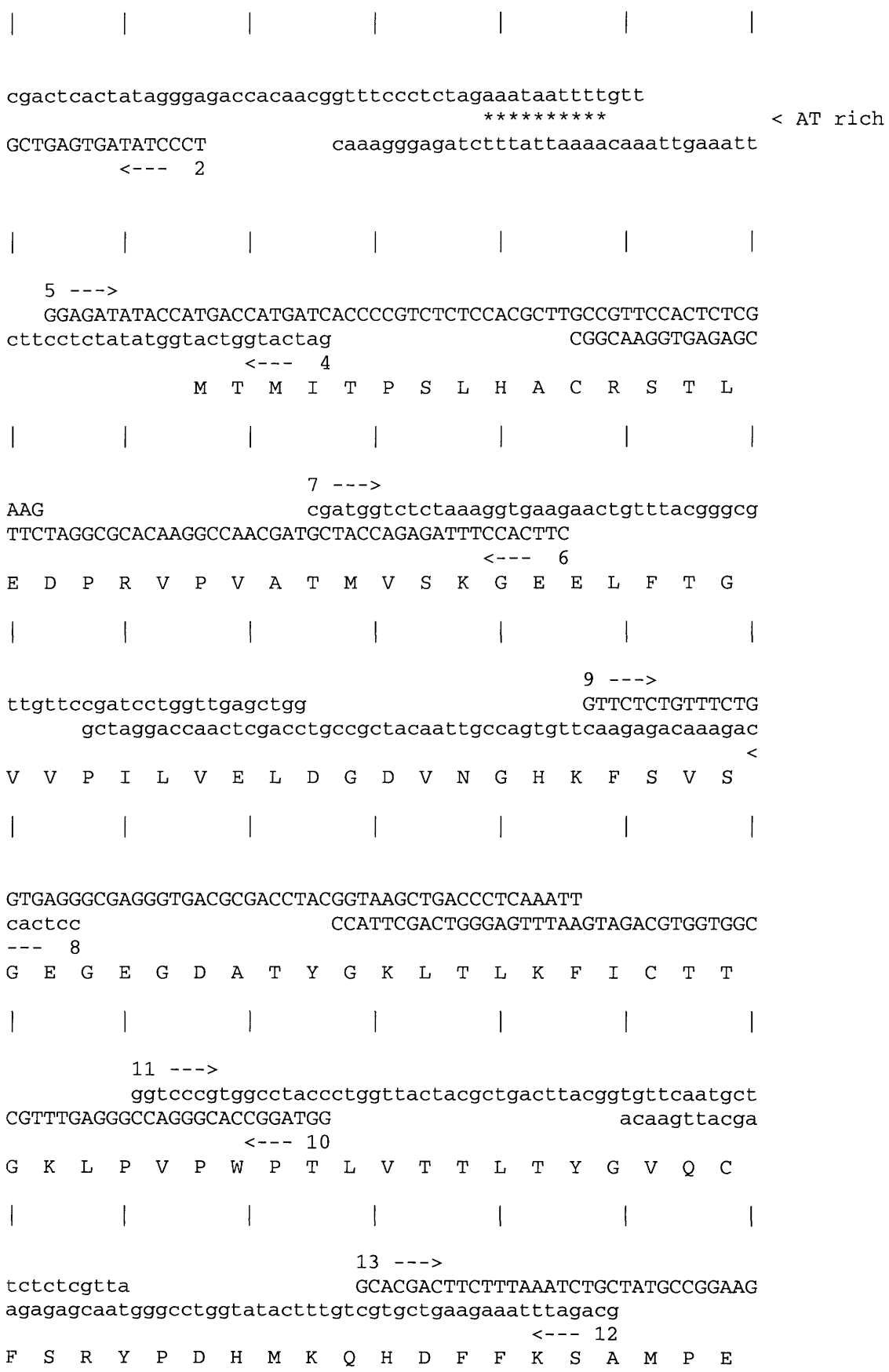
30 oligonucleotides need to be synthesized

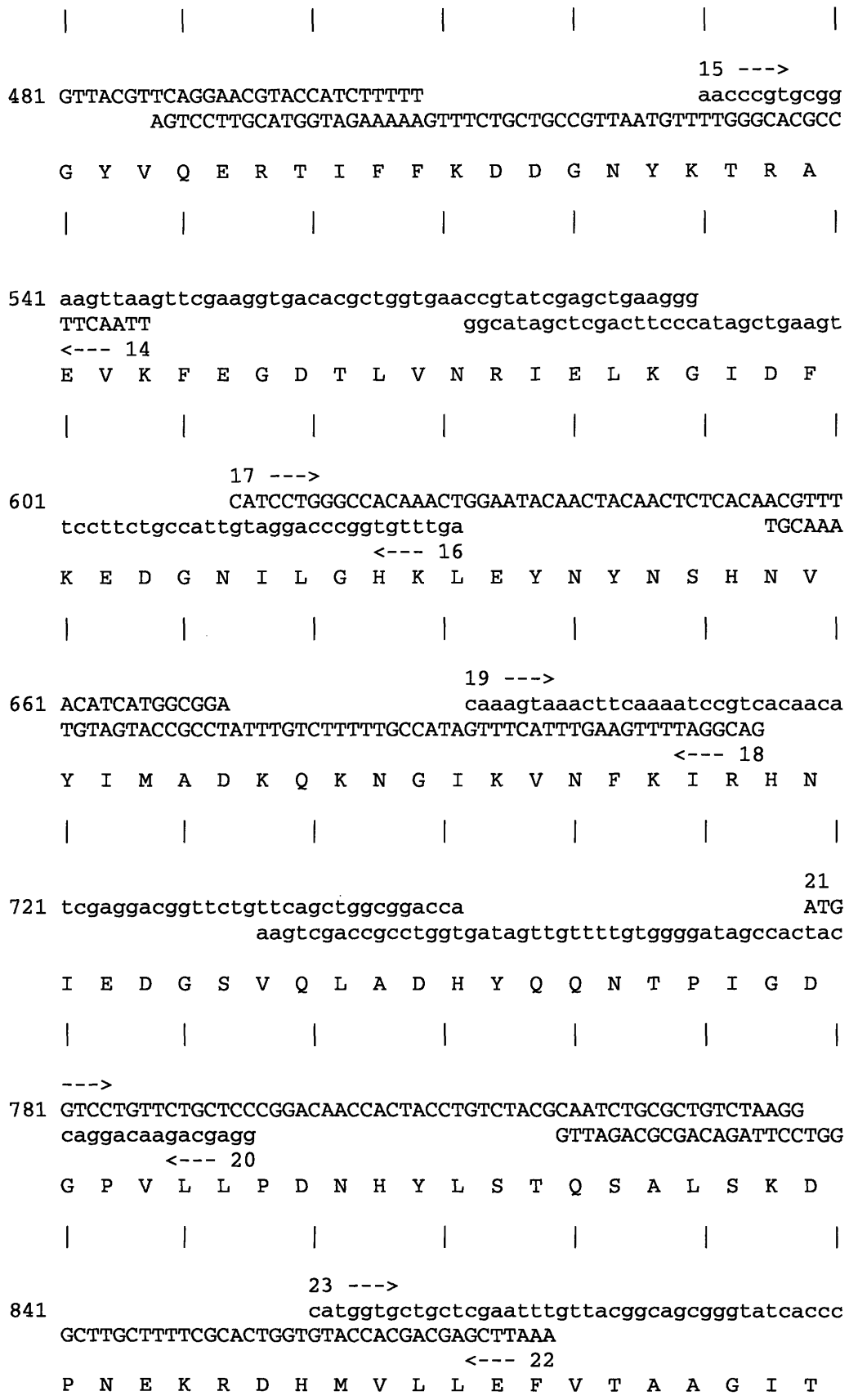
---

1 GATGCCGGCCACGATGCGTCCGG 23
2 AGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGC 58
3 ATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAA 58
4 CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGAAACCGTTGTG 58
5 GTTTAACTTTAAGAAGGAGATATACCATGGTTTCCAAAGGCGAGGAAAATAACATGGC 58
6 CATACGAACTTTGAAACGCATGAACTCTTTGATGATCGCCATGTTATTTTCCTCGCCT 58
7 TCATGCGTTTCAAAGTTCGTATGGAAGGCTCTGTTAACGGTCACGAATTTGAAATCGA 58
8 TGAAAACCTTCGTACGGACGGCCCTCGCCCTCACCTTCGATTTCAAATTCGTGACCGT 58
9 CGTCCGTACGAAGGTTTTCAAACCGCTAAACTGAAAGTTACCAAGGGTGGTCCGCTGC 58
10 GCCGTAGGTGAACTGCGGAGACAGGATGTCCCACGCAAACGGCAGCGGACCACCCTTG 58
11 CCGCAGTTCACCTACGGCTCTAAGGCGTACGTCAAACACCCAGCAGACATCCCTGACT 58
12 GCTCCCACTTGAAGCCTTCTGGGAAAGACAGTTTGAAGTAGTCAGGGATGTCTGCTGG 58
13 AGAAGGCTTCAAGTGGGAGCGTGTTATGAACTTCGAGGATGGTGGCGTGGTTACCGTT 58
14 TGTAGATAAATTCACCGTCTTGCAGAGAAGAGTCCTGCGTAACGGTAACCACGCCACC 58
15 TGCAAGACGGTGAATTTATCTACAAGGTTAAACTGCGCGGTACCAACTTCCCGTCTGA 58
16 CTTCCCAACCCATCGTTTTTTTCTGCATAACTGGGCCATCAGACGGGAAGTTGGTACC 58
17 AAAAAAACGATGGGTTGGGAAGCGTCTTCTGAACGTATGTACCCGGAAGATGGTGCCC 58
18 GCCGTCTTTGAGCTTCAGACGCATTTTGATCTCACCCTTCAGGGCACCATCTTCCGGG 58
19 GTCTGAAGCTCAAAGACGGCGGTCACTACACCTCTGAAGTTAAGACTACCTATAAAGC 58
20 GTACGCGCCTGGCAGCTGAACCGGCTTTTTGGCTTTATAGGTAGTCTTAACTTCAGAG 58
21 GCTGCCAGGCGCGTACATCGTTGGTATCAAACTCGACATCACTTCTCACAACGAGGAC 58
22 ACCTTCGGCACGTTCGTACTGTTCAACAATCGTGTAGTCCTCGTTGTGAGAAGTGATG 58
23 GTACGAACGTGCCGAAGGTCGCCACTCTACCGGTGGTATGGATGAACTGTACAAATGA 58
24 TTAATGATGATGATGATGATGAGAACCCCCCCCTCATTTGTACAGTTCATCCATACCA 58
25 GGTTCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCC 58
26 GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCACTAGTAACGGCCGCCAGTGTGCTG 58
27 TAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCA 58
28 AAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGG 58
29 TAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAG 58
30 GGCGACCACACCCGTCCTGTGGATATCCGGATATAGTTCC 40

eGFP

```
-----------------------------------------------------------------
                    PARAMETERS FOR TRIAL 11
 ---------------------------------------------------------------
              Total Size Of Gene ......... 1160 nt
              Protein Residues ........... 264
              Mutatable Residues ......... 255
              Fixed Nucleotides .......... 395 nt
              Oligo Size ................. 60 nt
              Annealing Temp ............. 58 +/- 1*C
              Oligo Concentration ........ 2.00E-8 M
              Sodium Concentration ....... 5.00E-2 M
              Mg2+ Concentration ......... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ........... 8 nt
              Mispriming Threshold ....... 8/18 (6 exact) nt

-------------------------------------------------------------------------------

 The DNA sequence #  11 is:
 ---------------------------------------------------------------
   1 GATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATA
  61 CGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAA
 121 GAAGGAGATATACCATGACCATGATCACCCCGTCTCTCCACGCTTGCCGTTCCACTCTCG
 181 AAGATCCGCGTGTTCCGGTTGCTACGATGGTCTCTAAAGGTGAAGAACTGTTTACGGGCG
 241 TTGTTCCGATCCTGGTTGAGCTGGACGGCGATGTTAACGGTCACAAGTTCTCTGTTTCTG
 301 GTGAGGGCGAGGGTGACGCGACCTACGGTAAGCTGACCCTCAAATTCATCTGCACCACCG
 361 GCAAACTCCCGGTCCCGTGGCCTACCCTGGTTACTACGCTGACTTACGGTGTTCAATGCT
 421 TCTCTCGTTACCCGGACCATATGAAACAGCACGACTTCTTTAAATCTGCTATGCCGGAAG
 481 GTTACGTTCAGGAACGTACCATCTTTTTCAAAGACGACGGCAATTACAAAACCCGTGCGG
 541 AAGTTAAGTTCGAAGGTGACACGCTGGTGAACCGTATCGAGCTGAAGGGTATCGACTTCA
 601 AGGAAGACGGTAACATCCTGGGCCACAAACTGGAATACAACTACAACTCTCACAACGTTT
 661 ACATCATGGCGGATAAACAGAAAAACGGTATCAAAGTAAACTTCAAAATCCGTCACAACA
 721 TCGAGGACGGTTCTGTTCAGCTGGCGGACCACTATCAACAAAACACCCCTATCGGTGATG
 781 GTCCTGTTCTGCTCCCGGACAACCACTACCTGTCTACGCAATCTGCGCTGTCTAAGGACC
 841 CGAACGAAAAGCGTGACCACATGGTGCTGCTCGAATTTGTTACGGCAGCGGGTATCACCC
 901 TGGGCATGGACGAACTCTACAAATAAGGGGGGGGTTCTCATCATCATCATCATCATTAAT
 961 AAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGGCTGCTAACAAAGC
1021 CCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG
1081 GGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCC
1141 ACAGGACGGGTGTGGTCGCC
 ---------------------------------------------------------------

 The oligonucleotide assembly is:
 ---------------------------------------------------------------
     1       10        20        30        40        50        60
     |       |         |         |         |         |         |

     1 --->                                          3 --->
   1 GATGCCGGCCACGATGCGTCCGGCGTAGAGGAT               aaattaata
                                 ******************             < repeat
                                                    ********* < AT rich
                      GCAGGCCGCATCTCCTAGCTCTAGAGCTAGGGCGCTTTAATTAT
```


61 cgactcactatagggagaccacaacggtttccctctagaaataattttgtt
****
< AT rich
GCTGAGTGATATCCCT caaagggagatctttattaaaacaaattgaaatt
<--- 2
5 --->
121 GGAGATATACCATGACCATGATCACCCCGTCTCTCCACGCTTGCCGTTCCACTCTCG
cttcctctatatggtactggtactag CGGCAAGGTGAGAGC
<--- 4
M T M I T P S L H A C R S T L
7 --->
181 AAG cgatggtctctaaaggtgaagaactgtttacgggcg
TTCTAGGCGCACAAGGCCAACGATGCTACCAGAGATTTCCACTTC
<--- 6
E D P R V P V A T M V S K G E E L F T G
9 --->
241 ttgttccgatcctggttgagctgg GTTCTCTGTTTCTG
gctaggaccaactcgacctgccgctacaattgccagtgttcaagagacaaagac
<
V V P I L V E L D G D V N G H K F S V S
301 GTGAGGGCGAGGGTGACGCGACCTACGGTAAGCTGACCCTCAAATT
cactcc CCATTCGACTGGGAGTTTAAGTAGACGTGGTGGC
--- 8
G E G E G D A T Y G K L T L K F I C T T
11 --->
361 ggtcccgtggcctaccctggttactacgctgacttacggtgttcaatgct
CGTTTGAGGGCCAGGGCACCGGATGG acaagttacga
<--- 10
G K L P V P W P T L V T T L T Y G V Q C
13 --->
421 tctctcgtta GCACGACTTCTTTAAATCTGCTATGCCGGAAG
agagagcaatgggcctggtatactttgtcgtgctgaagaaatttagacg
<--- 12
F S R Y P D H M K Q H D F F K S A M P E


15 --->
481 GTTACGTTCAGGAACGTACCATCTTTTT aacccgtgcgg
AGTCCTTGCATGGTAGAAAAAGTTTCTGCTGCCGTTAATGTTTTGGGCACGCC
G Y V Q E R T I F F K D D G N Y K T R A
541 aagttaagttcgaaggtgacacgctggtgaaccgtatcgagctgaaggg
TTCAATT ggcatagctcgacttcccatagctgaagt
<--- 14
E V K F E G D T L V N R I E L K G I D F
17 --->
601 CATCCTGGGCCACAAACTGGAATACAACTACAACTCTCACAACGTTT
tccttctgccattgtaggacccggtgtttga TGCAAA
<--- 16
K E D G N I L G H K L E Y N Y N S H N V
19 --->
661 ACATCATGGCGGA caaagtaaacttcaaaatccgtcacaaca
TGTAGTACCGCCTATTTGTCTTTTTGCCATAGTTTCATTTGAAGTTTTAGGCAG
<--- 18
Y I M A D K Q K N G I K V N F K I R H N
21
721 tcgaggacggttctgttcagctggcggacca ATG
aagtcgaccgcctggtgatagttgttttgtggggatagccactac
I E D G S V Q L A D H Y Q Q N T P I G D
---> 781 GTCCTGTTCTGCTCCCGGACAACCACTACCTGTCTACGCAATCTGCGCTGTCTAAGG
caggacaagacgagg GTTAGACGCGACAGATTCCTGG
<--- 20
G P V L L P D N H Y L S T Q S A L S K D
23 --->
841 catggtgctgctcgaatttgttacggcagcgggtatcaccc
GCTTGCTTTTCGCACTGGTGTACCACGACGAGCTTAAA
<--- 22
P N E K R D H M V L L E F V T A A G I T

```
    |       |       |       |       |       |       |

                                  25 --->
901 tgggcatggacgaactcta           TCTCATCATCATCATCATCATTAAT
                                    ******************     < repeat
                             ********                      < GC rich
                                                  ****** < AT rich
     cccgtacctgcttgagatgtttattccccccccaagagtagtagtagtagtagtaatta
                                                  <--- 2
    L  G  M  D  E  L  Y  K  X

    |       |       |       |       |       |       |


961 AAAAGGGCGAATTCCAGCACACTGGCGGCCGTTAC
                         ********                      < GC rich
    ****                                               < AT rich
    t                GTGACCGCCGGCAATGATCACCTAGGCCGACGATTGTTTCG
    4

    |       |       |       |       |       |       |

    27 --->
1021 ccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttgg
    GGCTTTCCTTCGACTCAAC                tattgatcgtattggggaacc
            <--- 26

    |       |       |       |       |       |       |

                  29 --->
1081              AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCC
                                          ************** < repeat
    ccggagatttgcccagaactccccaaaaaacgactttcc
                                <--- 28

    |       |       |       |       |       |       |


1141 ACAGGACGGGTGTGGTCGC
        CTGCCCACACCAGCGG
                <--- 30

 ------------------------------------------------------------

 The total codon usage score ..........    0.000
 The total length score ...............    0.000
 The total melting temperature score ...   0.002
 The total repeat score ...............    1.862
 The total pattern score ..............    0.000
 The total mispriming score ...........    0.000
 The total AT content score ...........    1.103
 The total GC content score ...........    0.276
```

The OVERALL score 3.243

DETAILED CODON FREQUENCY REPORT
[Codon, AA, Frequency, # of times used in coding sequence]

TTT	F	0.29	4	TCT	S	0.32	11	TAT	Y	0.35	1	TGT	C	0.39	0
TTC	F	0.71	8	TCC	S	0.27	1	TAC	Y	0.65	10	TGC	C	0.61	3
TTA	L	0.03	0	TCA	S	0.05	0	TAA	X	0.63	1	TGA	X	0.35	0
TTG	L	0.06	0	TCG	S	0.07	0	TAG	X	0.08	0	TGG	W	1.00	1
CTT	L	0.06	0	CCT	P	0.11	3	CAT	H	0.30	1	CGT	R	0.64	8
CTC	L	0.08	7	CCC	P	0.02	0	CAC	H	0.70	9	CGC	R	0.33	0
CTA	L	0.01	0	CCA	P	0.15	0	CAA	Q	0.19	4	CGA	R	0.01	0
CTG	L	0.77	16	CCG	P	0.72	10	CAG	Q	0.81	4	CGG	R	0.01	0
ATT	I	0.34	0	ACT	T	0.29	3	AAT	N	0.17	1	AGT	S	0.05	0
ATC	I	0.66	13	ACC	T	0.54	11	AAC	N	0.83	12	AGC	S	0.24	0
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	13	AGA	R	0.01	0
ATG	M	1.00	8	ACG	T	0.13	6	AAG	K	0.22	7	AGG	R	0.00	0
GTT	V	0.40	15	GCT	A	0.28	3	GAT	D	0.46	4	GGT	G	0.51	15
GTC	V	0.14	2	GCC	A	0.16	0	GAC	D	0.54	15	GGC	G	0.43	7
GTA	V	0.20	1	GCA	A	0.24	1	GAA	E	0.75	12	GGA	G	0.02	0
GTG	V	0.27	2	GCG	A	0.32	6	GAG	E	0.25	5	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	7
10% - 14%	11
15% - 19%	5
20% - 24%	14
25% - 29%	14
30% - 34%	17
35% - 39%	16
40% - 44%	7
45% - 49%	4
>= 50%	169

Total Codons Used = 264

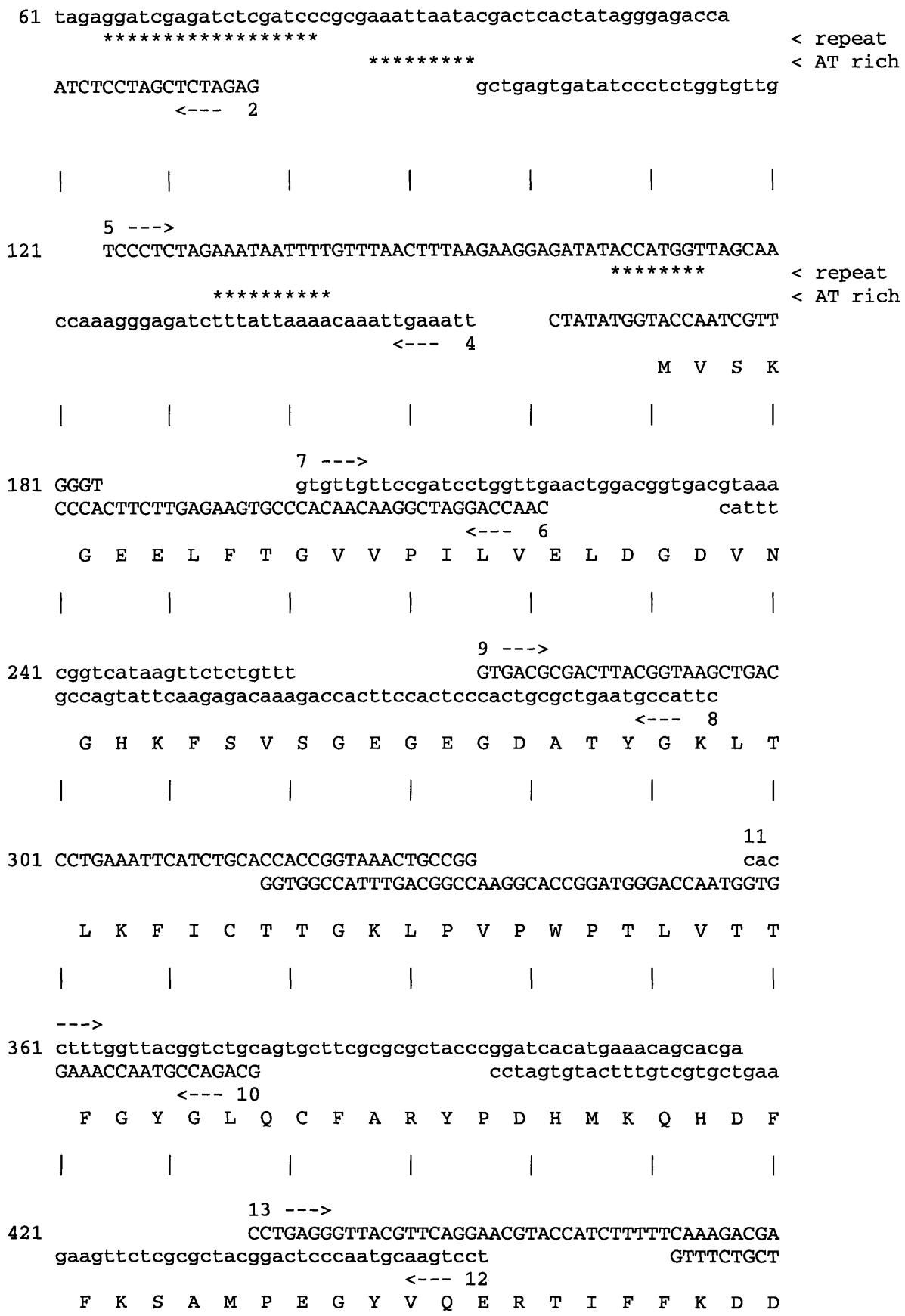
Tm Range	# of Overlaps
<55	0
55	0
56	1
57	15
58	13
59	0
60	0
61	0
62	0
>=63	0

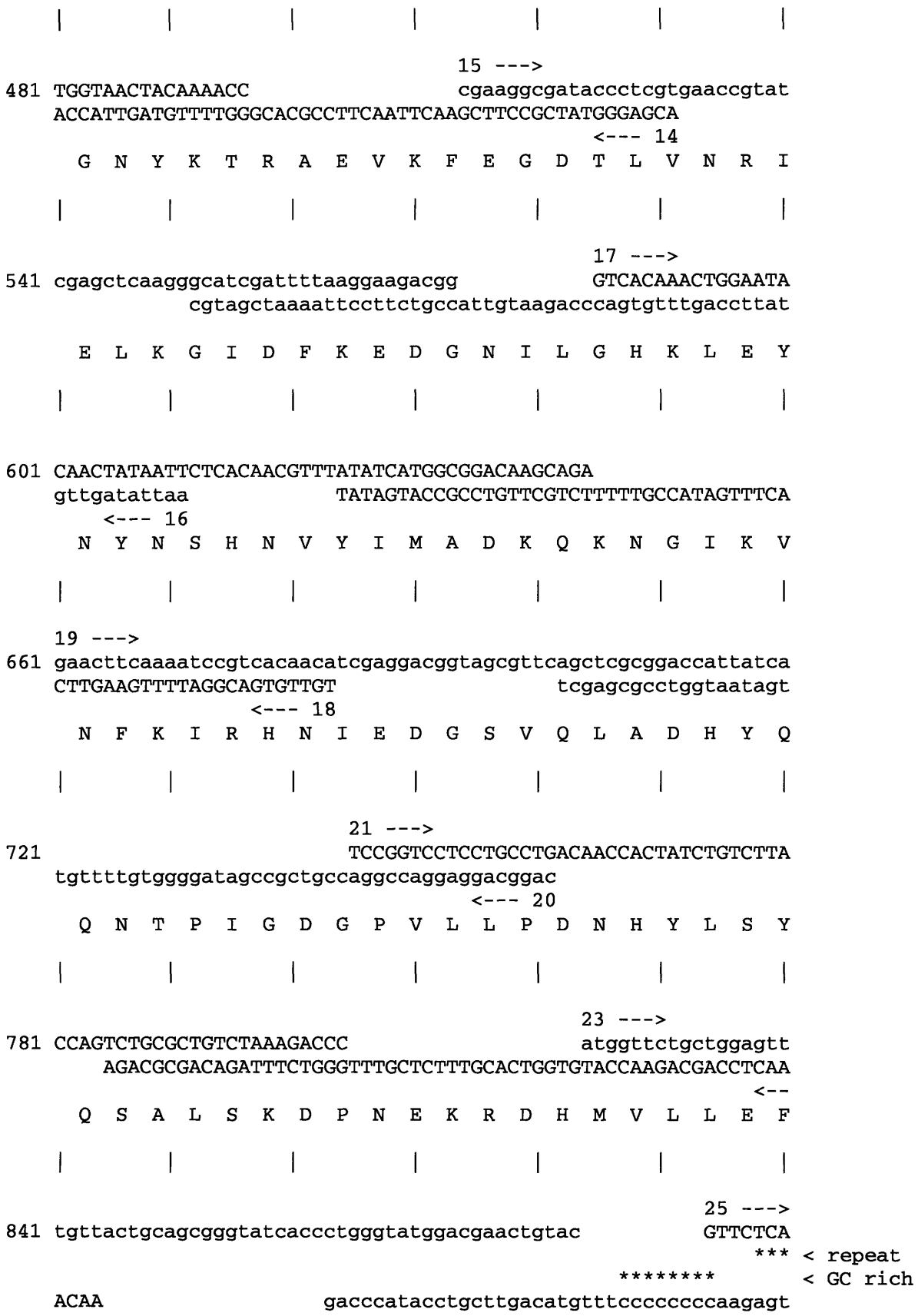
```
-------------------------------------
 Tm Range =  2.4

 Ovrlap Len Range  # of Oligos
-------------------------------------
      <16            1
       16            3
       17            1
       18            7
       19            4
       20            3
       21            5
       22            0
       23            2
       24            0
       25            2
     >=26            1
-------------------------------------
 Lowest Overlap = 15

 Length Range   # of Oligos
-------------------------------------
      <49            2
     49-50           0
     51-52           0
     53-54           0
     55-56           0
     57-58           0
     59-60          28
     61-62           0
     63-64           0
     65-66           0
     67-68           0
      >=69           0
-------------------------------------
 Longest =  60

 There are   4 repeats greater than  8 nt:
-----------------------------------------------------------------
PR Pos1 =   30   Pos2 =   30   Size = 18   Seq1 = GGATCGAGATCTCGATCC
                                          Seq2 = GGATCGAGATCTCGATCC
DR Pos1 =  938   Pos2 =  941   Size = 16   Seq1 = TCATCATCATCATCAT
                                          Seq2 = TCATCATCATCATCAT
PR Pos1 = 1127   Pos2 = 1127   Size = 12   Seq1 = ATATCCGGATAT
                                          Seq2 = ATATCCGGATAT
PR Pos1 = 1133   Pos2 = 1133   Size =  8   Seq1 = GGATATCC
                                          Seq2 = GGATATCC
-----------------------------------------------------------------


         Sequence Patterns Screened (As Supplied By User)
-----------------------------------------------------------------
                      None found
-----------------------------------------------------------------

       30 oligonucleotides need to be synthesized
```

```
 1 GATGCCGGCCACGATGCGTCCGGCGTAGAGGAT  33
 2 TCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACG  60
 3 AAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTT  60
 4 GATCATGGTCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGAAAC  60
 5 GGAGATATACCATGACCATGATCACCCCGTCTCTCCACGCTTGCCGTTCCACTCTCGAAG  60
 6 CTTCACCTTTAGAGACCATCGTAGCAACCGGAACACGCGGATCTTCGAGAGTGGAACGGC  60
 7 CGATGGTCTCTAAAGGTGAAGAACTGTTTACGGGCGTTGTTCCGATCCTGGTTGAGCTGG  60
 8 CCTCACCAGAAACAGAGAACTTGTGACCGTTAACATCGCCGTCCAGCTCAACCAGGATCG  60
 9 GTTCTCTGTTTCTGGTGAGGGCGAGGGTGACGCGACCTACGGTAAGCTGACCCTCAAATT  60
10 GGTAGGCCACGGGACCGGGAGTTTGCCGGTGGTGCAGATGAATTTGAGGGTCAGCTTACC  60
11 GGTCCCGTGGCCTACCCTGGTTACTACGCTGACTTACGGTGTTCAATGCTTCTCTCGTTA  60
12 GCAGATTTAAAGAAGTCGTGCTGTTTCATATGGTCCGGGTAACGAGAGAAGCATTGAACA  60
13 GCACGACTTCTTTAAATCTGCTATGCCGGAAGGTTACGTTCAGGAACGTACCATCTTTTT  60
14 TTAACTTCCGCACGGGTTTTGTAATTGCCGTCGTCTTTGAAAAAGATGGTACGTTCCTGA  60
15 AACCCGTGCGGAAGTTAAGTTCGAAGGTGACACGCTGGTGAACCGTATCGAGCTGAAGGG  60
16 AGTTTGTGGCCCAGGATGTTACCGTCTTCCTTGAAGTCGATACCCTTCAGCTCGATACGG  60
17 CATCCTGGGCCACAAACTGGAATACAACTACAACTCTCACAACGTTTACATCATGGCGGA  60
18 GACGGATTTTGAAGTTTACTTTGATACCGTTTTTCTGTTTATCCGCCATGATGTAAACGT  60
19 CAAAGTAAACTTCAAAATCCGTCACAACATCGAGGACGGTTCTGTTCAGCTGGCGGACCA  60
20 GGAGCAGAACAGGACCATCACCGATAGGGGTGTTTTGTTGATAGTGGTCCGCCAGCTGAA  60
21 ATGGTCCTGTTCTGCTCCCGGACAACCACTACCTGTCTACGCAATCTGCGCTGTCTAAGG  60
22 AAATTCGAGCAGCACCATGTGGTCACGCTTTTCGTTCGGGTCCTTAGACAGCGCAGATTG  60
23 CATGGTGCTGCTCGAATTTGTTACGGCAGCGGGTATCACCCTGGGCATGGACGAACTCTA  60
24 TATTAATGATGATGATGATGATGAGAACCCCCCCCTTATTTGTAGAGTTCGTCCATGCCC  60
25 TCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTTAC  60
26 CAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCACTAGTAACGGCCGCCAGTG  60
27 CCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG  60
28 CCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTAT  60
29 AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGC  60
30 GGCGACCACACCCGTC  16
```

eYFP

```
------------------------------------------------------------
                  PARAMETERS FOR TRIAL 44
------------------------------------------------------------
              Total Size Of Gene ..... 1125 nt
              Protein Residues ....... 239
              Mutatable Residues ..... 232
              Fixed Nucleotides ...... 429 nt
              Oligo Size ............. 60 nt
              Annealing Temp ......... 61 +/- 1*C
              Oligo Concentration .... 2.00E-8 M
              Sodium Concentration ... 5.00E-2 M
              Mg2+ Concentration ..... 2.00E-3 M
              Codon Frequency Threshold .. 10%
              Repeat Threshold ....... 8 nt
              Mispriming Threshold ... 8/18 (6 exact) nt

------------------------------------------------------------------------------

 The DNA sequence #  44 is:
 ------------------------------------------------------------
   1 CGGTCACGCTTGGGACTGCCATAGGCTGGCCCGGTGATGCCGGCCACGATGCGTCCGGCG
  61 TAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAAC
 121 GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAGCAA
 181 GGGTGAAGAACTCTTCACGGGTGTTGTTCCGATCCTGGTTGAACTGGACGGTGACGTAAA
 241 CGGTCATAAGTTCTCTGTTTCTGGTGAAGGTGAGGGTGACGCGACTTACGGTAAGCTGAC
 301 CCTGAAATTCATCTGCACCACCGGTAAACTGCCGGTTCCGTGGCCTACCCTGGTTACCAC
 361 CTTTGGTTACGGTCTGCAGTGCTTCGCGCGCTACCCGGATCACATGAAACAGCACGACTT
 421 CTTCAAGAGCGCGATGCCTGAGGGTTACGTTCAGGAACGTACCATCTTTTTCAAAGACGA
 481 TGGTAACTACAAAACCCGTGCGGAAGTTAAGTTCGAAGGCGATACCCTCGTGAACCGTAT
 541 CGAGCTCAAGGGCATCGATTTTAAGGAAGACGGTAACATTCTGGGTCACAAACTGGAATA
 601 CAACTATAATTCTCACAACGTTTATATCATGGCGGACAAGCAGAAAAACGGTATCAAAGT
 661 GAACTTCAAAATCCGTCACAACATCGAGGACGGTAGCGTTCAGCTCGCGGACCATTATCA
 721 ACAAAACACCCCTATCGGCGACGGTCCGGTCCTCCTGCCTGACAACCACTATCTGTCTTA
 781 CCAGTCTGCGCTGTCTAAAGACCCAAACGAGAAACGTGACCACATGGTTCTGCTGGAGTT
 841 TGTTACTGCAGCGGGTATCACCCTGGGTATGGACGAACTGTACAAAGGGGGGGGTTCTCA
 901 TCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTG
 961 GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAA
1021 TAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGA
1081 GGAACTATATCCGGAGCGACTCCCACGGCACGTTGGCAAGCTCGA
 ------------------------------------------------------------

 The oligonucleotide assembly is:
 ------------------------------------------------------------
     1       10        20        30        40        50        60
     |       |         |         |         |         |         |

     1 --->                                                  3 --
   1 CGGTCACGCTTGGGACTGCCATAGGCTGGCCCG                      ggcg
                      CGGTATCCGACCGGGCCACTACGGCCGGTGCTACGCAGGCCGC


     |       |         |         |         |         |         |

     ->
```


61 tagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggagacca
< repeat
< AT rich
ATCTCCTAGCTCTAGAG gctgagtgatatccctctggtgttg
<--- 2
5 --->
121 TCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAGCAA
< repeat
< AT rich
ccaaagggagatctttattaaaacaaattgaaatt CTATATGGTACCAATCGTT
<--- 4
M V S K
7 --->
181 GGGT gtgttgttccgatcctggttgaactggacggtgacgtaaa
CCCACTTCTTGAGAAGTGCCCACAACAAGGCTAGGACCAAC cattt
<--- 6
G E E L F T G V V P I L V E L D G D V N
9 --->
241 cggtcataagttctctgttt GTGACGCGACTTACGGTAAGCTGAC
gccagtattcaagagacaaagaccacttccactcccactgcgctgaatgccattc
<--- 8
G H K F S V S G E G E G D A T Y G K L T
11
301 CCTGAAATTCATCTGCACCACCGGTAAACTGCCGG cac
GGTGGCCATTTGACGGCCAAGGCACCGGATGGGACCAATGGTG
L K F I C T T G K L P V P W P T L V T T
--->
361 ctttggttacggtctgcagtgcttcgcgcgctacccggatcacatgaaacagcacga
GAAACCAATGCCAGACG cctagtgtactttgtcgtgctgaa
<--- 10
F G Y G L Q C F A R Y P D H M K Q H D F
13 --->
421 CCTGAGGGTTACGTTCAGGAACGTACCATCTTTTTCAAAGACGA
gaagttctcgcgctacggactcccaatgcaagtcct GTTTCTGCT
<--- 12
F K S A M P E G Y V Q E R T I F F K D D


15 --->
481 TGGTAACTACAAAACC cgaaggcgataccctcgtgaaccgtat
ACCATTGATGTTTTGGGCACGCCTTCAATTCAAGCTTCCGCTATGGGAGCA
<--- 14
G N Y K T R A E V K F E G D T L V N R I
17 --->
541 cgagctcaagggcatcgattttaaggaagacgg GTCACAAACTGGAATA
cgtagctaaaattccttctgccattgtaagacccagtgtttgaccttat
E L K G I D F K E D G N I L G H K L E Y
601 CAACTATAATTCTCACAACGTTTATATCATGGCGGACAAGCAGA
gttgatattaa TATAGTACCGCCTGTTCGTCTTTTTGCCATAGTTTCA
<--- 16
N Y N S H N V Y I M A D K Q K N G I K V
19 --->
661 gaacttcaaaatccgtcacaacatcgaggacggtagcgttcagctcgcggaccattatca
CTTGAAGTTTTAGGCAGTGTTGT tcgagcgcctggtaatagt
<--- 18
N F K I R H N I E D G S V Q L A D H Y Q
21 --->
721 TCCGGTCCTCCTGCCTGACAACCACTATCTGTCTTA
tgttttgtggggatagccgctgccaggccaggaggacggac
<--- 20
Q N T P I G D G P V L L P D N H Y L S Y
23 --->
781 CCAGTCTGCGCTGTCTAAAGACCC atggttctgctggagtt
AGACGCGACAGATTTCTGGGTTTGCTCTTTGCACTGGTGTACCAAGACGACCTCAA
<--
Q S A L S K D P N E K R D H M V L L E F
25 --->
841 tgttactgcagcgggtatcaccctgggtatggacgaactgtac GTTCTCA
*** < repeat
******** < GC rich
ACAA gacccatacctgcttgacatgtttccccccccaagagt

```
    - 22
      V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K

    |        |        |        |        |        |        |


 901 TCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTT
     ****************                                          < repeat
                                              ********         < GC rich
                  **********                                  < AT rich
     agtagtagtagtagtaattatt                GTGTGACCGCCGGCAATGATCAC
                    <--- 24


    |        |        |        |        |        |        |

                      27 --->
 961                  gcccgaaaggaagctgagttggctgctgccaccgctgagcaa
     CTAGGCCGACGATTGTTTCGGGCTTTCCTTCGACTCA                  tcgtt
                                   <--- 26


    |        |        |        |        |        |        |

                                      29 --->
1021 taactagcataacccctt                TCTTGAGGGGTTTTTTGCTGAAAGGA
     attgatcgtattggggaaccccggagatttgcccagaactccccaaaaaacgact
                                                    <--- 28


    |        |        |        |        |        |        |


1081 GGAACTATATCCGGAGCGACTCCCACGGCACGTT
                     CTGAGGGTGCCGTGCAACCGTTCGAGCT
                                        <--- 30

 ------------------------------------------------------------

 The total codon usage score ..........     0.000
 The total length score ...............     0.000
 The total melting temperature score ...    0.000
 The total repeat score ...............     1.493
 The total pattern score ..............     0.000
 The total mispriming score ...........     0.000
 The total AT content score ...........     1.138
 The total GC content score ...........     0.284
               The OVERALL score ......     2.916


                DETAILED CODON FREQUENCY REPORT
  [ Codon, AA, Frequency, # of times used in coding sequence ]
 ------------------------------------------------------------

 TTT F 0.29   4  TCT S 0.32   6  TAT Y 0.35   4  TGT C 0.39   0
 TTC F 0.71   9  TCC S 0.27   0  TAC Y 0.65   8  TGC C 0.61   2
```

TTA	L	0.03	0	TCA	S	0.05	0	TAA	X	0.63	0
TTG	L	0.06	0	TCG	S	0.07	0	TAG	X	0.08	0
CTT	L	0.06	0	CCT	P	0.11	4	CAT	H	0.30	2
CTC	L	0.08	5	CCC	P	0.02	0	CAC	H	0.70	7
CTA	L	0.01	0	CCA	P	0.15	1	CAA	Q	0.19	2
CTG	L	0.77	16	CCG	P	0.72	5	CAG	Q	0.81	6
ATT	I	0.34	1	ACT	T	0.29	2	AAT	N	0.17	1
ATC	I	0.66	11	ACC	T	0.54	11	AAC	N	0.83	12
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	12
ATG	M	1.00	6	ACG	T	0.13	1	AAG	K	0.22	8
GTT	V	0.40	13	GCT	A	0.28	0	GAT	D	0.46	4
GTC	V	0.14	1	GCC	A	0.16	0	GAC	D	0.54	14
GTA	V	0.20	1	GCA	A	0.24	1	GAA	E	0.75	10
GTG	V	0.27	2	GCG	A	0.32	8	GAG	E	0.25	6

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	5
10% - 14%	6
15% - 19%	4
20% - 24%	19
25% - 29%	10
30% - 34%	16
35% - 39%	17
40% - 44%	3
45% - 49%	4
>= 50%	155

Total Codons Used = 239

Tm Range	# of Overlaps
<58	0
58	0
59	0
60	19
61	10
62	0
63	0
64	0
65	0
>=66	0

Tm Range = 2.0

Ovrlap Len Range	# of Oligos
<18	4
18	2
19	2
20	4

21	8
22	1
23	3
24	0
25	2
26	0
27	1
>=28	0

Lowest Overlap = 16

Length Range	# of Oligos
<50	2
50-51	0
52-53	0
54-55	0
56-57	0
58-59	0
60-61	28
62-63	0
64-65	0
66-67	0
68-69	0
>=70	0

Longest = 60

There are 3 repeats greater than 8 nt:

```
PR Pos1 =   65   Pos2 =   65   Size = 18   Seq1 = GGATCGAGATCTCGATCC
                                           Seq2 = GGATCGAGATCTCGATCC
PR Pos1 =  167   Pos2 =  167   Size =  8   Seq1 = ACCATGGT
                                           Seq2 = ACCATGGT
DR Pos1 =  898   Pos2 =  901   Size = 16   Seq1 = TCATCATCATCATCAT
                                           Seq2 = TCATCATCATCATCAT
```

Sequence Patterns Screened (As Supplied By User)

None found

30 oligonucleotides need to be synthesized

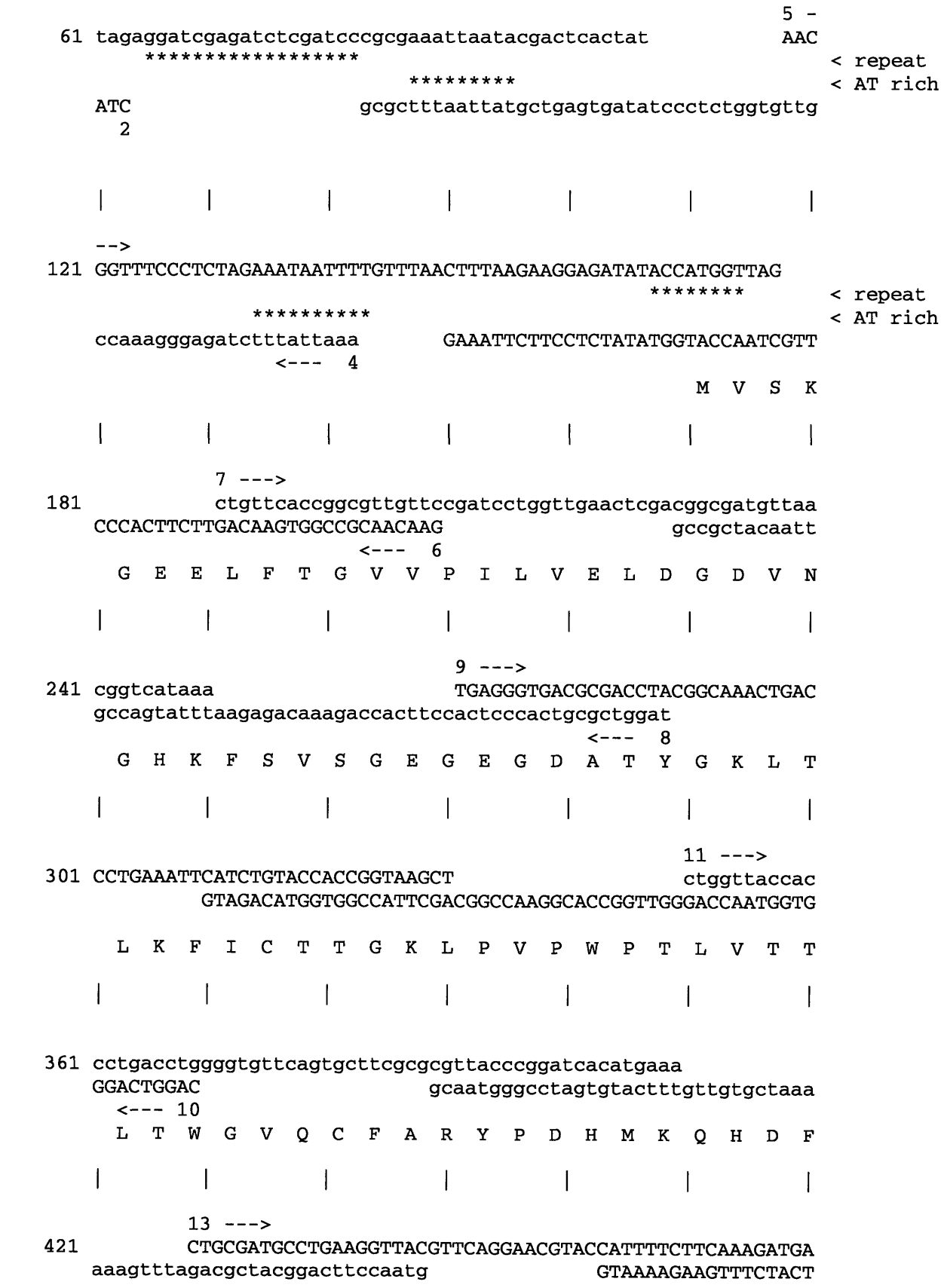
```
 1 CGGTCACGCTTGGGACTGCCATAGGCTGGCCCG  33
 2 GAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGGCCAGCCTATGGC  60
 3 GGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCA  60
 4 TTAAAGTTAAACAAAATTATTTCTAGAGGGAAACCGTTGTGGTCTCCCTATAGTGAGTCG  60
 5 TCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAGCAAGGGT  60
 6 CAACCAGGATCGGAACAACACCCGTGAAGAGTTCTTCACCCTTGCTAACCATGGTATATC  60
 7 GTGTTGTTCCGATCCTGGTTGAACTGGACGGTGACGTAAACGGTCATAAGTTCTCTGTTT  60
 8 CTTACCGTAAGTCGCGTCACCCTCACCTTCACCAGAAACAGAGAACTTATGACCGTTTAC  60
 9 GTGACGCGACTTACGGTAAGCTGACCCTGAAATTCATCTGCACCACCGGTAAACTGCCGG  60
10 GCAGACCGTAACCAAAGGTGGTAACCAGGGTAGGCCACGGAACCGGCAGTTTACCGGTGG  60
```

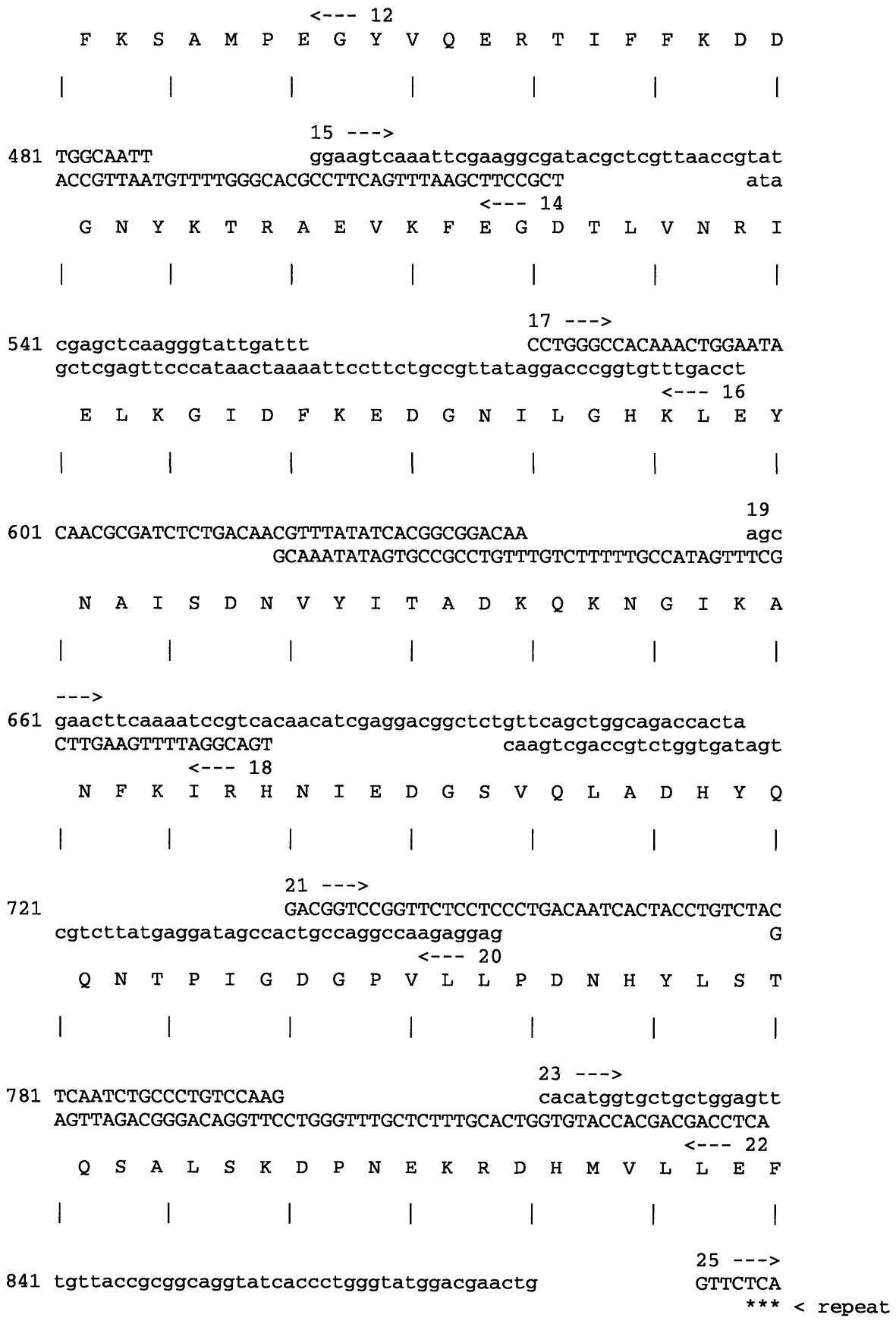
11 CACCTTTGGTTACGGTCTGCAGTGCTTCGCGCGCTACCCGGATCACATGAAACAGCACGA 60
12 TCCTGAACGTAACCCTCAGGCATCGCGCTCTTGAAGAAGTCGTGCTGTTTCATGTGATCC 60
13 CCTGAGGGTTACGTTCAGGAACGTACCATCTTTTTCAAAGACGATGGTAACTACAAAACC 60
14 ACGAGGGTATCGCCTTCGAACTTAACTTCCGCACGGGTTTTGTAGTTACCATCGTCTTTG 60
15 CGAAGGCGATACCCTCGTGAACCGTATCGAGCTCAAGGGCATCGATTTTAAGGAAGACGG 60
16 AATTATAGTTGTATTCCAGTTTGTGACCCAGAATGTTACCGTCTTCCTTAAAATCGATGC 60
17 GTCACAAACTGGAATACAACTATAATTCTCACAACGTTTATATCATGGCGGACAAGCAGA 60
18 TGTTGTGACGGATTTTGAAGTTCACTTTGATACCGTTTTTCTGCTTGTCCGCCATGATAT 60
19 GAACTTCAAAATCCGTCACAACATCGAGGACGGTAGCGTTCAGCTCGCGGACCATTATCA 60
20 CAGGCAGGAGGACCGGACCGTCGCCGATAGGGGTGTTTTGTTGATAATGGTCCGCGAGCT 60
21 TCCGGTCCTCCTGCCTGACAACCACTATCTGTCTTACCAGTCTGCGCTGTCTAAAGACCC 60
22 AACAAACTCCAGCAGAACCATGTGGTCACGTTTCTCGTTTGGGTCTTTAGACAGCGCAGA 60
23 ATGGTTCTGCTGGAGTTTGTTACTGCAGCGGGTATCACCCTGGGTATGGACGAACTGTAC 60
24 TTATTAATGATGATGATGATGATGAGAACCCCCCCCTTTGTACAGTTCGTCCATACCCAG 60
25 GTTCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTT 60
26 ACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCACTAGTAACGGCCGCCAGTGTG 60
27 GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTT 60
28 TCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCT 60
29 TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAGCGACTCCCACGGCACGTT 60
30 TCGAGCTTGCCAACGTGCCGTGGGAGTC 28

Cerulean

```
------------------------------------------------------------
                PARAMETERS FOR TRIAL 44
------------------------------------------------------------
            Total Size Of Gene ..... 1125 nt
            Protein Residues ....... 239
            Mutatable Residues ..... 232
            Fixed Nucleotides ...... 429 nt
            Oligo Size ............. 60 nt
            Annealing Temp ......... 61 +/- 1*C
            Oligo Concentration .... 2.00E-8 M
            Sodium Concentration ... 5.00E-2 M
            Mg2+ Concentration ..... 2.00E-3 M
            Codon Frequency Threshold .. 10%
            Repeat Threshold ....... 8 nt
            Mispriming Threshold ... 8/18 (6 exact) nt
```

```
The DNA sequence #  44 is:
------------------------------------------------------------
   1 CGGTCACGCTTGGGACTGCCATAGGCTGGCCCGGTGATGCCGGCCACGATGCGTCCGGCG
  61 TAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAAC
 121 GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAGCAA
 181 GGGTGAAGAACTGTTCACCGGCGTTGTTCCGATCCTGGTTGAACTCGACGGCGATGTTAA
 241 CGGTCATAAATTCTCTGTTTCTGGTGAAGGTGAGGGTGACGCGACCTACGGCAAACTGAC
 301 CCTGAAATTCATCTGTACCACCGGTAAGCTGCCGGTTCCGTGGCCAACCCTGGTTACCAC
 361 CCTGACCTGGGGTGTTCAGTGCTTCGCGCGTTACCCGGATCACATGAAACAACACGATTT
 421 TTTCAAATCTGCGATGCCTGAAGGTTACGTTCAGGAACGTACCATTTTCTTCAAAGATGA
 481 TGGCAATTACAAAACCCGTGCGGAAGTCAAATTCGAAGGCGATACGCTCGTTAACCGTAT
 541 CGAGCTCAAGGGTATTGATTTTAAGGAAGACGGCAATATCCTGGGCCACAAACTGGAATA
 601 CAACGCGATCTCTGACAACGTTTATATCACGGCGGACAAACAGAAAAACGGTATCAAAGC
 661 GAACTTCAAAATCCGTCACAACATCGAGGACGGCTCTGTTCAGCTGGCAGACCACTATCA
 721 GCAGAATACTCCTATCGGTGACGGTCCGGTTCTCCTCCCTGACAATCACTACCTGTCTAC
 781 TCAATCTGCCCTGTCCAAGGACCCAAACGAGAAACGTGACCACATGGTGCTGCTGGAGTT
 841 TGTTACCGCGGCAGGTATCACCCTGGGTATGGACGAACTGTACAAAGGGGGGGGTTCTCA
 901 TCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTG
 961 GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAA
1021 TAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGA
1081 GGAACTATATCCGGAGCGACTCCCACGGCACGTTGGCAAGCTCGA
------------------------------------------------------------

The oligonucleotide assembly is:
------------------------------------------------------------
    1       10        20        30        40        50        60
    |       |         |         |         |         |         |

    1 --->                                        3 --->
  1 CGGTCACGCTTGGGACTGCC                          cgatgcgtccggcg
      AGTGCGAACCCTGACGGTATCCGACCGGGCCACTACGGCCGGTGCTACGCAGGCCGC
                                                           <---

    |       |         |         |         |         |         |
```


5 -
61 tagaggatcgagatctcgatcccgcgaaattaatacgactcactat AAC
< repeat
< AT rich
ATC gcgctttaattatgctgagtgatatccctctggtgttg
2
-->
121 GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAG
< repeat
< AT rich
ccaaagggagatctttattaaa GAAATTCTTCCTCTATATGGTACCAATCGTT
<--- 4
M V S K
7 --->
181 ctgttcaccggcgttgttccgatcctggttgaactcgacggcgatgttaa
CCCACTTCTTGACAAGTGGCCGCAACAAG gccgctacaatt
<--- 6
G E E L F T G V V P I L V E L D G D V N
9 --->
241 cggtcataaa TGAGGGTGACGCGACCTACGGCAAACTGAC
gccagtatttaagagacaaagaccacttccactcccactgcgctggat
<--- 8
G H K F S V S G E G E G D A T Y G K L T
11 --->
301 CCTGAAATTCATCTGTACCACCGGTAAGCT ctggttaccac
GTAGACATGGTGGCCATTCGACGGCCAAGGCACCGGTTGGGACCAATGGTG
L K F I C T T G K L P V P W P T L V T T
361 cctgacctggggtgttcagtgcttcgcgcgttacccggatcacatgaaa
GGACTGGAC gcaatgggcctagtgtactttgttgtgctaaa
<--- 10
L T W G V Q C F A R Y P D H M K Q H D F
13 --->
421 CTGCGATGCCTGAAGGTTACGTTCAGGAACGTACCATTTTCTTCAAAGATGA
aaagtttagacgctacggacttccaatg GTAAAAGAAGTTTCTACT


12 <---
F K S A M P E G Y V Q E R T I F F K D D
15 --->
481 TGGCAATT ggaagtcaaattcgaaggcgatacgctcgttaaccgtat
ACCGTTAATGTTTTGGGCACGCCTTCAGTTTAAGCTTCCGCT ata
14 <---
G N Y K T R A E V K F E G D T L V N R I
17 --->
541 cgagctcaagggtattgattt CCTGGGCCACAAACTGGAATA
gctcgagttcccataactaaaattccttctgccgttataggacccggtgtttgacct
16 <---
E L K G I D F K E D G N I L G H K L E Y
19
601 CAACGCGATCTCTGACAACGTTTATATCACGGCGGACAA agc
GCAAATATAGTGCCGCCTGTTTGTCTTTTTGCCATAGTTTCG
N A I S D N V Y I T A D K Q K N G I K A
--->
661 gaacttcaaaatccgtcacaacatcgaggacggctctgttcagctggcagaccacta
CTTGAAGTTTTAGGCAGT caagtcgaccgtctggtgatagt
18 <---
N F K I R H N I E D G S V Q L A D H Y Q
21 --->
721 GACGGTCCGGTTCTCCTCCCTGACAATCACTACCTGTCTAC
cgtcttatgaggatagccactgccaggccaagaggag G
20 <---
Q N T P I G D G P V L L P D N H Y L S T
23 --->
781 TCAATCTGCCCTGTCCAAG cacatggtgctgctggagtt
AGTTAGACGGGACAGGTTCCTGGGTTTGCTCTTTGCACTGGTGTACCACGACGACCTCA
22 <---
Q S A L S K D P N E K R D H M V L L E F
25 --->
841 tgttaccgcggcaggtatcaccctgggtatggacgaactg GTTCTCA
*** < repeat

```
                                                  ********       < GC rich
                          gggacccatacctgcttgacatgtttcccccccaagagt

      V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K

     |        |        |        |        |        |        |


 901 TCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTT
     ****************                                        < repeat
                                                 ********   < GC rich
                   **********                                < AT rich
     agtagtagtagtagtaatta                GTGTGACCGCCGGCAATGATCAC
                  <--- 24

     |        |        |        |        |        |        |

                      27 --->
 961                  gcccgaaaggaagctgagttggctgctgccaccgctgagcaa
     CTAGGCCGACGATTGTTTCGGGCTTTCCTTCGACTCA                  tcgtt
                                    <--- 26

     |        |        |        |        |        |        |

                                       29 --->
1021 taactagcataacccctt                TCTTGAGGGGTTTTTTGCTGAAAGGA
     attgatcgtattggggaaccccggagatttgcccagaactccccaaaaaacgact
                                                  <--- 28

     |        |        |        |        |        |        |


1081 GGAACTATATCCGGAGCGACTCCCACGGCACGTT
                     CTGAGGGTGCCGTGCAACCGTTCGAGCT
                                      <--- 30

 ------------------------------------------------------------

 The total codon usage score ..........     0.000
 The total length score ...............     0.000
 The total melting temperature score ...    0.000
 The total repeat score ...............     1.493
 The total pattern score ..............     0.000
 The total mispriming score ...........     0.000
 The total AT content score ...........     1.138
 The total GC content score ...........     0.284
               The OVERALL score ......     2.916


                 DETAILED CODON FREQUENCY REPORT
  [ Codon, AA, Frequency, # of times used in coding sequence ]
 ------------------------------------------------------------
```

TTT	F	0.29	3	TCT	S	0.32	7	TAT	Y	0.35	2	TGT	C	0.39	1
TTC	F	0.71	9	TCC	S	0.27	1	TAC	Y	0.65	7	TGC	C	0.61	1
TTA	L	0.03	0	TCA	S	0.05	0	TAA	X	0.63	0	TGA	X	0.35	0
TTG	L	0.06	0	TCG	S	0.07	0	TAG	X	0.08	0	TGG	W	1.00	2
CTT	L	0.06	0	CCT	P	0.11	3	CAT	H	0.30	1	CGT	R	0.64	6
CTC	L	0.08	5	CCC	P	0.02	0	CAC	H	0.70	7	CGC	R	0.33	0
CTA	L	0.01	0	CCA	P	0.15	2	CAA	Q	0.19	2	CGA	R	0.01	0
CTG	L	0.77	16	CCG	P	0.72	5	CAG	Q	0.81	6	CGG	R	0.01	0
ATT	I	0.34	2	ACT	T	0.29	2	AAT	N	0.17	4	AGT	S	0.05	0
ATC	I	0.66	11	ACC	T	0.54	13	AAC	N	0.83	8	AGC	S	0.24	1
ATA	I	0.01	0	ACA	T	0.05	0	AAA	K	0.79	15	AGA	R	0.01	0
ATG	M	1.00	5	ACG	T	0.13	2	AAG	K	0.22	5	AGG	R	0.00	0
GTT	V	0.40	15	GCT	A	0.28	0	GAT	D	0.46	7	GGT	G	0.51	14
GTC	V	0.14	1	GCC	A	0.16	1	GAC	D	0.54	12	GGC	G	0.43	8
GTA	V	0.20	0	GCA	A	0.24	2	GAA	E	0.75	11	GGA	G	0.02	0
GTG	V	0.27	1	GCG	A	0.32	8	GAG	E	0.25	5	GGG	G	0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	5
10% - 14%	6
15% - 19%	9
20% - 24%	13
25% - 29%	8
30% - 34%	17
35% - 39%	18
40% - 44%	8
45% - 49%	7
>= 50%	148

Total Codons Used = 239

Tm Range	# of Overlaps
<58	0
58	0
59	0
60	14
61	15
62	0
63	0
64	0
65	0
>=66	0

Tm Range = 1.9

Ovrlap Len Range	# of Oligos
<17	1
17	3

18	3
19	3
20	5
21	6
22	1
23	1
24	2
25	1
26	1
>=27	1

Lowest Overlap = 16

Length Range	# of Oligos
<49	2
49-50	0
51-52	0
53-54	0
55-56	0
57-58	0
59-60	28
61-62	0
63-64	0
65-66	0
67-68	0
>=69	0

Longest = 60

There are 3 repeats greater than 8 nt:

```
PR Pos1 =   65   Pos2 =   65   Size = 18   Seq1 = GGATCGAGATCTCGATCC
                                           Seq2 = GGATCGAGATCTCGATCC
PR Pos1 =  167   Pos2 =  167   Size =  8   Seq1 = ACCATGGT
                                           Seq2 = ACCATGGT
DR Pos1 =  898   Pos2 =  901   Size = 16   Seq1 = TCATCATCATCATCAT
                                           Seq2 = TCATCATCATCATCAT
```

Sequence Patterns Screened (As Supplied By User)

None found

30 oligonucleotides need to be synthesized

```
 1 CGGTCACGCTTGGGACTGCC  20
 2 CTACGCCGGACGCATCGTGGCCGGCATCACCGGGCCAGCCTATGGCAGTCCCAAGCGTGA  60
 3 CGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTAT  60
 4 AAATTATTTCTAGAGGGAAACCGTTGTGGTCTCCCTATAGTGAGTCGTATTAATTTCGCG  60
 5 AACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTTAG  60
 6 GAACAACGCCGGTGAACAGTTCTTCACCCTTGCTAACCATGGTATATCTCCTTCTTAAAG  60
 7 CTGTTCACCGGCGTTGTTCCGATCCTGGTTGAACTCGACGGCGATGTTAACGGTCATAAA  60
 8 TAGGTCGCGTCACCCTCACCTTCACCAGAAACAGAGAATTTATGACCGTTAACATCGCCG  60
```

```
 9 TGAGGGTGACGCGACCTACGGCAAACTGACCCTGAAATTCATCTGTACCACCGGTAAGCT  60
10 CAGGTCAGGGTGGTAACCAGGGTTGGCCACGGAACCGGCAGCTTACCGGTGGTACAGATG  60
11 CTGGTTACCACCCTGACCTGGGGTGTTCAGTGCTTCGCGCGTTACCCGGATCACATGAAA  60
12 GTAACCTTCAGGCATCGCAGATTTGAAAAAATCGTGTTGTTTCATGTGATCCGGGTAACG  60
13 CTGCGATGCCTGAAGGTTACGTTCAGGAACGTACCATTTTCTTCAAAGATGATGGCAATT  60
14 TCGCCTTCGAATTTGACTTCCGCACGGGTTTTGTAATTGCCATCATCTTTGAAGAAAATG  60
15 GGAAGTCAAATTCGAAGGCGATACGCTCGTTAACCGTATCGAGCTCAAGGGTATTGATTT  60
16 TCCAGTTTGTGGCCCAGGATATTGCCGTCTTCCTTAAAATCAATACCCTTGAGCTCGATA  60
17 CCTGGGCCACAAACTGGAATACAACGCGATCTCTGACAACGTTTATATCACGGCGGACAA  60
18 TGACGGATTTTGAAGTTCGCTTTGATACCGTTTTTCTGTTTGTCCGCCGTGATATAAACG  60
19 AGCGAACTTCAAAATCCGTCACAACATCGAGGACGGCTCTGTTCAGCTGGCAGACCACTA  60
20 GAGGAGAACCGGACCGTCACCGATAGGAGTATTCTGCTGATAGTGGTCTGCCAGCTGAAC  60
21 GACGGTCCGGTTCTCCTCCCTGACAATCACTACCTGTCTACTCAATCTGCCCTGTCCAAG  60
22 ACTCCAGCAGCACCATGTGGTCACGTTTCTCGTTTGGGTCCTTGGACAGGGCAGATTGAG  60
23 CACATGGTGCTGCTGGAGTTTGTTACCGCGGCAGGTATCACCCTGGGTATGGACGAACTG  60
24 ATTAATGATGATGATGATGATGAGAACCCCCCCCTTTGTACAGTTCGTCCATACCCAGGG  60
25 GTTCTCATCATCATCATCATCATTAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTT  60
26 ACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCCACTAGTAACGGCCGCCAGTGTG  60
27 GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTT  60
28 TCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCT  60
29 TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAGCGACTCCCACGGCACGTT  60
30 TCGAGCTTGCCAACGTGCCGTGGGAGTC  28
```