Novel Frontier in Gene Regulation via Expression of T7 Bacteriophage RNA polymerase in Yeast

Inbal Vaknin

Novel Frontier in Gene Regulation via Expression of T7 Bacteriophage RNA polymerase in Yeast

Research Thesis

In partial fulfilment of the requirements for the
Degree of Master of Science in Biotechnology & Food Engineering

Inbal Vaknin

Submitted to the Senate of the Technion - Israel Institute of Technology
Adar aleph month 5776, Haifa, March 2016

The research thesis was done under the supervision of
Asst. Prof. Roee Amit in the department of Biotechnology and Food Engineering

The generous financial help of the Technion- Israel Institute of Technology is gratefully acknowledged.

Table of Contents

List of figures

List of Tables

Abstract

Gene expression systems in eukaryotes encompass gene transcription, mRNA editing, translation and post-translational modifications as crucial key steps for protein synthesis. Synthetic expression systems are no different from the natural systems, but nonetheless can act as simplified models in order to elucidate unknown areas in gene regulation. Therefore, developing such a synthetic expression system embedded in yeast cells, and based on T7 RNA polymerase (T7 RNAP), can potentially carve new paths to study gene regulation. Successful attempts to express the T7 RNAP have been carried out in yeast and also in other higher eukaryotic cells. However, no protein product was detected, due to the lack of a critical editing step in the target mRNA: G-capping on the 5' end of the mRNA. Hence, a capping enzyme, coupled to T7 RNAP activity, is needed for the translation of the T7-generated transcripts in yeast.

I decided to combine the T7 RNAP and a viral capping complex called D1-D12 capping complex, from the *Vaccinia* virus, in order to translate the green fluorescent protein (GFP) target gene. D1-D12 complex is a heterodimer enzyme: D1 is the large subunit and is responsible for the mRNA capping, while the small subunit (D12) plays a role as a stabilizer for D1 activity. It was also found out that this complex is a transcription factor in the *Vaccinia* lifespan. Two systems (plasmid and integration based systems) were designed in order to investigate the effect on GFP expression. In addition, for each system two 3' untranslated regions (UTRs) for the GFP transcription (i.e. a yeast terminator and native T7 terminator) were tested and compared in order to evaluate the terminators effect on the gene expression.

The current results show that in a synthetic expression system lacking the capping complex, T7 RNAP is expressed and active in the yeast nucleus. Accordingly, high levels of GFP mRNA were detected in the cells. However, GFP fluorescence levels were only somewhat elevated as compared to the wild type. When incorporating D1-D12 complex into the system, the GFP fluorescence levels were increased dramatically, especially in cells where the GFP's terminator is derived from a yeast source. An astonishing finding in this synthetic system is the high GFP fluorescence levels when T7 RNAP is absent, while D1 and D12 are expressed in the yeast. This implies that perhaps GFP transcription initiates from a weak yeast cryptic promoter, located somewhere upstream the GFP gene. I suggest that the presence of D1-D12 complex facilitates transcription, by stabilizing the yeast RNAP at the cryptic promoter site.

List of symbols and abbreviations:

Symbols:

L- Liter

ml- milliliter

µl- microliter

g- gram

mg- milligram

µg- microgram

M- Molar

mM- millimolar

µM- micromolar

Δ- gene deletion

Abbreviations:

OD- Optical Density

RNAP- RNA polymerase

yeGFP- Yeast enhanced green fluorescent protein

DNA- Deoxyribo nucleic acid

RNA- Ribonucleic acid

mRNA- messenger RNA

bp- base pairs

his- histidine

leu- leucine

met- methionine

mRNA- messenger RNA

ura- uracil

WT- wild type

PCR- Polymerase Chain Reaction

Ter- terminator/termination signal

cas- cassette

SDW- sterilized deionized water

1. Introduction

The regulation of both natural and synthetic gene expression systems is a complex process that occurs at the transcriptional, post-transcriptional and translational levels in all living organisms and in yeast in particular [1]. Moreover, despite being highly regulated, gene expression systems are inherently stochastic, or "noisy", and are characterized by "bursty" transcription (i.e. RNA synthesis occurs in "pulses"). Therefore, a population of clonal cells can exhibit substantial phenotypic variation, even though its members carry the isogenic expression system. This biological phenomenon was both described theoretically and characterized experimentally in bacteria by Elowitz *et al.* in 2002. In particular, simultaneous expression of cyan and yellow fluorescent proteins, regulated by the same promoter at two separate sites in the *E. coli*'s genome, showed that expression of the two genes may become uncorrelated in individual cells because of noise in transcription. As a result, the cyan/yellow intensity ratio differs from cell to cell [2]. Similar studies were done in yeast [3] and higher eukaryotic cells [4].

In order to substantially advance our understanding of the regulation of gene expression, novel experimental and theoretical approaches are needed. One type of strategy that can be used to study gene regulation is inspired by a synthetic biology approach. Such an approach can reveal unknown mechanisms by simplifying the natural gene regulatory systems: breaking them into individual biological parts, characterizing them and using these parts to construct synthetic biological systems. In other words, we can assemble synthetic biological circuits, which do not exist naturally, from interchangeable natural parts. By doing so, we can investigate the natural phenomena of gene regulation in all commonly-investigated organisms: bacteria, yeast, mammalian cells and whole organisms [5, 6].

1.1 Natural gene transcription in yeast

In eukaryotes, protein production encompasses three major steps: gene transcription, mRNA editing and translation (some of the proteins also undergo a fourth step, a post-translation modification). Gene transcription depends on assembly of transcription initiation factors and RNA polymerase II (RNA pol II) at the TATA box element in the core promoter, known as preinitiation complex. In general, RNA pol II is responsible for the transcription of protein-coding genes and non-coding RNA (ncRNA) molecules. A yeast core promoter (100-200bp long) includes two main

elements [7, 8]: a transcription start site (TSS), a TATA box, which is found 40-120bp upstream the TSS. Upstream to the core promoter is the 'upstream activating sequence' (UAS) element, which contains a regulatory sequence that enhances transcription by binding of transcription factors and provides the overall strength of the promoter [9, 10]. For the most part, a yeast RNA pol II promoter, typically stretches hundreds of base pairs and includes the UAS region and the core promoter [11]. A schematic structure of a typical RNA pol II promoter is depicted in Figure 1. Gene transcription in yeast is triggered upon the binding of RNA pol II to the core promoter with the help of additional transcription factors (pre-initiation complex) bound in close proximity. This results in promoter melting (20bp downstream TATA) and scanning for the TSS by the RNA pol II, which undergoes phosphorylation on its C-terminal domain (CTD) [11, 12].

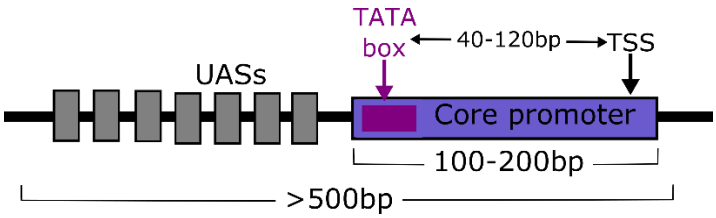


Figure 1: General structure of yeast RNA pol II promoter.

The RNA pol II CTD acts as a processing platform for RNA biogenesis, affecting transcription initiation, polyadenylation and G-capping. The CTD tail consists of 52 repeats of the consensus heptapeptide 'Tyr^{1}-Ser^{2}-Pro^{3}-Thr^{4}-Ser^{5}-Pro^{6}-Ser^{7}' [13] and serves as a docking site for nuclear proteins. The docking site is responsible for the RNA processing. The phosphorylation pattern of the CTD tail determines which nuclear factors bind the CTD platform and coordinates between different processes that the mRNA will undergo during the different stages of transcription [14].

When bound to the core promoter, the RNA Pol II is dephosphorylated. After promoter melting, transcription commences and the CTD is phosphorylated in Ser^{5} positions on the CTD. Once the 5' tail of the RNA emerges (about 20-30 nucleotides) from the holoenzyme [15], the RNA pol II pauses and a complex of several capping enzymes bind to the RNA pol II CTD and add a G-cap to the nascent transcript. This cap is needed for translation initiation and mRNA stability (natural G-capping process in yeast is explained broadly below) [12]. After capping, the RNA pol II CTD becomes massively phosphorylated at Ser^{2} and Ser^{5} positions (hyperphosphorylation) and the RNA elongation resumes [13]. Upon reaching the end of the gene, to a polyadenylation

signal, the transcript is cleaved and a polyadenosine (polyA) tail is added (~200 adenosines) at the 3' end by the polyA polymerase Pap1. As with capping, the RNA cleavage is triggered by changes in the phosphorylation pattern of the CTD (Ser^5 positions are de-phosphorylated)[14, 16]. The polyadenylation signal sequence is a consensus AAUAAA sequence, usually located in the beginning of the terminator and the polyadenylation signal halts the rest of the RNA transcription. Meanwhile, RNA pol II terminates the transcription and is released from the DNA template [13, 17, 18]. The polyA tail has several key roles: (1) it is required for nuclear export to the cytoplasm, (2) it prevents and stabilizes the mRNA degradation by RNases and (3) it facilitates translation initiation by mRNA circularization mediated by interactions with the G cap and polyA binding proteins [16, 18, 19]. Further explanation on transcription termination is given in subsection 1.4.

1.2 T7 RNA polymerase transcriptional activity in various types of cells

In this section I review what has been already done in the past regarding T7 RNAP expression in different hosts.

1.2.1 *E. coli* bacteria

Most of the studies regarding T7 RNAP expression systems have been carried out in *E. coli*, for recombinant protein production purposes in laboratories, industry and pharmaceutical uses. The well-known BL21 (DE3) strain and its derivatives are highly popular strains for heterologous protein overproduction [20]. Naturally, T7 bacteriophage is targeted against *E. coli* bacterium, making it easier to plan such an expression system [21, 22]. This system is known as the "pET system" and includes the following: the T7 RNAP gene, controlled by the inducible *lac* promoter at a genomic site within the *E. coli* genome and a *lac*I repressor that binds to lac promoter and inhibits T7 RNAP gene transcription. T7 RNAP gene transcription is triggered by external addition of isopropyl β-D-thiogalactoside (IPTG), an inducer that binds *lac*I repressor and relieves the repression. In addition, a gene of interest (the target gene) is regulated by a T7 promoter, an RBS (ribosome binding site) element and a T7 termination signal. For the most part, the target gene is cloned into a plasmid [22]. T7 RNAP exclusively recognizes its promoter and transcribes the target gene [23]. Interestingly, it was found out that T7 RNAP activity is eight times higher than *E. coli*'s RNAP, allowing higher gene transcription yield. The host's translational machinery binds to the RBS element and synthesizes the desired protein in bulk [22].

1.2.2 Mammalian cells

Early expression studies of T7 RNAP in *ex-vivo* mammalian cells began in parallel with preliminary applied expression systems studies in *E. coli* [21, 22, 24].
In the first published study in mammalian cells [24], T7 RNAP was expressed using the *Vaccinia* virus in two fashions: 1) T7 RNAP gene was cloned to a plasmid, and 2) T7 RNAP gene was integrated into the *Vaccinia* genome. In both fashions, the inserted gene was controlled by a *Vaccinia* promoter and was transcribed in mammalian cells by the vaccinia RNA polymerase. The target gene was inserted into a plasmid, containing a T7 promoter and a terminator. Then, the plasmid was transfected to the virus-infected CV-1 cells. T7 RNAP activity levels were higher when the T7 RNAP gene was integrated to the *Vaccinia* genome than when it was inserted into a plasmid, implying that the recombinant virus is more stable than the plasmid. The target gene was expressed by the T7 RNAP and was translated in the CV-1 cell line and target protein levels were detected. Since the *Vaccinia* virus is cytoplasmic, T7 RNAP gene is expressed in the cytoplasm by the *Vaccinia* RNA polymerase and remains cytoplasmic. In addition the virus RNA-modifying enzymes are located in the cytoplasm as well, thus T7-generated transcripts potentially could be processed in the cytoplasm (e.g. G-capped), regardless the host's nuclear RNA-modifying enzymes. As a result, the translation of the T7-generated transcripts by the host's ribosomes is possible.

To increase the target gene translation levels in the cells, this system was improved by introducing the target gene into a second recombinant virus. This improved the yield of target gene by 14, as compared to the plasmid-based system, which is less stable system with a relatively low transfection rate [25]. Despite the improvement in transcript levels, analysis of T7-generated transcripts revealed that only 5-10% are capped, possibly due to formation of a 5' stem loop that might interfere with several editing processes: 1) transcript capping by the *Vaccinia* capping enzymes, 2) ribosome binding to the mRNA and 3) ribosome scanning for the start codon, meaning that translation rate was not optimized [26]. To overcome the low percentage of capping, Moss and his collaborators attempted to confer a cap-independent mechanism to translate the target gene by adding a 5' untranslated region (UTR) from *Encephalomyocarditis* virus (EMCV), which contains an internal ribosomal entry site (IRES) within the UTR, allowing cap-independent translation of the target gene [27,

28]. This improved the target gene expression levels by 7 folds, compared to the system lacking the IRES [27].

The thorough and extensive research done in mammalian cells led to the successful development of a transient gene expression system, based on T7 RNAP and *Vaccinia* virus in mammalian cells.

1.2.3 Yeast cells

Attempts to develop a T7-based gene expression have been also carried out in *Saccharomyces cerevisiae* yeast cells. Unlike mammalian cells, yeast cells favor cap-dependent translation, because the 5' cap structure and the 3' polyA end are both required for translation initiation and direct binding of the 40S ribosomal subunits to the mRNA 5' end recruited by cap-binding proteins [29, 30]. There are a handful of examples, in which IRES elements dominate the translation in yeast, instead of a 5' cap structure: 1) 5' non-coding region HAP4 (transcription factor) and TFIID (TATA box binding protein) can be used *in vitro* as IRESs in di-cistronic mRNA translation [31]. 2) An intergenic region from cricket paralysis virus (CrPV) can mediate cap-independent translation when the eukaryotic initiation factor 2 (eIF2) and initiator $tRNA^{met}$ amounts are low in the cells [32]. 3) 5' UTR sequences from invasive growth genes were shown to be effective in cap-independent translation under environmental stress conditions [33]. These cases require special conditions and/or mutations in order to enhance the cap-independent translation of a target gene,which additionally challenges this already complex expression system. Hence, most of the studies performed on T7-generated transcripts in yeast were done under the assumption that translation in yeast is cap-dependent.

Three main questions have been addressed, regarding T7 RNAP expression in yeast cells by previous studies: 1) What is the subcellular localization of the T7 RNAP protein? 2) Do the target gene transcripts accumulate in the cell due to T7 RNAP activity? 3) Do the transcripts undergo proper editing and efficient translation?

To address the first question and in order to increase the possibility to direct the T7 RNAP to the nucleus, Benton and his collaborators [34] fused to T7 RNAP a nuclear localization signal (NLS) of large T-antigen from the simian virus 40 (SV40), under the assumption that T7 RNAP does not contain an NLS for yeast, thus remaining chiefly cytoplasmic. This assumption was tested in mammalian cells using microinjection and transient expression of the T7 RNAP in conjunction with immunofluorescence and was

proved correct [35]. It was shown that SV40 NLS directs the T7 RNAP to the yeast nucleus and target gene transcription by T7 RNAP is increased by 5-10 folds, compared to cytoplasmic T7 RNAP. However, translation of T7-generated transcripts was not detected, even though substantial amounts of target gene transcripts were accumulated in the cells. Moreover, it was demonstrated that T7 RNAP recognizes its specific terminator in yeast cells: the target mRNA's size was equal to the gene's size when T7 terminator was present downstream to the target gene, whereas the lack of T7 terminator resulted in considerably longer transcripts. Regarding the failure in translation, Benton *et al* proposed several plausible reasons for it: the transcripts were not capped, or not polyadenylated, or not exported to the cytoplasm or the 5' end of the mRNA interfered with translation initiation [34].

To address the third question, namely mRNA editing and translation, several studies were done in order to shed light on the specific elements in the RNA needed for translation of T7-generated transcripts in yeast. In one comparative study [29], capped and non-capped single stranded RNA molecules, derived from the luciferase gene, were introduced to yeast cells. The RNAs were synthesized *in vitro* by T7 RNAP in the presence and absence of a 5' cap analog. Luciferase activity was up to ~34 higher for the capped mRNAs than for the non-capped case. Nevertheless, non-capped mRNAs were also translated, but at significantly lower rates, thus reducing luciferase activity. One may infer that a G-cap element at the 5' end of the transcript is essential for efficient and high translation in yeast [29].

In a second comparative study [36], in order to characterize the proper 3' end formation of T7-generated transcripts in yeast, two types of 3' UTRs (polyadenylation or termination signals) of GFP target gene were compared: the native T7 terminator and yeast pol II polyadenylation signal from *TDH3* gene. It was observed previously that yeast RNA pol II may contribute to the formation of 3' end (polyA tail), but is not indispensable for this process, unlike the capping process that is tightly coupled to RNA pol II [37]. For the native T7 termination signal, the GFP mRNAs were not cleaved nor were they polyadenylated and were found to be nuclear. For the *TDH3* polyadenylation signal, the GFP mRNAs were cleaved, polyadenylated and cytoplasmic. Overall, T7-generated transcripts can undergo effective cleavage and addition of adenosine residues the CTD-less T7 RNAP. However, no GFP protein was detected in the system, due to lack of a G cap on the transcripts. This indicates that T7 RNAP cannot recruit the yeast enzymatic capping complex to its transcripts [36].

It should be noted that T7 RNAP was also expressed in yeast mitochondria [38]. As expected, T7 RNAP was enzymatically active in this organelle, where T7 RNAP was fused to a mitochondrial import signal. The mitochondrial target gene under T7 promoter was transcribed but was not translated by the mitochondrial ribosomes [38]. To summarize, translation of the T7-generated transcripts evidently requires the following transcript components: a G-cap structure at the 5' end and a polyA tail at the 3' end. The significance of these components for effective translation and protein synthesis is described in subsections 1.3 and 1.4.

1.3 Natural G-capping in yeast

G-capping occurs selectively on nascent RNA pol II transcripts, at their 5' end. The cap protects and stabilizes the newly-transcribed RNA from degradation by nucleases. It also has several other important roles: providing a binding site for proteins that direct the mature mRNA to the cytoplasm, initiation of mRNA translation and enhancement of mRNA splicing. From these functions, one can understand the great significance of G-capping on eukaryotic cell growth [13, 39]. Soon after transcription initiation, when the transcribed mRNA consists of only 20-30 nucleotides [15] and is accessible to modifications, a G-cap is added to the 5' end in three sequential enzymatic reactions [40]: (1) RNA 5′-triphosphatase (RTase) removes the first phosphate from the first nucleotide in the nascent RNA (2) guanylyltransferase (GTase) adds a guanosine residue (GTP) in an inverted position to the first nucleotide and finally, (3) N_7G-methyltransferase (MTase) methylates the inverted guanosine at position 7 [16]. This process is triggered by a unique phosphorylation pattern of the Ser^5 repeats in the CTD, which recruits the yeast capping enzymes [14]. It was discovered that transcripts generated by CTD-less RNA pol II are inefficiently capped [41]. The capping enzymes physically associate with the CTD tail, allowing the capping process to begin. Once the RNA is capped, the capping enzymes are released from the CTD and RNA pol II moves from the initiation to the elongation mode, while the CTD's phosphorylation pattern changes to a fully-phosphorylated pattern (hyperphosphrylation) at Ser^5 and Ser^2 positions [13, 42].

1.4 Transcription termination in yeast

Transcription termination is the least understood and least characterized mechanism among the three transcription steps (i.e. initiation, elongation and termination) [43]. Termination is a crucial process for survival and development of any

organism. Transcription of a gene ends when the RNA polymerase ceases the RNA synthesis and is released from the DNA template and the nascent RNA is parted from the RNA polymerase. In general, RNA pol II has two main discriminated pathways to end gene transcription: the NNS pathway for termination of non-coding RNA (ncRNA) genes and the CPF-CF pathway for termination of mRNA-coding genes [17]. Here I specifically refer the CPF-CF termination pathway of RNA pol II in yeast. The CPF-CF pathway involves components from the cleavage and polyadenylation factor (CPF) complex and cleavage factor (CF) complex [17].

In general, during the transcription elongation, Ser2 and Tyr1 positions in the RNA pol II CTD tail are highly phosphorylated. Towards the end of transcription, the CTD is dephosphorylated at Tyr1 position, yet it remains predominantly phosphorylated at Ser2 position. This phosphorylation pattern ensures the recruitment of cleavage and polyadenylation factors to the transcript, then the nascent RNA is cleaved and adenosines are added to the RNA 3' tail) [17].

In detail, upon transcription of the polyA site, cleavage and polyadenylation factors are recruited to the transcript's 3' UTR and cause the RNA pol II to pause temporarily on the DNA template. Next, the RNA is cleaved and polyadenylated at the polyA site by the endonuclease Ysh1, which binds to the CPF-CF complex (a structure made of 20 polypeptides). Then, polyadenylation is catalyzed by a polyA polymerase (pap1). Finally, the newly 3' processed mRNA molecule is escorted to the cytoplasm by polyA-binding proteins, either Pab1 or Nab1 proteins. Besides promoting nuclear export, this polyA-binding also provides protection from 3' degradation [17].

1.5 The virus bypass solution

Viruses have adapted throughout evolution in response to changes created by the host cells, in order to defend themselves from viral attacks. One of the mechanisms adopted by some viruses is an independent mRNA capping mechanism. One such virus is the *Vaccinia* virus, which belongs to the poxvirus family. This family is notable for their ability to infect vertebrates and insects [44]. It has the ability to replicate in the cytoplasm of its host, regardless of the host's nuclear replication machinery. Therefore, the *Vaccinia* virus has its own transcription machinery. In order to translate the cytoplasmic virus' mRNAs, the transcripts must be capped. As was noted before, the G-capping process occurs in the nucleus, thus the viral transcripts cannot be capped by the host capping enzymes. Because of this, the *Vaccinia* virus encodes a capping

machinery to acquire 5' G caps to its transcripts. This capping machinery is composed of two subunits, called D1 and D12 [45]. D1 is the large subunit (844 residues, 97kDa) and has the capping activity, subdivided into three active domains, which are responsible for: removal of the first phosphate from the 5' end of the mRNA (RTase), capping of the 5' with guanosine (MTase) and cap methylation (GTase). The D12 subunit is smaller (287 residue, 33 kDa) and does not have any capping abilities, but it has a great stabilizing effect on the MTase activity of D1. Without this association with the D12 subunit, the MTase activity decreases by 30-50 times, compared to the intact MTase-D12 complex [46].

This viral native solution to cap its transcripts in the cytoplasm, outside the nucleus where the host's capping enzymes are located, enables the virus to be independent, regarding the mRNA editing after transcription. As a result the viral mRNA molecules can be translated by the host's translational machinery in the cytoplasm.

Interestingly, besides the capping role, the cytoplasmic D1-D12 capping complex plays a role as a transcription factor in transcription termination of the viral early genes [47] and in transcription initiation of the viral intermediate genes [48].

In this work, we address the problem of translation of T7-generated transcripts by combining the T7 RNAP transcription activity together with D1-D12 capping complex. We therefore hypothesize that D1-D12 complex will cap the T7-generated transcripts, allowing more efficient translation of it in the yeast's cytoplasm.

2. Research Objectives

In this thesis, I focus on the development of a synthetic system, based on the T7 RNAP and viral proteins expressed *in vivo* in the yeast *Saccharomyces cerevisiae*. In order to successfully develop such a system, I have set the following research objectives:

2.1 Design and development of a synthetic expression system, based on T7 RNAP, to transcribe a target gene (a fluorescent protein) in yeast cells, including its translation by the yeast's translation machinery.

2.2 Successful transplantation in yeast of human viral elements for gene transcription and post-transcriptional modifications, needed for efficient translation, using "a bottom-up" approach:

2.2.1 First, I would like to identify the irreducible components needed for independent transcriptional and post-transcriptional modifications in yeast of a T7-generated transcript. In particular, I will test whether the D1-D12 capping complex can contribute to the synthetic system as a capping modulator. Moreover, different 3' UTR sequences will be tested and compared from mammalian and yeast sources.

2.2.2 Second, constructing and characterizing a minimal T7 RNAP-based synthetic gene expression circuit, while taking into account the following components: the 5' UTR sequence, polyadenylation signal, type of terminator (T7 termination signal or yeast terminator) for the target gene and the type of promoter for each component (inducible or constitutive).

3. Materials and Methods

3.1 Reagents, growth media and kits

Enzymes:

All enzymes (restriction enzymes, ligases and polymerases) were purchased from New England Biolabs (NEB).

Bacterial growth media:

Luria-Bertani (LB): 1%Bacto™ Tryptone (Becton Dickinson), 0.5% Bacto™ Yeast Extract (Becton Dickinson), 1%NaCl™ (Merck).

For agar plates: 1.5% Bacto™ Agar (Becton Dickinson).

Super Optimal Broth (SOB): 2% Bacto™ Tryptone (Becton Dickinson), 0.058% NaCl™ (Merck), 0.5% Bacto™ Yeast Extract (Becton Dickinson) and 0.019% Potassium Chloride (Merck).

For recovery after bacterial transformation, the following materials were added to SOB: 1% 1M $MgSO_4$ (Merck), 1% 1M $MgCl_2$ (Merck) and 2% 1M D-(+)-Glucose (Sigma-Aldrich).

Antibiotic- Antibiotic for bacterial selection: 0.1% Ampicillin (Sigma-Aldrich).

Yeast growth media:

Yeast Extract Peptone-Dextrose (YEPD): 1% Bacto™ Yeast Extract (Becton Dickinson), 2% Bacto™ Peptone (Becton Dickinson), 2% D-(+)-Glucose (Sigma-Aldrich).

Synthetic Defined (SD): 0.17% Difco™ Yeast Nitrogen Base w/o Amino Acids and Ammonium Sulfate (Becton Dickinson), 5% Ammonium Sulfate (Merck), 0.14% Yeast Synthetic Drop-out medium Supplements without histidine, leucine, tryptophan and uracil (Sigma-Aldrich). Carbon source- varies, depending on the experiment's purpose: 2% D-(+)-Glucose (Sigma-Aldrich), 2% or 1% D-(+)-Raffinose pentahydrate (Alfa Aesar). For inductive medium: 0.5% or 2% D-(+)-Galactose (Acros Organics).

For agar plates: 1.5% Bacto™ Agar (Becton Dickinson).

Amino acids: Added as supplements to the yeast growth media to a final concentrations of: 20mg/L L-Histidine, 80mg/L L-Leucine and 20mg/L Uracil. All were purchased from Sigma-Aldrich.

Kits:

- NucleoSpin Plasmid Easy Pure Kit (Macherey-Nagel) for plasmidial DNA extraction and purification.

- Wizard® SV Gel and PCR Clean-Up system (Promega) for DNA purification from gels and in-vitro enzymatic reactions.
- Wizard® Genomic DNA Purification Kit (Promega) for both bacterial and yeast genomic DNA isolation.
- Hylab Taq Ready Mix (2X) for bacterial colony PCR.
- Thermo-Fisher DreamTaq PCR Master Mix (2X) for yeast colony PCR.
- TURBO DNA-free^{TM} Kit (Ambion by life technologies) for removal of DNA contamination after total RNA isolation from yeast cells.
- High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) for cDNA synthesis.
- Fast SYBR® Green Master Mix Kit (Applied Biosystems) to perform real-time PCR.

3.2 Bacterial strain and growth conditions

For cloning purposes, the bacterial strain that was used throughout this study was the *E. coli* TOP10 strain (Genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$). Purchased from Invitrogen Corporation (Life Technologies), Carlsbad, CA, USA. After plasmid introduction to the bacteria, the TOP10 cells were grown on LB plates at 37°C with appropriate antibiotics.

3.3 Yeast strain

To test our synthetic biological system described previously, we used the engineered S288C laboratory strain BY4741 (Genotype: MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). By4741 strain was generously given by Yoav Arava, Technion – Israel Institute of Technology, Haifa, Israel.

3.4 Vectors

pUG34 vector- A 6.3kb shuttle vector used for both *E.coli* and yeast *S. cerevisae*. It was kindly obtained from Orit Hermesh, University of Tübingen, Germany. The original pUG34 map is shown below, in Figure 2A. The expression cassette of T7 RNAP (and other derivatives, will be explain extensively below) was constructed into the pUG34 plasmid. Thus, this vector was greatly modified during this project in three sequential steps: First, insertion of simian virus 40 nuclear localization signal (SV40 NLS) and an

MCS linker upstream to yeast enhanced green fluorescent protein (yeGFP) gene. Second, substitution of MET17 promoter in positively regulated GAL1 promoter. Lastly, T7 RNAP gene was added in between the yeGFP and the NLS sequences. These cloning steps were performed using the Gibson Assembly method [49] and verified by sequencing. The resulted plasmid was named pGSVT7e (shown in Figure 2B), also an SV40 NLS deficient plasmid, named pGT7e (Figure 2C), was generated by reverse PCR on the pGSVT7e vector. pGSVT7e and pGT7e vectors were generated in order to prove that the SV40 NLS imports the T7 RNAP protein into the cell nucleus, where it should transcribe its target gene, under T7 promoter.

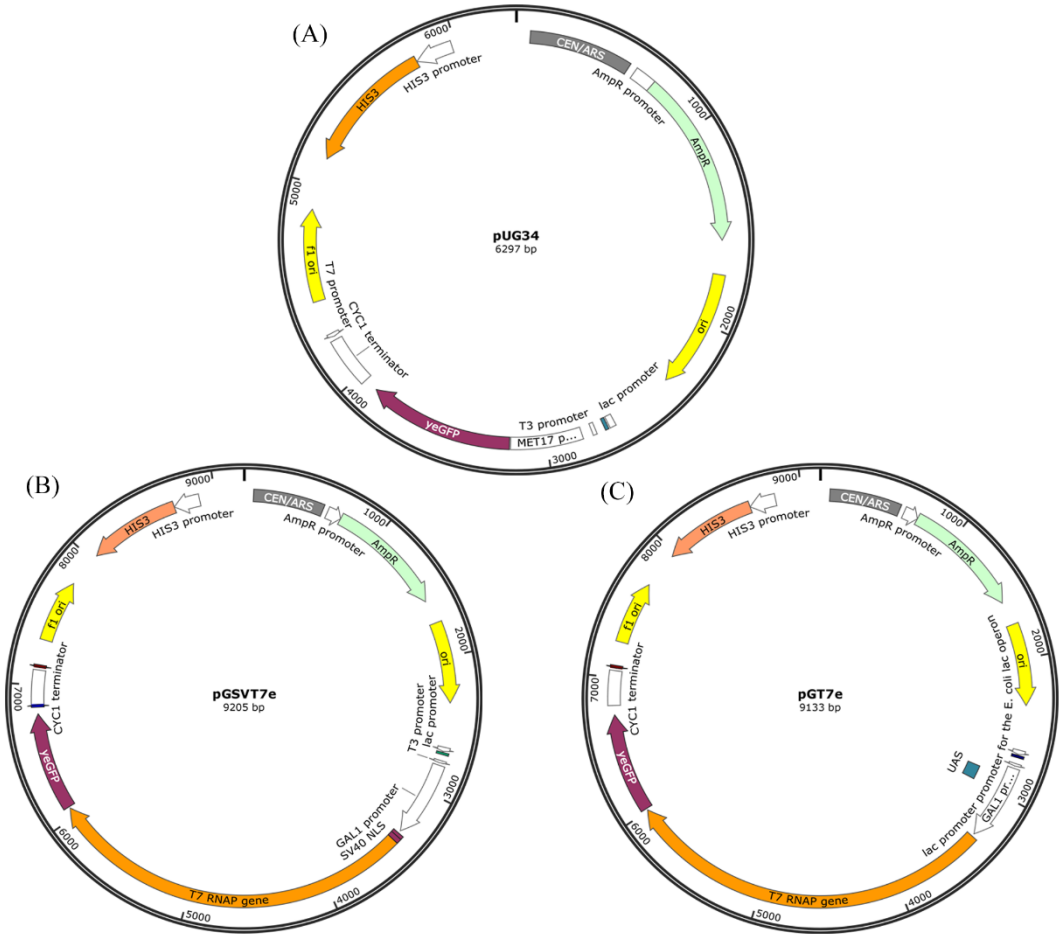


Figure 2: pUG34 derivatives

(A) *Original pUG34 map. The reporter's cassette contains: yeGFP reporter gene expression is controlled by the negatively regulated MET17 promoter and terminated by CYC1 terminator. In presence of growing concentrations of methionine, yeGFP expression is repressed.* ***(B)*** *pGSVT7e map. yeGFP expression is regulated by the GAL1 promoter, meaning that in presence of galactose the expression is induced. SV40 NLS imports the T7 RNAP- yeGFP fusion protein into the cell nucleus.* ***(C)*** *pGT7e map. T7 RNAP-yeGFP fusion protein is expressed by the GAL1 promoter, lacking the SV40 NLS.*

Three other vectors were created, based on pUG34 backbone: $pGSVT_7$, $pGSVD_{12}$-T_7 and $pGSVD_{12}$. The first vector express T7 RNAP gene, the second generates the D12-T7 fusion protein while the latter generates solely the D_{12} and considered as negative control (without T7 RNAP) for the target gene's expression (GFP). The vectors' maps are shown in Figure 3. Table 1 summarizes the genes used for this plasmid's cloning and their source.

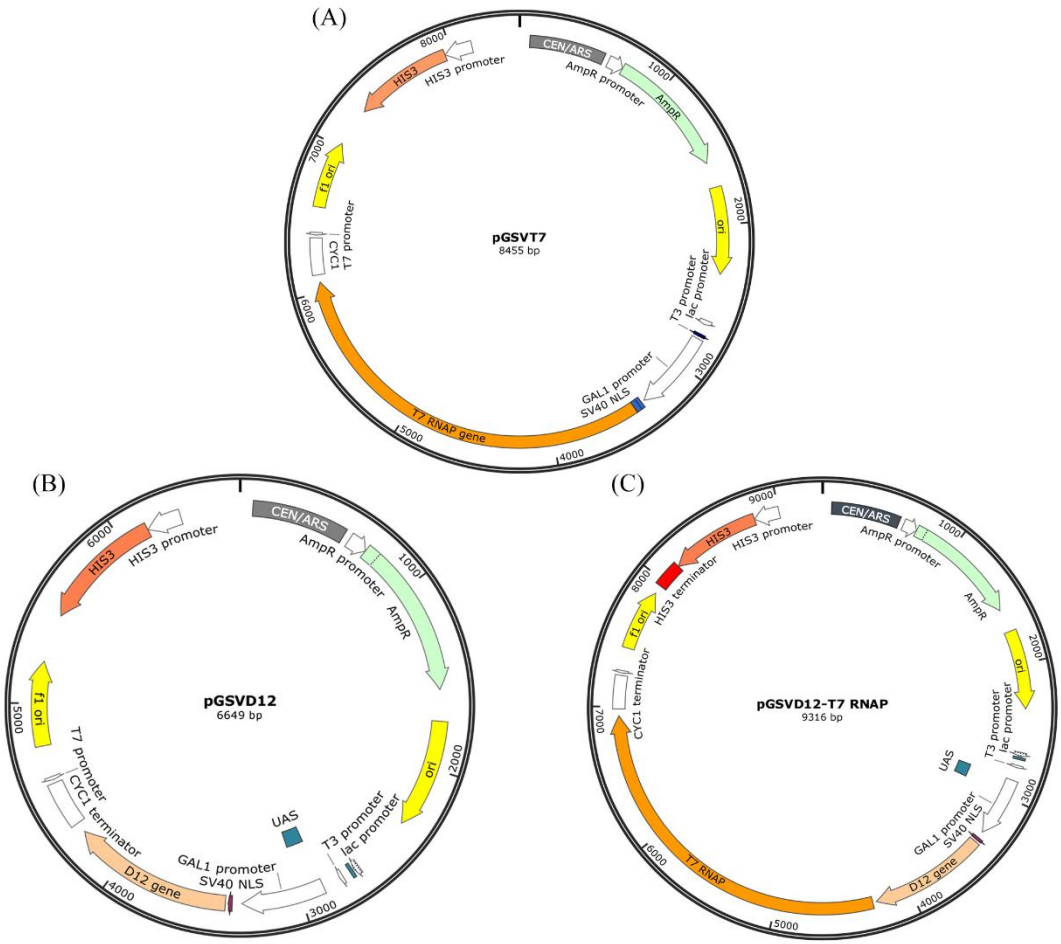


Figure 3: Derivatives of T7 RNAP expression vectors.

All three expression vectors have identical yeast regulatory elements: GAL1 promoter and CYC1 terminator. The vectors differ in the expressed genes. ***(A)*** *Expression of T7 RNAP as single protein, without D12 subunit in the system.* ***(B)*** *Expression of the D12 small subunit from the Vaccinia capping complex as a single protein, without T7 RNAP in the system.* ***C)*** *Expression of the D12-T7 RNAP as a fusion protein.*

#	Gene	Cloning method	Source of sequence
1	GAL1 promoter	Gibson assembly	Kindly provided by Yoav Arava's lab, Techion, Israel
2	SV40 NLS	Gibson assembly	[34]
3	T7 RNA Polymerase	Gibson assembly	*E. coli* BL21(DE3)
4	D12	Restriction enzymes	*Vaccinia* virus. Gene was kindly provided by Stephen Cusack's lab, France

Table 1: Genes cloned into pUG34 derevatives.

p416-met25-(atg)mcp-mCherry vector- A 6.5kb shuttle vector used for both *E.coli* and yeast *S. cerevisae*. This vector was kindly provided by Mordechai Choder, Technion – Israel Institute of Technology, Haifa, Israel. p416 vector's map is shown in Figure 4A. Table 2 summarizes the genes used for this plasmid's cloning and their source.

#	DNA fragments	Source of sequence	#	DNA fragments	Source of sequence
1	T7 promoter	[29]	7	D1	*Vaccinia* virus. Gene was kindlycer provided from Stephen Cusack's lab, France
2	T7 terminator	[23]	8	SV40 NLS	[34]
3	Triplex	[50, 51]	9	CYC1 terminator	pUG34 plasmid
4	ADH2 5' UTR	*Pichia stipites* yeast	10	TEF1 terminator	*Saccharomyces cerevisiae*
5	yeGFP	pUG34 plasmid	11	TPS1 terminator	*Saccharomyces cerevisiae*
6	ADH1 promoter	*Saccharomyces cerevisiae*			

Table 2: DNA fragments cloned into the p416 derevatives.

This vector has several derivatives (will be explained in greater detail below), all have a common cassette comprising of: T7 promoter upstream to the target gene (yeGFP) with an ADH2 5' UTR and one out of three terminators: 1) Triplex-T7 terminator (Figure 4B) 2) yeast CYC1 terminator (represented in Figure 4C) and 3) yeast TEF1 terminator (map not shown). The selection of the best terminator is explained in detail in the results section. The vector also has a D1 expression cassette, which includes the

constitutive ADH1 promoter, D1 gene and CYC1 terminator (Figure 4B). Later on this terminator was substituted to TPS1 terminator (map not shown). These plasmids were constructed using the Gibson assembly method.

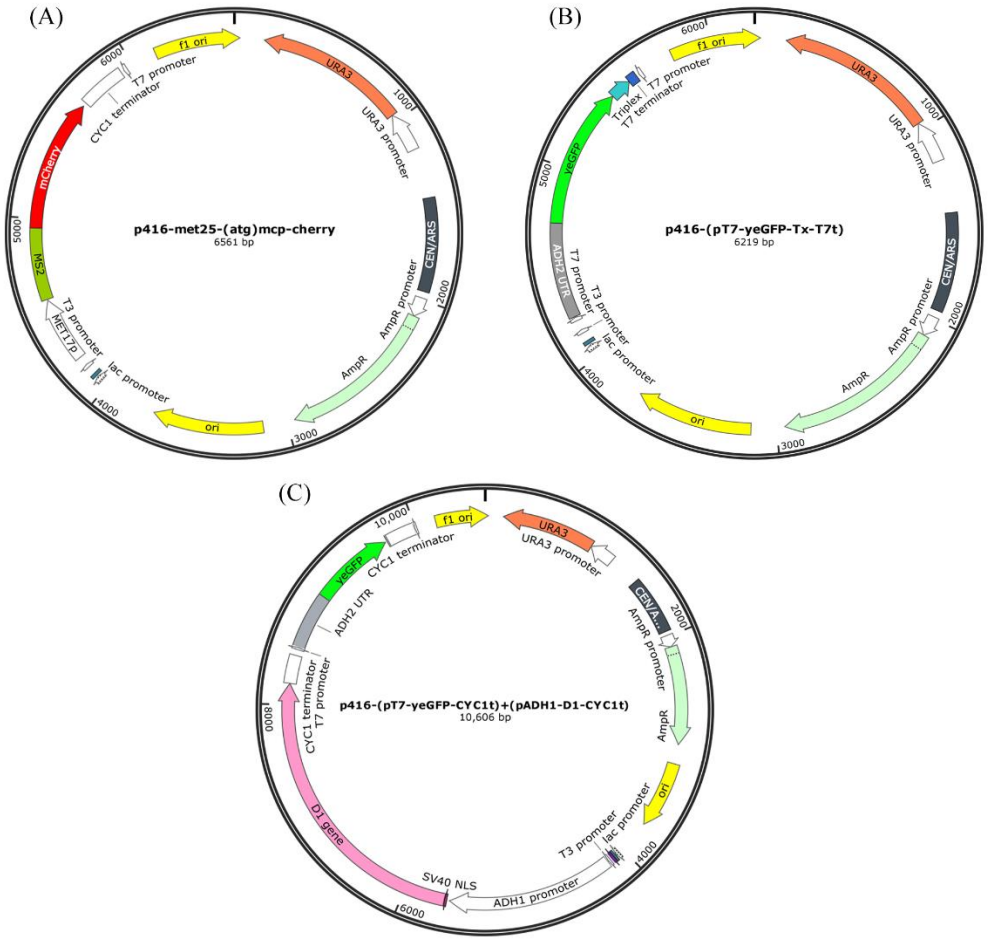


Figure 4: p416 derivatives.

(A)*Original p416 vector. MS2-mCherry gene expression is regulated by MET17 promoter and terminated by CYC1 terminator.* ***(B)*** *The target gene generator plasmid-yeGFP expression cassette includes T7 promoter, an ADH2 5' UTR and a triplex sequence and native T7 terminator.* ***(C)*** *The target gene generator plasmid- yeGFP expression cassette includes T7 promoter, an ADH2 5' UTR and CYC1 terminator. Also D1 subunit is expressed under the constitutive yeast promoter ADH1.*

pUC19 vector- A small bacterial vector, 2690bp in size that was used as an intermidiate cloning vector for the integrated fragments to yeast cells (will be explained in detail at the chromosomal integration section). This vector is commonly used in Roee Amit's lab for cloning in bacteria (map not shown).

3.5 Yeast growth conditions and transformation procedure

Growth conditions:

BY4741 strain was grown appropriate medium plate at 30°C overnight. A 10 ml starter from one grown colony was made, the starter grew overnight before diluting the cells in a fresh liquid medium, for inductive conditions: a medium contained galactose was added to the cells. Afterwards, the cells grew until OD_{600}~0.4-0.8 for yeast transformation or until OD_{600}~0.6 for other purposes.

Yeast transformation: Yeast cells were grown as described above and were harvested by centrifugation for 4 minutes at 4000rpm, washed with 20ml sterilized deionized water. Then, 1 ml of lithium acetate 0.1M (LiAc) was added, centrifugation for 3 minutes at 3000rpm, pellet was re-suspended in 0.1 lithium acetate. Subsequently, for each 100µl cells, the cells were centrifuged and 40ul sdw, 36ul 1M LiAc, 5ul of boiled single stranded DNA (10mg/ml salmon sperm ssDNA, D7656, Sigma-Aldrich), 4ul of vector (for episomal transformation) or linear fragment (for genome integration) and 240ul 50% poly ethylene glycol were added. Cells were incubated at 30°C for 30 minutes and then transferred to 42°C for 15 minutes. After the heat shock at 42°C, cells were spun down for 3 minutes at 3000 rpm and washed with sdw. Lastly, the cells were plated on appropriate plates and incubated at 30°C for at least two days.

3.6 Fluorescence microscopy experiments

Induced yeast cells were grown as described previously until cells were reached a mid-log phase. Cells were grown in 2% galactose and 1% raffinose containing media.

Prior observation under the microscope yeast cells were stained with a DAPI dye (4′,6-Diamidine-2′-phenylindole dihydrochloride, D9542, Sigma-Aldrich), according to a quick DAPI staining protocol: cells were fixed in a 3.7% formaldehyde solution at room temperature for 2 hours. After fixation, cells were spun down at 4°C, 2000rpm for 2 minutes and were washed twice with PBSX1 solution. 300µl of PBSX1 were added to the cells and a 70% concentration was reached by adding 700µl absolute ethanol. After

a 40-minute incubation on ice, the chilled cells were spun down as previously and were re-suspended in 1 ml of PBSX1. Light sonication was carried out (7 pulses at 1.5 output and 35% duty cycle) while keeping the cells on ice. Equal volumes of cells and DAPI (stock concentration is 100ng/ml) were mixed in a separate tube and kept on ice until observation. Meanwhile, slides were prepared according to the protocol described by Young *et al.* [52]. In brief, 1.5% (weight/volume) low-melt agarose (SeaPlaque™ GTG™ Agarose) was added to Phosphate Buffered Saline and dissolved by microwave. After a few minutes of cooling, 1 ml of agarose was pipetted onto a 22-mm^2 cover glass slide. A second cover glass was placed on top of the agarose to create an agarose sandwich in between the two slides. These pads were then left to solidify at room temperature for 30 minutes. Once the cells were ready for imaging, the pads were uncovered and cut into smaller pieces, using a sterile scalpel. 3-4 µl of cells were pipetted on each piece and left to dry at room temperature for 20 minutes. Next, pads were flipped onto a cover glass–bottom dish, with the yeast sandwiched between the agarose pads and the cover glass. Lastly, cells were taken for observation to the inverted Nikon ECLIPSE Ti microscope (purchased from Agentek).

Image visualization was performed by an Andor Xion-Ultra EMCCD camera for obtaining the images presented in this thesis. Acquisition software: NIS-Elements Microscope Imaging Software. Images were edited for color enhancement, contrast and brightness in PhotoShop.

3.7 Chromosomal Integration to the yeast genome

In order to eliminate a possible homologous recombination between the two used plasmids, one cassette was integrated to the yeast genome. There are 4 cassettes, each of them includes either the D12, T7 RNAP, D12-T7 RNAP fusion protein, or T7 RNAP-yeGFP fusion protein, all with SV40 NLS regulated by the GAL1 promoter and CYC1 terminator. The latter of which was designed as a positive control for the integration success, for its analysis- cells were observed under the microscope (as depicted in "microscopy experiments" sub-section). HIS3 chromosomal locus, located at chromosome XV, was decided as the integration locus in the genome. PCR reactions were performed in order to amplify the cassette and HIS3 overlapping regions at the cassette's tails. Following this, Gibson assembly reaction was carried out into PstI digested pUC19 vector. A resulting fragment for example is shown in Figure 5. All fragments were confirmed by sequencing. The fragments were amplified using PCR,

resulting in linear fragments, which were individually transformed to WT component BY4714 yeast cells, as done previously (depicted in "yeast transformation" sub-section). After plating, the successfully transformed yeast cells were able to grow on plates without histidine. The colonies were verified by colony PCR with appropriate primers specifically for the insert, the procedure is described in the "yeast colony PCR" sub-section. A single positive colony was picked for the rest of the cloning procedure.

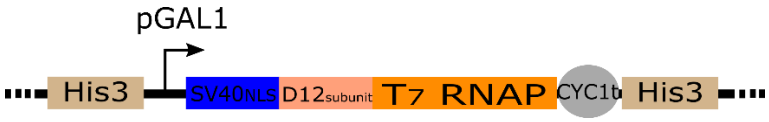


Figure 5: A schematic architecture of the D12-T7 RNAP fragment.
The fragment is flanked by HIS3 complementary sequences for chromosomal integration into HIS3 locus. HIS3 completion confers the yeast the ability to grow on histidine-free plates, once the fragment integrates to the genome. The black arrow denotes the GAL1 promoter, followed by an SV40 NLS and the D12-T7 RNAP fusion protein. CYC1 terminator is represented by a circular shape.

After genome integration to the cells, a second transformation was performed and p416 derivatives vectors were introduced to yeast clone that was previously created. The cells were plated on medium plates without histidine and uracil amino acids.

3.8 Yeast colony PCR

Screening for positive yeast colonies after the genome integration was performed by colony PCR. First, cell wall and inner membranes were disrupted by SDS treatment: a medium-size yeast colony was picked, mixed with 30µl of 0.2% SDS and vortexed vigorously for 15 seconds. Cells were heated to 90°C for 5 minutes and gradually cooled down afterwards to room temperature. After the DNA extraction, cell debris was spun down by microfuge for 1 minute and 1µl was taken for the colony PCR reaction. The crude DNA was stored at -20°C. The colony PCR procedure was carried out following the "Dream Taq PCR Master Mix" manual with an addition of 2µl of 25% Triton X-100. After the PCR reaction, the products were run on 1% agarose gel and positive yeast clones were determined by the amplicon's length.

3.9 Real Time PCR –Relative quantification

To prove T7 RNAP's ability to transcribe its target gene (GFP) in yeast cells under the T7 promoter, we performed real time PCR experiments on purified RNA samples for several yeast clones, induced and uninduced. Overnight grown yeast cells were diluted to OD_{600} 0.1 in 20ml of fresh SD medium with 2% raffinose (non-inductive medium)

or 1% raafinose+2% galactose (inductive medium), both media were supplemented with amino acids. The cells were re-grown to logarithmic phase (OD_{600}~0.6) at 30°C, 250 rpm. A list of the yeast clones, which were analyzed in real-time PCR, is shown in Table 3. Each clone had three biological repeats, in order to obtain statistically significant results in the analysis step.

Yeast clone	**Vector 1 (pUG34 der.)**		**Vector 2 (p416 der.)**	
	D12	**T7 RNAP**	**D1 cas.+ter. signal**	**GFP cas.+ter. signal**
WT	-	-	-	-
T7 RNAP- GFP	-	+	-	+ Tx-T7t
D12-T7 RNAP fusion, GFP+D1	+	+	+CYC1	+CYC1

Table 3: Yeast clones used for real-time PCR experiments.

RNA extraction:

For total RNA isolation, 10ml of cells with medium were spun down (4000 rpm, 4 minutes at 4°C), washed with sdw and re-suspended with 500µl hot phenol lysis buffer (10mM Tris pH 7.5, 10mM EDTA, 0.5% SDS). Equal volume of acidic phenol (P4682, Sigma-Aldrich) was added and samples were incubated at 65°C for 1 hour. Samples were then centrifuged (17,000 g, 10 min at 4°C) and the upper aqueous phase was collected, mixed with equal amount of acid - equilibrated phenol-chloroform (5:1) (P1944, Sigma-Aldrich) and phases were separated again by centrifugation (17,000 g, 10 min at 4°C). The upper aqueous phase was collected, mixed with equal volume of chloroform, centrifuged as above, and collected once more. RNA was precipitated by adding 1 ml 100% cold ethanol and 50µl of 3M sodium acetate (pH=5.2) and incubated overnight at -20°C. Samples were then centrifuged at 17,000 g for 30 min at 4°C and the RNA-containing pellet was washed with 80% ice cold ethanol, centrifuges again at 17,000 g for 15 min at 4°C and re-suspended in ultra-pure water. The concentrations of the samples were determined and 1200ng from each sample, in duplicate, were treated with DNase I (life technologies, TURBO DNA-free kit, AM1907) in a 20µl reaction volume, to avoid any unwanted DNA residues in subsequent reactions. Following the previous treatment, 500ng RNA was examined by 1% denaturing agarose gel to identify

the 18S and 28S ribosomal RNA subunits to assess the RNA integrity. Lastly, the RNA samples were stored at -80C or immediately to subsequent DNaseI reactions.

DNaseI treatment:

To avoid DNA contamination in the total RNA samples in subsequent reactions, 1200ng isolated RNA were subjected to DNaseI treatment using the TURBO DNA-free kit for 30 minutes at 37°C (catalog number: Cat# AM1907, Ambion by life technologies).

Reverse transcription:

To generate cDNA from total RNA, we used the High capacity cDNA reverse transcription kit (Cat# 4368814, Applied Biosystems by life technologies). For this reaction, duplicate of 400ng DNA-free RNA were taken from each sample. The components: RNA, RNase inhibitor, Reverse transcriptase, random primers, dNTP mix and RT buffer (X10) were added accordingly to the manufacturer manual to a final volume of 20µl.

The PCR instrument was programmed as follows:

Step 1 – 25°C, 10 minutes

Step 2 – 37°C, 120 minutes

Step 3 – 85°C, 5 minutes

Step 4 – 4°C

Real time PCR experiment:

After cDNA synthesis, Real time PCR of 5-fold serial dilutions for each inspected gene was performed using SYBR-mix (Applied Biosystems) and a standard curve was generated for the genes. Table 4 lists the analyzed genes in RT-PCR, their designed primer pairs and primer efficiencies. Generally, the primers were designed in primer express software and checked for alignments in BLAST (NCBI), unless otherwise indicated.

Gene name	**Primer pair**	**Primer efficiency**
ACT1 (endogenous control)*	F: GGAAATCACCGCTTTGGCTC R: AACCACCAATCCAGACGGAG	102.7%
T7 RNAP	F: CCTTGCGTTCTGCTTTGAGT R: GACCACCTACCTCATCTCGG	101.6%
Yeast enhanced GFP	F: CGGTGAAGGTGAAGGTGATG R: CGAAAGTAGTGACTAAGGTTGGC	104.7%
D1 subunit	F: ATCATCGACTGGCAGTTTGCTA R: TACCTTGCCTCCAGAAGCAGTT	95.6%
D12 subunit	F: TAAAGCGGACGCCGTAGTTG R: GTCGAAACACGTCGAAGGTTAAC	95.4%

Table 4: Primer pairs used in real-time PCR.

***Primer pair was planned according to life technologies recommendations.*

The standard curve for the ACT1 primers was obtained from the WT strain. The standard curves for the rest of the primers were obtained from the induced samples, strains that potentially express all the genes, particularly the target gene (yeGFP). All examined genes were normalized to the housekeeping gene ACT1.

To ensure accuracy in the real-time PCR assay, three replicates were performed from each sample. Also, no-template and no-RT controls were analyzed for each real-time PCR run.

Appropriate volumes of cDNA, primers and SYBR mix (, applied Biosystems) were added to MicroAmp® Fast Optical 96-well Reaction Plate with Barcode (0.1 ml) (applied Biosystems) and analyzed in the QuantStudio 12K flex real-time PCR system (Life Technologies). The instrument was programmed as follows:

50°C 2min

95°C 10min
95°C 15sec
60°C 1min
} X40 cycles

A different program for the dissociation stage was set as followed:

95°C 15sec

58°C 30sec

95°C 15sec

The generated results from the real time PCR were analyzed using Microsoft Excel 2013 and MATLAB software.

3.10 Flow cytometry

Cloned yeast cells were inoculated overnight (30°C, 250rpm) in SD media with 2% raffinose and appropriate amino acids supplementations. In the morning, the cells were diluted to OD_{600} of 0.1 and incubated at 30°C and 250rpm. When the cells reached OD_{600} of ~0.6, they were kept at 4°C until analysis. Afterwards, cells were spun down at 3000rpm for 3 minutes and re-suspended in PBSX1 on ice. After re-suspension, the cells were analyzed in flow cytometer: the GFP expressing cells were analyzed using BD^{TM} LSR II flow cytometer: excitation at 488nm in conjunction with 530/30 emission filter. Results were analyzed and graphs were generated using the FlowJo software and MATLAB.

A list of yeast clones is shown (Table 5), which describes all cloned yeasts and vector type of each clone (i.e. genes expressed and terminator types).

Yeast clone	Vector 1 (pUG34 der)		Vector 2 (p416 der)	
	T7 RNAP	D12	D1 cas.+ter. signal	GFP cas.+ter. signal
WT	-	-	-	-
T7,GFP-Tx-T7t	+	-	-	+Tx-T7t
T7	+	-	-	-
GFP-Tx-T7t	-	-	-	+Tx-T7t
T7-D12,D1-CYC1t,GFP-Tx-T7t	+	+	+CYC1t	+Tx-T7t
T7-D12,GFP-CYC1t, D1-CYC1t	+	+	+CYC1t	+CYC1t
T7,GFP-Tx-T7t,D1-CYC1t	+	-	+CYC1t	+Tx-T7t
T7-D12, GFP-Tx-T7t	+	+	-	+Tx-T7t
T7-D12, GFP-CYC1t, D1-CYC1t	+	-	+CYC1t	+CYC1t
D12,GFP-CYC1t,D1-CYC1t	-	+	+CYC1t	+CYC1t
GFP-Tx-T7t,D1-CYC1t	-	-	+CYC1t	+Tx-T7t
GFP-CYC1t,D1-CYC1t	-	-	+CYC1t	+CYC1t

Table 5: Yeast clones, containing the two-plasmid based system, analyzed in the flow cytometer.

A list of integrated yeast cells and transformed target genes is shown below (Table 6).

Clone name	Genome integrated cassette (His3 locus)		Target gene (yeast enhanced GFP) generator plasmid	
	T7 RNAP	D12 protein	D1 cas. + terminatio n signal	GFP cas. + Termination signal
WT	-	-	-	-
gT7,D1-CYC1t,GFP-CYC1t	+	-	+CYC1t	+CYC1t
gD12,D1-CYC1t,GFP- CYC1t	-	+	+CYC1t	+CYC1t
gT7,D1-TPS1t,GFP-TEF1t	+	-	+TPS1t	+TEF1t
gD12,D1-TPS1t, GFP-TEF1t	-	+	+TPS1t	+TEF1t
gD12	-	+	-	-
gT7	+	-	-	-
D1-TPS1t,GFP-TEF1t	-	-	+TPS1t	+TEF1t

Table 6: Yeast clones, containing the integration-based system, analyzed in the flow cytometer.

4. Results

4.1 T7 RNAP cellular location in the yeast cells

Nuclear localization signals (NLSs) are widely used to target various proteins (e.g. T7 RNAP) into the cell nucleus in higher eukaryotes in general and yeast cells in particular. SV40 large T antigen NLS from the simian virus was successfully shown previously to direct T7 RNAP to the yeast nucleus [32, 34]. Therefore, in a plasmid, I cloned the SV40 NLS at the N-terminus of T7 RNAP gene and fused a reporter GFP at the C-terminus, downstream to the inducible GAL1 promoter. The plasmid was introduced to the BY4741 strain. A ΔSV40 NLS was cloned as well, as negative control for T7 RNAP nuclear import.

Upon induction with galactose, fixation of the cells and DAPI staining, cells were visualized under the fluorescent microscope (Figure 6). The figure shows that T7 tagged with the SV40 NLS are clearly localized to the nucleus as compared with non-tagged T7, which is spread throughout the cell.

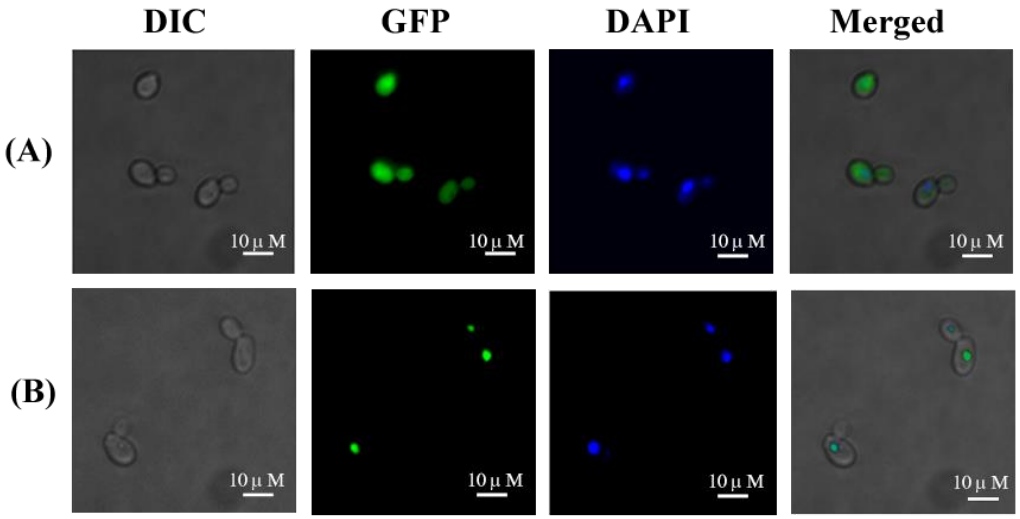


Figure 6: T7 RNAP's cellular location in yeast.
A) *ΔSV40 NLS yeast cells: GFP-tagged T7 RNAP protein is localized in the cytoplasm.* ***B)*** *Cells express GFP-tagged T7 RNAP fused to SV40 NLS: the localization focuses principally in the nucleus, indicating that SV40 NLS does mediate nuclear import of T7 RNAP-GFP fusion protein.*

4.2 Minimal two-plasmid system

In order to produce a protein product from a T7-RNAp treanscript, I first tested the following minimal design made of two vectors: in the first (Figure 3A) regulates,

the *SV40-T7 RNAP* geneis encoded downstream of a GAL1 promoter, while the other vector (Figure 4B) contains a T7 promoter upstream to the GFP target gene and a T7 terminator fused to a sequence called "triplex" [51]. This sequence forms at the 3’ RNA tail a triple helix, that was shown to inhibit degradation in mammalian cells [50, 51]. Since the formation of a triple helix is a sequence-dependent phenomenon, I assumed that triplex formation will occur in yeast cells as well. Therefore, placement of this sequence in the 3’UTR should result in a prolonged life-time of the target mRNA in yeast cells. Furthermore, an *ADH2* 5' UTR was added upstream the gene's location in order to mimic the yeast's native expression cassettes. Both vectors were transformed to component yeast cells, according to the Lithium Acetate technique. The goal of this design was to test the minimal component system, lacking the D1-D12 capping complex, by measuring the GFP fluorescence, derived from non-capped T7-generated transcripts.

4.2.1 Real-time PCR: T7 RNAP and GFP mRNA levels

In principal, T7 RNAP expression is induced only by external galactose addition to the yeast media. Hence, non-induced cells (i.e. absence of galactose in the growth media) should not express T7 RNAP, and as a result the target gene should not be transcribed.

Induced and non-induced cells were first examined by measurement of T7 RNAP and GFP mRNA relative levels using real-time PCR, normalized to a housekeeping gene, the endogenous protein ACT1. ACT1 had a stable expression at the tested growth conditions. As shown in Figure 7, upon induction T7 RNAP expression levels were ~x5 elevated than the non-induced cells. A surprising finding was that non-induced cells do express T7 RNAP in low amounts. Therefore, I concluded that in absence of galactose, the GAL1 promoter had a residual low-level activity (i.e. basal activity), which implies that even in the non-induced conditions a functional the T7 RNAP protein product exists in low titers.

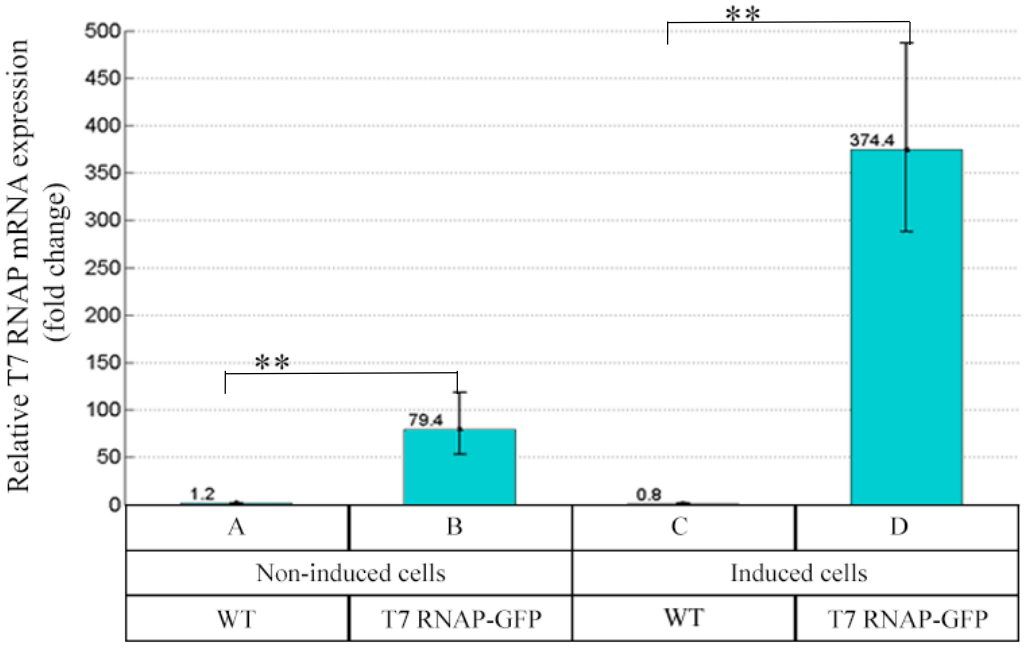


Figure 7: Relative T7 RNAP expression levels.

*For inductive and non-inductive conditions, cells containing T7 RNAP and GFP genes were compared to the WT strain, which lacks the T7 RNAP and GFP genes. Data were normalized to ACT1 reference gene. The real-time PCR analysis for each cell type was based on three independent biological repeats, **PV<0.01.*

As for the GFP mRNA levels, shown in Figure 8, similarly to T7 RNAP mRNA levels, both induced and non-induced cells have accumulated GFP transcripts. These results are consistent with the T7 RNAP presence in the cells, for both cases. Interestingly, induced cells expressed only twice as much target mRNAs, as compared with the non-induced cells. This ratio is markedly different than the x5 ratio observed for the T7 RNA production, and indicates that induced cells have reached saturation in T7-RNAP activity. Given these results, the non-inductive conditions were found to be sufficient for the GFP fluorescence quantification in this system.

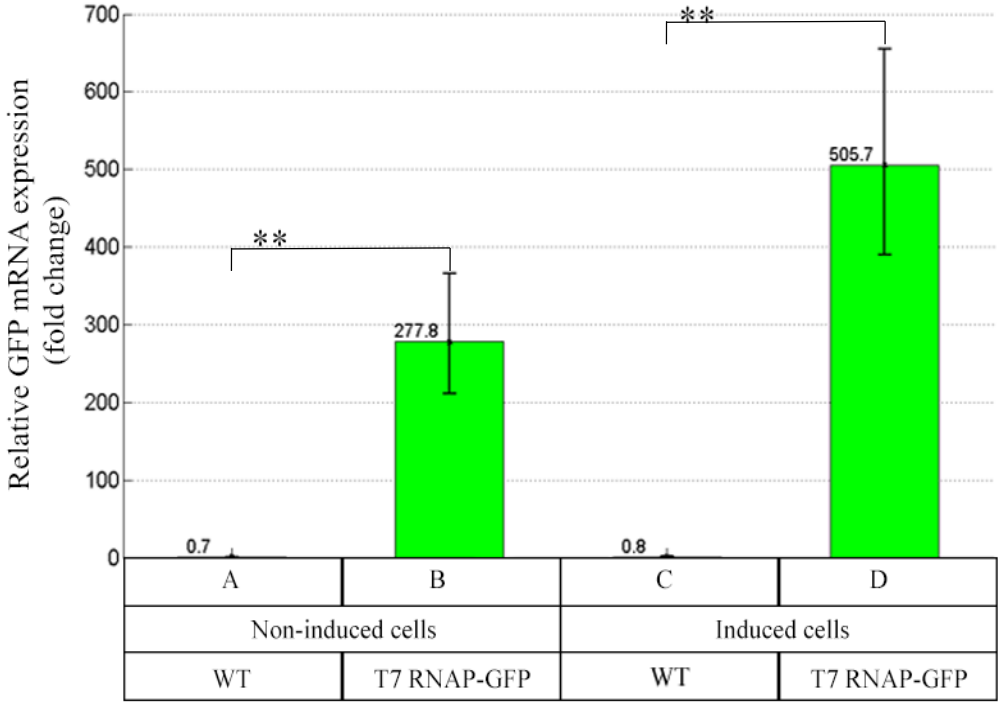


Figure 8: Relative GFP expression levels.

*For inductive and non-inductive conditions, cells containing the T7 RNAP and GFP genes were compared to the WT strain, which lacks the T7 RNAP and the target gene GFP. Data were normalized to ACT1 reference gene. The real-time PCR analysis for each cell type was based on three independent biological repeats, **PV<0.01.*

4.2.2 GFP fluorescence in the minimal two-plasmid system

In order to check if a protein product was generated using this design, I searched for a fluorescent reporter signal using a flow-cytometer in cells grown under non-inductive growth conditions, as mentioned before (Figure 9).

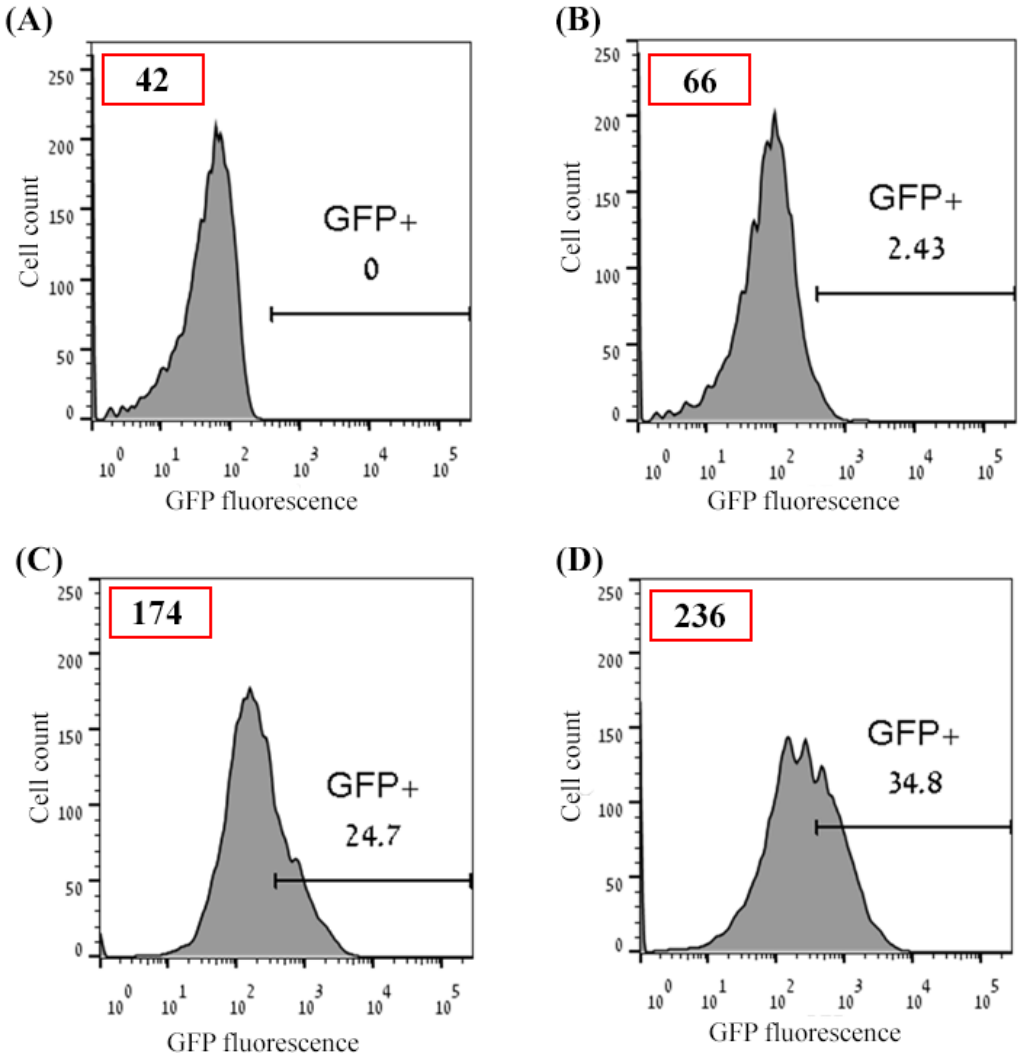


Figure 9: GFP fluorescence in different yeast clones.

For each yeast clone, a representive histogram for the GFP fluorescence is shown. ***A)*** *WT strain- lacking T7 RNAP and GFP genes.* ***B)*** *Cells expressing T7 RNAP only.* ***C)*** *Cells conntaining the GFP expression cassette.* ***D)*** *Cell containing both T7 RNAP and GFP expression cassettes. The flow cytometry analysis is based on three independent biological repeats for each clone. The median value (in arbitrary units- a.u.) for each histogram is denoted in the red-outlined rectangles. Positives cells are ranged equally in all histograms. The number beneath the "GFP+" represents the GFP-positive cell percentage in the analysed population.*

The WT strain (Figure 9A) depicts the GFP autofluorescence of the cells under non-inductive conditions, with a median value of 42. Cells expressing only the T7 RNAP have similar fluorescence median value to the WT cells (Figure 9B). A slight increase was observed in the median fluorescence of cells containing the GFP expression cassette, as shown in Figure 9C, despite the fact that no known yeast promoter is located upstream the target gene. Finally, I observed relatively high

percentage of positive cells (34.8%), shown in Figure 9D, when both expression cassettes (i.e. T7-RNAP and pT7-GFP)were present in the cells. The data shows a distinctly higher median fluorescence expression level as compared with with the WT control. However, the small difference recorded as compared to the strain containing the GFP expression cassette without the T7 RNAP, is indicative of other processes, that may be involved with GFP expression and may be independent of pT7 pathway.
Consequently, I concluded that while the basic initial design may be promising, there was certainly room for improvement.

4.3 The two-plasmid system with the capping components

To the minimal system design described in subsection 4.2 (i.e. T7-RNAP, pT7-GFP-triplex-T7 terminator), I added the D1 and D12 G-capping genes (Figure 3B and Figure 4C). This was based on the hypothesis that if the D1-D12 proteins were added to the cells, non-capped RNA such as the one generated by T7 RNAP will be G-capped, resulting in increased levels of GFP expression as compared with non-capped constructs. To do so, the small D12 subunit was fused to the T7 RNAP gene (Figure 3C) and regulated by GAL1 promoter, whilst the large D1 subunit was cloned in a separate expression cassette into the target gene vector, under a different promoter, the constitutive yeast ADH1 promoter (as shown in Figure 4C). As negative controls, D12 was expressed lacking the T7 RNAP gene and vice versa (Figure 4A and Figure 4B). In addition, to compare the effect on GFP fluorescence, two termination signals for the target gene transcription were tested and were compared to each other. The first termination signal was the native T7 terminator, fused to the triplex, which we used in the initial system. The second terminator was the yeast terminator from the *CYC1* gene, a commonly used yeast terminator in expression systems.

4.3.1 Real-time PCR: D1, D12 and GFP mRNA levels

In order to examine the D1 and D12-T7 RNAP fusion mRNA levels in non-induced cells, compared to the WT strain, a real time PCR experiment was carried out. Figure 10 depicts D1 and D12-RNAP mRNA levels as compared with the WT strain on the left (purple) and right (burgundy) respectively. In both cases the strains expressing the constructs exhibit a significant RT-PCR signature, while the WT control is essentially null. Interestingly, the D1 mRNA level is larger by about x4 than the level of D12 mRNA, fold change of ~ 1032 as oposed to ~233, respectively.

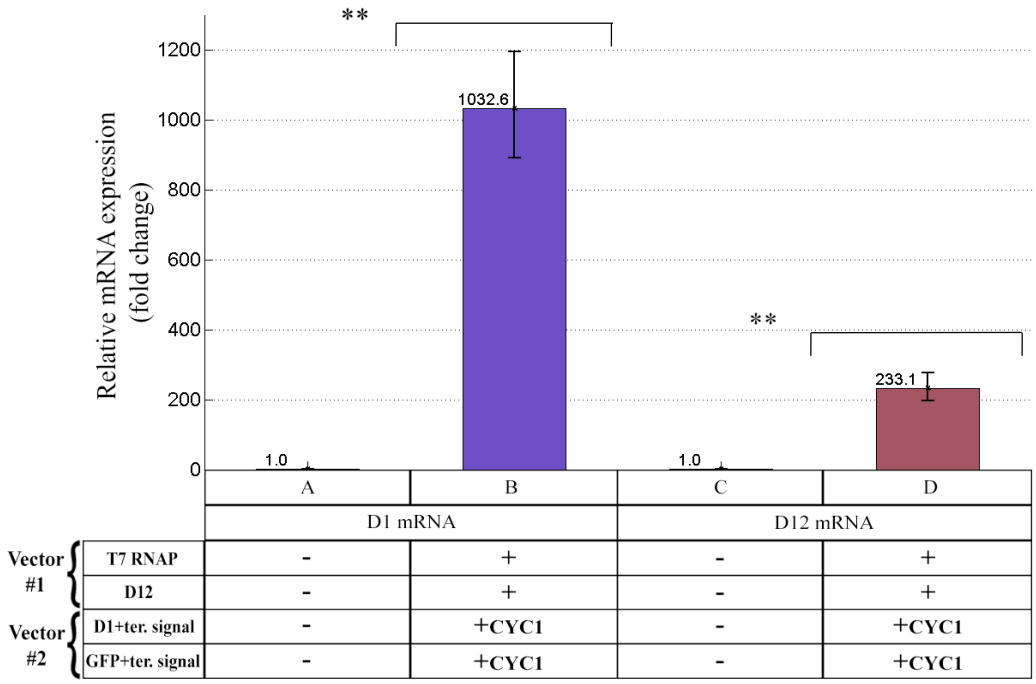


Figure 10: Relative D1 and D12 mRNA levels in the whole two-plasmid system.

*The table is coupled to the bar graph and depicts the genes found in each yeast clone. T7 RNAP and D12 genes are expressed by vector #1, while D1 and GFP genes are expressed by vector #2. For example, (B) represents a yeast clone expressing the T7 RNAP and D12 proteins, and contains the target gene vector: GFP expression cassette and D1 expression cassette, both end with CYC1 terminator. Data were normalized to ACT1 reference gene. The real-time PCR analysis for each cell type was based on three independent biological repeats, **PV<0.01.*

Next, I measured the GFP-fold change in the non-induced cells via real time PCR. The GFP mRNA levels were compared to the WT strain (i.e. containing no GFP or T7-RNAP expression cassettes). I hypothesized that the relative GFP reporter expression will be slightly higher than the GFP expression of the minimal two-plasmid system, described in subsection 4.2.1 (Figure 8), due to the modification of the GFP terminator to CYC1 termination signal, and perhaps the addition of the capping complex might improve the target mRNA stability in the cells.

Given my usage of the same house-keeping gene as base-line for both system, I could utilize this technique to quantitatively compare the mRNA levels in both strains. The results for the real time PCR measurement on the modified system are shown in Figure 11. When compared to the results shown in Figure 8 (non-induced cells), we do not observe a slight increase in mRNA levels, but rather a significant fold change (x16), which indicates that GFP mRNA is much more abundant in these cells as compared with the original design (subsection 4.2).

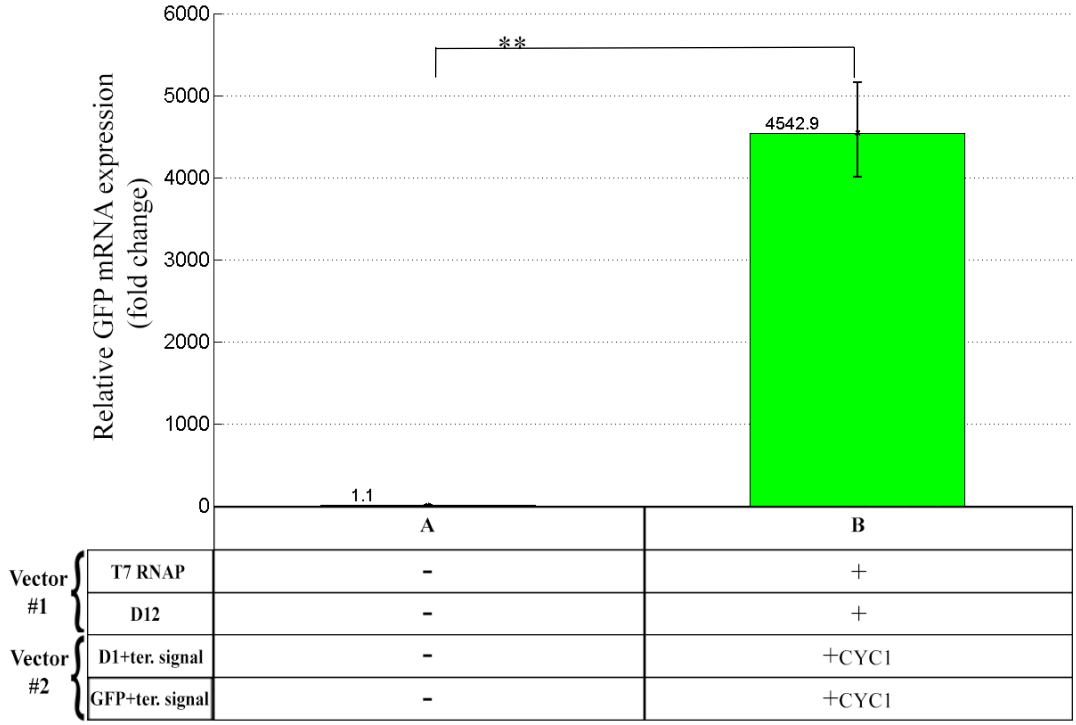
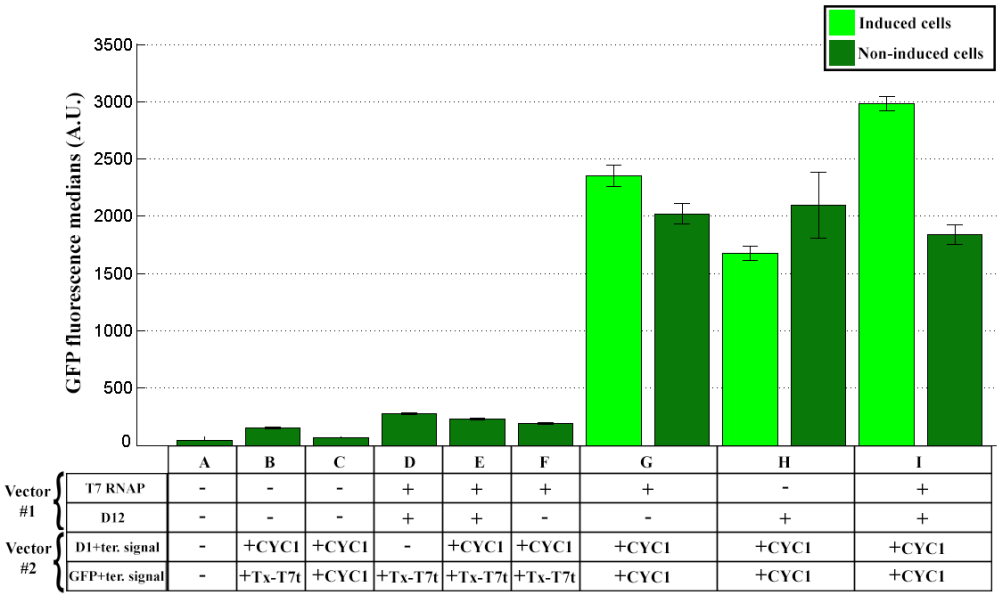


Figure 11: GFP mRNA levels in the whole two-plasmid system.

*The table is coupled to the bar graph and depicts the genes found in each yeast clone. T7 RNAP and D12 genes are expressed by vector #1, while D1 and GFP genes are expressed by vector #2, as in Figure 10. Data were normalized to ACT1 reference gene. The real-time PCR analysis for each cell type was based on three independent biological repeats, **PV<0.01.*

4.3.2: GFP fluorescence in the two-plasmid system with the capping components

Quantification of GFP fluorescence was done by analyzing induced and non-induced cells, containing the whole two-plasmid system, in the flow cytometer. A complete list of the yeast cloned and analyzed in the flow cytometer is presented in Table 5.



		A	B	C	D	E	F	G	H	I
Vector #1	T7 RNAP	-	-	-	+	+	+	+	-	+
	D12	-	-	-	+	+	-	-	+	+
Vector #2	D1+ter. signal	-	+CYC1	+CYC1	-	+CYC1	+CYC1	+CYC1	+CYC1	+CYC1
	GFP+ter. signal	-	+Tx-T7t	+CYC1	+Tx-T7t	+Tx-T7t	+Tx-T7t	+CYC1	+CYC1	+CYC1

Figure 12: GFP fluorescence of yeast containing the two-plasmid system.

The table is coupled to the bar graph and depicts the genes found in each yeast clone. T7 RNAP and D12 genes are expressed by vector #1, while D1 and GFP genes are expressed by vector #2. For example, bar B represents a non-induced yeast clone lacks the T7 RNAP and D12 proteins, but contains the target gene vector: GFP expression cassette (ends with triplex-T7 terminator) and D1 expression cassette (ends with CYC1 terminator).

The results are presented in Figure 12. When only the target gene vector is present in the system, regardless the termination signal, concurrently expressing D1 protein as well, GFP fluorescence is not observed and is similar to the WT strain (Figure 12, bars A-C). In addition, the triplex-T7 terminator element contributes very little to the cells' fluorescence (bars D-F). However, GFP fluorescence increased dramatically when GFP is terminated with CYC1 termination signal when T7 RNAP is present in the system (Figure 10, bars G and I). These results are consistent with the x16 fold change difference in intracellular mRNA titer observed for these systems in the RT-PCR measurement. Surprisingly, the fluorescence did not decline in a system lacking T7 RNAP (Figure 12, bar H). Finally, the induction condition seems to confer only a very slight improvement in GFP expression (Bars G and I), but there is not a definite trend (Bar H).

Based on these observations, it would seem that: first, CYC1 terminator facilitates mRNA export to the cytoplasm. Second, GFP is transcribed by other RNA polymerase, obviously a yeast's type, and as a results the GFP mRNAs can undergo efficient translation.

4.4 The integration-based system with the capping components

In parallel to the two-plasmid system experiments, another approach to express T7 RNAP was tested in order to reduce the number of vectors introduced to the yeast cells, and in order to eliminate possible recombination between the vectors. Therefore, three integration cassettes were introduced separately to the yeast cells: 1) expression of D12 subunit only, 2) expression of T7 RNAP only and 3) D12-T7 RNAP fusion protein, as depicted in Figure 5. The cassettes were sequence-verified by sequencing and were integrated into *HIS3* genomic locus, as detailed in subsection 3.7. Positive clones were screened by genomic DNA extraction and a routine yeast colony PCR, as described in subsection 3.8. However, a clone for the integration of the cassette D12-T7 RNAP fusion has yet to be found. In addition, in order to examine our synthetic system with other terminators, the termination signals for D1 subunit and the GFP target gene were substituted from CYC1 to TPS1 and TEF1 yeast terminators, respectively.

4.4.1 GFP fluorescence in the integration-based system with the capping components

The quantification of GFP fluorescence was carried out by analyzing non-induced cells, containing the integration-based system, in the flow cytometer. I hypothesized that the integration approach of T7 RNAP and D12 genes into the yeast genome might perform better than the two-plasmid based system, in terms of GFP fluorescence and cellular energy balance, as the T7 RNAP protein production would not be as vast as in the two-plasmid based system.

Figure 13 presents the GFP fluorescence for non-induced cells, containing the integration based system.

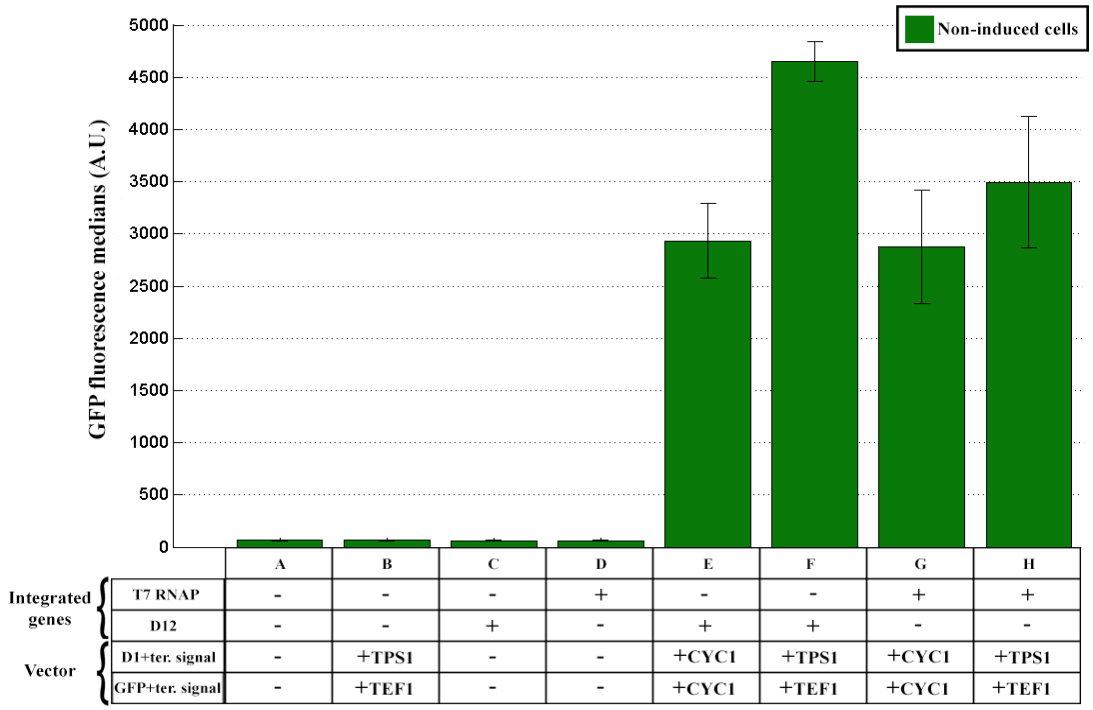


Figure 13: GFP fluorescence of yeast containing the integration based system.
The table is coupled to the bar graph and depicts the genes found in each yeast clone. T7 RNAP and D12 genes are cloned within the yeast genome, while D1 and GFP genes are expressed by the target gene vector. For example, bar H represents a yeast clone lacks the D12 protein, but express T7 RNAP and contains the target gene vector: GFP expression cassette (ends with TPS1 terminator) and D1 expression cassette (ends with TEF1 terminator).

As shown in Figure 13, clones lacking either the reporter vector (containing also D1 expression cassette) or the integrated genes (bars B, C and D), present the same fluorescence as the WT strain (bar A). Cloned yeast cells, containing the target gene vector with D1 expression cassette, and D12 or T7 RNAP gene (bars E-H), displayed high GFP fluorescence levels. Bars E and F represent yeast cells without the T7 RNAP, and again the cells highly express the GFP target gene. These results, together with the one shown in Bar B, indicate that GFP transcription in these designs can likely occur by a native yeast RNA polymerase (likely RNA pol II) found in the cells, provided that D12 and possibly D1 are present in the system. This is consistent with the findings of the two-plasmid system (Figure 12, bars C and H). Moreover, the absence of D12 in the system slightly reduces, if any, the GFP fluorescence (bars G and H). Lastly, as for

the terminators' substitution (bars E-F and G-H), there is no apparent effect on GFP fluorescence for non-induced cells.

5. Discussion

This thesis focused on the development of a synthetic gene expression system, based on T7 bacteriophage RNA polymerase and viral proteins expressed *in vivo* in the yeast *Saccharomyces cerevisiae*. The synthetic biology approach we took to develop such a synthetic system, enabled us to "plug and play" with natural biological parts from several entities (e.g. yeast, viruses etc.) and create a new system, which had not existed previously. Hence, this system was likely to function in a different way than one can predict.

Previous studies, regarding T7 RNAP expression in *Saccharomyces cerevisiae*, taught us what was missing in order to translate T7-generated transcripts. While transcription of an RNA molecule from a T7 promoter is possible and considered as an easy task for scientists, the translation of RNA entails sophistication, creativity and profound knowledge on the natural mechanisms of the host such as mRNA editing, gene expression, in addition to each biological element used to create this synthetic system. I designed, based on present knowledge, two synthetic gene expression systems to ultimately translate the T7-generated transcripts, which included all the necessary components for effective transcription, post-transcriptional modifications, and translation of the target gene. The first was a plasmid based system, in which the T7 RNAP, alone or together with the capping complex, had multiple copies in the cell. In the second system, the T7 RNAP and/or D12 genes were integrated to a certain locus in the yeast genome.

Based on fluorescence microscopy and initial quantitative PCR experiments and analysis, the expression and activity of a functional T7 RNAP in yeast cells were confirmed. T7 RNAP was directed to the nucleus by the well-known SV40 large T NLS element fused to it. Based on previous studies [34, 53], it was proven that in yeast cells, the SV40 large T NLS can mediate nuclear import of many proteins, and T7 RNAP is among them. The GFP mRNA, transcribed by the T7 RNAP, was not able to generate a protein product, hence no significant fluorescence signature was detected. This finding supports past studies in yeast [29, 36, 38], where T7-generated transcripts were observed without a discernable protein amounts as well.

Another implication from our results is that non-induced cells express sufficient T7 RNAP proteins in order to transcribe the GFP target gene. Therefore, this means that in the case of T7 RNAP expression, which is a very active RNAP with high

promoter activity, one can control the T7 RNAP expression with a weak constitutive yeast promoter, such as TEF1 promoter. TEF1 promoter has weak basal activity level at different glucose concentrations [54]. In our synthetic system, TEF1 and other similar-strength promoters are ideal as a regulating promoters for T7 RNAP and D1-D12 capping complex expression cassettes.

When the basic T7 gene expression plasmid-based system was augmented by the D1-D12 5' G-capping enzymatic complex, functional GFP was detected in the cloned cells, at the mRNA as well as at the protein levels. It is possible that the D1-D12 capping complex may have contributed unexpectedly to some of my results. In particular, when T7 RNAP was not present in the cloned strains, while D1 and D12 were present, GFP fluorescence was detected. Moreover, T7 RNAP in concert with D1 without the presence of D12 also generated high fluorescence levels. However, in both the plasmid and integration systems, when only the target plasmid, containing the D1 and GFP expression cassettes, was present in the cells (i.e. without T7 RNAP or D12) no fluorescence was detected. In addition, based on the flow cytometry analysis, the triplex-T7 terminator element contributes very little to the cells' fluorescence, implying that the T7-generated transcripts are nuclear and the triplex-T7 terminator sequence cannot promote nuclear export, as opposed to the CYC1 terminator. This supports Dower and Rosbash's findings [36], explained in subsection 1.2.3 in the introduction. Taken together, these results imply several things. First, the results infer that CYC1 terminator facilitates mRNA export to the cytoplasm. Second, GFP is likely also transcribed by other yeast RNA pol II, and as a result the GFP mRNAs can undergo efficient translation (detailed below in depth). Third, the most important difference inferred on expression was due to the choice made in 3'UTR or terminator sequence. In particular, the total mRNA detected in non-induced cells (Figure 8 and Figure 11) implies that the choice of terminator mostly effects the degradation rate of RNA. This could be due to inefficient export outside of the nucleus for the triplex-terminated mRNA and to a higher sensitivity to break down by RNases, or some combination of both. The current experiments do not allow us to distinguish between the different scenarios.

The integration-based system shows a similar picture to the plasmid-based system, that is to say GFP fluorescence elevates even when T7 RNAP is absent in a system, containing D1-D12 or D12 by itself. Although the clone containing the integrated D12-T7 RNAP fusion is yet to be positively-screened, I have no reason to

believe, based on the previous results, that a different picture will emerge. In addition, the continuation of the flow cytometry and real-time PCR experiments is indispensable in order to deeply understand the synthetic system and how it exactly behaves in the cloned cells.

The current results imply that GFP transcription may have been initiated from some cryptic promoter, or from a normally inactive TATA-TSS combination that got turned on in the presence of at least two of three of the D1, D12, and T7 components. Specifically, the yeast RNA pol II may have bound loosely to some weak pre-initiation complex or transcriptional start site in either the *ADH2* 5' UTR or the T7 promoter of target gene generator plasmid for both the two-plasmid and integration based systems. This weak binding events by themselves are probably not sufficient to induce expression, as the RNAP-promoter complex is likely short-living. However, in the presence of D1-D12 (or somehow with D1 and T7 RNAP) the RNAP-promoter complex may be stabilized for a sufficiently long time to commence transcription and to facilitate G-capping of the transcripts and as a result, to enable subsequent effective translation. Therefore, I suspect that D1-D12 capping complex might potentially have the ability to aid the yeast RNA pol II to bind a cryptic initiator element upstream to the GFP gene, due to the fact that in the absence of T7 RNAP and D1-D12, the GFP fluorescence remains basal. Alternatively, a transcribing T7 RNA polymerase in concert with D1, may be sufficiently strong to keep histones off the target DNA facilitating the binding of a RNA Pol II complex, which can in turn transcribe at the "heals" of the upstream processive T7 RNAP.

To provide further support for this hypothesis, a literature search shows that the D1-D12 capping complex also plays a roll as a transcription factor in transcription initiation and termination of the viral genes. In addition, the D1-D12 complex is thought to be a part of the *Vaccinia* RNAP itself, playing a direct role in the transcription process [47, 48]. Thus, it may be possible that these allegedly secondary rolls of the D1-D12 complex can play a larger role in transcription than originally thought. Additionally, taking a second look at my T7 promoter and *ADH2* 5' UTR regions, I was able to identify the consensus TATA box 'tataaata', at the end of the *ADH2* 5' UTR. This TATA box corresponds with the canonical TATA box: TATAWAWR (where: W=A or T, R=A or G) [11]. In general, The TATA box is the assembly site for the RNA pol II machinery at the core promoter region [55], while the TSS is found in native promoters at a distance of 40–120 bp downstream of the TATA box [7]. Perhaps some potential

weak initiator sequences (TSSs) are located 40-120bp downstream from the identified TATA box. This configuration, together with D1-D12 complex, presumably stabilizing the assembling RNA pol II, may enable the GFP transcription. To resolve this complex issue, further testing must be done by the 5' RACE or RNA-seq techniques. These will reveal the TSS and will help us to investigate in depth the cryptic initiator element, from which GFP is transcribed.

Although the plasmid based system might seem to be easier to clone and can be introduced to the cell simultaneously, the preferred system design is the integration based system. One can say that in order to minimize the total number of vectors in the cell and to reduce the enormous energetic burden on the cells, the preferred design system is the integration based system, in which the T7 RNAP and the capping complex are expressed from a single copy in a specific and well-known locus within the genome. This will hopefully, lessen the energetic overload of excessive amounts of T7 RNAP, D1 and D12 proteins.

Ultimately, I prefer that the T7 RNAP expression will be tightly coupled to D1-D12 expression. Therefore, these genes can be regulated under the same promoter in different expression cassettes, or as a three-component fusion protein, in order to minimize any undesired capping events on the yeast RNA molecules. In addition, it will prevent over-expression of one of the components, as seen in our results, the relative mRNA levels for D1 and D12-T7 RNAP fusion were different from each other.

Another necessary, but missing, part of the system, to this day, is the effect of D1 subunit absence on the GFP fluorescence. D1 subunit is the large subunit in the capping complex and it has the ability to cap the 5' mRNA, while D12 subunit is a stabilizer unit for D1 activity [46]. ΔD1 yeast cells might display a decrease in GFP fluorescence. Alternatively, D12 subunit might be the key player by the recruitment of a yeast RNA polymerase to the target gene proximity, triggering the GFP transcription. In order to address these hypotheses, ΔD1 yeast clone needs to be created and tested in the flow cytometry as well as analysis in quantitative PCR.

Another possible step that can be done, in order to improve our understanding on this synthetic system, is to exchange the 5' UTR sequence upstream the GFP target gene. The 5' UTR affects the mRNA in terms of translation rate, RNA-protein interactions, transcription and degradation rate [56]. Consequently, one can understand that different 5' UTRs may have different effects on the GFP translation rate, and accordingly the GFP fluorescence.

To conclude this thesis, I developed a synthetic gene expression system in yeast cells, based on viral proteins and designated to complete the central dogma of biology from bottom to top: DNA→ RNA→protein, with minimal number of biological components. I succeeded translating the GFP target gene and although there are many new questions remained unanswered, I began to comprehend the synthetic system I created and its unnatural behavior in yeast cells.

6. References

[1] Hartemink, A.J., “Reverse engineering gene regulatory networks,” *Nature Biotechnology* , vol. 23, pp. 554-555, 2005.

[2] Elowitz, M. B.; Levine, A. J.; Siggia, E. D.; Swain, P. S., "Stochastic Gene Expression in a Single Cell," *Science*, vol. 297, pp. 1183-1186, 2002.

[3] Raser, J. M.; O'shea, E. K.;, "Control of stochasticity in eukaryotic gene expression," *Science*, vol. 304, pp. 1811-1814, 2004.

[4] Chubb, J. R.; Trcek, T.; Shenoy, S. M.; Singer, R. H.;, "Transcriptional Pulsing of a Developmental Gene," *Current Biology*, vol. 16, pp. 1018-1025, 2006.

[5] Weiss, R., Andrianantoandro, E., Basu, S., Karig, D.K., “Synthetic biology: new engineering rules for an emerging discipline,” *Molecular systems biology* , vol. 2006, pp. 1-14, 2006.

[6] Sismour, A. M., Benner, S.A., “Synthetic biology,” *Nature Reviews Genetics*, vol. 6, pp. 533-543, 2005.

[7] Redden, H.; Alper, H. S.;, "The development and characterization of synthetic minimal yeast promoters," *Nature communications*, vol. 6, pp. 1-9, 2015.

[8] S. T. Smale and J. T. Kadonaga, "The RNA polymerase II core promoter," *Annual reviews biochemistry*, vol. 72, p. 449–479, 2003.

[9] Hahn, S.;, "Structure and mechanism of the RNA polymerase II transcription mahinery," *Nature Structural & Molecular Biology*, vol. 11, pp. 394-403, 2004.

[10] Hahn, S.; Young, E. T.;, "Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators.," *Geenetics* , vol. 189, pp. 705-736, 2011.

[11] Lubliner, S.; Regev, I.; Lotan-Pompan, M.; Edelheit, S.; Weinberger, A.; Segal, E.;, "Core promoter sequence in yeast is a major determinant of expression level," *Genome Research*, vol. 25, p. 1008–1017, 2015.

[12] Hampsey, M.; Woychik, N. A.;, "The RNA Polymerase II Machinery: Structure Illuminates Function," *Cell*, vol. 108, p. 453–463, 2002.

[13] Reinberg, D., Orphanides, G., "A Unified Theory of Gene Expression," *Cell*, vol. 108, p. 439–451, 2002.

[14] Greenleaf, A. L.; Phatnani, H. P.;, "Phosphorylation and functions of the RNA polymerase II CTD," *Genes & Development*, vol. 20, p. 2922–2936, 2006.

[15] Rasmussen, E. B.; Liss, J. T.;, "In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes," *PNAS*, vol. 90, pp. 7923-7927, 1993.

[16] Shatkin, A. J.; Manley, J. L.;, "The ends of the affair: Capping and polyadenylation," *Nature Structural Biology*, vol. 7, pp. 838-842, 2000.

[17] Libri, D.; Porrua, O.;, "Transcription termination and the control of the transcriptome: why, where and how to stop," *Nature Reviews*, vol. 16, pp. 190-202, 2015.

[18] Proudfoot, N. J.; Furger, A.; Dye, M. J.;, "Integrating mRNA Processing with Transcription," *Cell*, vol. 108, pp. 501-515, 2002.

[19] Sonenberg, N.; Richter, J. D.;, "Regulation of cap-dependent translation by eIF4E inhibitory proteins," *Nature*, vol. 433, pp. 477-480, 2005.

[20] K. Terpe, "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems," *Applied Microbiology and Biotechnology,* vol. 72, pp. 211-222, 2006.

[21] Tabor, S.; Richardson, C. C.;, "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes," *PNAS,* vol. 82, pp. 1074-1078, 1985.

[22] W. F. Studier and B. A. Moffattf, "Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes," *Journal of Molecular Biology,* vol. 189, pp. 113-130, 1986.

[23] J. J. Dunn and F. W. Studier, "Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements," *Journal of Molecular Biology,* vol. 166, pp. 477-535, 1983.

[24] Moss, B., Fuerst, T.R., Niles, E.G., Studier, F.W., "Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase," *PNAS* , vol. 83, pp. 8122-8126, 1986.

[25] Fuerst, T. R.; Earl, P. L.; Moss, B.;, "Use of a Hybrid Vaccinia Virus-T7 RNA Polymerase System for Expression of Target Genes," *Molecular and Cellular Biology,* vol. 7, pp. 2538-2544, 1987.

[26] Fuerst, T. R.; Moss, B.;, "Structure and stability of mRNA synthesized by vaccinia virus-encoded bacteriophage T7 RNA polymerase in mammalian cells: Importance of the 5′ untranslated leader," *Journal of Molecular Biology,* vol. 206, pp. 333-348, 1989.

[27] Elroy-Stein, O.; Fuerst, T. R.; Moss, B.;, "Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5' sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system," *PNAS,* vol. 86, pp. 6126-6130, 1989.

[28] Elroy-Stein, O.; Moss, B.;, "Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells," *PNAS,* vol. 87, pp. 6743-6747, 1990.

[29] Russell, P. J.; Hambidge, S. J.; Kirkegaard, K.;, "Direct introduction and transient expression of capped and non-capped RNA in Saccharomyces cerevisiae," *Nucleic Acids Research,* vol. 19, pp. 4949-4953, 1991.

[30] Preiss, T.; Hentze, M. W.;, "Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast.," *Nature,* vol. 392, pp. 516-520, 1998.

[31] Iizuka, N.; Najita, L.; Franzusoff, A.; Sarnow, P.;, "Cap-Dependent and Cap-Independent Translation by Internal Initiation of mRNAs in Cell Extracts Prepared from Saccharomyces cerevisiae," *Molecular and Cellular Biology,* vol. 14, pp. 7322-7330, 1994.

[32] Thompson, S. R.; Gulyas, K. D.; Sarnow, P.;, "Internal initiation in Saccharomyces cerevisiae mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element," *PNAS,* vol. 98, pp. 12972-12977, 2001.

[33] Gilbert, W. V.; Zhou, K.; Butler, T. K.; Doudna, J. A.;, "Cap-Independent Translation Is Required for Starvation-Induced Differentiation in Yeast," *Science,* vol. 317, pp. 1224-1227, 2007.

[34] Benton, B. M.; Eng, W.; Dunn, J. J.; Studier, F. W.; Sternglanz, R.; Fisher, P. A.;, "Signal-Mediated Import of Bacteriophage T7 RNA Polymerase into the

Saccharomyces cerevisiae Nucleus and Specific Transcription of Target Genes," *Molecular and Cellular Biology,* vol. 10, pp. 353-360, 1990.

[35] Dunn, J. J.; Krippl, B.; Bernstein, K.; Westphal, H.; Studier, F. W.;, "Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus.," *Gene,* vol. 68, pp. 259-266, 1988.

[36] Dower, K.; Rosbash, M.;, "T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export," *RNA,* vol. 8, pp. 686-697, 2002.

[37] McNeil, J. B.; Agah, H.; Bentley, D.;, "Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast," *Genes and Development,* vol. 12, pp. 2510-2521, 1998.

[38] Pinkham, J. L.; Dudley, A. M.; Mason, T. L.;, "T7 RNA Polymerase-Dependent Expression of COXII in Yeast Mitochondria," *Molecular and Cellular Biology,* vol. 14, pp. 4643-4652, 1994.

[39] Buratowski, S.; Fresco, L. D.;, "Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but iis not required for mRNA splicing," *RNA,* vol. 2, pp. 584-596, 1996.

[40] Shuman, S.;, "What messenger RNA capping tells us about eukaryotic evolution," *Nature reviews,* vol. 3, pp. 619-625, 2002.

[41] McCracken, S.; Fong, N.; Rosonina, E.; Yankulov, K.; Brothers, G.; Siderovski, D.; Hessel, A.; Foster, S.; Shuman, S.; Amgen EST Program; Bentley, D. L.;, "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II," *Genes and Development,* vol. 11, p. 3306–3318, 1997.

[42] Cho, E. J.; Kobor, M. S.; Kim, M.; Greenblatt, J.; Buratowski, S.;, "Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain," *Genes and Development,* vol. 15, p. 3319–3329, 2001.

[43] Richard, P.; Manley, J. L.;, "Transcription termination by nuclear RNA polymerases," *Genes & Development,* vol. 23, pp. 1247-1269, 2009.

[44] Harrison , S. C.; Alberts, B.; Ehrenfeld, E.; Enquist, L.; Fineberg, H.; McKnight, S. L.; Moss, B.; O'Donnell, M.; Ploegh, H.; Schmid, S. L.; Walter, K. P.; Theriot, J.;, "Discovery of antivirals against smallpox," *PNAS,* vol. 101, p. 11178–11192, 2004.

[45] Shuman, S.; Surks, M.; Furneaux, H.; Hurwitz, J.;, "Purification and Characterizationo f a GTP-Pyrophosphate Exchange Activity from Vaccinia Virions," *The Journal of Biological Chemistry,* vol. 255, pp. 11588-11598, 1980.

[46] Higman, M. A.; Christen, L. A.; Niles, E. G.;, "The mRNA (Guanine-7-)methyltransferase Domain of the Vaccinia Virus mRNA Capping Enzyme," *The Journal of Biological Chemistry,* vol. 269, pp. 14974-14981, 1994.

[47] Shuman, S.; Broyles, S. S.; Moss, B.;, "Purification and characterization of a transcription termination factor from vaccinia virions.," *The Journal of Biological Chemistry,* vol. 262, pp. 12372-12380, 1987.

[48] Vos, J. C.; Sasker, M.; Stunnenberg, H. G.;, "Vaccinia virus capping enzyme is a transcription initiation factor," *The EMBO Journal,* vol. 10, pp. 2553-2558, 1991.

[49] Gibson, D. G.; Smith, H. O.; Hutchison, C. A.; Venter, J. C.; Merryman, C.;, "Chemical synthesis of the mouse mitochondrial genome," *Nature Methods,* vol. 7, pp. 901-905, 2010.

[50] Nissim, L.; Perli, S. D.; Fridkin, A.; Perez-Pinera, P.; Lu, T. K.;, "Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells," *Molecular Cell,* vol. 54, pp. 1-13, 2014.

[51] Wilusz, J. E.; JnBaptiste, C. K.; Lu, L. Y.; Kuhn, C. D.; Joshua-Tor, L.; Sharp, P. A.;, "A triple helix stabilizes the 3' end of long noncoding RNAs that lack poly(A) tails," *GENES & DEVELOPMENT,* vol. 26, pp. 2392-2407, 2012.

[52] Young, J. w.; Locke, J. C.W.; Altinok, A.; Rosenfeld, N.; Bacarian, T.; Swain, P. S.; Mjolsness, E.; Elowitz, M.;, "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy," *Nature Protocols,* vol. 7, pp. 80-88, 2012.

[53] Muhua, L.; Adames, N. R.; Murphy, M. D.; Shields, C. R.; Cooper, J. A.;, "A cytokinesis checkpoint requiring the yeast homologue of an APC-binding protein," *Nature,* vol. 393, pp. 487-491, 1998.

[54] Partow, S.; siewers, V.; Bjorn, S.; Nielsen, J.; Maury, J.;, "Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae," *Yeast,* vol. 27, p. 955–964, 2010.

[55] Basehoar, A. D.; Zanton, S. Z.; Pugh, B. F.;, "Identification and Distinct Regulation of Yeast TATA Box-Containing Genes," *Cell,* vol. 116, p. 699–709, 2004.

[56] Dvir, S.; Velten, L.; Sharon, E.; Zeevi, D.; Carey, L. B.; Weinberger, A.; Segal, E.;, "Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast," *PNAS,* pp. 2792-2801, 2013.

תקציר

בתאים איאוקריוטים, מערכות ביטוי גנים כוללות מספר תהליכים: שעתוק של הגן מדנ"א לרנ"א, עיבודו של הרנ"א, תרגומו לחלבון ומודיפיקציות שלאחר התרגום שנעשות על החלבון (התהליך האחרון לא מתרחש בחלק מהמקרים). מערכות ביטוי סינתטיות, שאינן קיימות בטבע, לא שונות ממערכות הביטוי הטבעיות. מערכות סינתטיות יכולות להוות מודלים ביולוגיים פשוטים, שישפכו אור נוסף על מה שאינו ידוע כיום בחקר בקרת הביטוי הגנטי ובנוסף, יכולות ללמד על תופעות ביולוגיות שקורות בטבע. ולכן, פיתוח מערכות סינתטיות מעין אלה בשמרים, המבוססות על הביטוי של האנזים T7 רנ"א פולימראז, יכולות ללמד אותנו על בקרת ביטוי הגנים. T7 רנ"א פולימראז שובט ובוטא בהצלחה בתאי שמרים ובתאים אנימליים ופעילותו הביולוגית נשמרה. אולם, לא זוהו מולקולות חלבון מרנ"א המטרה בתאים. הסיבה לכך הינה שרנ"א המטרה, הנוצר מ-T7 רנ"א פולימראז, לא עבר עיבוד של הוספת cap לקצה ה-5' בשמרים. ולכן, על מנת לתרגם את רנ"א המטרה לחלבון, יש צורך באנזים המקטלז את ה-capping. לצורך כך, הוחלט לבטא קומפלקס capping ויראלי (מוירוס ה-*Vaccinia*), ביחד עם ה-T7 רנ"א פולימראז, הנקרא D1-D12, בכדי לתרגם את החלבון הפלורסנטי הירוק (GFP). קומפלקס זה הינו הטרו-דימרי, כלומר, המורכב משתי יחידות: תת היחידה הגדולה D1, המוסיפה cap לקצה ה-5' של הרנ"א. תת היחידה הקטנה D12 הינה חלבון המייצב את פעילותו של D1. נמצא שקומפלקס זה גם מעורב בשעתוק של גנים ויראליים, כלומר הינו "גורם שעתוק", השולט על בקרת הגנים במהלך מחזור חייו של וירוס ה-*Vaccinia*.

על מנת לחקור את השפעתם של החלבונים T7 רנ"א פולימראז והקומפלקס D1-D12 על ביטוי חלבון המטרה GFP בשמרים, תוכננו שתי מערכות: האחת מבוססת על וקטורי ביטוי, בה T7 רנ"א פולימראז, חלבוני ה-capping וגן המטרה GFP משובטים בוקטורי ביטוי. המערכת השנייה מבוססת על אינטגרציה של T7 רנ"א פולימראז וחלבון D12 לגנום של השמר, בעוד ה-GFP ו-D1 מבוטאים מוקטור ביטוי. יתר על כן, נבדקו גם רצפי טרנמינציה שונים, הנמצאים בסוף גן המטרה ותפקידם לסיים את השעתוק של הגן. השפעתם של רצפים אלה נבחנה ע"י מדידת הפלורסנציה של ה-GFP.

נמצא כי במערכת הוקטוריאלית, המורכבת מ-T7 רנ"א פולימראז (ללא חלבוני capping), T7 רנ"א פולימראז נמצא פעיל בגרעין השמר וזוהו רמות גבוהות של רנ"א המטרה בתאים. כאשר נמדדה רמת הפלורסנציה של ה-GFP, היא נמצאה גבוהה יותר מהפלורסנציה של זן הבר, שאינו מכיל את המערכת. לאחר הוספת הקומפלקס D1-D12 למערכת הסינתטית, זוהתה עלייה משמעותית ברמות הפלורסנציה של ה-GFP, בעיקר בתאים בהם רצף הטרמינציה לשעתוק של ה-GFP היה מהשמר. באופן מעניין, בתאים שמבטאים את הקומפלקס D1-D12, אך אינם מבטאים את ה- T7 רנ"א פולימראז רמת הפלורסנציה של ה-GFP נמצאה גבוהה ודומה לרמת הפלורסנציה בתאים שכן ביטאו את T7 רנ"א פולימראז ואת הקומפלקס D1-D12. תוצאה זו מעלה את השאלה האם רנ"א פולימראז שמרי מסוגל לשעתק את גן המטרה מפרומוטר נסתר בעל פעילות שעתוק חלשה ונמוכה, הנמצא בסמוך למיקומו של גן המטרה. תשובה אפשרית לשאלה זו היא שכנראה והקומפלקס D1-D12 מאפשר את ייצובו של רנ"א פולימראז השמרי על הפרומוטר הנסתר ובכך גדל הסיכוי להתחלה מוצלחת של שעתוק גן המטרה, שבסוף מתורגם לחלבון.

חזית חדשה בבקרת גנים באמצעות ביטוי של רנ"א פולימראז של הבקטריופאג' T7 בשמרים

ענבל וקנין