Rapid Assembly of Standardized and Non-Standardized Biological Parts

Alexander Power

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of:

MSc. Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine

Faculty of Medicine

University of Ottawa

Abstract

A primary aim of Synthetic Biology is the design and implementation of biological systems that perform engineered functions. However, the assembly of double-stranded DA molecules is a major barrier to this progress, as it remains time consuming and laborious. Here I present three improved methods for DNA assembly. The first is based on, and makes use of, BioBricks. The second method relies on overlap-extension PCR to assemble non-standard parts. The third method improves upon overlap extension PCR by reducing the number of steps and the time it takes to assemble DNA, Finall, I show how the PCR-based assembly methods presented here can be used, in concert, with in vivo homologous recombination in yeast to assemble as many as 19 individual parts in one step. These methods will also be used to assemble an incoherent feedforward loop, gene regulatory network.

Acknowledgements

I would like to thank Dr. Mads Kaern for his support and mentorship throughout this project. Additionally, thanks are due to the 2011 iGEM team (Ian Roney, Zuhair Sayed, Wilson Lam, and Matt Orton) for their work with the BioBricks based Assembly method. Daniel Jedrysiak (Dan J) deserves special acknowledgement for his invaluable contributions towards the development of the BioBricks-based assembly method; it is hard to imagine this project coming to fruition without his conceptual contributions. I would also like to thank everyone in the Kearn lab for their friendship, in addition to their tolerance with my endless (and probably exhausting) appetite for debate. Finally, I would like to thank my partner Michelle and my daughter Sagan for being awesome and keeping me on track.

Funding for this project was provided by NSERC (Discovery Grant 313172-2011) and the Government of Ontario Ministry of Economic Development and Innovation (Early Research Award ER07-04-164)

Table of Contents

List of Tables

List of Figures

Chapter 1

Synthetic Biology

1.1 Introduction

A frequently articulated goal of Synthetic Biology is the design and construction of new biological parts and systems; such as, gene regulatory networks, biosynthetic pathways, and bio-sensors[1, 26, 64, 55]. The field was galvanized at the turn of the millennium by the publication of two short letters in the journal *Nature* [25, 30] describing the construction of two artificial gene regulatory networks (GRNs) in the bacterium *Escherichia coli*. The first GRN was a bi-stable genetic switch in which the mutual inhibition of two transcriptional repressors, either TetR and LacI, or TetR and λ-cI (Figure 1.1), enabled long-term epigenetic memory. This toggle switch allowed

downstream gene expression to be turned on or off in response to environmental stimuli (small molecule inducers). The second GRN used the same genetic components to encode an oscillatory network of repressors (repressilator), resulting in the periodic rise and fall in the expression of each transcription factor (Figure 1.2).

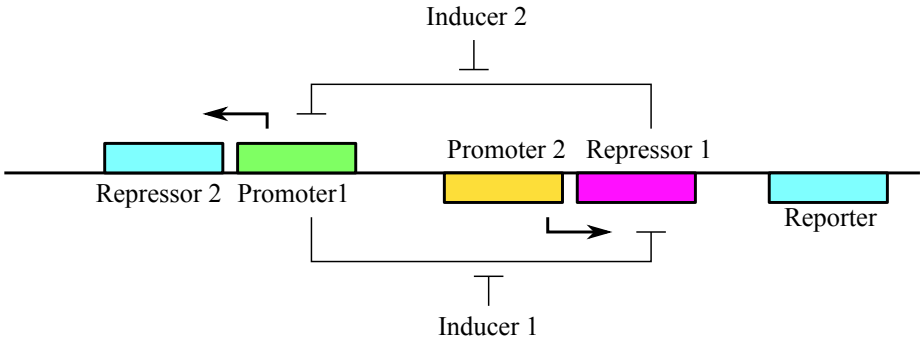

Figure 1.1: Repressor 1 represses expression from Promoter 1 which drives expression of Repressor 2. Repressor 2 is expressed from Promoter 1 and represses expression of Repressor 1 from Promoter 2. Repressors 1 & 2 can are inhibited by small molecule Inducers 1 and 2 respectively.

While the past decade has seen significant expansion of Synthetic Biology principles and applications, the first two *Nature* papers embody the most significant foundational principles. Firstly, the studies demonstrated that GRNs could be engineered to enable complex biological functions based on a forward design strategy and function could be accurately predicted using computational models. This model-aided design of synthetic GRNs has remained an active area of research, with notable

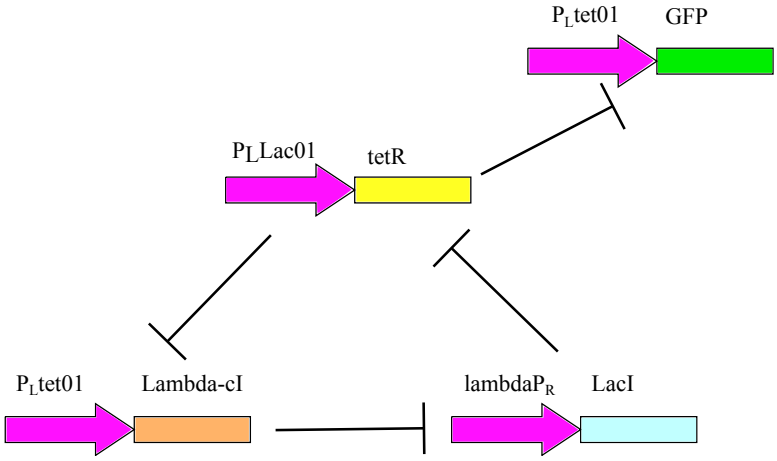

Figure 1.2: LacI expressed from λP_R represses tetR expressed from $P_L$lac01. TetR prepresses λcI expressed from $P_L$tet01, and λcI represses LacI. GFP levels expressed from $P_L$tet01 oscillate inversely with TetR levels.

successes[22, 28, 21]. Secondly, the studies demonstrated that very different biological functions could be encoded by GRNs that share the same, reusable "parts" but differ in the way that the parts are interconnected with one another. That is, altered network topology or *motif* results in altered function. The successful demonstration of these principles resulted in significant enthusiasm for the application of traditional engineering approaches to biology[55, 13, 23, 14].

The adoption of model-based design and standardized parts have accelerated the implementation of synthetic GRNs. An idealized Synthetic Biology Product Development Cycle (PDC) starts with a desired network function (e.g., memory or oscilla-

tions), then uses computer models to predict the most appropriate network architecture to achieve this function. Next, the network components (transcription factors, promoters, *etc.*) with the required functional characteristics are assembled into a prototype, and delivered into the desired organism. In general, each of the steps in the PDC represents a set of decisions about which technologies or design principles to use. These decisions, in turn, reflect an analysis of factors, including: cost, speed, reliability, intellectual property, and availability. An important task, therefore, is to shorten the PDC by making each step in the cycle faster and more robust, using low-cost reagents and accessible techniques.

While the implementation of the PDC model has resulted in a significant increase in the number of synthetic GRNs, each step remains laborious and fraught with pitfalls[22, 11]. The foremost obstacle for GRN development is a paucity of reliable methods to rapidly assemble multiple DNA fragments. This is especially true as DNA assembly impacts multiple steps in the PDC. In addition to making the assembly steps laborious and time-consuming, it also increases the cost of the parts characterization that is an absolute prerequisite for accurate model-based designs.

In the following sections I will review methods and approaches that have been developed to accelerate the assembly of multiple DNA molecules into larger constructs. Specifically, I will discuss general approaches and specific methods that bear directly on my thesis research. BioBrick-based DNA assembly methods, PCR-based DNA assembly, and homologous recombination of DNA in yeast will all be addressed. Each of these three approaches are used to deploy methods presented in this thesis.

1.2 Standardized Biological Parts

Synthetic Biologists recognized very early in the development of the field that a major barrier to advancement is the paucity of diverse genetic elements that can be "mixed-and-matched" to meet a particular set of design constraints. To date, the most successful non-commercial attempt at addressing this problem is the Registry of Standard Biological Parts (or simply the Registry). Based out of the Massachusetts Institute of Technology (M.I.T.) the Registry comprises a physical repository of biological parts, called BioBricks, that conform to standards aimed at facilitating sharing and idempotent composability. The Registry currently maintains and distributes over

7,000 BioBricks, and sequence information for more than 20,000 parts. The Registry has borrowed significantly from the open-source software initiative, in that it has made sequence and other part information, as well as detailed protocols, available free of charge.

The first of the BioBrick standards was designed by Tom Knight and named Assembly Standard 10 (AS-10)[41]. Collectively, the standards amount to a set of rules that specify sequence constraints. For example, AS-10 specifies that each BioBrick be framed by five endonuclease restriction sites (EcoRI, XbaI, SpeI, PstI and NotI), and that these sequences are not permitted to appear elsewhere in a part. Since 2003, four additional Assembly Standards have been introduced. Each new standard aims to address some constraint or shortcoming of previous standards[53, 3, 62, 29] (Figure 1.3), but can largely be viewed as variants of the original standard.

Standard	Description	Advantages	Disadvantages
BBF RFC 10	This is the original BioBrick assembly standard in additon to being the standard to which most parts in the Registry conform	Well tested and documented; large and still growing list of available parts	cannot be used directly for protein engineering due to 8bp, frameshift inducing scar
BBF RFC 20	An 8bp cut site overlapping PstI, otherwise identical to BBF RFC 10	Improved the ability for multiple enzymes to be used in single buffer. Unlike PstI SbfI has 100% activity in buffer #4(NEB)	The same disadvantages as BBF RFC 10
BBF RFC 23	Single nucleotides between restriction sites and parts removed from BBF RFC 10	Resulting 6 bp scar enables protein fusion	Care must be taken to prevent forming Dam-methylation seqeunces when assembling parts
BBF RFC 25	Addition of NgoMI restriction site downstream of XbaI in prefix sequence; addition of AgeI restriction site upstream fo SpeI in suffix sequence.	Enables protein fusion by maintaining frame of assembled parts; retains composability with BBF RFC 10; benign residues in scar	Requires more mutagenesis to generate parts; which is already labourious.
BBF RFC 21	Parts are prefixed by EcoRI-atg-BglI; suffixed by BamHI-taa-XhoI	Enables in-frame assembly; benign residue scar	BglII and BamHI cannot be heat-inactivated; not compatible with current 3A assembly method; incompatible with BBa format; incompatible with BioFusion Format

Figure 1.3: Comparison of currently existing BioBrick assembly standard. BBF RFC (BioBricks Foundation Request For Comments)

The rapid fusion and amplification of double stranded DNA molecules to form novel sequences is a cornerstone of Synthetic Biology, and a primary motivation for establishing the Registry. AS-10 was intended to facilitate DNA assembly by simplifying part design[41]. The 5'-end of each part is prefixed with the endonuclease restriction sites for EcoRI, NotI, and XbaI; while being suffixed on its 3'-end by the restriction sites for SpeI, NotI and PstI (Figure 1.4). Single bases were chosen to flank the XbaI

and SpeI restriction sites to prevent the formation of the GATC sequence following ligation with other parts. This was done to avoid the formation of Dam methylation sites in *E. coli*. Dam methylation of the adenine residue in the sequence GATC renders the XbaI restriction site resistant to cleavage by XbaI[51].

Figure 1.4: The assembly standard 10 part is flanked by the restriction sites EcoRI, NotI and XbaI on its 5' end and SpeI, NotI and PstI on its 3' end. Note the single nucleotide 'spacers' flanking the XbaI and SpeI restriction sites.

Idempotence is a central design feature of BioBricks. Simply put, the assembly of two BioBricks yields a new BioBrick (Figure 1.5); this allows for parts to be continually added to make increasingly complex products. For example, to assemble BioBricks A & B into the standard vector pSB1A3, A is digested with both EcoRI and SpeI, BioBrick B is digested with XbaI and PstI, while the backbone is digested with EcoRI and PstI. All three digested products are then mixed together and ligated. Ligation into the vector results in the regeneration of the BioBrick prefix and suffix. The junction between the two BioBricks contains a mixed XbaI/SpeI (TCTAGT) site that is not digestible by any of the BioBrick standard endonucleases. The result is a

new BioBrick that can be composed with any other AS-10 BioBrick.

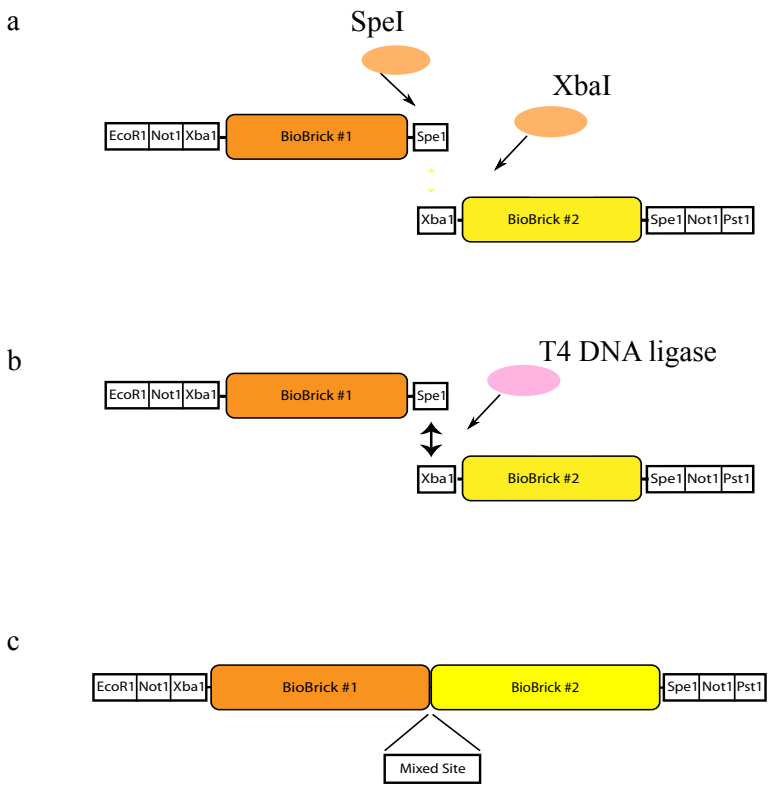

Figure 1.5: BioBrick Idempotence for AS-23: Row A; cleavage of BioBricks 1 & 2 by SpeI and XbaI respectively. Row B; the ligation of the complimentary overhangs generated by the previous digestion step. Row C; ligation of parts 1 & 2 regenerates the original BioBrick overhangs while leaving a digestion resistant 6 bp scar (ACTAGA).

The design of AS-10 is elegant in its simplicity, and most parts in the Registry adhere to this standard. However, AS-10 is not without its shortcomings. The nucleotides flanking the XbaI and SpeI sites results in an 8 bp ligation "scar" at the junction between assembled BioBricks. This scar introduces a frame-shift, and so, prevents AS-10 being employed for protein fusion. Assembly Standard 23 (AS-23)

was proposed as a solution to this problem and simply eliminates the nucleotides flanking the XbaI and SpeI recognition sequences, thus enabling in-frame translation of the fusion protein by reducing the scar to 6 bps[53].

In an effort to speed the assembly of BioBrick pairs, Shetty *et al.* [58] devised the Triple-Ligation (or 3A) cloning method. This method employs a positive/negative selection strategy to improve cloning efficiency (Figure 1.6). While triple ligation protocols have been described elsewhere[60], Shetty *et al.* improved on its implementation. Method 3A, relies on the simultaneous use of three standard plasmids (pSB1A3 (ampicillin^r), pSB1C3 (chloramphenicol^r) and pSB1T3 (tetracycline^r)) to clone two BioBricks together in one reaction. The plasmid pSB1A3 harbours the p1010 BioBrick, a ccdB Killer gene expression cassette which targets DNA gyrase in *E. coli*[9]. This cassette is lethal to most laboratory strains of *E. coli* and must be propagated in specialized strains, such as DB3.1 (Invitrogen Inc.). The BioBricks A & B are harboured in pSB1C3 and pSB1T3 respectively. Plasmid pSB1A3(p1010) is digested with EcoR1 and PstI, pSB1C3(A) is digested with EcoRI and SpeI, while pSB1T3(B) with XbaI and PstI. The three reaction products are subsequently pooled together

and assembled with T4-DNA ligase. The reaction product is used to transform a p1010 sensitive *E. coli* strain. The amp^r cassette in pSB1A3 provides positive selection for the assembled construct; whereas p1010 provides negative selection against reassembly of pSB1A3(p1010).

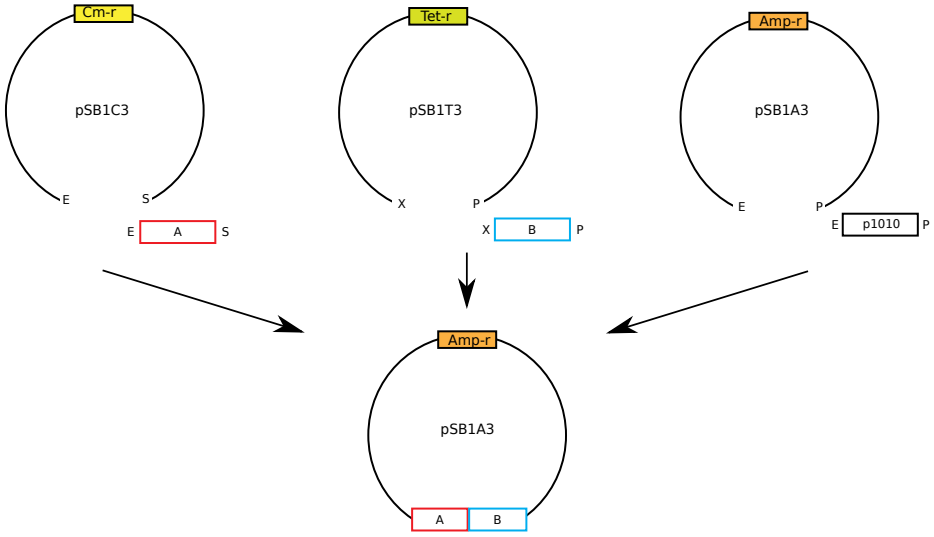

Figure 1.6: Method 3A combines positive and negative selection to improve cloning efficiency when assembling two BioBricks. Part A is digested with EcoRI & SpeI, part B is digested with XbaI & PstI, part p1010 is digested with EcoRI & PstI. When these three products are mixed and ligated, as shown, only the ligation product pSB1A3(A-B) can confer ampicillin resistance when used to transform *E. coli*. Tet-r denotes tetracycline resistance cassettes, Cm-r denotes chloramphenicol resistance cassette.

Method 3A enables the assembly of two BioBricks in two days, and significantly reduces the number of clones that need to be screened for errors. However, its reliance on endonucleases, as with all current ASs, invariably introduces a number of shortcomings. Foremost is the sequence constraints imposed by the inclusion of restriction

sites, necessitating frequent rounds of site directed mutagenesis before new parts can be used with this method. For example, the commonly used yeast tryptophan auxotrophic marker *TrpI*[61] contains three such *illegal* sequences that must be eliminated before it can be used as a standard BioBrick.

This problem becomes more and more severe as the length of the desired constructs increases. Additionally, incomplete digestion and errors introduced in the subsequent ligation lowers the success rate[65], imposing significant screening and replication of cloning attempts. Because the digestion/ligation-based methods are sequential, they consume significant resources when a large number of parts need to be assembled. Many of these problems can be circumvented by using assembly methods that are based on high fidelity polymerase chain reaction (PCR) or homologous recombination.

What's more, chromosomal instability in *E. coli* can lead to the deletion of tandem repeat sequences[10]. This can be extremely vexing when attempting to clone eukaryotic sequences in *E. coli* as repeats of this type are common in yeast terminator sequences.

1.3 PCR-based Assembly Methods

The fusion of DNA molecules by overlap extension PCR (oePCR) was first described by Horton *et al.*, in 1989 [38]. In this work, the authors assembled a 972 bp gene from four separate amplicons, encoding a fusion protein from two disparate mouse class-I Major Histocompatibility genes. The way this was done is illustrated in Figure 1.7. First, two introns from the first gene (H-2L), and two alpha helical regions from the second gene (H-2K) were amplified by PCR. The primers were designed such that the amplicons had complementarity for one another. The four fragments (AB, CD, EF, and FH) were then assembled to give two parts AD, and EH. Each of these dimers was assembled by mixing the requisite monomers together with terminal primers for amplification in a PCR reaction. During the annealing step, the 3'-ends of the complementary regions from each part was able to act as a primer. The resulting AD, and EH dimers were then amplified by the terminal primers. The successful assembly of the AD and EH dimers into the AH tetramer was performed in the same fashion.

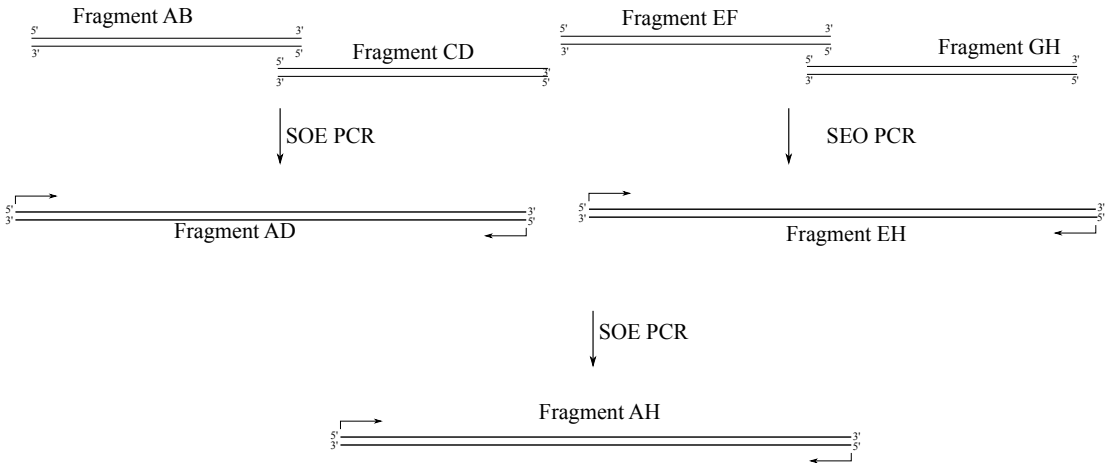

Figure 1.7: Overlap extension PCR performed by Horton *et al.*[38]. The first step sees the assembly of fragments AB & CD into fragment AD. In parallel, EF & GH are assembled into EH. The process is iterated in the final step to assemble AD & EH into the final molecule AH. Note: Horton *et al.* originally referred to this method as Splicing by Overlap Extension (SOE)

Strategies for the oePCR of more than two parts generally fall into two categories: 1) designing custom homologous regions with high binding specificity; such as high GC content linkers[27, 17] and 2) selecting regions within a sequence with convergent (usually high) annealing temperatures[45]. The first method requires that each part to be assembled is flanked by some standard sequence, precluding full freedom in sequence selection[27, 17]. Freedom in sequence selection is essential to our goals, making this approach inadequate. The second method involves selecting regions in a sequence that are best suited to being assembled with one another based on a set of annealing constraints. For example a 1000 bp sequence could be broken up into

8 parts (usually single-stranded oligonucleotides) whose complimentary overhangs all have a similar (usually high) annealing temperature. Design algorithms also try to avoid predicted secondary structures (such hairpin loops and G-C clamps) or spurious binding targets[45]. This second approach is well suited to DNA synthesis from overlapping oligonucleotides, as complimentary regions can be selected nearly arbitrarily; however, the broad constraints make its execution intractable.

In Chapter 2, I will demonstrate how an oePCR approach can be used in conjunction with AS-23 to rapidly assemble a gene expression cassette from six individual parts. In Chapter 3, I will further demonstrate a two-step procedure, wherein, oePCR is used to completely eliminate the need for restriction digest and ligation. Notably, I will show how introducing a *pre-cycling step* without primers can greatly increase efficiency. This was first demonstrated by Shevchuck et al. (2004)[59], who assembled a 10.8 kb construct from four initial parts in three PCR steps (see Figure 1.8). The initial steps in Shevchuck's method mirror those in the original Horton method. Four fragments (A, B, C and D) were amplified using primers that introduced homology to adjacent parts and split into two separate reactions (A-B, and C-D). To increase

efficiency, each pair is put through a heat cycle program without primers using an annealing temperature calculated for the complimentary sequence shared by the parts. Correspondingly, the polymerase uses the complementary overhang to initiate the synthesis of full-length fused dimers, which, can subsequently be assembled into the final A-D tetramer by iterating the process. Only after this step are terminal primers; i.e., forward A primer and reverse D primer, added to increase the yield of the final sequences.

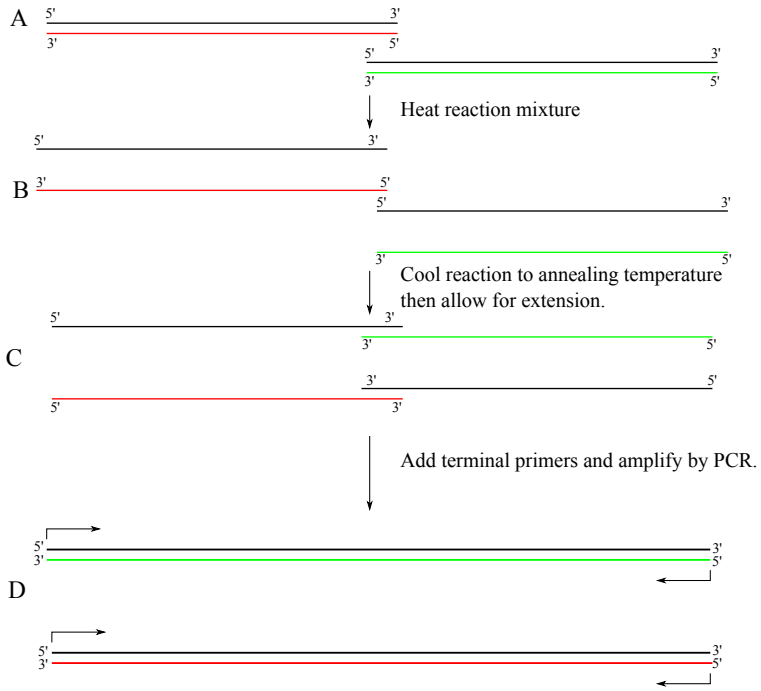

Figure 1.8: Pre-cycling method used by Shevchuck *et al.*[59] to improve assembly of long DNA fragments. Row A; overlapping parts are mixed together. Row B; double stranded molecules are melted in heat cycler (98°C). Row C; complimentary regions of adjacent sequences are allowed to anneal and serve to prime polymerization. Row D; assembled double-stranded sequences are amplified using terminal primers.

Recently, Gibson *et al.*[35] introduced a "chew-back" method to join multiple long single-stranded DNA molecules that have terminal homology. I mention this method because it represents a method that can be used to assemble very long molecules from multiple mid-length DNA molecules constructed using the PCR-based methods that I have developed.

In the Gibson assembly method (Figure 1.9) the DNA molecules to be assembled are pooled in the same reaction vessel with linearized destination vector. The termini of the linear backbone are homologous to the termini of the fully assembled construct. The key element of the method is that the reaction mixture contains T5 exonulease, which converts the double-stranded ends of each DNA fragment into single-stranded complementary overhangs. Complimentary overhangs are then able to anneal to one-another; and the single stranded gaps are filled by DNA polymerase.

The Gibson method has successfully been used to assemble multiple parts into a construct with a final length of 583 kb; in addition to a 5 kb construct from multiple parts with as little as 40 bp of terminal homology. However, the method is not well-suited for the assembly of short sequences (*i.e.*, sequences less than 300 bp in length)

as the action of the T5 exonuclease may result in the complete digestion of such fragments.

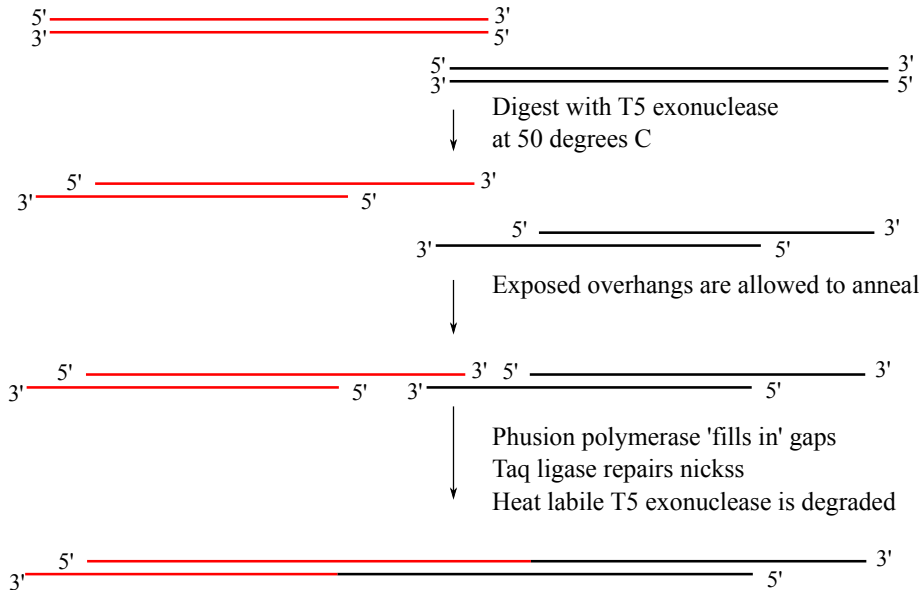

Figure 1.9: The Gibson Method [35] starts with overlapping molecules which are digested with a 5'-to-3' exonuclease. The resulting complimentary overhangs are allowed to anneal and single stranded gaps are filled by Phusion DNA polymerase. In the final step Taq ligase is used to repair single-strand nicks left between joined molecules.

1.4 Assembly by Homologous Recombination

The ability of *Saccharomyces cerevisiae* to efficiently integrate linear DNA molecules into its genome *via* homologous recombination has been known for some time, and has been used to make various and very extensive mutant libraries[32, 12]. In 1997, Larionov *et al.*[44] demonstrated that the high efficiency of homologous recombination

in yeast can also be exploited to assemble stable plasmids. This Transformation Associated Recombination (TAR) cloning, which is depicted in Figure 1.10, involves the co-transformation of yeast with linear DNA fragments containing a minimum of 60 bp homology with one another at their termini. Homology is shared with the terminal regions of a linearized plasmid. This plasmid backbone contains both a yeast origin of replication and a auxotrophic selection marker. The authors co-transformed yeast with the human BRCA2 gene and a linearized yeast artificial chromosome (YAC). One end of the YAC had 669 bp homology for the promoter region of BRCA2 and the other had 308 bp homology for the end of BRCA2 (Figure 1.10). The 97 kb long circular product was subsequently isolated and confirmed to be the YAC(BRCA2) construct. By 2003, Noskov *et al.*[50] had further refined TAR cloning to isolate and clone a 300 kb human genomic fragment.

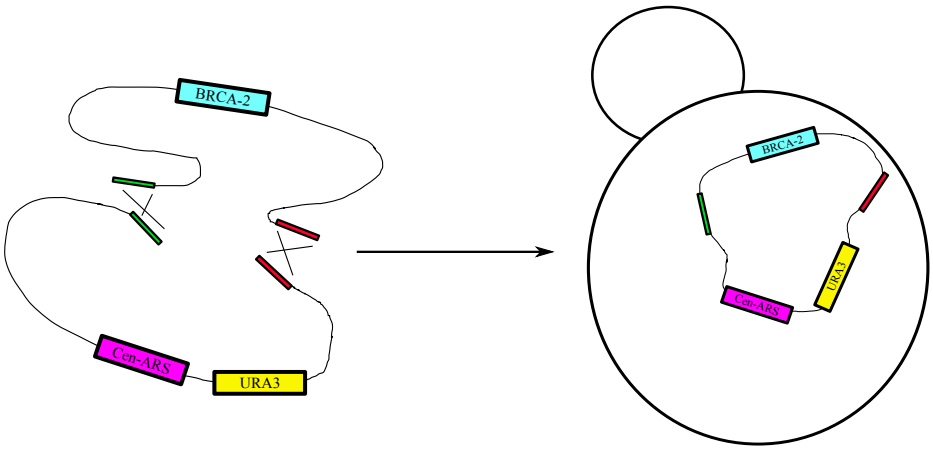

Figure 1.10: Transformation Associated Recombination (TAR) cloning of the human BRCA2 gene[44]. The linearlized YAC contains a Cen-ARS and Ura3 cassette. The green termini represent the 668 bp homologous BRCA2 promoter region and the red termini represent the 308 bp homology for the BRCA2 terminator region.

TAR cloning has also been used in Synthetic Biology applications. Two papers published in late 2008 showcased how yeast's prowess at homologous recombination could be used to assemble DNA molecules of prodigious size and complexity[57, 34]. One demonstrated the simultaneous construction of a synthetic D-xylose utilization pathway and synthetic zeanthin biosynthesis pathway[57]. The pathways were assembled by transforming yeast with the eight linear parts, each with at least 200 bp terminal homology to its neighbouring part. The fragments themselves ranged in length from roughly 2.1 kb to 3.4 kb, with the final assembled pathways being ∼19 kb in length. In the second paper, researchers at the J. Craig Venter Institute further pushed the

boundary of DNA assembly *via* homologous recombination by constructing a complete *Mycoplasma genitallium* genome in yeast[34]. This 590 kb genome was assembled by co-transforming yeast spheroplasts with 25 overlapping fragments ranging from 17 kb to 35 kb in length.

In Chapter 3, I will show how the PCR-based methods I have developed can be used to rapidly prepare DNA fragment for assembly by homologous recombination in *S. cerevisiae*.

Moreover, in Chapter 4, I demonstrate that a standard ssDNA/LiAc based yeast transformation protocol is sufficiently efficient to enable the simultaneous transformation of yeast with multiple fragment.

Due to the unambiguous benefits provided by the rapid amplification and assembly of DNA molecules, there has been a significant amount of research done in recent years towards improving or obviating traditional molecular cloning methods [48, 34, 50, 57, 33]. Currently available cloning technologies are diverse; however, most continue to rely on the conventional molecular cloning methods of digestion, ligation and transformation[22, 3, 31]. Despite this effort there remains a need for

rapid, reliable and affordable DNA assembly methods.

1.5 Research Objectives

The overarching goal of my thesis research is to implement improved methods for the rapid assembly of DNA parts into functional transcriptional devices in budding yeast. My research effort in this respect was motivated by three hypotheses:

1. Standardized biological parts can be assembled without time-consuming sequential ligation and digestion in a two-step process that combines existing BioBrick Assembly Standards and overlap extension PCR.

2. The requirement of standardization arising from the use of restriction endonucleases can be completely eliminated by relying entirely on PCR-based assembly.

3. Homologous recombination in yeast can be exploited to assemble constructs composed of a large number of non-standardized DNA fragments.

The Specific Aims of my research were to confirm these hypotheses. This work is described in Chapters 2, 3 and 4, respectively.

Chapter 2

A Method for Rapid *In Vitro* Assembly of BioBricks

2.1 Background

This chapter illustrates how BioBricks from the Registry of Standard Biological Parts can be assembled without the use of *E. coli* to propagate and amplify cloned DNA. We developed this method in response to inconsistent results we experienced using available BioBrick assembly protocols. Previously, we had adopted the BioBrick standard AS-23 in conjunction with the Triple-Ligation method (Figure 1.6)[58]. Our previous commitment to AS-23, coupled with the large number of available parts conforming to this standard informed the decision to develop an alternate assembly method that

preserved reverse compatibility with AS-23. Ultimately, we abandoned this method in favour of PCR based methods I subsequently developed; however, the method detailed in this chapter retains utility for laboratories committed to digestion/ligation based assembly methods and the BioBrick standards.

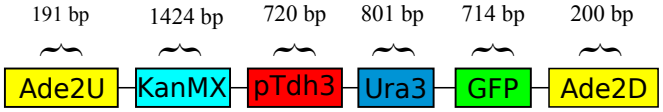

Figure 2.1: 4.05 kb Proof-of-principle construct with KanMX selection marker, pTdh3 constitutive promoter, Ura3-GFP fusion protein, and Ade2U & Ade2D. The Ade2U and Ade2D sequences target the construct for integration into the ADE2 locus. The Ade2D sequence also contains a 3'-UTR and terminator sequence for the Ura3-GFP CDS .

As a proof-of-principle, we designed a 4.05 kb construct comprised of six parts for integration into the yeast genome (Figure 2.4). This construct was built from individual parts conforming to AS-23: Ade2U, KanMX, pTdh3, Ura3 coding sequence (CDS), yEGFP, Ade2D. The terminal Ade2U and Ade2D BioBricks are homologous to the Ade2 promoter and terminator, respectively; KanMX is a G418 resistance cassette[12]; pTDH3 is a constitutive yeast promoter; Ura3CDS is the CDS of the URA3 gene minus the stop codon (TAA) obtained from the plasmid pRS406[61]; and yEGFP is a yeast-codon-optimized green fluorescent protein[24].

The terminal Ade2U and Ade2D BioBricks allow for the homologous recombination of the finished construct into the yeast genome. The long ∼200 bp regions were selected to confer high transformation efficiency[50], compared to the 40bp-60bp homology typically introduced with primers. This design also allows for the reuse of the terminal primers in the final assembly step. The Ade2 locus was chosen for integration so as to capitalize on the auxotrophic selection conferred by Ade2 disruption[66]. Yeast without a functional Ade2 gene are easily identified by their pink/red colonies, when grown aerobically on medium without adenine supplementation. Finally, the phenotypic effects of Ade2 auxotrophy is largely mitigated when growth medium is supplemented with adenine[66, 68].

The Ura3-GFP fusion protein was chosen both to show fusion-protein assembly, and for facile screening and confirmation. The BY4742 yeast strain[12] that was transformed with this construct does not have a functional Ura3 gene, allowing for functional screening on uracil deficient medium. Finally, yEGFP allows for flow-cytometric based colony screening.

2.2 Results

The first step in the assembly of this proof-of-principle construct was the amplification of individual parts. The six parts were each amplified with primers that added overhangs corresponding to BioBricks Assembly Standard 23 (Figure 1.4). Figure 2.2a shows the agarose-gel electrophoretic confirmation of the 6 amplicons used in this assembly. Following gel-electrophoretic confirmation each monomer was purified using gel-purification kits from Bio Basic Inc. (Markham, ON)

The addition of the BioBrick ends with primers required that we adopt a two step amplification program where the calculated primer annealing temperature changes over the course of the heat-cycler program to reflect the change in the available binding sequence. Specifically, a lower annealing temperature was used for the first 10 cycles to allow for the binding of the primer sequence minus the BioBrick overhang. After 10 cycles the remaining 15 cycles are carried out with an increased annealing temperature as the pool of molecules that possess the full length binding sequence come to include the BioBrick ends.

The next step was to prepare BioBrick pairs for assembly into heterodimers. Each

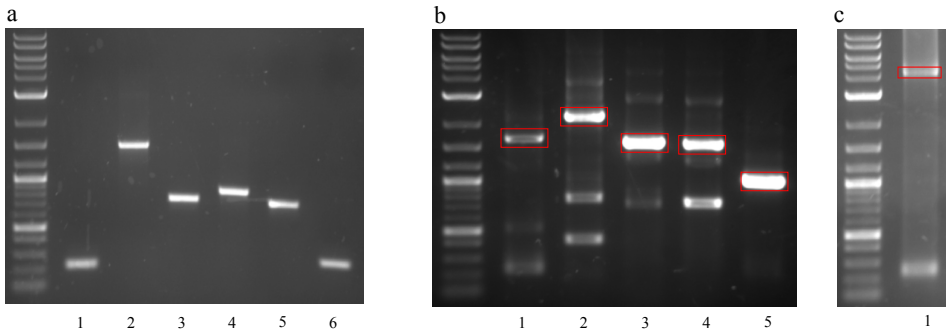

Figure 2.2: a) confirmation of monomer amplification on 1% agarose gel. Lanes 1-6 are Ade2U (191 bp), KanMX (1424 bp), pTdh3 (720 bp), Ura3 (801 bp), yEGFP (714 bp) and Ade2D (200bp) respectively. b) Confirmation of dimer amplification with 1% agarose gel, red boxes show correct bands. Lanes 1-5 are Ade2U-KanMX (1615 bp), KanMX-pTdh3 (2144 bp), pTdh3-Ura3 (1521 bp), Ura3-GFP (1515 bp), and GFP-Ade2D (914 bp). c) Confirmation of final assembly of 4.05 kb proof-of-principle construct on 0.5% agarose gel. All images use a 2 kb DNA ladder (NEB).

pair of neighbouring monomers (A & B) were digested with SpeI and XbaI respectively. Each digestion reaction was carried out by adding 10 ng of the amplified DNA to a 10 μL reaction containing bovine serum albumin (BSA), NEB buffer$^{\#}$4, 1 unit of either SpeI or XbaI and sterile double-distilled water. The reaction is incubated at 37°C (either in a water bath, heat cycler or incubator) for 20 min followed by a 20 minute incubation in a heat-cycler at 80°C, to inactivate the restriction enzyme. Digestion with SpeI and XbaI results in complimentary overhangs which form a mixed restriction site when ligated to one another. The resulting mixed site is resistant to subsequent digestion with either SpeI or XbaI, and forms the in-frame 6 bp scar characteristic of AS-23. This ligation step also results in the formation of the homodimers

A-A and B-B, which are cleavable by SpeI and XbaI respectively (Figure 2.3).

Following inactivation of the digestion reaction, 4 ng (4 μL from each of the digestion reactions) of BioBricks A and B are mixed together with 1 μL 10X NEB buffer $^{\#}4$, ATP (to 0.1mM), 0.25 ng T4-DNA ligase (NEB) and double distilled water to a final volume of 10 μL. This reaction mixture is then incubated at room temperature for 5 min, followed by a 20 min incubation at 80°C to deactivate the ligase.

As previously stated, the ligation step results not only in the desired A-B heterodimer but also in the B-B and A-A homodimers. Before PCR amplifying the ligation product, the reaction was subjected to another digestion reaction to remove any unwanted homodimers. This time both 1 unit of SpeI and XbaI each are used to cleave A-A and B-B respectively; where a unit is defined by the manufacturer (NEB) a the amount of enzyme required to digest 1 μg of Adenovirus-2 DNA in 1 hour at 37°C in a 50 μl reaction volume. The user is able to add both XbaI and SpeI directly to the ligation reaction mixture. This is made possible by the use of NEB Buffer $^{\#}4$ supplemented with 0.1 mM ATP, in place of the T4-DNA ligase buffer. This mixture is incubated at 37°C for 20 min followed by 20 min at 80°C to deactivate the endonu-

cleases. The digestion of the homodimers is important for the following PCR step as both the A-A and B-B dimers can serve as templates for amplification. In fact, A-A and B-B homodimers outnumber A-B in the reaction mix.

We made additional attempts to optimize this method; specifically, by attempting to reduce homodimer formation. In order for T4-DNA ligase to join cleaved molecules, a 5' phosphate must be present on the downstream residue of the molecule to be joined. We reasoned that removing the 5' phosphate from one of the monomer pairs before ligation would eliminate the formation of one of the homodimers and obviate the need for one of the restriction enzyme digests in the following steps. For example, treating BioBrick A with a phosphatase to remove the 5' phosphate group should prevent the formation of the A-A dimer during the ligation step. We implemented this regime by treating one of the two parts to be ligated with Antarctic phosphatase (NEB), prior to ligation. Unfortunately, we were unable to see any difference in performance in subsequent PCR steps. Further, this additional step that not only added time to the protocol and afforded more opportunities for handling errors.

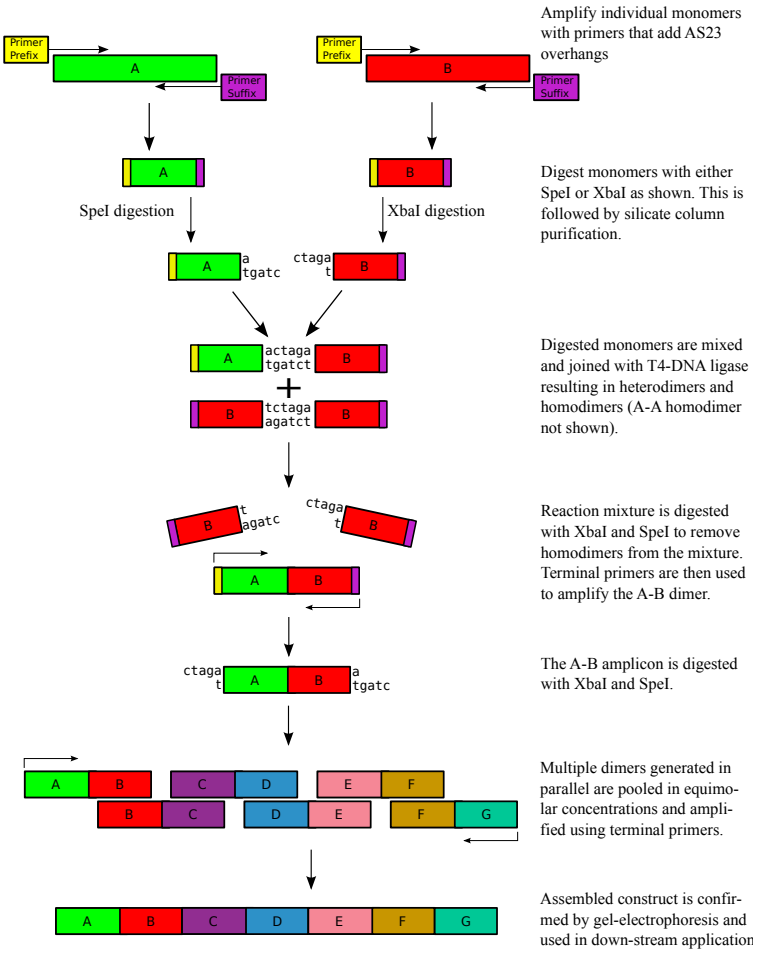

Figure 2.3: Step-by-step illustration of the assembly of standard biological parts used in this chapter.

Eliminating homodimers in the previous step allows the above reaction mixture to be used as a template for the PCR amplification of the A-B fragment. Figure 2.2b shows the results of this procedure for the amplification of the dimers Ade2U-KanMX6, KanMX6-pTdh3, pTdh3-Ura3, Ura3-GFP and GFP-Ade2D. Each of these

dimers constitutes a new BioBrick, with the BioBrick overhangs being re-constituted by the re-use of the primers used in the initial BioBrick amplification (Figure 2.3).

Each new BioBrick has homology to its neighbour corresponding to a complete BioBrick monomer (Figure 2.3); making them well suited to overlap-extension PCR. However, it is first necessary to remove the reconstituted BioBrick termini. This is done by an additional digestion step with SpeI and XbaI. Following this digestion, the BioBricks heterodimers are pooled and terminal primers corresponding to the Ade2U and Ade2D parts are used to assemble and amplify the final construct via PCR (Figure 2.3).

Figure 2.2c shows the amplified and assembled final construct. I used 100 ng of the final amplicon to transform 10^8 BY4742 yeast cells. Following a 4 hour recovery in YPD, 20μl ($\sim$ 2.0 x 10^6 cells) of the transformation mixture were plated directly on YPD-agar^{G418}. The KanMX6 cassette confers resistance to the eukaryotic antibiotic G418. Additionally, the terminal regions homologous to the Ade2 promoter and terminator targets the amplicon for integration into the Ade2 locus. Completely eliminating the Ade2 gene, yields red-colonies (when grown aerobically), and adenine

auxotrophy. The red colouration allows the user to ignore white colonies that may result from the integration of the KanMX6 cassette into undesired loci.

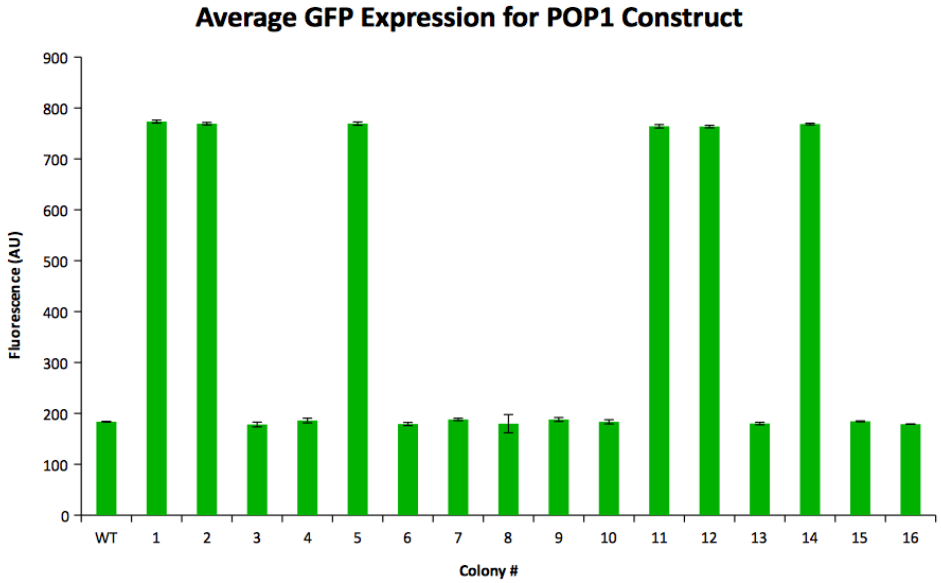

Figure 2.4: Flow cytometric confirmation of GFP fluorescence in proof-of-principle *vs* wildtype BY4742 yeast. Columns 1-16 represent average GFP fluorescence for individual colonies. Cells were interrogated using a 488 nm LASER and fluorescent emission was filtered through FITC filter set. Error bars indicate Standard Error or the Mean (S.E.M.). S.E.M was calculated as S.E.M. = $\frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated. A minimum of 10 000 cells were measured for each sample.

Following the two day incubation 31 red colonies were visible on the YPD-agar^{G418} plate. Sixteen red colonies were selected and re-plated on SM-$agar^{-ura}$. Since the strain being used (BY4742) lacks a functional Ura3 gene (resulting in uracil auxotrophy), only cells with a correctly assembled and integrated Ura3 gene should grow; all 16 colonies grew. In this construct, Ura3 forms a fusion protein with yEGFP; allowing for confirmation by GFP fluorescence. Only 6 of the 16 colonies screened

yielded GFP fluorescence (Figure 2.4). Sequencing data of the 16 colonies revealed a frame-shift mutation in the Ura3-GFP fusion protein. Interestingly, the offending mutation appeared immediately adjacent the SpeI-XbaI ligation site in all 10 colonies.

2.3 Discussion

This BioBrick assembly method introduces several improvements over previously existing methods. Foremost, this method is faster than standard BioBrick cloning methods. This improved speed of assembly arises from two features: the ability to parallelize assembly steps, and the obviation of *E. coli* as an intermediate organism for plasmid propagation and amplification.

The time-saving benefits of abandoning *E. coli* extend beyond merely not having to wait for colonies to grow; recombination-associated chromosomal instability in *E. coli* can generate errors and deletions in plasmids following transformation[10]. Specifically, mono-nucleotide repeats similar to those found in eukaryotic terminators are targeted for excision. This can result in a significant reduction in efficiency; as accounting for this effect can mean repeating cloning experiments and screening large

numbers of colonies.

The method we show here allows for the parallelization of dimer assembly, reducing the number of steps required to assemble complex molecules made up of many parts. This degree of parallelization is difficult to replicate with traditional cloning methods. Typically, parts must be assembled one at a time, with each part queued behind the part to which it will be assembled. Even when design features such as those used in the triple ligation method (Figure 1.6) allow for a binary-type assembly strategy, factors such as the aforementioned chromosomal instability slow progress.

To draw the distinction between the time required to clone the proof-of-principle construct in this chapter with standard cloning methods, a quick comparison is instructive. Whereas, the method presented in this chapter results in a transformation ready product in 1 day, the triple ligation method requires a minimum of 6 days to generate the same product. Again, in practice, we have found the triple-ligation method to be capricious and the 6 days estimate for the assembly of construct-1a is very optimistic. As with many of the serial cloning steps, a positive transformant is not found, thus, the transformation and DNA extraction must be repeated. Together

these steps require overnight . coli colony growth followed by a 12 hour growth period for DNA extraction.

Another potential benefit of this method is the specificity observed when assembling dimers into the final construct. While the assembly of monomers into dimers may results in multiple off-target amplicons (see non-specific bands in Figure 2.2b), the assembly of dimers into multimers seldom results in multiple non-specific bands(Figure 2.2c).

While this method is an improvement over standard BioBrick assembly standards, it remains laborious. The number of steps involved provides several opportunities for human error. Similarly, the nature of the steps make this method impractical to automate; specifically, the requirement to perform gel-extraction is problematic.

Even though significant effort was made to develop a method that was independent of *E. coli*, it should be noted that this method retains reverse compatibility with AS-23. For example, all of the parts submitted to the Registry by the UOttawa iGEM team were cloned into *E. coli* by digesting the final assembled product with EcoR1 and PstI and ligating it into the pSB1C3 plasmid.

2.4 Methods

2.4.1 Extraction of Yeast Genomic DNA

Individual yeast colonies were inoculated into 10 mL of YPD in a sterile 15 mL snap-top tube. Tubes were placed in a 30°C shaker at 200 RPM, overnight. The next day, overnight inoculate was transferred to a 50 mL Falcon tube and centrifuged for 2 min at 3,000 RCF. Supernatant was then removed and the pellet was resuspended in 500 μL sterile double-distilled H_2O (ddH_2O). The resuspended pellet was then transferred to a 1.5 micro-centrifuge tube and centrifuged at 10,000 RCF for 30 seconds and the supernatant removed. The pellet was then resuspended in 200μL breaking buffer; followed by the addition 300 mg of acid-washed glass beads and 200μL phenol:chloroform:isoamyl alcohol (25:24:1). This mixture was centrifuged for 5 minutes at 13,000 RCF and the aqueous phase was transferred to a second micro-centrifuge tube; the contents of the first tube were discarded. 430μL of TE buffer and 1μL of 32μg/μL of RNaseA were added to the aqueous solution and the mixture was incubated for 5 minutes at 37°C. Following incubation, 10μL ammonium acetate and 1 mL of 100% ethanol were added, and mixed by inversion. This mixture was centrifuged

for an additional 3 minutes at 13,000 RCF; followed by the removal of supernatant. The tube was then air-dried for 30 minutes followed by resuspension in 100μL TE buffer. This protocol typically yields between 200 ng/μL and 1000 ng/μL of genomic DNA.

2.4.2 PCR amplification of monomers

2.4.2.1 Primer Design Annealing Consideration

Table 2.1 lists all primers used for the assembly of POP-1. The monomers KanMX, pTdh3, Ura3 and yEGFP were all amplified using primers that add the sequences corresponding the BioBrick AS-23. In this example the final product was used for yeast transformation and thus did not need to be flanked by AS-23 sequences; therefore, for Ade2U the forward primer did not contain any BioBrick ends; while the reverse primer added the AS-23 suffix sequence. Likewise, the reverse primer for Ade2D did not add the AS-23 sequence; while the forward primer added the AS-23 prefix sequence.

A two-step PCR program was used to amplify all BioBricks. To calculate annealing temperatures for BioBrick primer sets the following procedure was used. The estimated melting temperature for the initial primer binding sequence (*i.e.*, excluding

the AS-23 suffix and prefix) was determined using the Integrated DNA Technologies online OligoAnalyzer 3.1. The numerical average for the melting temperatures was taken for primer pairs (forward and reverse) and 5°C was subtracted from this value. The resultant number was the annealing temperature used for the first 10 PCR cycles. The same procedure was performed for the full length primer sequence (*i.e.*, including the BioBrick overhangs) and this temperature was used for the annealing temperature for the remaining 16 cycles.

In the case of mixed primer sets (*e.g.*, Ade2U and Ade2D) the averaging procedure was done for the initial binding regions. This temperature was selected for all 26 PCR cycles.

Primer	**Primer Sequence**	$\mathbf{T}_m1[^\circ\mathbf{C}]$	$\mathbf{T}_m2[^\circ\mathbf{C}]$
Ade2U_FP	ttcgaaaagttgcctagtttc	51.2	51.2
BB_Ade2U_RP	aaggctgcagcggccgctactagttcaatcatgacgtaaga	48.3	68.7
BB_KanMX_FP	ccttgaattcgcggccgcttctagaagcttgcctcgtccccg	60.8	72.1
BB_KanMX_RP	aaggctgcagcggccgctactagttttgtaattaaaacttagattagattgc	49.3	66.4
BB_pTdh3_FP	ccttgaattcgcggccgcttctagatacgtaagggagttagaatc	48.2	66.7
BB_pTdh3_RP	aaggctgcagcggccgctactagttttgtttgtttatgtgtg	48.0	67.7
BB_Ura3_FP	ccttgaattcgcggccgcttctagacacagcttttcaattcaatt	47.2	67.1
BB_Ura3/stop_RP	aaggctgcagcggccgctactagtgttttgctggccgcatc	55.3	73.0
BB_GFP_FP	ccttgaattcgcggccgcttctagaatgtctaaaggtgaaga	48.4	66.7
BB_GFP/stop_RP	aaggctgcagcggccgctactagttttgtacaattcatccatac	45.3	67.6
BB_Ade2D_FP	ccttgaattcgcggccgcttctagataatatataagtttattg	48.7	61.6
Ade2D_RP	cctcggttctgcattgagcc	58.7	58.7

Table 2.1: Primers used to amplify BioBrick monomers, and for the amplification of dimers in preparation for assembly.

2.4.2.2 Parts Sourcing

The Ade2U, Ade2D, and pTdh3 were amplified from template derived from yeast genomic DNA. All Genomic DNA was extracted from the WT BY4742 yeast. The yEGFP BioBrick was amplified from the plasmid pRS4D1 courtesy the laboratory of James Collins at the Boston University; yEGFP does not contain a stop codon. The Ura3 CDS was amplified from the plasmid pRS416[61]; the Ura3 CDS was also amplified without a stop codon.

2.4.2.3 Amplification Protocol

Monomer template, primers, dNTPs and buffer were added to a thin-walled PCR tube in concentration stipulated in Table 2.3. When a plasmid was being used as template for the reaction, 0.1 ng/μL of DNA was used was used. Conversely, when genomically extracted DNA was used as template, 20-30 ng/μL of DNA was used. PhusionHF DNA polymerase was used for all reactions. Because of the proof-reading properties of this polymerase, it is always added to the reaction mixture last; just before being put in the thermal cycler. Finally extension time for plasmid based template assumed

a 15 second extension time per 1 kb of final amplicon length; this time was doubled for genomic DNA template. Table 2.2 shows the heat cycler program for amplifying monomers

PCR Reaction Mixture
5X HF buffer
DNA template (concentration varies by template source)
200 μM dNTPs
1 unit/50μl Phusion DNA polymerase
0.5 μM/primer
(ddH_2 to final volume)

Table 2.2: PCR mixture guidelines

Step	**Action**	**Time**
1	98°C	30 seconds
2	98°C	10 seconds
3	annealing temp1	15 seconds
4	72°C (extension temp.)	variable
5	return step 2 for 10 cycles	
6	98°C	10 seconds
7	annealing temp2	15 seconds
8	72°C (extention temp)	variable
9	return to step 6 for 15 cycles	
10	72°C	3 minutes
11	4°C 5 min	

Table 2.3: PCR program for the amplification of BioBrick monomers

2.4.3 Assembly and Purification of Dimers

To illustrate dimer assembly one dimer will be used as an example. Monomers (Ura3-CDS and GFP) were gel-extracted from a 1% agarose gel using Bio Basics Inc. gel extraction kit. Manufacture's instructions were followed. Ura3 and GFP were digested

with SpeI and XbaI respectively, followed by purification with Bio Basic Inc's PCR column purification kit. This step removes restriction enzymes and DNA molecules shorter than 60 bp in length. Next 1ng/μL of digested Ura3 and GFP are mixed together with 0.25 ng/μL T4-DNA ligase (NEB) in a 10 μL reaction and incubated at room temperature for 5 minutes; followed immediately by a 20 minute incubation at 80°C to inactivate T4-DNA-ligase. This mixture was then re-digested with XbaI and SpeI for 30 min., and again purified with a PCR purification kit. 5μ/L of this product is then used as template for PCR amplification. The primers BB_Ura3/stop_FP and BB_GFP/stop_RP and PCR program parameters were selected as per Table 2.3. The reaction product was extracted from a 1% agarose gel using a gel extraction kit and re-digested with XbaI and SpeI. A final purification step was performed on the reaction mixture with PCR purification kits.

2.4.4 PCR based-Assembly of Final Construct from Dimers

Dimers prepared as described in the previous section were pooled in a 50 μL reaction at a concentration of approximate 1 ng/μL. The terminal primers (Ade2U_FP & Ade2D_RP) were added as per Table 2.2.The program used to amplify and assemble

the parts is shown in Table 2.4. Assembly was confirmed on a 1% agarose gel.

Step	Action	Time
1	98°C	30 sec
2	98°C	10 sec
3	annealing temp	15 sec
4	72°C (extension temp.)	60 sec
5	return to step 2 for 26 cycles	
6	72°C	3 min
7	4°C	5 min

Table 2.4: Heat cycler program for the assembly and amplification of dimers into final constructs

2.4.5 Yeast Transformation

All yeast in this experiment were transformed using ssDNA/LiAc based transformation, as per Gietz, (2007)[36]. Following transformation, yeast were allowed to incubate for 6 hours in 1 mL YPD. Incubation was performed at 30°C, on a shaker at 200 RPM. After 6 hours, 100 μL of the media was plated directly onto YPD-agar+G418. The 6 hour incubation allows for the direct plating on G418 medium obviating the need for replica plating procedures.

2.4.6 Flow Cytometry, and Data Analysis

Flow cytometry data was analysed with Kaluza Flow Cytometry Analysis v1.2. A manual gate was used to select the WT cell on both forward scatter and side scatter data. This gate was then applied to all experimental data. Cells were interrogated

using a 405 nm and a 488 nm laser. yEBFP was excited by the 405 nm laser, and fluorescence was detected with the FL6 detector through a 485 DLP filter. yEGFP was excited by the 488 nm laser and fluorescence was detected by the FL1 detector through a FITC/GFP filter. Figures for flow cytometry were made using Microsoft Excel for Mac 2008.

Colonies selected for flow cytometric analysis were inoculated overnight in SM-complete and grown to stationary phase. Four hours before analysis, 3 mL of SM-complete was inoculated with the overnight inoculate at an O.D. 600 of 0.04. This was grown at 30°C at 200 RPM on an orbital shaker. After four hours, cells were diluted 1:10 with ddH_2O, and analysed immediately by flow cytometry. All flow cytometry was performed using a Beckman Coulter Cyan flow cytometer.

Figures displaying flow-cytometry data were generated using Microsoft Excel for Mac:2008.

Chapter 3

PCR-based Assembly Methods

3.1 Background

The method developed in Chapter 2 is an improved method of DNA assembly, when compared to the sequential approach used to assemble BioBricks; however, it can still be significantly improved upon. Notably, the use of endonucleases introduces undesirable "scars" at the junctions between parts, and enforces sequence constraints that may require extensive modifications of individual parts before they can be used. To address these problems, and to further reduce the time, cost, and number of steps required, I examined the possibility of using overlap extension PCR (oePCR) instead of digestion and ligation.

For exploratory and proof-of-principle experiments, I designed a construct (Figure 3.4) similar to that used in Chapter 2, with a Trp1 CDS instead of a Ura3 CDS. This CDS contains an "illegal" restriction site (XbaI) rendering it incompatible with BioBrick assembly standards. In the same way that uracil was used to demonstrate functionality of the assembly in Chapter 2, tryptophan can be used to demonstrate functionality of the second POP construct as the expression of the Trp1 CDS confers prototrophy to tryptophan.

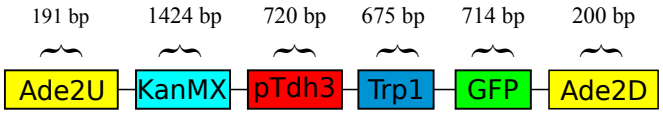

Figure 3.1: Proof-of-principle construct with TrpI CDS in place of Ura3. This 3.9 kb construct is otherwise identical to the proof-of-principle construct in Figure 2.1

I successfully implemented two assembly methods based on overlap extension PCR. The results obtained with these methods are presented in Section 3.2.1 and Section 3.2.2, respectively. The first variant is a direct analogue of the method described in Chapter 2; oePCR is used to fuse monomers into overlapping dimers, which are then assembled into the final construct by a second round of PCR. The second variant relies on a modified reaction solution coupled with a modified pre-cycling protocol

adapted from Shevchuck *et al.*[59] to fuse monomers into multimers instead of dimers. Although the dimer-based method proved to be useful for the assembly of constructs containing between four and six monomers, we found in exploratory experiments that the efficiency of the dimer assembly rapidly decreases as the number of heterodimers increases (Appendix A). This limitation could be addressed by implementing an iterative PCR procedure (see e.g,. [38, 59]). However, errors introduced by PCR accumulate in a non-linear fashion with the number of cycles[54, 45, 5, 15] making a strategy involving a high number of PCR rounds undesirable from a quality control perspective. Therefore, I turned my attention to increasing the number of monomers that could be assembled in a single PCR reaction. Iterative PCR assembly is also limited by the size limitations imposed by the polymerase processivity. An alternative, non-PCR method to assemble multimers into larger constructs is presented in Chapter 4.

3.2 Results

3.2.1 Dimer-based Assembly Method

The dimer-based PCR assembly method starts with the amplification of monomers using primers designed to introduce 25-30 bp of homology to the 5-end of each part (Figure 3.2) for the 3'-end of the neighbouring part it will be linked to in the final construct. For example, to join the two parts Ade2U and KanMX6 from the proof-of-principle construct, each of these parts was amplified as shown in Figure 3.2. Following the initial amplification, each monomer is purified and then mixed together with a set of terminal primers to create dimers. This can be performed in parallel and results in the creation of heterodimers spanning the full length of the desired assembly (Figure 3.2).

The six monomeric parts of the proof-of-principle construct were amplified and confirmed by agarose-gel electrophoresis (Figure 3.3a) as per Methods 3.4.3.1. Following gel confirmation of monomer amplicons, each amplicon was purified using silicate column purification kits. The PCR-amplified monomers were assembled into dimers using oePCR in reaction mixtures containing 5% v/v dimethyl sulphoxide (DMSO).

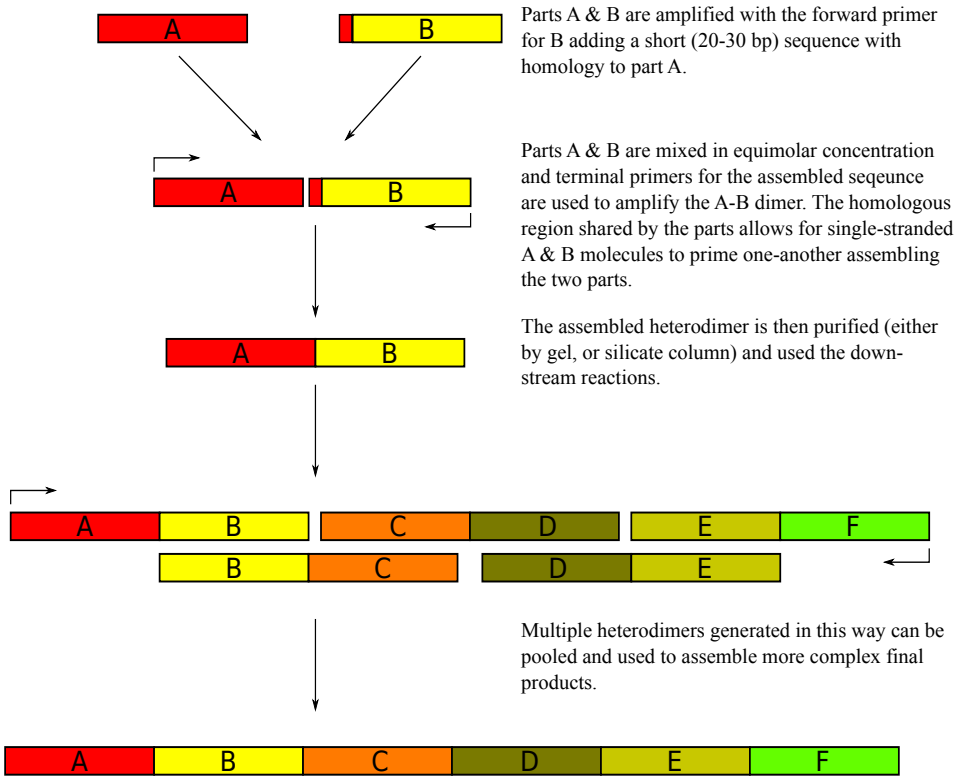

Figure 3.2: Step-by-step assembly of dimers and multimers using oePCR as described in this chapter.

Monomer pairs were pooled at approximately 1ng/μl(length of part in kb) each in a PCR reaction tube. For example, in the case of a 1 000 bp part 1ng/μl is used in the reaction mixture; conversely, for a 200 bp part 0.2 ng/μl is added. This template concentration was chosen as it is similar to previously described concentrations[59]. It is likely that other concentrations would also work with this protocol, however we did not pursue this line of experiments.

During initial experiments, I found that simply pooling monomers with short (25-30 bp) homology at their terminal ends did not consistently yield the correctly sized

bands, presumably because homologous regions introduced by the primers showed a significant disparity in estimated melting temperatures with terminal primers used to amplify the assembled dimer. Dimethyl sulphoxide (DMSO) has been previously shown to permit the use of lower annealing temperatures while improving amplification specificity by preventing secondary structure formation, and increasing binding stringency in single stranded DNA molecules[19, 39]. Moreover, the addition of 5% v/v DMSO to a PCR reaction mixture was not shown to adversely affect the reaction, while concentrations greater than 10% increased error rates[2]. I performed several experiments to evaluate the utility of adding DMSO to PCR reactions (see Appendix A).

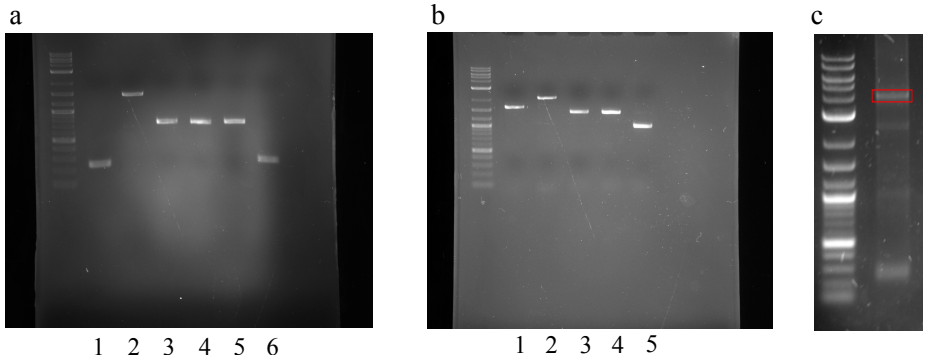

Figure 3.3: a) PCR amplicons of monomers for proof-of-principle construct on 1% agarose gel. Lanes 1-6 correspond to Ade2U, KanMX, pTdh3, TrpI, yEGFP, and Ade2D respectively. b) Heterodimers assembled by oePCR visualized on a 1% agarose gel. Lanes 1-5 correspond to Ade2d-KanMX, KanMX-pTdh3, pTdh3-TrpI, TrpI-GFP and GFP-Ade2D respectively. c) Fully assembled, ∼3.9 kb construct (highlighted in red box), 0.5% agarose gel..

The dimers obtained in PCR reactions where 5% DMSO was added were visualized by agarose-gel electrophoresis (Figure 3.3b). The lack of non-specific amplification in the dimer assembly stands contrast to the digestion/ligation method from Chapter 2, which produced multiple undesired bands (see Figure 2.2b). Because of the lack of non-specific amplification, dimers were purified with silicate-column kits rather than the more involved gel extraction and purification required for the digestion/ligation method.

The full-length proof-of-principle construct was assembled as described in Chapter 2 using a PCR reaction mixture containing overlapping dimers at equimolar concentrations ($\sim$1ng/μl*(length each dimer in kb)) and the terminal primers of the most upstream and downstream monomers (Figure 3.2). The assembled final construct was then confirmed by agarose gel electrophoresis (Figure 3.3c). Next, 100 ng of the silicate column purified PCR product was used to transform 10^8 BY4742 yeast cells. After a 4 hour recovery period ($\sim$2.0 x 10^6) were plated on YPD-agar^{G418}, and incubated for 48 hours. From this plate, 15 red colonies were selected for flow cytometry analysis, to evaluate GFP fluorescences. Each of the 15 selected colonies were first

re-streaked on YPD-agar^{G418} plates and grown for 24 hrs in an incubator at 30°C. Individual colonies from the 15 re-streaked colonies were used to inoculate 3 ml of SM complete in preparation for flow cytometry, as detailed in Methods (Section 2.4.6). Thirteen of the fifteen selected colonies were positive for GFP expression (Figure 3.4).

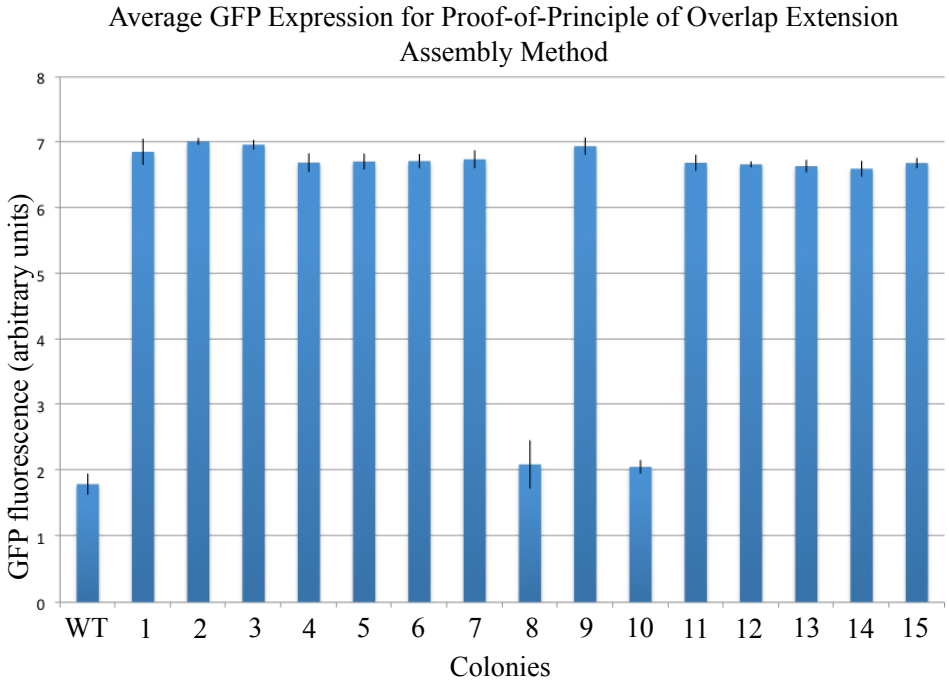

Figure 3.4: Flow cytometric confirmation showing mean GFP fluorescence of POP2 construct *vs* wild-type BY4742 yeast. Cells were interrogated using a 488 nm LASER and fluorescent emission was filtered through FITC filter set. Colonies 1-15 were grown overnight to stationary phase in liquid SM and used to inoculate 1 ml of SM at an OD 600 of 0.04. Following 4 hours of growth of an orbital shaker at 30°C, cells were diluted 1:10 and measured immediately. Error bars indicate Standard Error or the Mean (S.E.M.). S.E.M was calculated as S.E.M. $= \frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated. A minimum of 5 000 cells were measured for each sample.

3.2.2 Single-step Assembly Method

The single-step PCR assembly method involves the pooling of monomers carrying short regions of homology between neighbouring pairs at their termini, and eliminates the multiple PCR reactions used in the dimer-based method for the pairwise fusion of monomers. Although the dimer-based method is effective, the lower number of separate PCR reactions and the lower overall number of PCR amplifications makes the single-step method attractive from both a cost-reduction and a quality control perspective. PCR induced errors can arise from the action of the polymerase[20, 67], incorrect primer binding (*e.g.* primer slippage)[20, 45], and hydrolysis as a result of the high temperatures experienced in the reaction[54].

Preliminary experiments (see Appendix B) showed that using standard PCR reaction conditions is not a robust general approach to fuse multiple monomers in a single step (see appendix). I found that two modifications resulted in greater success. The first, is to minimize the impact of disparity in annealing temperatures and reduce the formation of secondary structures by including 5% DMSO in the PCR reaction mixture. The second, is to run several amplification cycles before adding terminal

primers. Previously, Shevchuk *et al.*[59] demonstrated that such a pre-cycling regime can significantly improve the yield of fusing two parts ∼10 kb long. I hypothesized that a similar approach might also improve the single-step assembly of multiple, much shorter monomers.

To test the single-step assembly method, I amplified six monomers with primers that establish 25-30bp regions of homology between neighbouring pairs, as described for the dimer-based method. These monomers were mixed in roughly equal concentrations, and added to a PCR reaction mixture containing 5% DMSO v/v. Figure 3.5 illustrates the step-by-step procedure for assembling multiple monomers in one reaction.

Parameter selection for the pre-cycling program was informed by several factors. Whereas Shevchuk *et al.*[59] used extension times needed to amplify only the longer of the two parts being assembled, I chose an extension time corresponding to the amplification of the full length construct. I also used 10 pre-cycles instead of the 20 cycles used by Shevchuk *et al.* I did this because I wanted to reduce the total number of amplification cycles, in order to reduce the opportunity for the introduction of errors. Moreover, the annealing temperature was chosen by selecting an annealing

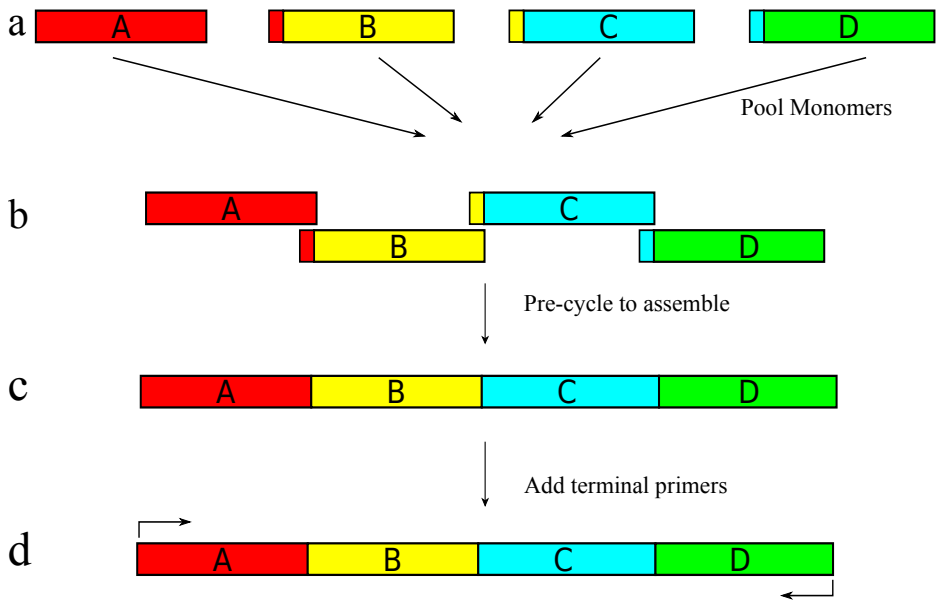

Figure 3.5: One-step assembly of multiple monomers into a polymer. a) Monomers are amplified with short (25-30 bp) overlaps to their neighbours. b) Pooled monomers are *pre-cycled* without primers to assemble the final construct. d) Terminal primers for the final construct are added and an additional 25 PCR cycles are performed.

temperature 2°C below the lowest T_m estimated for each of the homologous regions obtained using the online calculator provided by Integrated DNA Technologies. As outlined in the previous section, the addition of DMSO to the reaction mixture enables the selection of a low annealing temperature in the face of disparate estimated annealing temperatures presented by multiple overlapping sequences. The full-length construct was amplified by adding terminal primers to the reaction tube immediately after the pre-cycling was completed, and subjecting the mixture to an additional 25 cycles. Appendix B illustrates a set of experiments showing the use of the pre-cycling

protocol to improve the performance of PCR based assembly.

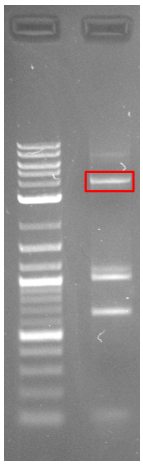

Figure 3.6: Results for the one-step assembly of 6 piece proof-of-principle construct. The red box highlights the 3.9 kb band corresponding to the correctly assembled product. Gel-electrophoresis was performed on a 1% agarose gel with a 2.0 kb ladder (NEB)

Figure 3.6 shows the agarose-gel electrophoretic image for the final assembly reaction. The ∼3.9kb band corresponding to the correctly assembled final construct (Figure 3.4) is visible, in addition to several smaller bands. Despite these contaminations, 100 ng of the column purified PCR reaction mixture was used directly to transform competent BY4742 yeast cells. A total of 15 red colonies were selected form after incubation at 30°C on G418 selective medium for two day. All clones were positive for GFP fluorescence (see Figure 3.7).

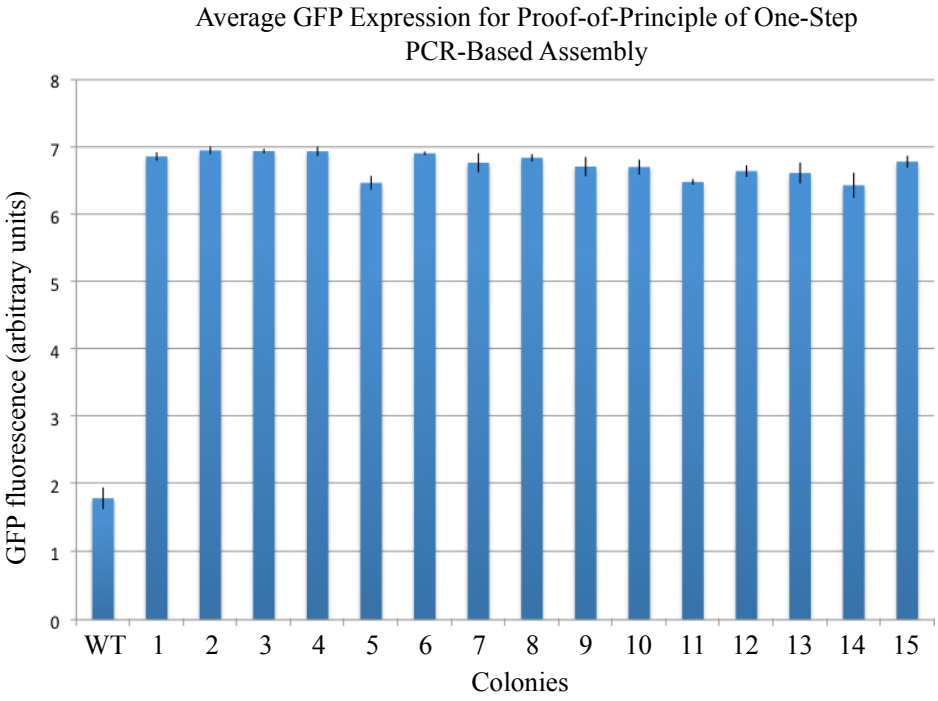

Figure 3.7: Flow cytometric confirmation of GFP fluorescence in proof-of-principle *vs* wildtype BY4742 yeast. Colonies 1-15 were assembled using the one-step PCR method. Error bars indicate Standard Error or the Mean (S.E.M.). S.E.M was calculated as S.E.M. $= \frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated. A minimum of 5 000 cells were measured for each sample.

3.3 Discussion

PCR-based assembly methods have a number of advantages over those using restriction endonucleas and ligation. Notably, PCR assembly allows for a further reduction in the production time and cost due to protocols with fewer and simpler steps. In this Chapter, I have demonstrated that both the dimer-based and the single-step methods can be used to assemble constructs ~4kb in length from up to six parts in a single day

without imposing sequence constraints in the form of *illegal* sites or introducing ligation "scars". It is noted, however, that the methods require the design and purchase of assembly-specific PCR primers, and thus introduces some additional costs.

The choice of which PCR-based assembly method to use depends on the specific application. Compared to the single-step method, the dimer-based method requires more amplifications cycles, which increases the likelihood of errors, and more reaction purifications and reagents, which drives up the overall assembly cost. The drawback of the single-step assembly is an apparent loss of specificity; evidenced by the off target bands visible in Figure 3.6, and the difficulty in designing primers that have homologous regions with sufficiently convergent T_m's. This problem, which can be addressed using temperature gradients in the dimer-based method, increases with the number of initial parts and the single-step assembly will not always work.

3.4 Methods

3.4.1 Strains, Transformations, Flow-Cytometry

BY4742-MATα (ATCC 201389) was used for all yeast experiments in this chapter. For transformation, flow-cytometry, DNA purification, and data analysis protocols see Section 2.4.

3.4.2 Primer Design and Monomer Sourcing

Table 3.1 shows all primers used for experiments in Chapter 3. Primers were designed in pairs for each monomer, with forward primers having complimentary to the reverse primer for the part it will be assembled with. Restated, the amplified monomer will have homology for its neighbour on its 5'-end. Conversely, the reverse primer only contains the sequence allowing it to bind to the sequence it is intended to amplify. Whenever possible, the T_m's for the forward primer and the initial binding region of the reverse primer were designed to be within 5°C of each other. This design results in primers with disparate T_m's once the initial template contains the homologous region added by the reverse primer. To mitigate this, 5% DMSO was added to all

reactions. The homologous regions added by the reverse primers were also designed to have convergent annealing temperatures with overhangs that would participate in the same reaction, when possible.

TrpI was amplified from the pRS404 plasmid[61], while the remaining monomers were sourced as in Chapter 2.

Primer	**Primer Sequence**	$\mathbf{T}_m1[^\circ\mathbf{C}]$	$\mathbf{T}_m2[^\circ\mathbf{C}]$
Ade2U_FP	ttcgaaaagttgcctagtttc	51.2	51.2
BM_Ade2U_RP	tcaatcatgacgtaagaaatg	48.3	68.7
BM_KanMX-A2Link_FP	catttcttacgtcatgattgacttgcctcgtccccgccgggtcac	60.8	72.1
BM_KanMX_RP	ctggatggcggcgttagtatcgaatcg	49.3	66.4
BM_pTdh3-tTefLink_FP	tactaacgccgccatccagtactagatacgtaagggagttagaatc	48.2	66.7
BM_pTdh3_RP	cgtgtgtttattcgaaactaagttcttg	48.0	67.7
BM_pTdh3-TrpLink_FP	aataaacacacataaacaaacaaaactagaatgtctgttattaatttcac	55.3	73.0
BM_GFP-TrpLink_RP	gaataattcttcacctttagacattctagttttcttagcatttttgacga	48.4	66.7
BM_GFP_FP	atgtctaaaggtgaagaattattc	45.3	67.6
BM_GFP_RP	gttttgtacaattcatccatac	45.3	67.6
BM_GFP-Ade2D-Link_FP	gtatggatgaattgtacaaaactaatatataagtttattgatatacttg	48.7	61.6
Ade2D_RP	cctcggttctgcattgagcc	58.7	58.7

Table 3.1: Primers used for the amplification of monomers

3.4.2.1 Amplification on Monomers

PCR Reaction Mixture
5X HF buffer
DNA template (concentration varies by template source)
200 μM dNTPs
1 unit/50μl Phusion DNA polymerase
0.5 μM/primer
5% DMSO v/v ddH_2O (to final volume of 25 μl)

Table 3.2: PCR mixture guidelines

Primers, DMSO, template, dNTPs, buffer and water were added to a thin-walled

PCR tube as in Table 3.2. Annealing temperatures were estimated using the IDT Oligo Designer webtool. The extension time was assumed to be 15 sec/kb for reactions using plasmid template, and 30 sec/kb when using genomic DNA as template. Following amplification and confirmation on a 1% agarose gel (2% gels were used to visualize amplicons shorter than 500 bp), reaction products were purified with PCR purification kits (Bio Basics). Table 3.3 shows the heat cycler program used for monomer amplification.

Step	**Action**	**Time**
1	98°C	30 sec
2	98°C	10 sec
3	annealing temp	15 sec
4	72°C (extension temp.)	variable
5	return to step 2 for 26 cycles	
6	72°C	3 min
7	4°C	5 min

Table 3.3: Heat cycler program for assembly of dimers by overlap extension PCR

3.4.3 Assembly and Amplification of Dimers

To assemble dimers: dNTPs, buffer, primers, water, DMSO and equimolar concentration of monomer were pooled in a PCR tube as per Table 3.2. An annealing temperature 2°C lower than the lowest estimated T_m for either primer and the complimentary region between the two monomers was chosen. Table 3.3 shows the heat

cycler program used.

3.4.4 One-step PCR assembly

To assemble multimers, equimolar concentrations of monomers were pooled with DMSO, template, dNTPs, buffer and water (Table 3.2), except without primers. This reaction mixture was then subjected to 10 cycles with an annealing temperature 2°C below the lowest estimated annealing temperature for a complimentary region between monomers. Extension time was chosen based on 15 sec/kb, and the length of the final product was used to calculate the chosen time.

Step	Action	Time
1	98°C	30 sec
2	98°C	10 sec
3	annealing temp1	15 sec
4	72°C (extension temp.)	60 sec
5	return to step 2 for 10 cycles	
6	add primers	
7	98°C	10 sec
8	annealing temp2	15 sec
9	72°C (extension temp.)	60 sec
10	return to step 2 for 20 cycles	
11	72°C	3 min
12	4°C	5 min

Table 3.4: Heat cycler program for the one-step assembly of multiple monomers

Following the PCA pre-cycling steps terminal primers (Table 3.4) for the fully assembled sequence were added to the reaction tube upon removing the tube from

the heat cycler briefly, and then returned to the heat cycler and an additional 20 cycles were performed. Following confirmation by agarose gel electrophoresis, reactions were purified with PCR purification kits (BioBasics).

Chapter 4

DNA Assembly by *In vivo* Homologous Recombination

4.1 Background

The assembly of large DNA molecules by PCR-based methods is ultimately limited by the fidelity of the polymerase and the methods presented in the previous Chapters have only been successful applied to constructs below ∼5kb.

To enable the assembly of constructs longer than ∼5kb, I have implemented a method in which medium-sized fragments, assembled using the methods from Chapter 3, are fused together by homologous recombination in yeast *S. cerevisiae*.

The high fidelity of homologous recombination in yeast has generated renewed

interest from Synthetic Biologists in using this organism as a vehicle to assemble DNA[18, 23]. Yeast transformed with linear DNA molecules possessing either homologous or complimentary termini are spontaneously fused into contiguous molecules defined by the individual parts[43, 33]. Correspondingly, DNA assembly can be performed by direct integration into the yeast genome, or by co-transforming the fragments with a linearized replicating vector[43, 48].

A primary challenge with DNA assembly by homologous recombination is to get each DNA fragment to be present in a sufficient number of copies in any given cell. The odds of this decreases dramatically with the number of fragments in the assembly; with a 1/100 chance that a given cell is transformed with a given fragment in sufficient quantity, the probability that this cell will carry 5, 10 and 20 different fragments in sufficient quantity are 10^{-10}, 10^{-20} and 10^{-40}, respectively. Correspondingly, currently described methods that bootstrap homologous recombination (see Chapter 1) rely on non-standard transformation protocols to achieve the desired transformation efficiency[44, 50, 57, 34].

In order continue to extend our ability to assemble multiple DNA fragments, I

developed a simplified strategy for homologous recombination-based DNA assembly, relying on a LiAc transformation protocol. The goal was to augment our PCR-based technologies to assembly DNA constructs up to 10kb in length, while relying on LiAc/ssDNA methods currently in use in laboratory.

As a proof-of-principle demonstration for this strategy, I designed a construct, depicted in Figure 4.1c, containing 19 monomeric parts and a total length of 9.34kb. Designed with simple confirmation as a central feature, the construct contains two drug selection cassettes, Nat^R and KanMX (ClonNat and G418 resistance markers respectively), two gain-of-function expression cassettes, His3 and Ura3, two fluorescent reporters, yEGFP and yEBFP, and adenine auxotrophy conferred by Ade2 gene deletion.

I devised an assembly strategy whereby the 19 monomers that make up the construct were amplified with overlapping termini using the one-step PCR method described in Chapter 3 (Figure 3.5b). Monomers were to be assembled into six multimers using one-step PCR assembly and subsequently used to transform BY4742 yeast. As with previously described TAR based methods (see Chapter 1), each multimer would

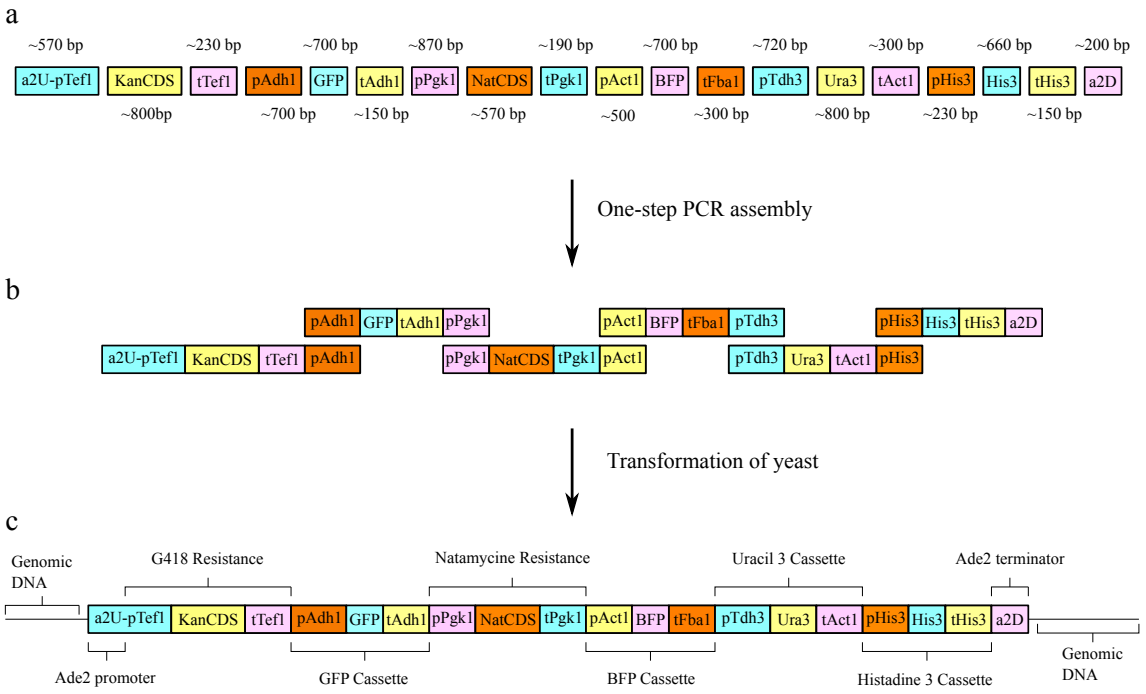

Figure 4.1: Proof-of-principle for DNA assembly by homologous recombination in yeast. a) Nineteen monomers were amplified with homologous ends added by primers. b) Monomers were assembled into six tetramers using the one-step PCR assembly described in 3.2.2, each tetramer shares homology over the full length of a monomer. c) All six tetramers are used to transform BY4742 yeast for assembly and integration into the Ade2 locus.

have homology with its neighbour to facilitate recombination after transformation. However, unlike previously describe strategies based on homologous recombination, ssDNA/LiAc based transformation was used[36].

4.2 Results

Before attempting the recombination-based assembly of the six tetramers, I performed a preliminary experiment to establish guidelines for concentrations of DNA to use in

the transformation. For this, I returned to the proof-of-principle construct used for the one-step method (Section 3.1). I amplified this construct from genomic DNA in two parts, overlapping at pTdh3, and purified each part. Then 200ng (approximately 6ng/μl) of each part were pooled and used to transform 10^8 BY4742 yeast using ssDNA/LiAc based transformation[36]. This concentration was chosen as it represents 2x the concentration of linear DNA used in previous experiments to transform the same locus. As described in Geitz *et al.* 2007, transformation was followed by a 2-4 hour recovery period at 30°C, in YPD, on an orbital shaker at 200 RPM. This recovery period is designed to ensure adequate antibiotic-resistance gene expression to allow plating directly onto antibiotic selection, such as G418 plates. The recovery period was followed by plating 200 μL of the transformation mixture directly onto YPD-agar ^{G418} plates, resulting in hundreds of red colonies. I selected 10 red colonies from this plate and tested them using flow cytometry. Each colony was restreaked as patches on YPD-agar ^{G418}. Each patch was used to inoculate 3 ml of SM-complete, which was then grown overnight to stationary phase. The overnight sample was then used to inoculate 1ml of SM-complete at an OD 600 of 0.04. This was then grown

for 4 hours at 30°C on an orbital shaker at 200 RPM, followed by a 1:10 dilution in sterile ddH_2O. Each of the 10 samples were immediately evaluated by flow cytometry; all 10 colonies demonstrated the anticipated GFP fluorescence (Figure 4.2).

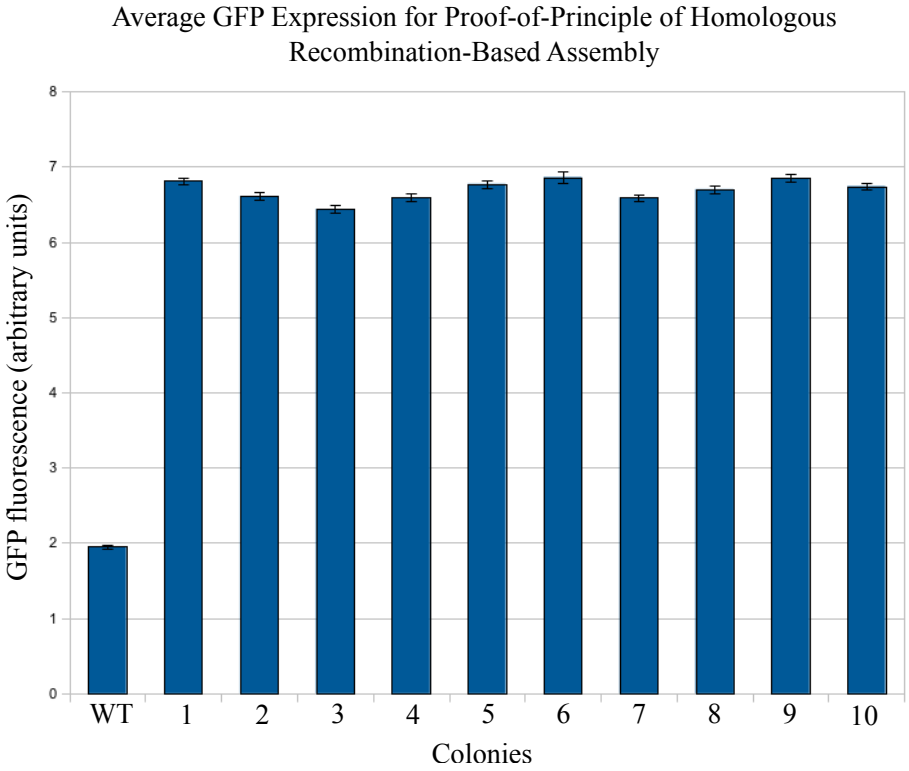

Figure 4.2: Flow cytometric confirmation of fluorescence of cells following co-transformation with two overlapping parts. Rows 1-10 show GFP fluorescence *vs.* wildtype BY4742 yeast. Error bars indicate standard error of the mean (S.E.M.) as calculated for one sample of each colony and a minimum of 5 000 events. S.E.M. = $\frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated.

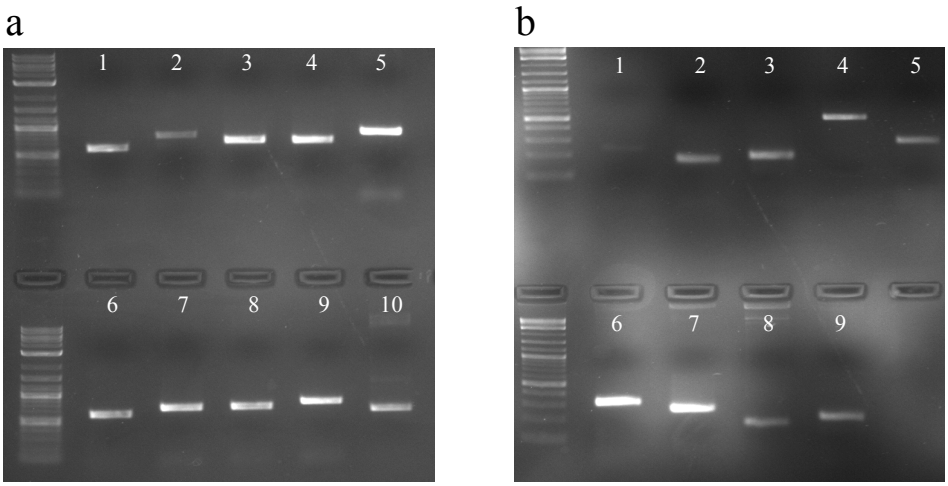

Figure 4.3: Agarose-gel confirmation of monomer amplification. a) Lanes 1-10 correspond to amplicons of ade2-pTef1 , KanCDS, pAdh1, GFP, pPgk1, NatCDS, BFP, pTdh3, Ura3 and His3 respectively. Electrophoresis was performed on a 1% agarose gel, with a 2 kb ladder (NEB). b) Lanes 1-9 correspond to amplicons of tTef1, tAdh1, tPgk1, pAct1, tFba1, tAct1, pHis3, tHis3 and Ade2D respectively. Electrophoresis was performed on a 2% agarose gel, with a 2 kb ladder (NEB).

Following the above success I proceeded with the assembly of the proof-of-principle construct in Figure 4.1. First the 19 monomers were amplified from their respective sources and confirmed by agarose-gel electrophoresis (Figure 4.3a & Figure 4.3b). Next, the nineteen monomers were assembled into the 6 tetramers illustrated in Figure 4.1 using the one-step PCR assembly method, and confirmed by gel electrophoresis (Figure 4.4).

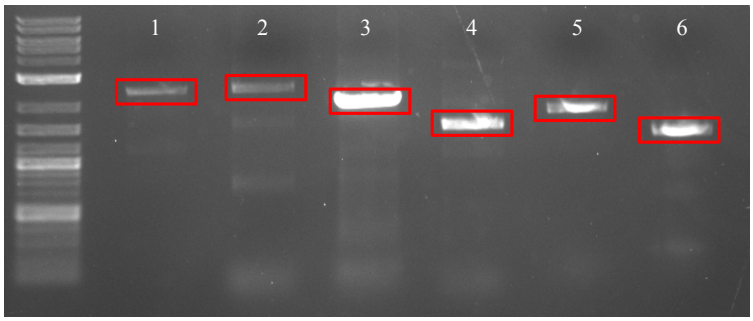

Figure 4.4: Agarose-gel confirmation of tetramer assembly. Electrophoresis was performed on a 1% agarose gel, with a 2 kb ladder (NEB). Lanes 1-6 correspond to tetramers 1-6 in Figure 3.5b.

Initial experiments indicated that the ∼200 ng/part concentration used in the previous experiments was insufficient to achieve the desired transformation efficiency. The first attempt to assemble the 6 tetramers by co-transformation yielded 2 red colonies when 2 x 10^8 cells were plated. This experiment was followed by another attempt where concentration was doubled to 400 ng/part, which again yielded 2 red colonies. At this point I did not posses sufficient quantities of each tetramer to perform the co-transformation experiment again; specifically, the pAdh1-GFP-tAdh1-pPgk1 construct. Following the re-assembly of this tetramer, I increased the concentration of each tetramer to 600ng/part. After increasing the concentration to 600ng/part the final volume of the pooled parts mixture exceeded the allowable volume for our transformation protocol. I proceeded, by precipitating the DNA and re-suspending in the

allowable 34 μl volume[36]. This mixture was used to transform BY4742 yeast, as in previous chapters.

Following a 4 hour recovery incubation in YPD (as described in previous chapters), 2 x 10^8 cells were plated directly onto YPD-agar^{G418}, followed by incubation at 30°C for 2 days. After two days of incubation a total of 11 red colonies were visible on the plate. All eleven colonies were selected and re-streaked on YPD-agar^{clonNat} to verify the correct assembly of the Nat^r cassette. All 11 colonies grew on clonNat plates and were subsequently re-streaked to SM-agar$^{-uracil}$, SM-agar$^{-histidine}$.

Next, colonies were evaluated for both BFP and GFP fluorescence *via* flow cytometry. Each of the 11 colonies were used to inoculate 11 tubes with 5ml each of SM-complete and grown overnight to stationary phase. The following day, all 11 samples were used to re-inoculate 1ml of SM-complete each at an OD 600 of 0.04. These samples were then grown for 4 hours at 30°C on a rotary shaker at 200 RPM. Figure 4.5 a & Figure 4.5b shows average GFP and BFP expression for all 11 colonies.

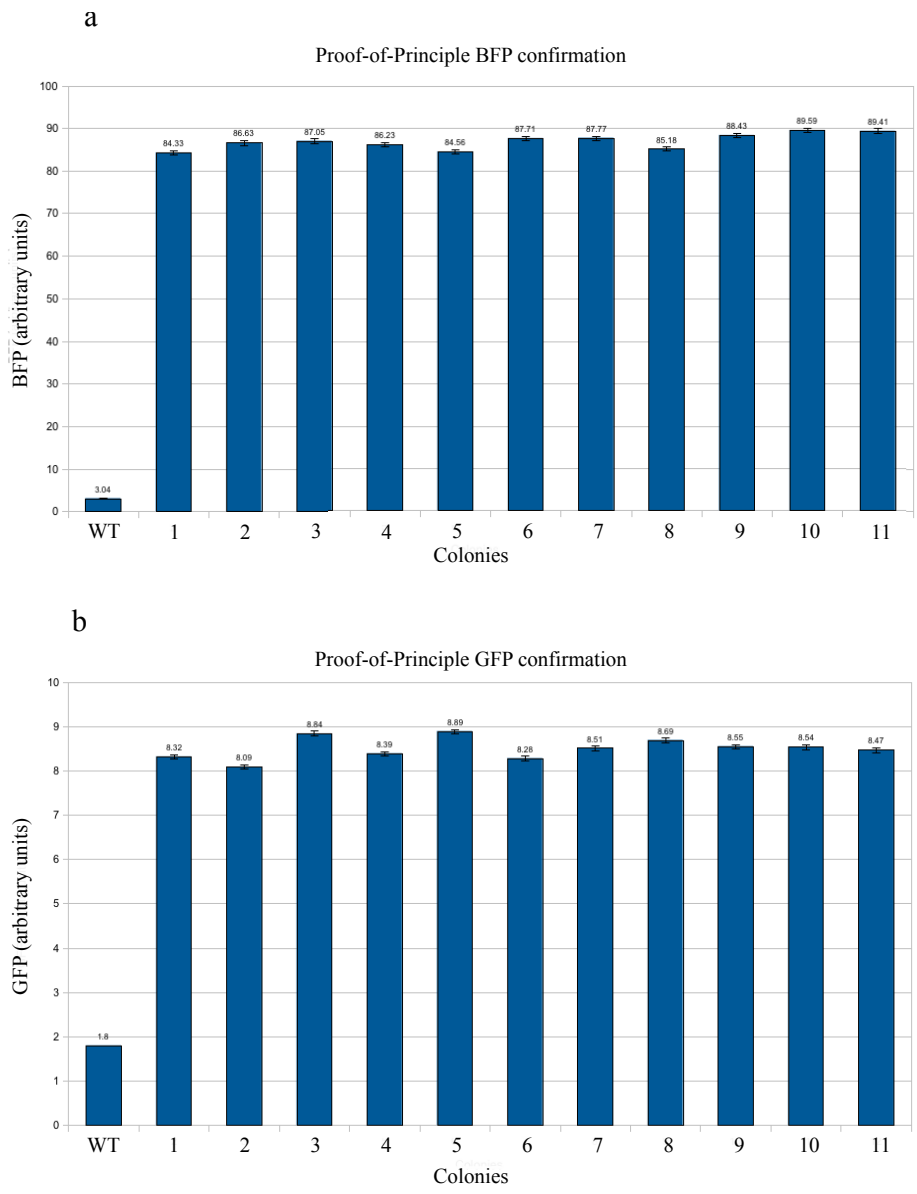

Figure 4.5: Flow cytometric confirmation of fluorescence. a) BFP fluorescence of proof-of-principle strain compared to wildtype BY4742 yeast. b) GFP fluorescence of proof-of-principle strain *vs.* wildtype BY4742 yeast. Error bars for both a & b indicate standard error for single inoculates with minimum of 5 000 cells evaluated. S.E.M. = $\frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated.

Finally, I performed PCR-based confirmation on one colony to establish that the

construct contained parts assembled in the correct order. Seven overlapping regions were chosen for amplification (see primers Table 4.1). The seven correctly sized amplicons support the conclusion that the construct was assembled correctly. These seven amplicons were visualized on a 1% agarose gel (Figure 4.6).

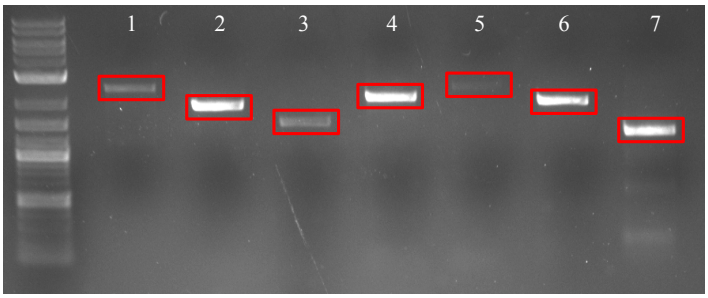

Figure 4.6: Agarose-gel confirmation of tetramer assembly. Electrophoresis was performed on a 1% agarose gel, with a 2 kb ladder (NEB). Lanes 1-7 correspond to tetramers 1-7 in Figure 3.5b

4.3 Discussion

Methods for the assembly of complex networks using *in vivo* recombination have been previously described[57, 34]; however, these methods rely on non-standard transformation protocols. In the case of the *M. genitallium* genome synthesis, 25 parts were assembled into a 590 kb genome by transforming yeast spheroplasts[34]. Spheroplasts are prepared by the enzymatic degradation of the yeast cell wall, yielding cells that are extremely transformation competent. However, spheroplasts can be challenging

to prepare, requiring delicate handling. Moreover, few laboratories currently have the need or the means for whole genome synthesis.

The monomers selected for assembly into tetramers were selected without sequence bias. That is, none of the parts were chosen so as to make their assembly easier or simpler. Monomers were deliberately selected in a way that reflects how promoters, coding sequences and terminator regions would be selected for their functional characteristics. In fact, sets of monomers were assembled despite being poorly suited for simultaneous PCR-based assembly. For example, the tetramer pAdh1-GFP-tAdh1-pPgk1 (Figure 4.1b) contained very disparate calculated annealing temperatures between primers; 67.8°C, and 49.3°C for the pAdh1 forward primer and pPgk1 reverse primer respectively.

In this Chapter, I have presented results demonstrating that multiple DNA fragments can be successfully assembled and integrated into the genome of *S. cerevisiae* using standard ssDNA/LiAc based methods. This method provides a simple, low-cost method for the construction of medium-sized DNA fragments comprising as many as 6 expression cassettes. This should be sufficient for many contemporary Synthetic Bi-

ology applications, including the construction of complex transcriptional regulatory networks. Furthermore, this approach complements the methods presented in Chapter 3; which can be used to rapidly generate the raw parts required for homologous recombination based assembly. Possible future extensions of the method includes the use of electroporation to further increase transformation efficiency[57], and the targeting of linearized autonomously replicating shuttle vectors to enable the transfer of DNA material to other organisms.

4.4 Methods

4.4.1 Amplification of Monomers and Assembly of Tetramers

All monomers were amplified and assembled into tetramers using the methods described in Chapter 3 (see Methods in Section 3.4). Table 4.1 shows all primers used in this experiment chapter for the assembly of the 19 part construct.

4.4.2 Transformation

The transformation protocol used in this chapter was performed as in previous chapters[36] with minor modification. Specifically, relatively high amounts of DNA were used;

Primer	**Primer Sequence**	**T_m1[°C]**
Ade2U_FP	ttcgaaaagttgcctagtttc	51.2
pTef1_RP	gttgtttatgttcggatgtgatgtgag	56.3
Kan_FP	cacatcacatccgaacataaacaaccatgggtaaggaaaagactc	64.7
Kan_RP	ttagaaaaactcatcgagcatcaaatg	54.2
tTef1_FP	tcatttgatgctcgatgagtttttctaatcagtactgacaataaaaag	61.7
tTef1_RP	ctggatggcggcgttagtatcgaatcg	62.5
pAdh1_FP	gatactaacgccgccatccagactagatccgggtgtacaatatggacttc	67.8
pAdh1_RP	tgtatatgagatagttgattgtatgcttgg	54.9
GFP_FP	ccaagcatacaatcaactatctcatatacaatgtctaaaggtgaagaattattcac	62.7
GFP_RP	ttatttgtacaattcatccataccatggg	55.5
tAdh1_FP	cccatggtatggatgaattgtacaaataagcgaatttcttatgatttatgattt	62.7
tAdh1_*RP*	agcgacctcatgctatacctgagaaagc	61.5
pPgk1_FP	aggtatagcatgaggtcgctcactagaactgtaattgcttttag	63.8
pPgk1_RP	tgttttatatttgttgtaaaaagtagataa	49.3
NatCDS_FP	tgttttatatttgttgtaaaaagtagataaatgggtaccactcttgacgacacgg	64.3
NatCDS_RP	ggggcagggcatgctcatgtagagcgcc	69.8
tPgk1_FP	ctacatgagcatgccctgcccctaaattgaattgaattgaaatcgatagatc	65.7
tPgk1_RP	aaaggcattaaaagaggagcgaatttt	56.0
pAct1_FP	cgctcctcttttaatgcctttatgagcaattaatgcacaacatttaacct	64.8
pAct1_RP	gttaattcagtaaattttcgatcttgg	51.4
BFP_FP	ccaagatcgaaaatttactgaattaacatgtccgaattgatcaaggaaaacatg	63.7
BFP_RP	gttcaatttgtgacccaattttgaaggc	57.7
tFba1_FP	cttcaaaattgggtcacaaattgaactaagttaattcaaattaattgatatag	60.2
tFba1_RP	gagctatcaaaaacgatagatcgattagg	55.3
pTdh3_FP	tcgtttttgatagctcaactagatacgtaagggagttagaatc	61.7
pTdh3_RP	cgtgtgtttattcgaaactaagttcttg	54.8
Ura3_FP	cgtgtgtttattcgaaactaagttcttgatgtcgaaagctacatataagg	63.0
Ura3_RP	gttttgctggccgcatcttctcaaata	59.9
tAct1_FP	tatttgagaagatgcggccagcaaaactaatctctgcttttgtgcgcgtatgt	68.7
tAct1_FP	atatgatacacggtccaatggataaac	54.9
pHis3_FP	gtttatccattggaccgtgtatcatattaattaacaacacagtcctttcccgcaattttc	65.8
pHis3_*RP*	ctttgccttcgtttatcttgcctgctc	60.2
His3_FP	gagcaggcaagataaacgaaggcaaagatgacagagcagaaagccctagt	68.7
His3_RP	tacataagaacacctttggtggaggg	58.3
tHis3_FP	ccctccaccaaaggtgttcttatgtagtgacaccgattatttaaagctgcag	66.9
tHis3_RP	gcgcgcctcgttcagaatgacacgtataga	64.5
ade2D_FP	tctatacgtgtcattctgaacgaggcgcgctaatatataagtttattgatatac	63.5
Ade2D_RP	gttgtttatgttcggatgtgatgtgag	56.3

Table 4.1: Primer list

500ng/tetramer, for a total of 3μg DNA. Transformed yeast were allowed to recover in YPD at 30∘C for 4 hrs, on an orbital shaker (200 RPM). Two days following plating on YPD^{G418}-agar, 11 colonies were re-streaked to SM-agar$^{-uracil}$ and SM-agar$^{-histidine}$, and YPD-agar^{clonNat}.

4.4.3 Flow Cytometry

Flow cytometry was performed on 11 transformants. Selected colonies were inoculated in 3 ml YPD$^{1\%adenine}$ and grown overnight. The next morning, cells were diluted in water to an OD-600 of 0.08, and interrogated by flow-cytometry.

Chapter 5

Conclusion

Synthetic Biology aims to apply engineering principles to the design and construction of synthetic biological systems. The design and construction of robust GRNs remains one of the primary objectives among Synthetic Biologists[49, 64], but Synthetic Biologist have forayed into high value molecule Biosynthesis[52], and Biofuels [16]. Unfortunately, a number of technological barriers continue to impede progress towards these goals. Foremost, the synthesis and assembly of double stranded DNA.

To date, the single largest effort towards improving the assembly of DNA and availability of parts has been the establishment of the Registry of Standard Biological Parts, and the annual iGEM competition. The Registry and the iGEM competition

have grown in lockstep; with 2012 seeing 191 teams swelling the Registry to over 7,000 BioBricks. The impact of BioBricks and the Registry has been significant, and there are several advantages to their use. As detailed in Chapter 1, idempotents, ease of sharing between laboratories, and a set of optimized protocols developed by an active community of iGEM participants are key features contributing to the success of the BioBrick platform. BioBricks are, however, not without their drawback. The most significant of these limitations being sequence constraints imposed by the standards themselves, and the reliance on *E. coli* as an intermediary organism.

Each BioBrick assembly standard relies on the re-use of specific restriction sites for their cloning, which in turn cannot appear within the part sequence (see Figure 1.3). In practice, this means that in order to generate a new BioBrick it is frequently necessary to perform multiple rounds of mutagenesis to eliminate these *illegal* restriction sites. This problem becomes particularly acute for researchers working with eukaryotic organisms, whose coding sequences are significantly longer than those of prokaryotes. This fact is reflected in the preponderance of prokaryotic BioBricks available through the Registry.

The continuous and precipitous drop in cost of oligonucleotide synthesis and, by extension, *de novo* double stranded DNA synthesis[15] hints at the full commodification of DNA synthesis. At first glance this resembles a solution to the assembly problem; instead of labouring over molecular cloning, *assembled* sequences could be synthesized on-demand by a service provider. However the nature of the complex systems being designed interrogated by Synthetic Biologists often require many iterations in order to arrive at the desired outcome[7, 55].

One of the strengths of our method is the ability to assemble short to medium length sequences. This makes it possible, for example, to assemble multiple synthetic promoters from subcomponents. Presciently, Integrated DNA Technologies recently began supplying short (sub-500 bp) linear sequences, called GBlocks. This represents a departure from how DNA synthesis has historically been marketed; i.e., synthesized sequences are delivered in a plasmid flanked by restriction sites. This move towards commercial synthesis companies supplying short linear DNA fragments bodes well for our technology, and will only broaden its utility. Finally, the facility with which our method manages shorts sequences allows it to compare well to other popular methods,

such as the Gibson Assembly method (Figure 1.9)[35]. While the Gibson method has proven itself reliable for the assembly of multiple long fragments, it has limitations in the assembly of sequences less than ∼200 bp.

The ability to easily manipulate short sequences with this method represents a decision based on the requirements in our laboratory, and alludes to a shortcoming of our PCR based assembly. Our method is not well suited to the assembly of multi-part constructs longer than 5 kb in length, which is precisely the type of assembly the aforementioned Gibson method[35] is so well suited to. For example, if the user aims to assemble multiple, long (over 10 kb), transformation ready sequences the Gibson method would be a good choice. Conversely, to assemble multiple short sequences for transformation without the use of *E.coli* our method might serve well.

Another strength of our technologies is the speed with which multimers can be assembled. A single user is able to generate multiple overlapping fragments for down-stream applications in a single day. In Chapter 4, I showed how this can be used in conjunction with homologous recombination to assembly 6 expression cassettes from 19 individual fragments. In the case of more typical assemblies comprised of

2-to-3 expression cassettes, the process has become trivial in one day. By contrast, traditional cloning methods and their descendants, such as the BioBrick assembly standards, require a minimum of 2 days to assemble two parts. Moreover, our PCR-based assembly methods could also be used to generate fragments for assembly using the Gibson Method, or TAR cloning into a replicating plasmid.

Speed is an obvious benefit to all PCR based methods, with assembly times measured in hours instead of days. However, PCR methods also share a set of shortcomings. The PCR amplification and assembly of low complexity sequences remains a challenge. Low complexity sequences must be dealt with differently, depending of whether they result from either a high GC or a high TA content. High GC content sequences are characterized by high melting temperatures and a tendency towards forming strong secondary structures

Conversely, sequences with high TA content are characterised by low melting and annealing temperatures. As a result, the PCR assembly and amplification of TA rich sequences often requires the use low annealing temperatures to ensure sufficient primer binding to allow for polymerizations[19, 63]. The low temperatures required for proper

annealing of TA rich sequences also favours secondary structure formation in peripheral sequences with a TA sequence bias. While adjuvants can be used to address the secondary structure problem, this approach is limited as it further reduces annealing temperatures for TA rich sequences. One method shown to improve PCR performance for TA rich sequences involves the selection of reduces temperatures for the extension phase of the PCR[63]. Su *et al.* 1996 suggest that the observed improvement in PCR performance is the result of reduced detachment of TA rich sequences following the initial primer binding event. However, this method requires longer extension times as the lower temperatures result in lower activity of the polymerase.

Finally, multiple repeat sequences such as tandem operators present challenges to PCR amplification. Identical, or near-identical, operator sequences can result in multiple binding targets, interfering with amplification specificity. One method that can be used to assemble sequences of this type is the co-transformation of overlapping oligonucleotides[33]. This method does not rely upon PCR, but instead relies on homologous recombination and gap-repair mechanisms in *S.cerevisiae* to assemble the sequence in question.

In conducting this research I made the decision to not perform intensive evaluation of transformation efficiency of *S. cerevisiae*. In the case of the PCR-based assembly methods presented in Chapters 2 and 3 the direct focus was on validation the pre-transformation assembly of DNA. Chapter 4 makes use of transformation associated recombination, with has been detailed elsewhere[43, 35, 57]. A number of factors have been shown to improve transformation efficiency. Length of homologous regions have been shown to impact the efficiency of transformation based assembly, with longer homology conferring greater efficiency[42, 50, 57] both into the genome and circularized plasmids. Conversely, local transformation efficiency has been shown to be negatively impacted by integration into open reading frames (ORFs) [37] and nucleosome occupancy[4]. Finally, global transformation efficiency is affected by the choice of transformation protocol[40, 36, 8].

Ian Roney (an honours students in our laboratory) and I used the methods presented here to construct a synthetic incoherent feedforward loop (iFFL) GRN (see Appendix C). The first of its kind in yeast. Preliminary characterization shows that this inducible network behaves in a manor characteristic of a iFFL[47, 6], generating

a rapid pulse of GFP expression following induction with β-estradiol. Building this network required the insertion of three discrete multi-cassette constructs into separate genomic loci.

In addition to using the PCR-based methods presented here to build the components for this network, we made use of yeast mating to assemble network subcomponents. Mating network sub-components to build completed networks allows for the combinatorial assembly of networks. For example, the user can build 5 variants of a network sub-component in one yeast mating-type and 5 variants of a different network sub-component in the complimentary mating type. Next, the 25 possible completed networks can be assembled through mating instead of individually constructing each of the 25 networks.

All of the constructs presented in this thesis were assembled for use in the budding yeast *S.cerevisiae*; however, this method is not limited in its application to yeast. As detailed above, parts generated using the PCR assembly methods here can be used as material for the Gibson Assembly method[35], or with the conceptually similar CPEC method[56]. Likewise, the BioBrick based method presented in Chapter 2 shows

how restriction sites can be added with terminal primers; differential primer selection makes would make this method compatible with a large number of restriction enzyme-based cloning methods. Similarly, TAR cloning in yeast can be used to generate circular plasmid which can in turn be used to transform other organisms. The *caveat* for the TAR cloning approach is that the final circularized plasmid will contain yeast replications of origin (such as an ARS) and selection markers.

Currently, the University of Ottawa has submitted a provisional patent application with Dr. Mads Kaern, Daniel Jedrysiak, and myself as the co-inventors. The application applies to the technologies surrounding the one-step overlap extension PCR assembly of multiple parts used in Section 3.2.2. We are continuing with this process and intend to submit a full patent application based on the work presented in this Thesis.

Appendix A

DMSO as a PCR Adjuvant

Figure A.1 illustrates the POP2 construct and its attempted PCR amplification in the absence of DMSO. The set of 6 reactions were performed using an annealing temperature gradient ranging from 55°C to 65°C. There is no discernible band at the desired 3.9 kb mark (indicated by the red line on the 2kb DNA ladder).

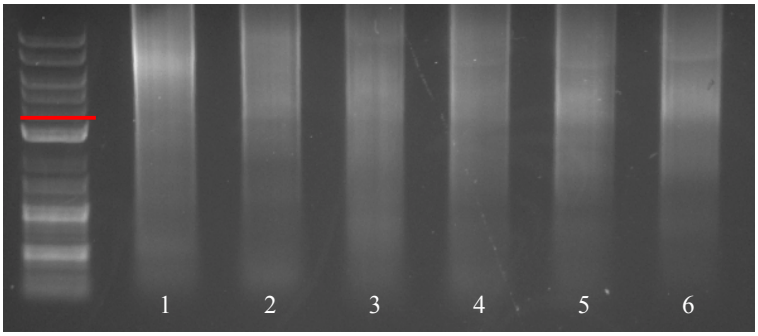

Figure A.1: This figure shows results for the attempted PCR based assembly of the POP2 construct (Figure 3.1). The reactions were performed using an annealing temperature gradient. Lane 1: 55.0°C, lane 2: 56.6° lane 3: 58.2° lane 4: 60.4° lane 5: 62.4° lane 6: 65.0°. The gel-electrophoresis was performed using a 1% agarose gel with a 2 kb ladder (NEB). The red band in the ladder indicates the anticipate position for a 3.9 kb construct.

Figure A.2 shows a DMSO concentration gradient from 1% to 5% v/v. In lane 1, the 5% DMSO v/v reaction resulted in a 3.9 kb band corresponding to the fully assembled POP2 construct. The annealing temperature selected for all 5 reactions was the same, at 50 °C.

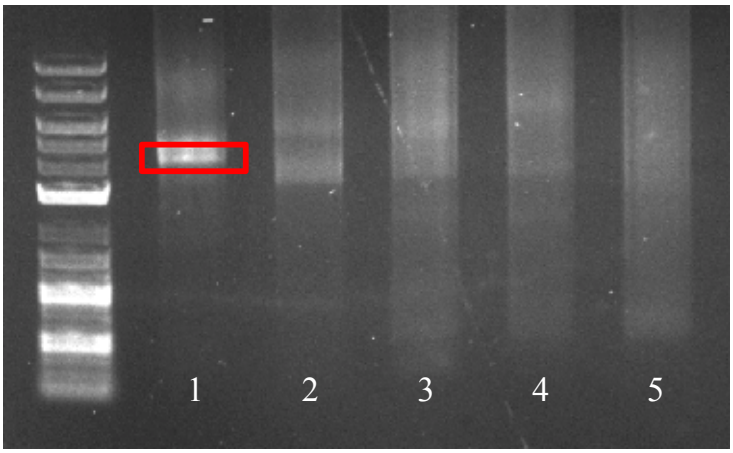

Figure A.2: PCR DMSO concentration gradient. Lane 1: 5% DMSO, Lane 2: 4% DMSO, Lane 3: 3% DMSO, Lane 4: 2% DMSO, Lane 5: 1% DMSO.

Figure A.3 shows a final application of DMSO to facilitate the assembly of multiple parts. This figure shows the attempted assembly and amplification of a $\sim$ 4.6 kb construct from 3 multimers with and without 5% DMSO. Both reactions were carried using a 60°C annealing temperature, with a 75 second extension time.

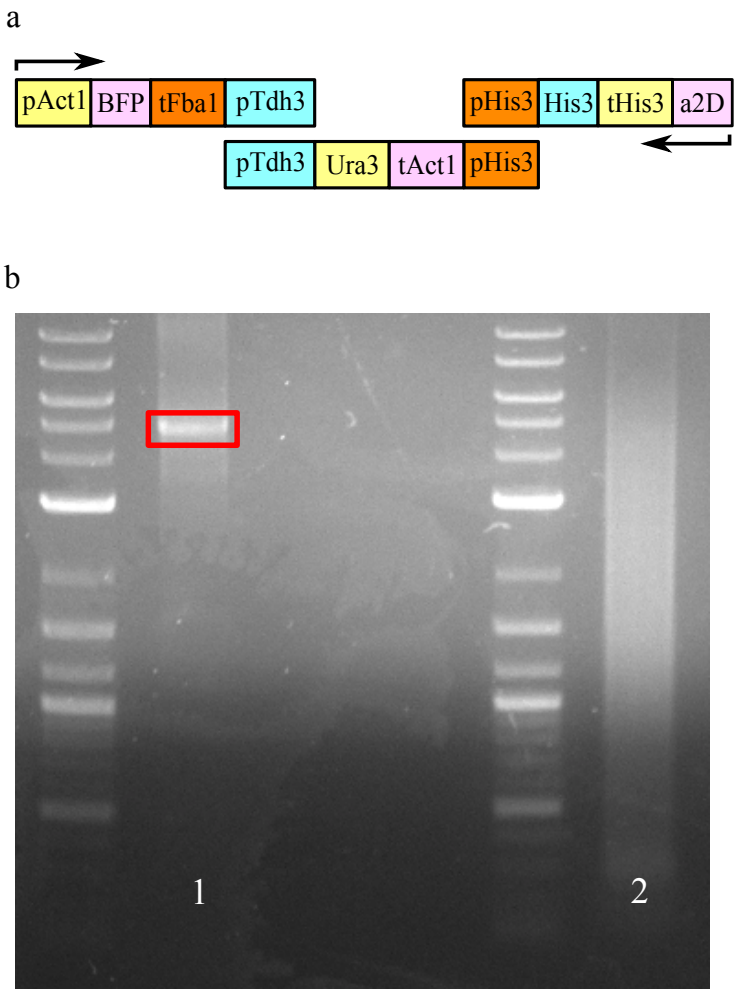

Figure A.3: a) depicts the three tetramers that I attempted to assemble with and without 5% DMSO. b) Lane 1 shows the assembly of the 3 multimers into the $\sim$ 4.6 kb construct. Lane 2 shows the same reaction attempted without a DMSO adjuvent.

Appendix B

Pre-Cycling Experiments

Figure B.1 illustrates four monomers that I attempted to assemble and amplify using standard PCR assembly methods. Approximately 1ng/μl x (length of part in kb) were pooled in a 50 μl PCR mixture. The PCRs were carried out with the addition of 0.5 μM primers and an extension time corresponding to the amplification of the full length construct (*i.e.*, 25 seconds). The reaction mixture included 5% DMSO in an attempt to improve amplification specificity. The reaction mixture was split into 10 thin walled PCR with 9 μl/tube. I performed two annealing-temperature gradients to ensure that the negative result of this assembly was not merely the result of an incorrectly selected annealing temperature. The two gradients span annealing

temperatures ranging from 45°C to 63°C. None of the selected annealing temperatures resulted in the desired 1.6 kb band.

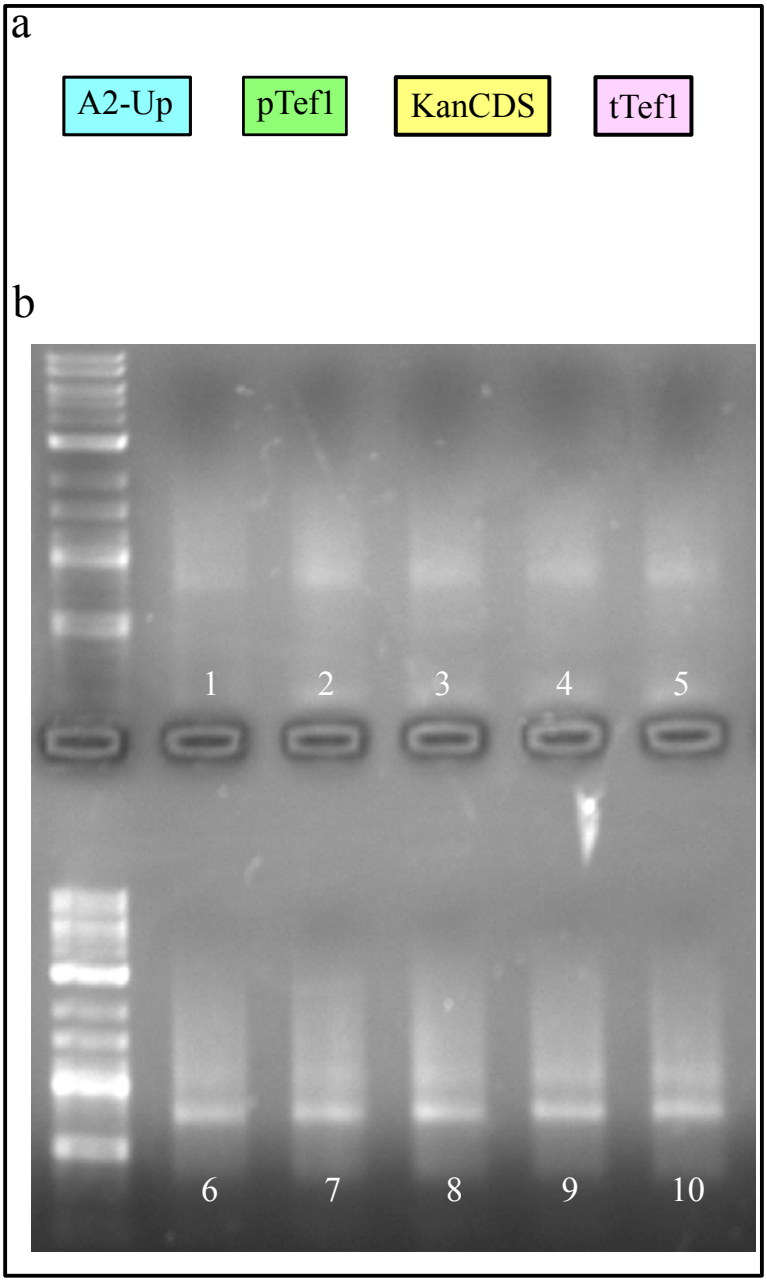

Figure B.1: a) depiction of 4 monomers subjected to PCR gradients reactions without pre-cycling, but with 5% DMSO b) two gradient reactions for the assembly of tetramers in 'a'. Lanes 1-5 show results for gradient reaction from 63.0°C to 56°C respectively. Lanes 6-10 show results for gradient reaction from 55°C to 45°C, respectively.

Figure B.2 compares the use of pre-cycling (10 cycles) for the assembly of 4, 5 and 6 monomers with a no pre-cycling protocol. The first 4 monomers of each of the three assemblies are the same monomers that were not successfully assembled using the gradient reactions in Figure B.1. Both pre-cycled and non pre-cycled reactions contained 5% DMSO. Both the pre-cycled and non pre-cycled reactions were subjected to the same annealing temperatures. The Selected annealing temperature was 50°C. This temperature is 2°C lower than the lowest estimated inter-monomer homologous region and primers, corresponding to the reverse primer sequence for the KanMX CDS amplicon (primer: U05). The heat cycler program for the pre-cycled reactions was halted after 10 cycles and terminal primers were added. The remaining 15 cycles were carried out as described in methods section 3.4.4.

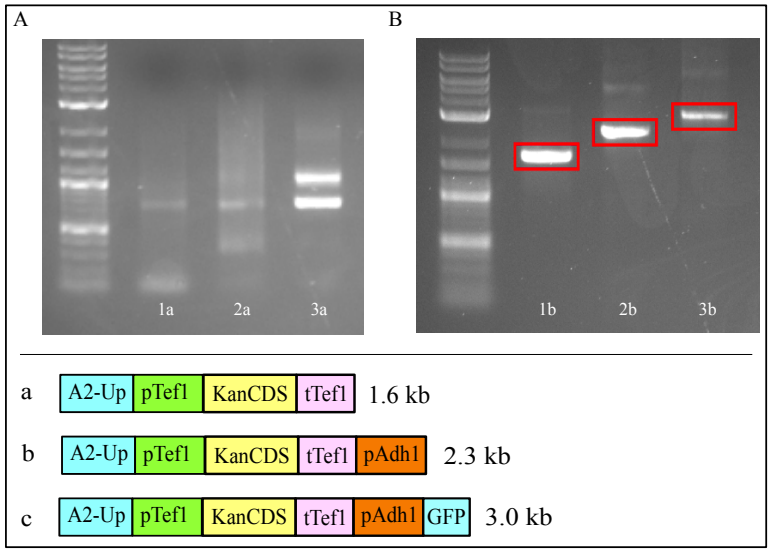

Figure B.2: A) This gel shows the results for the attempted assembly of a, b and c respectively; without the use of pre-cycling. Lanes 1a, 2a and 3a correspond to the constructs a, b and c respectively B) This gel shows the results for the attempted assembly of a, b and c respectively; with the use of the pre-cycling regime. Lanes 1b, 2b and 3b correspond to the constructs a, b and c respectively. Both gel-images were generated on a 1% agarose gel using a 2 kb ladder (NEB)

Following this success with the pre-cycling method in assembling 4, 5 and 6 monomers I attempted to use the same pre-cycling regime to assemble more than 6 monomers. Figure B.3 shows the attempted assembly of 5-10 monomers in a single reaction. The same annealing temperature parameters and monomer concentration parameters were used for these reactions as for those in Figure B.2B. The extension times chosen were based on 15sec/kb of fully assembled construct. This figure shows the gradual decrease in band intensity as the number of assembled monomers increases to 9. The lane corresponding to the assembly of 10 monomers shows no visible band of the

correct length.

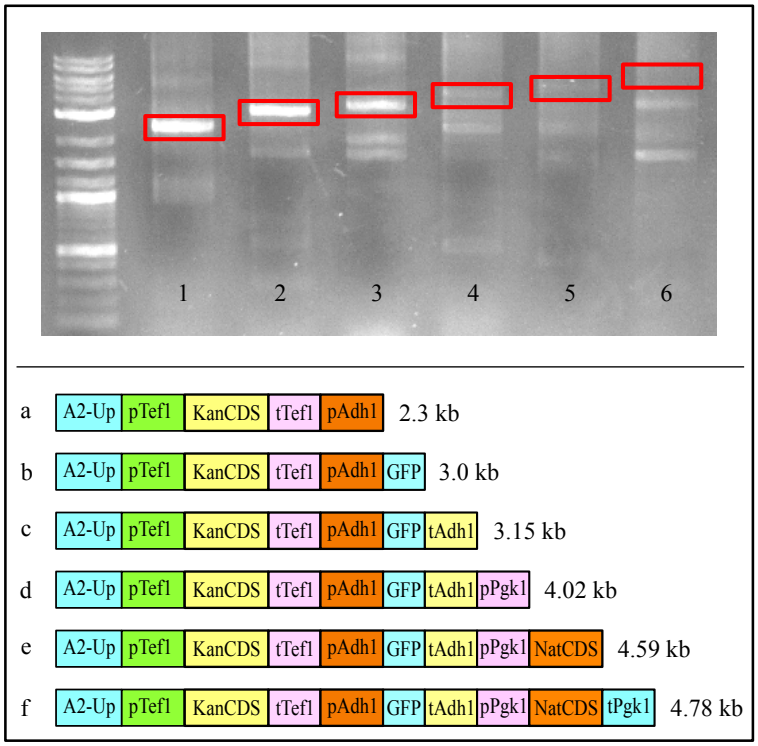

Figure B.3: Results in gel-lanes 1-6 correspond to constructs a-f respectively. This figure was generated using a 0.5% agarose gel with a 2 kb ladder (NEB). Red boxes indicate where the band for each construct should appear.

Appendix C

Incoherent Feedforward Loop

Ian Roney and Myself used both PCR and homologous recombination based assembly methods to construct an iFFL GRN (Figure C.1). The network consists of the β-estradiol induced GEV transactivator[46] which is a recombinant protein made up of an *S.cerevisiae* Gal4 DNA binding domain (Gal4dbd), linked to the ligand binding domain of the human estradiol receptor-2, and a VP-16 activation domain. The constitutively expressed GEV is targeted to two versions of the Gal1 promoter. The first is the native Gal1 promoter which was inserted into the Ade4 locus of BY4741(MATa). The second target for GEV is the pGalTX[24] promoter inserted into the Ade4 locus of BY4742(MATα). This synthetic promoter consists of a Gal1 promoter with

two Tet-operators immediately downstream of the TATA box. These Tet-operators serve as targets for repression by the TetR-BFP fusion protein expressed from the Ade2 locus. The BY4741 and BY4742 strains were mated together and selected on YPD-agar$^{G418\&clonNat}$

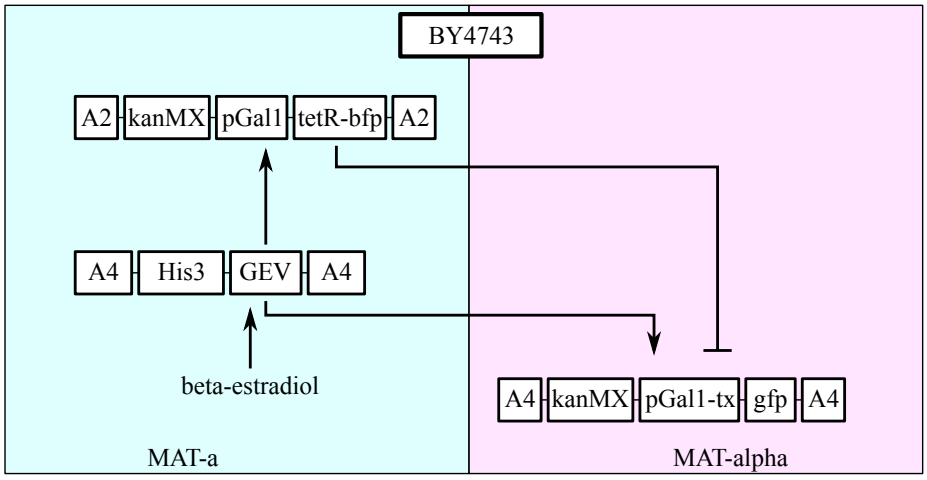

Figure C.1: Illustration of iFFL GRN as mated together. The GEV trans-activator is responsive to the input β-estradiol resulting in the expression of both TetR-BFP and GFP. TetR-BFP, in turn, represses expression from the pGalTX promoter. TetR-BFP repression can be relieved by the doxycycline analogue anhydrotetracycline (ATc).

Figure C.2 shows the responsiveness of the iFFL GRN to β-estradiol and ATc. In the presence of 200 nmol β-estradiol both BFP and GFP expression are detectable $vs.$ WT diploid cells. In the presence of 200 nmol β-estradiol and 100 ng/μl ATc TetR-BFP repression is relieved allowing for high GFP expression compared to WT diploid cells. This network demonstrates no detectable leakage as the uninduced network is

indiscernible from WT cells.

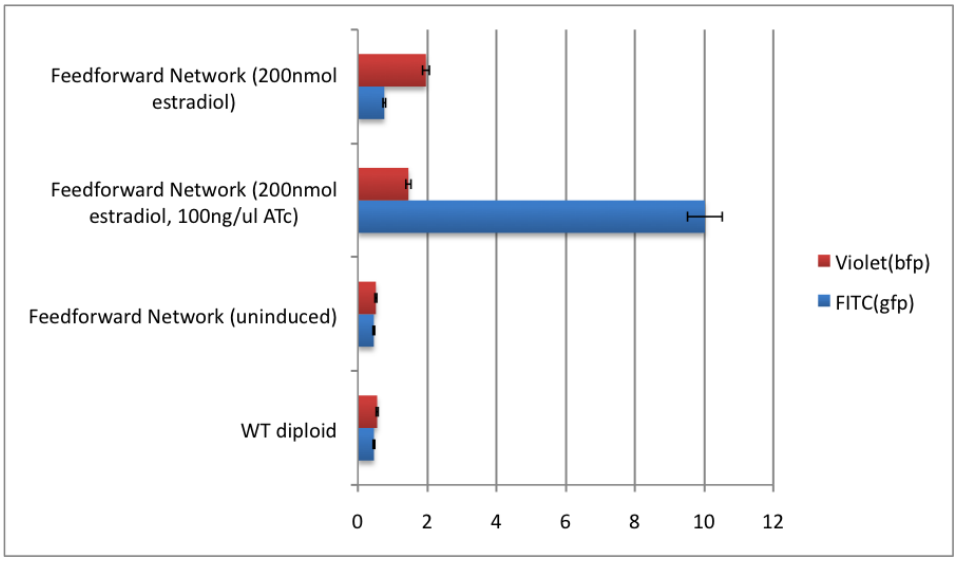

Figure C.2: A single re-streaked colony was used to inoculate 3 separate tube with in one of 3 conditions: uninduced (SM-complete), induced (SM-complete with 200 nmol β-estradiol), and induced and repressed (SM-complete with 200 nmol β-estradiol with 100ng/mul ATc). All samples were grown overnight at 30$\circ$C on a rotary shaker (200 RPM) to stationary phase and then used to inoculate 1 ml SM (at an OD 600 of 0.04) under the same conditions as the overnight inoculates. Following 4 hours of incubation, cells were diluted 1:10 with ddH_2O and evaluated *via* flow cytometry). Five thousand cells were evaluated for each sample, and S.E.M. was calculated as S.E.M. = $\frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated.

Figure C.3 shows preliminary results of the iFFL GRN generating a GFP pulse. A six hour time course following β-estradiol induction. Pulsative gene expression has been previously demonstrated in prokaryotic synthetic iFFL GRNs[6, 47], in addition to being predicted by computational models. This preliminary experiment was only carried out for 6-hours and steady-state was not reached; TetR-BFP expression is still rising and GFP is still decreasing.

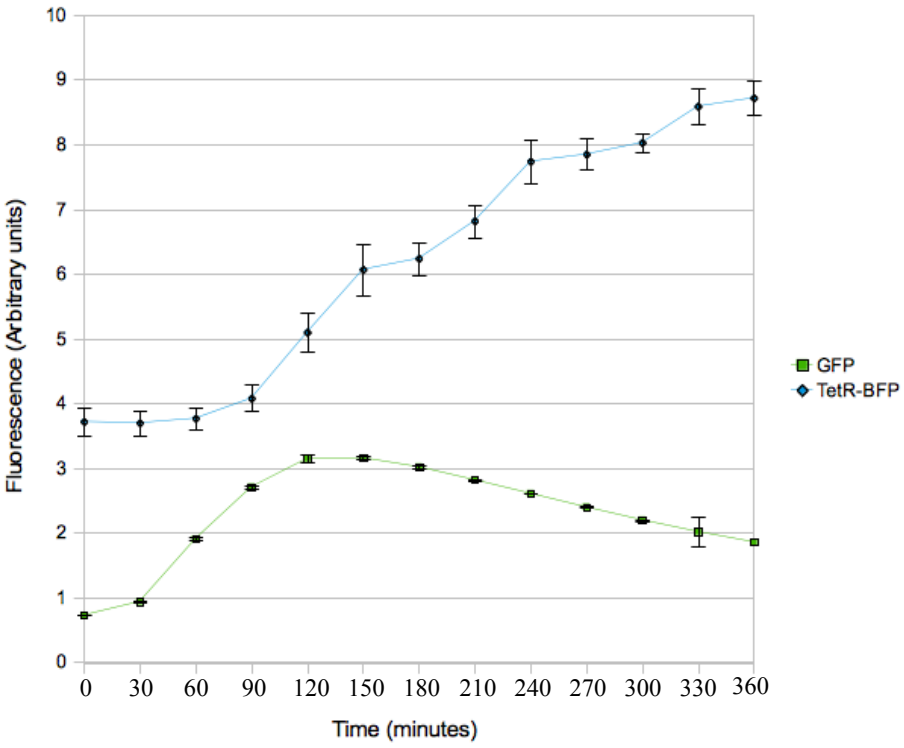

Figure C.3: The same strain tested in Figure C.2 was used to inoculate 3 ml liquid SM-complete and grown overnight to stationary phase. Overnight samples were used to inoculate 1 ml of liquid SM-complete with 200 nmol β-estradiol, to an OD 600 of 0.04. The inoculated sample was grown for 6 hours at 30°C on an orbital shaker at 200 RPM. Cells were harvested every 30 min and evaluated by flow-cytometry, 5 000 events were recorded for each sample. Error bars indicate S.E.M., S.E.M. was calculated as S.E.M. = $\frac{\sigma}{\sqrt{n}}$; where σ = standard deviation, and n = number of cells interrogated.

Bibliography

[1] C M Agapakis and P A Silver. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Mol. BioSyst.*, 5(7):704–13, Jun 2009.

[2] H Amelia, M Ling, K Ma, and H Yu. Investigating the effects of dmso on pcr fidelity using a restriction digest-based method. *Journal of Experimental Microbiology adn Immunology*, 14:161–164, Jul 2010.

[3] J C Anderson, J E Dueber, M Leguia, G C Wu, J A Goler, A P Arkin, and J D Keasling. Bglbricks: A flexible standard for biological part assembly. *Journal of biological engineering*, 4(1):1, Dec 2010.

[4] E Aslankoohi, K Voordeckers, H Sun, A Sanchez-Rodriguez, E Van Der Zande,

K Marchal, and K. J Verstrepen. Nucleosomes affect local transformation efficiency. *Nucleic Acids Research*, 40(19):9506–9512, Oct 2012.

[5] D Bang and G Church. Gene synthesis by circular assembly amplification. *Nature Methods*, 5(1):37–39, Jan 2008.

[6] S Basu, R Mehreja, S Thiberge, M T Chen, and R Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of the Sciences*, 101(17):6355–60, Apr 2004.

[7] G Batt, B Yordanov, R Weiss, and C Belta. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics*, 23(18):2415–22, Sep 2007.

[8] L Benatuil, J. M Perez, J Belk, and C.-M Hsieh. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering Design and Selection*, 23(4):155–159, Apr 2010.

[9] P Bernard, P Gabant, E M Bahassi, and M Couturier. Positive-selection vectors using the f plasmid ccdb killer gene. *Gene*, 148(1):71–4, Oct 1994.

[10] M Bichara, J Wagner, and I B Lambert. Mechanisms of tandem repeat instability in bacteria. *Mutat Res*, 598(1-2):144–63, Jun 2006.

[11] B A Blount, T Weenink, and T Ellis. Construction of synthetic regulatory networks in yeast. *FEBS Letters*, pages 1–10, Feb 2012.

[12] C B Brachmann, A Davies, G J Cost, E Caputo, J Li, P Hieter, and J D Boeke. Designer deletion strains derived from saccharomyces cerevisiae s288c: a useful set of strains and plasmids for pcr-mediated gene disruption and other applications. *Yeast*, 14(2):115–32, Jan 1998.

[13] D M Camacho and J J Collins. Systems biology strikes gold. *Cell*, 137(1):24–6, Apr 2009.

[14] R carlson. Laying the foundations for a bio-economy. *Systems and Synthetic Biology*, 1(3):109–117, Aug 2007.

[15] R carlson. The changing economics of dna synthesis. *Nat Biotechnol*, 27(12):1091–1094, Nov 2009.

[16] J M Carothers, J A Goler, and J D Keasling. Chemical synthesis using synthetic biology. *Curr Opin Biotechnol*, 20(4):498–503, Jul 2009.

[17] K Cha-Aim, T Fukunaga, H Hoshida, and R Akada. Reliable fusion pcr mediated by gc-rich overlap sequences. *Gene*, 434(1-2):43–49, Mar 2009.

[18] A A Cheng and T K Lu. Synthetic biology: an emerging engineering discipline. *Annu Rev Biomed Eng*, 14:155–78, Dec 2012.

[19] N Chester and D R Marshak. Dimethyl sulfoxide-mediated primer tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Analytical Biochemistry*, 209(2):284–90, Feb 1993.

[20] M J Czar, J C Anderson, J S Bader, and J Peccoud. Gene synthesis demystified. *Trends Biotechnol*, 27(2):63–72, Jan 2009.

[21] D A Drubin, J C Way, and P A Silver. Designing biological systems. *Genes & Development*, 21(3):242–54, Jan 2007.

[22] T Ellis, T Adie, and G Baldwin. Dna assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.*, 3(2):109, Jan 2011.

[23] T Ellis, X Wang, and J J Collins. Gene regulation: hacking the network on a sugar high. *Mol Cell*, 30(1):1–2, Apr 2008.

[24] T Ellis, X Wang, and J.J Collins. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol*, 27(5):465–71, Apr 2009.

[25] M Elowitz and S Leibler. a synthetic oscilatory network of transcriptional regulators. *Nature*, 403:1–4, Nov 2000.

[26] D Endy. Foundations for engineering biology. *Nature*, 438(7067):449–453, Nov 2005.

[27] U Frey, H Bachmann, J Peters, and W Siffert. Pcr-amplification of gc-rich regions: 'slowdown pcr'. *Nat Protoc*, 3(8):1312–1317, Jul 2008.

[28] A E Friedland, T. K Lu, X Wang, D Shi, G Church, and J J Collins. Synthetic gene networks that count. *Science*, 324(5931):1199–1202, May 2009.

[29] M Galdzicki, C Rodriguez, D Chandran, H.M Sauro, and J Gennari. Standard biological parts knowledgebase. *PLoS ONE*, 6(2):e17005, Feb 2011.

[30] TS Gardner, CR Cantor, and J J Collins. Construction of a genetic toggle switch in esscherichia coli. *Nature*, 403:339–342, Nov 2000.

[31] F Geu-Flores, H H Nour-Eldin, M T Nielsen, and B A Halkier. User fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple pcr products. *Nucleic Acids Research*, 35(7):e55–e55, Mar 2007.

[32] G Giaever, A M Chu, L Ni, C Connelly, L Riles, S Véronneau, S Dow, A Lucau-Danila, K Anderson, B André, A P Arkin, A Astromoff, M El-Bakkoury, R Bangham, R Benito, S Brachat, S Campanaro, M Curtiss, K Davis, A Deutschbauer, KD Entian, P Flaherty, F Foury, DJ Garfinkel, M Gerstein, D Gotte, U Güldener, JH Hegemann, S Hempel, Z Herman, D Jaramillo, DE Kelly, SL Kelly, P Kötter, D LaBonte, DC Lamb, N Lan, H Liang, H Liao, L Liu, C Luo, M Lussier, R Mao, P Menard, SL Ooi, JL Revuelta, CJ Roberts, M Rose, P Ross-Macdonald, B Scherens, G Schimmack, B Shafer, DD Shoemaker, S Sookhai-Mahadeo, RK Storms, JN Strathern, G Valle, M Voet, G Volckaert, CY Wang, TR Ward, J Wilhelmy, EA Winzeler, Y Yang, G Yen, E Youngman, K Yu, H Bussey, JD Boeke, M Snyder, P Philippsen, R Davis, and M Johnston. Functional pro-

filing of the saccharomyces cerevisiae genome. *Nature*, 418(6896):387–91, Jul 2002.

[33] D G Gibson. Synthesis of dna fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research*, pages 1–7, Sep 2009.

[34] D G Gibson, G A Benders, K C Axelrod, J Zaveri, M A Algire, M Moodie, M G Montague, J C Venter, H O Smith, and C A Hutchison. One-step assembly in yeast of 25 overlapping dna fragments to form a complete synthetic mycoplasma genitalium genome. *Proceedings of the National Academy of the Sciences*, 105(51):20404–9, Dec 2008.

[35] D G Gibson, L Young, R Y Chuang, J C Venter, C A Hutchison, and H O Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–5, Apr 2009.

[36] R D Gietz and R H Schiestl. High-efficiency yeast transformation using the liac/ss carrier dna/peg method. *Nat Protoc*, 2(1):31–4, Dec 2007.

[37] K Gjuracic, E Pivetta, and C V Bruschi. Targeted dna integration within different

functional gene domains in yeast reveals orf sequences as recombinational cold-spots. *Mol Genet Genomics*, 271(4):437–46, Apr 2004.

[38] R M Horton, H D Hunt, S N Ho, J K Pullen, and L R Pease. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, 77(1):61–8, Apr 1989.

[39] M Jensen, M Fukushima, and R Davis. Dmso and betaine greatly improve amplification of gc-rich constructs in de novo synthesis. *PLoS ONE*, 5(6):e11024, Jun 2010.

[40] Shigeyuki Kawai, Wataru Hashimoto, and Kousaku Murata. Transformation of saccharomyces cerevisiaeand other fungi: Methods and possible underlying mechanism. *biobugs*, 1(6):395–403, Nov 2010.

[41] T Knight. Idempotent vector design for standard assembly of biobricks. pages 1–11, May 2003.

[42] P Koren, I K Svetec, P T Mitrikeski, and Z Zgaga. Influence of homology size

and polymorphism on plasmid integration in the yeast cyc1 dna region. *Curr Genet*, 37(5):292–7, Apr 2000.

[43] N Kouprina and V Larionov. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast saccharomyces cerevisiae. *Nat Protoc*, 3(3):371–7, Dec 2008.

[44] V Larionov, N Kouprina, G Solomon, J C Barrett, and M A Resnick. Direct isolation of human brca2 gene by transformation-associated recombination in yeast. *Proc Natl Acad Sci USA*, 94(14):7384–7, Jul 1997.

[45] L S Larsen, C D Wassman, G Hatfield, and R H Lathrop. Computationally optimised dna assembly of synthetic genes. *International Journal of Bioinformatics Research and Applications*, 4(3):324–36, Dec 2008.

[46] J F Louvion, B Havaux-Copf, and D Picard. Fusion of gal4-vp16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*, 131(1):129–34, Sep 1993.

[47] S Mangan, S Itzkovitz, A Zaslaver, and U Alon. The incoherent feed-forward

loop accelerates the response-time of the gal system of escherichia coli. *Journal of Molecular Biology*, 356(5):1073–81, Mar 2006.

[48] C Merryman and D G Gibson. Methods and applications for assembling large dna constructs. *Metabolic Engineering*, 14(3):196–204, Apr 2012.

[49] N Nandagopal and M Elowitz. Synthetic biology: Integrated gene circuits. *Science*, 333(6047):1244–1248, Sep 2011.

[50] V N Noskov, M Koriabine, G Solomon, M Randolph, J C Barrett, S H Leem, L Stubbs, N Kouprina, and V Larionov. Defining the minimal length of sequence homology required for selective gene isolation by tar cloning. *Nucleic Acids Research*, 29(6):E32, Mar 2001.

[51] B R Palmer and M G Marinus. The dam and dcm strains of escherichia coli–a review. *Gene*, 143(1):1–12, May 1994.

[52] P Peralta-Yahya, F Zhang, S Del Cardayre, and J D Keasling. Microbial engineering for the production of advanced biofuels. *Nature*, 488(7411):320–328, Aug 2012.

[53] I Phillips and P A Silver. A new biobrick assembly strategy designed for facile protein engineering. pages 1–6, Apr 2006.

[54] E Pienaar, M Theron, M Nelson, and H J Viljoen. A quantitative model of error accumulation during pcr amplification. *Computional Biological Chemistry*, 30(2):102–11, Mar 2006.

[55] P Purnick and R Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, Jun 2009.

[56] J Quan and J Tian. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE*, 4(7):e6441, Jul 2009.

[57] Z Shao, H Zhao, and H Zhao. Dna assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Research*, 37(2):e16–e16, Nov 2008.

[58] R Shetty, D Endy, and T Knight. Engineering biobrick vectors from biobrick parts. *Journal of biological engineering*, 2(1):5, Jan 2008.

[59] N Shevchuk, A Bryksin, V Nusinovich, F Cabello, M Sutherland, and S Ladisch.

Construction of long dna molecules using long pcr-based fusion of several fragments simultaneously. *Nucleic Acids Research*, 32(2):19e–19, Jan 2004.

[60] M L Siegal, D A Petrov, and D De Aguiar. Triple-ligation strategy with advantages over directional cloning. *Biotech.*, 21(4):614, 616, 618–9, Sep 1996.

[61] R S Sikorski and P Hieter. A system of shuttle vectors and yeast host strains designed for efficient manipulation of dna in saccharomyces cerevisiae. *Genetics*, 122(1):19–27, Apr 1989.

[62] S C Sleight, B A Bartley, J A Lieviant, and H M Sauro. In-fusion biobrick assembly and re-engineering. *Nucleic Acids Research*, 38(8):2624–2636, May 2010.

[63] X Z Su, Y Wu, C D Sifri, and T E Wellems. Reduced extension temperatures required for pcr amplification of extremely a+t-rich dna. *Nucleic Acids Research*, 24(8):1574–5, Apr 1996.

[64] M Tigges and M Fussenegger. Recent advances in mammalian synthetic biology-design of synthetic transgene control networks. *Curr Opin Biotechnol*, 20(4):449–60, Jul 2009.

[65] J Tong, W Cao, and F Barany. Biochemical properties of a high fidelity dna ligase from thermus species ak16d. *Nucleic Acids Research*, 27(3):788–94, Jan 1999.

[66] R A Woods, D G Roberts, T Friedman, D Jolly, and D Filpula. Hypoxanthine: Guanine phosphoribosyltransferase mutants in saccharomyces cerevisiae. *Molecular Gene Genetics*, 191:407–412, Dec 1983.

[67] A S Xiong, R H Peng, J Zhuang, J G Liu, F Gao, J M Chen, Z M Cheng, and Q H Yao. Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress. *Biotechnology Advances*, 26(2):121–34, Dec 2008.

[68] B J M Zonneveld and A Van Der Zanded. The red ade mutants of kluyveromyces lactis and their classification by complementation with cloned ade1 and ade2 genes from saccharomyces cerevisiae. *Yeast*, 11:823–827, Nov 1995.