UNIVERSITY
of
GLASGOW

TRANSLATION, FOLDING AND FUNCTION OF BACTERIOPHAGE T4 DNA TOPOISOMERASE GENE *60*

Mona-Lissa-Khaled Halablab

Submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

June 2022

ABSTRACT

Biotechnology enables the design and assembly of customized biological systems to deliver solutions to challenges in medicine, agriculture, sustainable energy production and industrial processes. One current objective within biotechnology, as well as synthetic biology, is to create tools that can be used to manipulate biological systems in a reliable and predictable manner. While many transcription-based and translation-based control devices have been reported, they are based on a limited repertoire of biological component types, and there is a need for new systems that can be used to implement more elaborate devices.

This work explored the potential of reprogramming translation for engineering new gene-regulatory tools. Chapter 1 reviews the standard process of protein synthesis and the numerous translation reprogramming mechanisms that exist in nature, focusing on the role of the RNA message and the nascent translated peptide in these processes. Previous efforts to reprogram translation are also covered, including an examination of potential opportunities for reprogramming translation that have not yet been explored. In Chapter 2, the materials and methods used throughout this work are presented.

In Chapter 3, a reporter system is developed for examining elements of bacteriophage T4 DNA topoisomerase gene *60* mRNA that reprogram protein synthesis by translational bypassing. It is demonstrated that the reporter system is a reliable tool for identifying and verifying previously characterized gene *60* translational bypassing elements and proposes the use of the sequences for reprogramming translation through RNA. In Chapter 4, an experimental framework is developed for engineering *de novo* translational bypassing devices (based on T4 gene *60* bypassing elements) and their application to regulate protein synthesis in bacteria. Evidence is presented that demonstrates that these engineered devices can regulate gene expression and control the relative stoichiometry of two major distinct protein outputs from one gene. The utility of these devices in expanding our synthetic capabilities to manipulate biological systems is explored, as well as revealing unforeseen details of the bypassing mechanism in T4 gene *60*.

A recent focus of synthetic biology has been the control of protein function in cells. The achievement of this goal has focused on split protein complementation approaches, in particular when the N- and C- terminal fragments are fused to inducible protein-protein interaction systems, that enable spatiotemporal control of gene expression. In Chapter 5, a previously predicted protein-protein interaction system between two T4 DNA topoisomerase subunit fragments is demonstrated and its application for generating functional split proteins is explored. Lastly, the concluding remarks on this work are presented in Chapter 6.

CONTENTS

ABBREVIATIONS

I) Chemicals

DNA	–	deoxyribonucleic acid
RNA	–	ribonucleic acid
X-gal	–	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
IPTG	–	isopropyl-β-D-thiogalactopyranoside
DTT	–	dithiothreitol
mRNA	–	messenger RNA
tRNA	–	transfer RNA
ATP	–	adenosine 5'-triphosphate
GTP	–	guanosine 5'-triphosphate
cAMP	–	cyclic adenosine monophosphate

II) Measurements

g	–	gram
mg	–	milligram
µg	–	microgram
ng	–	nanogram
µl	–	microliter
ml	–	milliliter
L	–	liter
M	–	molar
mM	–	millimolar
µM	–	micromolar
%	–	per cent
‰	–	per thousand
V	–	voltage
S	–	svedberg
bp	–	base pair
kbp	–	kilo base pair
kDa	–	kilo dalton
µm	–	micrometer
nm	–	nanometer

rpm	−	rotations per minute
°C	−	degree Celsius
λ	−	wavelength

III) Antibiotics

Ap	−	ampicillin
Km	−	kanamycin
Cm	−	chloramphenicol

IV) Miscellaneous

~	−	approximately
α	−	alpha
β	−	beta
ϕ	−	phi
MBP	−	maltose-binding protein
GFP	−	green fluorescent protein
RFP	−	red fluorescent protein
RBS	−	ribosome binding site
Rf	−	relative fluorescence

I declare that the research reported in this thesis is my own and original work, except where explicit reference is made to the contribution of others, and has not been submitted for any other degree.

Printed Name: Mona-Lissa-Khaled Halablab

Signature: *M. Halablab*

ACKNOWLEDGEMENTS

I thank University of Glasgow for my PhD scholarship.

I thank Dr. Sean Colloms and Prof. Cheryl Woolhead, my supervisors, for their advice and enthusiasm regarding this project. I also thank Donald Campbell for keeping everything running smoothly in the laboratory.

Thanks to Profs. J. Christie, B. Hallet and C. Woolhead, and Drs. S. Colloms, P. Fogg, H. Bracken and J. Provan, for their kind gifts of plasmids and bacterial strains.

Thanks to Prof. W. M. Stark for our collaboration on the size exclusion chromatography project and equipment, and J. Southall for microscale thermophoresis experiments. Special thank is extended to Dr. Elaine Huston, as well as to every member at the Davidson and Bower Building for their help and cooperation at work.

I thank Prof. W. M. Stark, Prof. H. D. Bernstein and Dr. R. Karnik, my Ph.D. Viva Examination committee, for reading my thesis and for their kind advices on improvements to it.

I thank my family for all their love and support. They have always encouraged me to strive for excellence, and most importantly, believed in me.

To my family, for their love and support.

CHAPTER ONE

INTRODUCTION

1.1 General introduction

The ribosome and its associated translation components are tasked with the synthesis of the entire cellular proteome. Despite the complexity of the process, the components of the translational system cooperate to achieve what appears as a simple program for reading the genetic information and synthesizing proteins with defined sequences. The simplicity in the program makes protein translation an attractive engineering target − edit the program and you alter protein expression. While diverse translation reprogramming or recoding mechanisms discovered to date challenge the simplistic view of protein synthesis, they open opportunities that can be exploited to expand the ability to regulate protein synthesis and gene expression for biotechnology and synthetic biology. This thesis investigates a translation reprogramming mechanism that has been previously described (Herr *et al.*, 2000), to examine possibilities for reprogramming translation through previously unexplored approaches.

1.2 Translation mechanism

All living cells and viruses use their genetic information to synthesize proteins, the main executors of biological function. According to the central dogma of molecular biology (Crick, 1970), the transfer of information from nucleic acid to protein is unidirectional and requires that the information is first transcribed from a DNA template into an RNA template. The RNA template is then decoded and translated to protein. Translation is carried out by the ribosome, a macromolecular assembly of protein and RNA consisting of a small subunit (30S in prokaryotes, 40S in eukaryotes) and a large subunit (50S in prokaryotes, 60S in eukaryotes), and in collaboration with a multitude of protein factors and components that coordinate protein synthesis following a well-defined program (Ramakrishnan, 2002). This program relies on decoding rules that are applied in succession to non-overlapping triplets of nucleotides, denoted codons, within messenger RNA (mRNA) (Crick *et al.*, 1961).

The code consists of 64 codons, which arise from all possible triplet combinations of the four standard RNA nucleotide monomers: uridine (U), cytidine (C), adenosine (A), and guanosine (G). The program defines each codon to either specify the incorporation of a single amino acid residue or the termination of protein synthesis. The AUG codon is also used to initiate protein synthesis. Because proteins are synthesized from 20 standard amino acid building blocks, there is redundancy in the code such that a single amino acid is specified by multiple codons (Crick, 1966) (Table 1.1). Three of the 64 codons (UAG, UAA and UGA) often signal termination of protein synthesis.

First position	Second position				Third position
	U	**C**	**A**	**G**	
U	Phe	Ser	Tyr	Cys	**U**
	Phe	Ser	Tyr	Cys	**C**
	Leu	Ser	Stop	Stop	**A**
	Leu	Ser	Stop	Trp	**G**
C	Leu	Pro	His	Arg	**U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	Gln	Arg	**A**
	Leu	Pro	Gln	Arg	**G**
A	Ile	Thr	Asn	Ser	**U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	Glu	Gly	**G**

Table 1.1. The genetic code. The genetic code is in standard form with the frequent codons used in *E. coli* genes underlined. Data from the Codon Usage Database of *E. coli* K-12 (OpenWetWare, 2014).

Generally, protein translation is executed in three basic processes: initiation, elongation, and termination (Figure 1.1). Protein synthesis in bacteria initiates mainly at AUG codons (and in some cases also GUG and UUG (Blattner *et al.*, 1997)) assisted by a Shine-Dalgarno (SD) sequence / ribosome binding site (RBS) and requires numerous initiation factors that recruit the components to prepare the translation machinery for elongation (Simonetti *et al.*, 2009). Initiation is a highly regulated step (Kozak, 2005), as the decision to initiate signifies a commitment to the synthesis of a protein. During the elongation phase, amino acids are delivered to the ribosome in the form of aminoacyl-tRNA, covalently linked by an ester bond to L-shaped transfer RNA (tRNA) molecules. In a ternary complex with elongation factor EF-Tu and GTP, aminoacyl-tRNAs bind to the ribosomal A-site (Pape *et al.*, 1998). If the ribosome detects proper pairing between the codon and tRNA anticodon, the aminoacyl-tRNA is selected and accommodation into the A-site occurs. Once the aminoacyl-tRNA is properly placed in the A-site, peptide bond formation occurs (Rodnina and Wintermeyer, 2001) (Figure 1.1 and Figure 1.2).

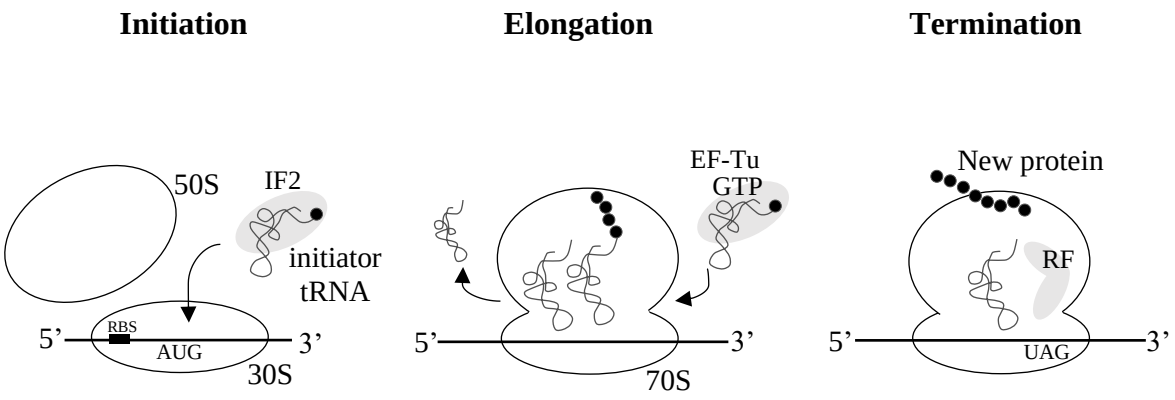


Figure 1.1. Stages of translation. Protein synthesis can be described in three basic stages: (1) Initiation that requires the initiator tRNA, multiple distinct initiation factors (IFs), and joining of the separate 30S and 50S subunits, (2) Elongation during which ribosomes migrate along the mRNA synthesizing the encoded polypeptide through consecutive decoding, peptide bond forming and translocation events, and (3) Termination in which the polypeptide is released from the ribosome at stop codons and the subunits dissociate from the mRNA.

At this stage of elongation, the translation machinery must move on the mRNA by one codon (or three nucleotides), to allow incorporation of the next amino acid residue. This process, termed translocation, involves hydrolysis of GTP by elongation factor EF-G that provides the energetic driving force for translocation of the translation apparatus, placing the elongating peptide linked to the tRNA from the ribosomal A-site to the P-site, ejection of the uncharged tRNA from the P-site and the next codon in the ribosomal A-site (Rodnina *et al.*, 1997; Tourigny *et al.*, 2013). The elongation cycle continues until the ribosome reaches a stop codon. Release factors with overlapping specificity, RF1 (UAG / UAA stop codon recognition) and RF2 (UGA / UAA) (Scolnik *et al.*, 1968), catalyze the hydrolysis of the newly synthesized polypeptide from the P-site tRNA and recycling factors disassemble the translation apparatus for reuse (Peske *et al.*, 2005).

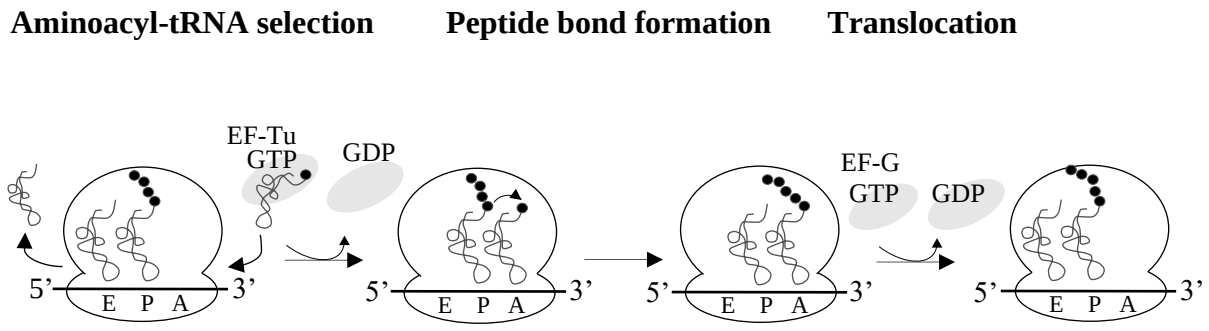


Figure 1.2. The elongation cycle. The ribosome has 3 binding sites for tRNAs across both subunits: (i) the A-site which accepts the incoming aminoacyl-tRNAs that transport amino acids to the ribosome, (ii) the P-site which holds the peptidyl-tRNA carrying the nascent peptide, and (iii) the E-site which holds the deacylated-tRNAs before they leave the ribosome. There are three steps in the elongation cycle. 1) A ternary complex of aminoacyl-tRNA, EF-Tu and GTP binds the ribosomal A-site. The ribosome selects the aminoacyl-tRNA based on codon-anticodon interactions, then induces GTP hydrolysis by EF-Tu and release tRNA in the ribosomal A-site (accommodation). 2) Peptide bond formation occurs, during which the peptide chain is transferred to the tRNA at the A-site. 3) The translation machinery translocates by one codon with the aid of EF-G to start another elongation cycle.

The fidelity of protein synthesis is ensured at all steps (Zaher and Green, 2009), beginning with charging of aminoacyl tRNAs by aminoacyl tRNA synthetases and an amino acid. These enzymes are responsible for correctly linking the free amino acids to their cognate carrier tRNA molecules (Ibba and Söll, 2000). Errors in aminoacylation, which would result in mis-incorporation of an amino acid at a codon, occur at very low rates between 10^{-4} - 10^{-5} due to the kinetic discrimination and editing capabilities of the enzymes (Jakubowski and Goldman, 1992). Error rates for the decoding process that are codon dependent are estimated to occur between 10^{-4} - 10^{-7} per amino acid incorporation (Garofalo *et al.*, 2019) owing to kinetic proofreading by the translation apparatus that discriminates between codons of similar composition during tRNA selection and accommodation.

Maintenance of translation reading frames is also necessary to ensure that the proper sequence is translated. An undirected shift in the reading frame during elongation would change the sequence of the entire downstream peptide and often results in the synthesis of a truncated, junk protein product. Errors in reading frame maintenance have been estimated at ~ 1 x 10^{-4} in *E. coli* (Jørgensen and Kurland, 1990). Maintenance is significantly influenced by the nature of the codons and tRNAs that are present in the P-site (Jenner *et al.*, 2010), as well as the modification status of the tRNAs (Urbonavicius *et al.*, 2001).

The last translation phase, termination, must be specific and efficient for stop codons to prevent unintended extension or truncation of the protein sequence. Unlike sense codons, stop codons are recognized through mRNA-protein interactions as opposed to RNA-RNA base-pairing interactions. However, similar in both, the signal for codon recognition must be transmitted to the ribosome's peptidyl transferase center (PTC) to promote the specific reaction. As release factors resemble tRNAs in shape and size (Nakamura and Ito, 2011; Youngman *et al.*, 2008), the structural mimicry agrees with their functional similarities. When recognition of a stop codon occurs, the release factor induces an active conformation of the PTC and positions a water molecule (using a conserved GGQ motif) for nucleophilic attack on the peptidyl-tRNA ester bond. The nucleotide context surrounding the stop codon in the mRNA has long been recognized to play an important role in the efficiency of termination at stop codons (Miller and Albertini, 1983). One mechanism to enhance the fidelity of termination could be to select for efficient stop codon recognition context.

1.3 Translational recoding

Faithful translation of the code is essential for ensuring accurate protein synthesis. However, in some cases, the translation apparatus is reprogrammed to read the genetic information in alternative ways. While the standard mode of translation predominates for most translation events, examples of non-canonical decoding that result in the expression of alternative protein products are plentiful (Atkins and Gesteland, 2010). The sequence diversity specified in mRNAs and nascent peptide chains was unsurprisingly exploited by nature to re-write translation and expand the coding capacity of single genes. Examples of such events include nascent peptide-mediated translational stalling, codon redefinition, programmed ribosomal frameshifting and translational bypassing (Baranov *et al.*, 2001).

1.3.1 Nascent peptide-mediated translational stalling

Previous structural studies (Milligan and Unwin, 1986; Yonath *et al.*, 1987; Frank *et al.*, 1995), showed that there is a tunnel running through the large subunit of the ribosome and that the polypeptides pass through it as they are synthesized. This tunnel, termed the ribosomal polypeptide exit tunnel, is primarily composed of RNA, but in addition has a constricted region created by two ribosomal proteins (L4 and L22) (Voss *et al.*, 2006). The notion that the tunnel is non-interactive was disproven by cryo-EM studies (Seidelt *et al.*, 2009), showing distinct contacts between the polypeptide and the ribosomal tunnel. Various polypeptides are now known to interact with these elements in the tunnel to regulate translation and gene expression (Ito and Chiba, 2013). The 2A peptide of the foot-and-mouth disease virus, for example, induces its own termination in a stop codon independent manner at a proline codon (Doronina *et al.*, 2008; Sharma *et al.*, 2012). Likewise, the secretion monitor (SecM) polypeptide of *E. coli* regulates the translation of the downstream co-transcribed *secA* gene in response to the secretion status of the cell (Oliver *et al.*, 1998). The SecM polypeptide interacts with the exit tunnel, causing the ribosome to temporarily pause elongation. Stalled complexes expose the ribosome binding site element for the downstream *secA* gene, causing up-regulation of this protein exporter (Nakatogawa and Ito, 2001). Interestingly, the SecM peptide stalling sequence functions autonomously and has been shown to block the translation of downstream reporters to which it is fused (Nakatogawa and Ito, 2002). Lastly, the gp60 peptide of bacteriophage T4 DNA topoisomerase gene *60* was shown to promote a translation reprogramming event known as bypassing (Weiss *et al.*, 1990; Samatova *et al.*, 2014; Agirrezabala *et al.*, 2017), that is the subject of chapters 3 and 4 of this thesis.

1.3.2 Codon redefinition

At the time of its elucidation, the genetic code was viewed as universal and unchangeable, a result of a 'frozen accident' (Crick, 1968). However, since then, a 'thaw' has set in and over 20 naturally occurring alternative genetic codes are now known (Knight *et al.*, 2001). Many of the changes involve stop codons UAG and UGA that are reassigned to specify atypical amino acids such as selenocysteine (Böck *et al.*, 1991) and pyrrolysine (Srinivasan *et al.*, 2002). This translation reprogramming event, termed codon redefinition, often occurs in a context-specific manner. This is true for selenocysteine incorporation at UGA stop codons which requires the aid of a *cis*-acting RNA element denoted SECIS (selenocysteine insertion sequence) element and additional *trans*-acting factors for efficient incorporation (Zinoni *et al.*, 1990) (Figure 1.3).

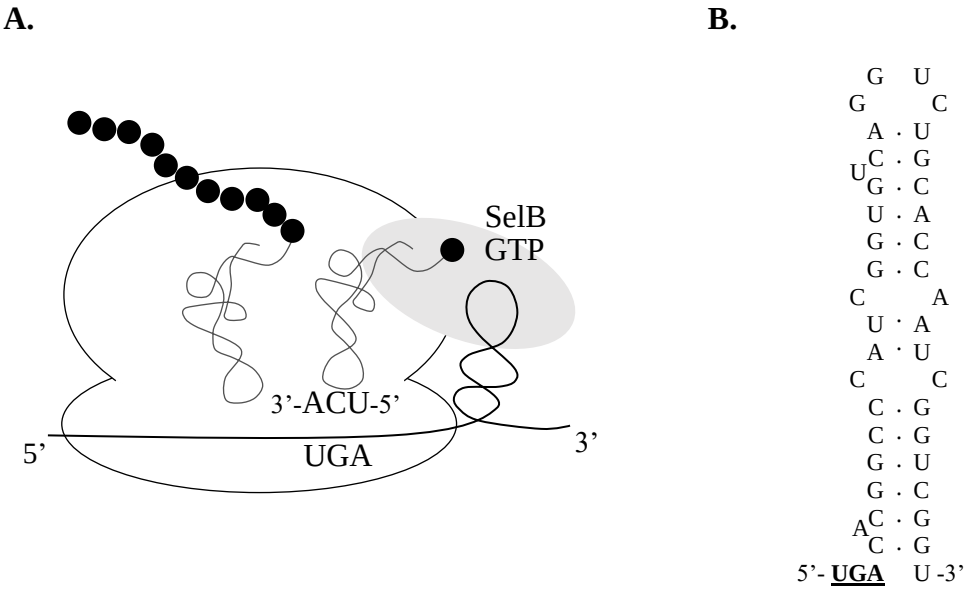


Figure 1.3. Bacterial selenocysteine insertion. A. UGA codons that specify selenocysteine are immediately followed downstream by a stem loop structure in the mRNA, denoted SECIS element, that recruits the selenocysteine elongation factor, SelB, tethered to selenocysteine aminoacyl-tRNA to be delivered to the ribosome. **B.** SECIS element of *E. coli* formate dehydrogenase H selenopolypeptide mRNA that induces stop codon redefinition (Hüttenhofer *et al.*, 1996).

Although a PYLIS element has been proposed to promote pyrrolysine incorporation at UAG stop codons (Théobald-Dietrich *et al*., 2005), the role of this structure is unclear and probably not necessary (Zhang *et al.*, 2005). The UAG stop codon may be ambiguous in organisms that utilize pyrrolysine. In contrast, selenocysteine (Sec), requires several factors for efficient incorporation at reprogrammed UGA codons. In addition to a downstream SECIS element, Sec incorporation relies on an EF-Tu-like elongation factor SelB. In bacteria, this factor is a bifunctional protein with one domain that tightly binds to Sec-tRNA^{Sec}, and a second domain that binds to the SECIS element to recruit the tRNA to the ribosome (Kromayer *et al.*, 1996; Fischer *et al.*, 2016). Interestingly, Sec is not directly charged to its tRNA, but instead synthesized from a serine charged intermediate Ser-tRNA^{Sec} (Itoh *et al.*, 2013). Although this tRNA appears to be a viable substrate for translation, its extended acceptor stem prevents it from binding to EF-Tu (Rudinger *et al.*, 1996) and the lack of the Sec residue reduces the binding affinity to SelB (Baron and Böck, 1991). Hence, serine incorporation at UGA codons is prevented.

1.3.3 Programmed ribosomal frameshifting

A common translation reprogramming event found in bacteria and eukaryotic retroviruses is -1 programmed ribosomal frameshifting (Farabaugh, 1996). In eukaryotes, this reprogramming mechanism was first described in retrovirus Rous sarcoma virus (Jacks and Varmus, 1985), where synthesis of the Gag-Pol protein was found to require a frameshift event to access the *pol* open reading frame. In most retroviruses, including HIV (Jacks *et al.*, 1988), frameshifting provides a way to tightly control the stoichiometry of Gag and Gag-Pol proteins, which is important for viral fitness (Biswas *et al.*, 2004).

As with selenocysteine recoding, both primary sequences, as well as downstream RNA structures, can promote this reprogramming event. These signals include a general ‘slippery site’ sequence X-XXY-YYZ (shown as codons in the zero frame where X, Y and Z can generally be any nucleotide) which is the site of the frameshift event, and a downstream RNA hairpin or pseudoknot structure at a precisely defined distance from the latter (Brierley, 1995; Giedroc and Cornish, 2009) (Figure 1.4).

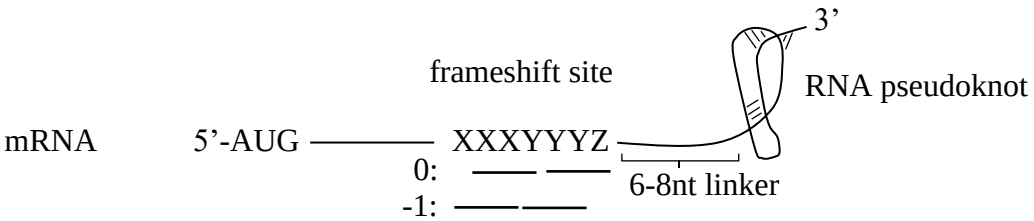


Figure 1.4. -1 programmed ribosomal frameshifting signals. The sequence of the heptanucleotide frameshift site is such that during re-pairing of the P- and A-site tRNAs in the new −1 reading frame, only the wobble codon-anticodon interaction is changed (“0” frame: X-XXY-YYZ to “−1” frame: XXX-YYY-Z).

Other frameshift events include +1 programmed ribosomal frameshifting which is often used in bacteria for regulatory purposes (Baranov *et al.*, 2002). One such example is the decoding of bacterial release factor 2 (RF2) mRNA, for which the product of its translational recoding, the RF2 protein, inhibits frameshifting and downregulates its own expression through a negative feedback loop (Craigen *et al.*, 1985; Craigen and Caskey, 1986). The mRNA sequence is CUU-UGA-C where frameshifting competes with termination since it involves peptidyl-tRNA dissociating from CUU and re-pairing at the overlapping UUU whose last base is the first base of the stop codon UGA (recognized by RF2). With high concentration of RF2, most ribosomes terminate at UGA, and at a low concentration of RF2, ribosomes shift frame and synthesize full-length active RF2 protein. The mRNA sequences involved in +1 frameshifting are simple, for example, the sequence CUU-AGG-C causes about 40% frameshifting when inserted into an mRNA in the yeast *Saccharomyces cerevisiae* (Farabaugh *et al.*, 2006).

1.3.4 Translational bypassing

Translational bypassing is another remarkable example of a reprogramming event in which the ribosome suspends translation at a certain site, and then resume translation downstream without decoding a block of intermediate nucleotides. This reprogramming event was first identified in expression of *E. coli* bacteriophage T4 DNA topoisomerase gene *60* (Huang *et al.*, 1988), and was later found in several yeast *Magnusiomycetes* mitochondrial genes (Lang *et al.*, 2014; Nosek *et al.*, 2015). Since these initial reports, numerous other gene *60*-derived synthetic gene constructs and others have been reported for bypassing (Weiss *et al.*, 1990; Maldonado and Herr, 1998; Herr *et al.*, 1999; Herr *et al.*, 2000; Herr *et al.*, 2001;

Gallant *et al.*, 2003; Gallant *et al.*, 2004; Herr *et al.*, 2004; Bucklin *et al.*, 2005; Wills *et al.*, 2008; Chen *et al.*, 2015).

The *cis*-acting signals that promote this reprogramming event can take various forms (Figure 1.5). Generally, two key features are necessary: (1) matched codons (denoted 'take-off' and 'landing' site codon) flanking the bypassed nucleotide sequence and (2) an RNA stimulatory structure, typically a stem loop, that begins with a stop codon (or unused codon in yeast) (Figure 1.5A). The wide range of variability in sequence, size and structure of the stimulatory elements that promote this translational event (Figure 1.5B), suggests that bypassing relies less on specific interactions and more on general engagement of the translational apparatus. In some cases, bypassing may also be influenced by auxiliary factors (Herbst *et al.*, 1994; Klimova *et al.*, 2019; O'Loughlin *et al.*, 2020).

Several studies have attempted to resolve the fundamental principles behind bypassing to construct a descriptive mechanistic model for this important reprogramming event. It was originally hypothesized that the role of the stimulatory stem loop structure was to provide mechanical tension on the elongating ribosome (Huang *et al.*, 1988), and evidence suggested that the stability of the structure corelated with bypassing activity (Weiss *et al.*, 1990; Wills *et al.*, 2008; Todd and Walter, 2013; Agirrezabala *et al.*, 2017).

Another study detailed the use of mass spectrometry (MS) to characterize and measure the stoichiometry of translation products of an mRNA containing bypass signals that were specific to T4 gene *60* (Wills *et al.*, 2008). Translation products characterized by MS were found to contain peptides that resulted from not just productive and ineffective bypass, but also from -7 backward bypassing (given that another matched 'landing' site was introduced upstream to the 'take-off' site). Ribosome trajectories corresponding to this position were further verified by protein sequencing of the resultant peptides. The results of this study are intriguing, suggesting that the ribosome makes unspecific large movements along the mRNA when encountering a bypass site, and settles on a 'landing' codon as a result of acceptable codon-anticodon pairing. However, other evidence suggests that the ribosome movement is directed by RNA structures flanking either end of the bypass site (Samatova *et al.*, 2014) and the nucleotide sequence of the bypass site (Weiss *et al.*, 1990; Wills *et al.*, 2008), indicating that the operative mechanism may be more subtle.

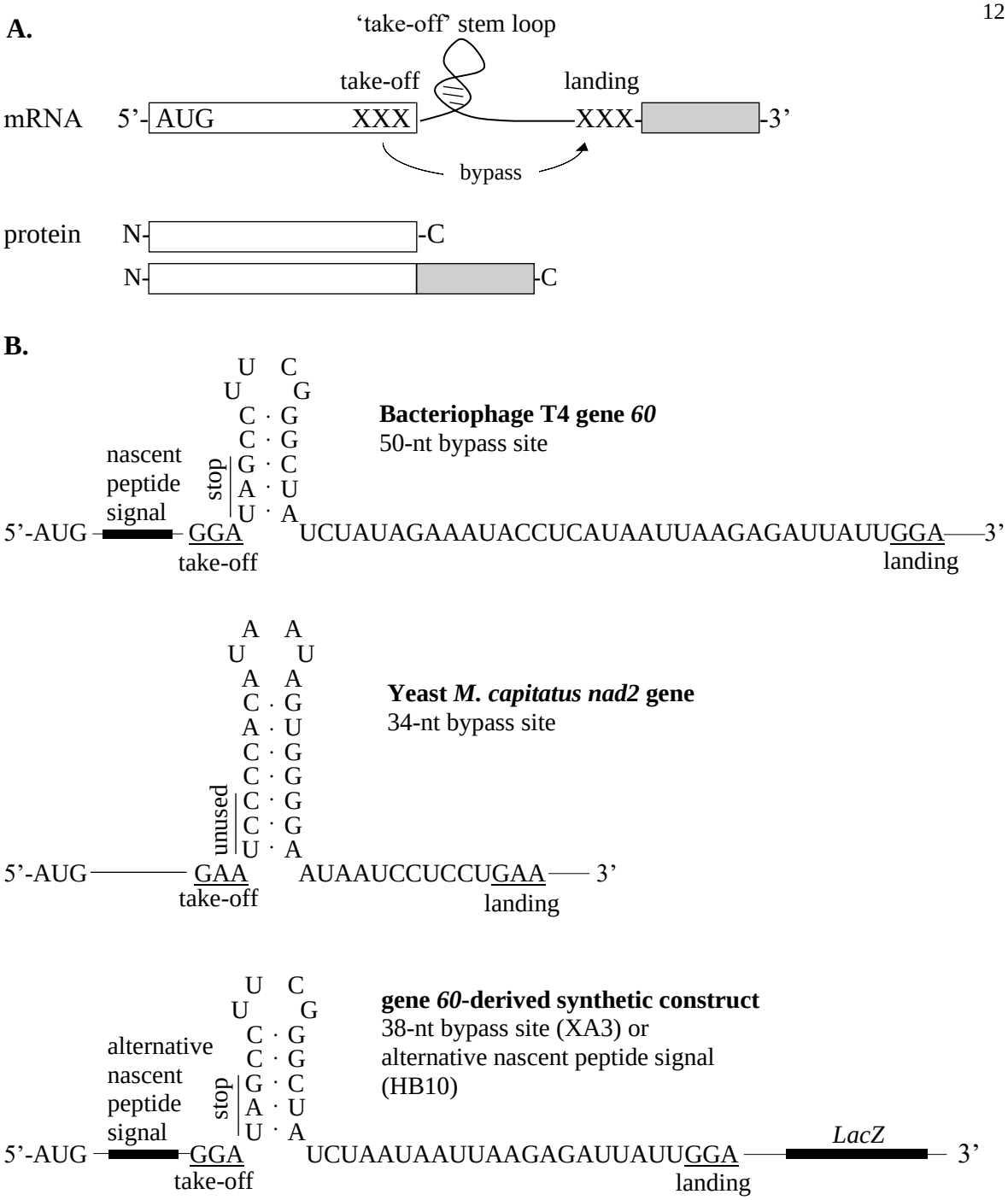


Figure 1.5. Diversity of *cis*-acting signals that promote translational bypassing. A. The bypassing event occurs when tRNAs occupy the 'take-off' site and is stimulated by the downstream RNA structure (above, a 'take-off' stem loop). The matched 'landing' site codon allows the translational apparatus to re-attach or 'land' on the mRNA after 'take-off' by codon-anticodon interactions and resume translation. Generally, this reprogramming event serves to enable the synthesis of two proteins from a single mRNA: one, a product of standard translation (or ineffective reprogramming), and the other a C-terminally extended product that results from productive bypassing. **B.** Translational bypassing signals in bacteriophage T4 gene *60* (Huang *et al.*, 1988; Weiss *et al.*, 1990), Yeast *M. capitatus nad2* gene (Nosek *et al.*, 2015) or a synthetic construct (Weiss *et al.*, 1990).

Recent cryo-EM experiments have provided some mechanistic insight into this reprogramming event (Agirrezabala *et al.*, 2017). This study resolved the structure of an *E. coli* ribosome complexed with gene *60* mRNA at its 'take-off' site. In the published structure, the stimulatory stem loop is found folded in the ribosomal A-site and the exit tunnel of the ribosome is found interrogating the nascent peptide. One immediate hypothesis regarding bypass is apparent from the structure. The nascent peptide causes the ribosome to pause elongation, presumably to favor stem loop formation in the ribosomal A-site. The stem loop, in turn, destabilizes the complex so that the kinetically unfavorable alternative decoding event can occur (Klimova *et al.*, 2019).

Another study monitored individual *E. coli* ribosome progression on gene *60* mRNA by means of an *in vitro* single-molecule fluorescence with zero-mode waveguide (smFRET-ZMW) technique (Chen *et al.*, 2015). This study revealed that gene *60* bypassing signals produced significant ribosome pausing at the take-off site, and that bypassing was not restricted to a gene-specific nascent peptide signal but instead could be replaced by SecM peptide stalling motif.

1.4 Reprogramming translation for biotechnology

Nature has already provided us with the means of redefining translation of the genetic information into protein. One of the most exciting opportunities is our ability to reprogram translation for biotechnology and synthetic biology to achieve expanded regulatory capabilities. Previous efforts to modify translation have focused on engineering the components of the translational system such as tRNAs and mRNAs to control gene expression and protein synthesis outcomes.

1.4.1 Unnatural amino acid mutagenesis

Amber suppressor tRNA technology has been the most successful strategy to site-specifically incorporate unnatural amino acids. Aminoacylation of suppressor tRNAs allows incorporation of the charged amino acid residue into proteins at UAG stop codons. This was first demonstrated with unnatural amino acid substrates linked *in vitro* to suppressor tRNAs by T4 RNA ligase (Noren *et al.*, 1989). The open reading frame of a protein of interest is site-specifically mutated to encode a UAG codon, then translated *in vitro* in the presence of the chemically charged suppressor tRNA. This method was also shown to work in *E. coli*

(Wang *et al.*, 2006), where engineered aminoacyl synthetase could recognize and charge suppressor tRNAs with an unnatural amino acid *in vivo*.

One valuable application of this technology is the incorporation of fluorescent amino acids into proteins *in vivo* (Wang *et al.*, 2006; Chatterjee *et al.*, 2013), that could potentially replace big fluorescent tag fusions used for protein localization in cells. While a great deal of research has been devoted to developing unnatural amino acid mutagenesis technologies (de la Torre and Chin, 2021), none of these approaches have taken advantage of the natural design of selenocysteine incorporation.

1.4.2 Reprogramming translation through RNA

Considering the various mRNA-based translation control mechanisms and translation recoding events that utilize signals contained within the mRNA (and inherently the encoded polypeptides), these mechanisms have great potential as devices to control and reprogram translation. However, few of these mechanisms have been adopted for synthetic biology. One example has been the development of RNA-based devices, called riboregulators, that turn protein synthesis on and off in response to a ‘trigger’ sequence (Ausländer *et al.*, 2014). Riboregulators contain an 'anti-RBS sequence' that binds to the RBS to form a stem loop, preventing initiation of protein synthesis. A 'trigger’ sequence interacts with and disrupts the hairpin, forming an alternative RNA structure that permits RBS–ribosome binding. Other diverse types of riboregulators, termed toehold riboregulators or switches, have also been made that can regulate translation of bacterial mRNA in response to the presence or absence of any desired 'trigger' RNA (Green *et al.*, 2014). Recently, toehold switches have been employed in the development of new SARS-CoV-2 detection approaches during the coronavirus pandemic (Park and Lee, 2021), and might prove useful in reprogramming bacterial chemotaxis for bioremediation purposes to seek and destroy a pollutant (Topp and Gallivan, 2007).

In addition to riboregulators, -1 programmed ribosomal frameshifting has been exploited as means to control gene expression in an inducible manner. It was recently demonstrated that -1 programmed ribosomal frameshifting signals could be used to tightly control the relative stoichiometry of two protein outputs from a single mRNA in eukaryotes and control a cell death module in yeast (Anzalone *et al.*, 2016), and recent developments have extended these -1 programmed ribosomal frameshifting approaches to application in

human cells (Matsumoto *et al.*, 2018). In a unique review (Dinman, 2019), the author suggests how these RNA-based regulatory signals can be employed like electronic circuit components such as resistors (a device having resistance to the passage of an electric current) to differentially engineer gene expression. Because recoding events are never 100% efficient, recoding signals decrease the fraction of ribosomes that pass through them and effectively decrease the amplitude of the signal output. Although significant progress has been made in these areas, other translation reprogramming mechanisms such as translational bypassing remain unexplored despite their potential to regulate gene expression and protein synthesis.

1.5 Conclusions

Protein translation is a dynamic process involving the interaction of large numbers of proteins and RNA components that collaborate to translate the genetic code. Nature has provided many examples in which the protein synthesis machinery can be re-programmed in a gene-specific manner to achieve extended regulatory capabilities. Modifications of the standard translational program holds great potential for engineering and manipulating biological systems by following a 'mimetic' approach to nature's reprogramming solutions.

1.6 Project aims

The aims of the project were:

I) Probe the elements of bacteriophage T4 gene *60* required for translational bypassing using *in vitro* translational systems and *in vivo* genetic assays

II) Create artificial 'hop' sequences that work in other genetic contexts and allow production of truncated and full-length proteins in varying stoichiometric ratios for biotechnological purposes

III) Investigate a predicted interaction between amino acids 473-516 of T4 topoisomerase subunit gp39 and amino acids 2-31 of T4 topoisomerase subunit gp60 using protein-protein interaction assays, that could be useful for holding partner proteins together more tightly for structural studies and other biotechnological application to tether proteins.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Bacterial strains. The derivatives of *Escherichia coli* used are listed in Table 2.1.

2.2 Plasmids. The plasmids used and constructed in this study are listed in Table 2.2.

2.3 Chemicals.

CHEMICALS	SOURCE
DNA	Integrated DNA Technologies, Inc. (IDT)
General chemicals, biochemicals and reagents	VWR, Sigma, Thermofisher, Biotium
Media	Formedium
Agarose	Bio-Rad, Thermofisher
IPTG	Melford
Antibiotics	Sigma
Radiochemicals	Perkin Elmer

2.4 Proteins.

Enzymes	New England BioLabs, Promega, Thermofisher
Antibodies	Abcam, LI-COR

2.5 Culture media.

L-Broth: 10 g tryptone, 5 g yeast powder extract, 10 g NaCl, made up to 1 liter in distilled water and autoclave sterilized at 121°C for 20 minutes.

L-Agar: as L-Broth with addition of 15 g/l agar prior to autoclaving.

2.6 Sterilisation. All growth media were sterilized by autoclaving at 121°C for 20 minutes. Supplements and buffer solutions were sterilized by filtration through 0.22 μm filter membranes.

2.7 Equations.

$C_1V_1 = C_2V_2$ (for dilutions) (C is concentration (M), V is volume (L))

Mass (g) = Concentration (M) x Volume (L) x Molecular weight (g/mol)

Table 2.1. Bacterial strains.

Strain	Genotype	Reference/source
DH5α	F$^-$ φ80 *lacZ*ΔM15 Δ(*lacZ*YA-*argF*) U169 *recA*1 *endA*1 *hsdR*17 (rk-, mk+) *phoA supE*44 *thi-1 gyrA*96 *relA*1 λ-	C. A. Woolhead
C41	F$^-$ *ompT hsdS*B (rB-, mB-) *gal dcm* (DE3)	Miroux and Walker, 1996/ C. A. Woolhead
BTH101	F$^-$ *cya*-99 *araD*139 *galE*15 *galK*16 *rpsL*1(strR) *hsdR*2 *mcrA*1 *mcrB*1	Karimova *et al.*, 2001/ P. C. M. Fogg
DS941	AB1157, but *recF*143 *lacZ*ΔM15 *lacI*^{q}	Bachmann, 1972; Summers and Sherratt, 1984/ S. D. Colloms
DS953	DS941, but *Sup*^{0}	Summers and Sherratt, 1984/ S. D. Colloms

Table 2.2. Plasmids.

Plasmid	Relevant information	Resistance	Reference/source
pTrc99A	Cloning vector, inducible Trc promoter, strong ribosome binding site	Ap	Amann *et al.*, 1988/ C. A. Woolhead
pMH20	pTrc99A, without SalI and PstI restriction sites in multiple cloning site	Ap	This study
pHop	pMH20, NcoI/XbaI gene fragment which encodes a His-tagged gene *60-iLOV* fusion	Ap	This study
pHopΔgap	pHop, SacI/SpeI gene fragment which encodes part of gene *60* with an exact deletion of the 50 nucleotide interruption in gene *60*	Ap	This study
pHop-L47	pHop, silent mutation TTA to CTG of gene *60* resume codon 47	Ap	This study
pHop-Stop	pHop, non-sense mutation TTA to TAA of gene *60* resume codon 47	Ap	This study
pGEM-4Z	Cloning vector, inducible T7 promoter, no ribosome binding site	Ap	Promega/ C. A. Woolhead
$pHop_{T7}$	pGEM-4Z, HindIII/XbaI gene fragment with an upstream ribosome binding site from pTrc99A which encodes a His-tagged gene *60-iLOV* fusion, two BamHI sites	Ap	This study
$pHop\Delta gap_{T7}$	$pHop_{T7}$, SacI/SpeI gene fragment which encodes part of gene *60* with an exact deletion of the 50 nucleotide interruption in gene *60*	Ap	This study
$pHop\text{-}L47_{T7}$	$pHop_{T7}$, silent mutation TTA to CTG of gene *60* resume codon 47	Ap	This study
$pHop\text{-}Stop_{T7}$	$pHop_{T7}$, non-sense mutation TTA to TAA of gene *60* resume codon 47	Ap	This study
$pHop_{T7}$-MBP	$pHop_{T7}$, one BamHI site, NcoI/BamHI gene fragment which encodes a His-tagged MBP fusion	Ap	This study

pHopΔgap$_{T7}$-MBP	pHopΔgap$_{T7}$, one BamHI site, NcoI/BamHI gene fragment which encodes a His-tagged MBP fusion	Ap	This study
pHop-L47$_{T7}$-MBP	pHop-L47$_{T7}$, one BamHI site, NcoI/BamHI gene fragment which encodes a His-tagged MBP fusion	Ap	This study
pHop-Stop$_{T7}$-MBP	pHop-Stop$_{T7}$, one BamHI site, NcoI/BamHI gene fragment which encodes a His-tagged MBP fusion	Ap	This study
pUC19	cloning vector, zero XhoI restriction site	Ap	Norrander *et al.*, 1983/ Thermofisher
pFM211 or pSWITCH1-PB	attachment sites *attP* and *attB* for ϕC31 serine integrase, two XhoI restriction sites, green fluorescent protein superfolder (GFP) gene, mCherry (RFP) gene	Km	Zhao *et al.*, 2019/ S. D. Colloms
pMFT-1	Bacteriophage ϕC31 serine integrase gene *int* (not codon optimized), one XhoI restriction site, constitutive promoter plus an inducible T7 promoter, strong ribosome binding site	Ap	M. Fares/ S. D. Colloms
pMFT-1 TAA	pMFT-1, two TAA stop codons in-frame to codon 327 in ϕC31 gene *int*	Ap	This study
pMFT-1 G327S	pMFT-1, G327S in ϕC31 gene *int*	Ap	This study
pHop1A	pMFT-1, 49 nucleotide untranslated sequence in-frame to codon 327 in ϕC31 gene *int*	Ap	This study
pHop1B	pMFT-1, 48 nucleotide untranslated sequence in-frame to codon 327 in ϕC31 gene *int*	Ap	This study
pHop1C	pMFT-1, 50 nucleotide untranslated sequence in-frame to codon 327 in ϕC31 gene *int*	Ap	This study
pHop2A	pMFT-1, 47 nucleotide untranslated sequence in-frame to codon 537 in ϕC31 gene *int*	Ap	This study
pHop2B	pMFT-1, 50 nucleotide untranslated sequence in-frame to codon 537 in ϕC31 gene *int*	Ap	This study

pHop3D	pMFT-1, 50 nucleotide-long sequence in-frame to codon 45 in ϕC31 gene *int*	Ap	This study
pHop1CΔ328-607	pHop1C, deletion of codon 328-607 in ϕC31 gene *int*	Ap	This study
pHop2BΔ538-607	pHop2B, deletion of codon 538-607 in ϕC31 gene *int*	Ap	This study
pHop1C-TCC	pHop1C, GGA to TCC substitution mutation in the 50 nucleotide untranslated sequence (landing codon)	Ap	This study
pHop1C-TCC-TCC	pHop1C-TCC, *GGA*327*TCC* (G327S) substitution in ϕC31 gene *int* (take-off codon)	Ap	This study
His_6-pMFT-1	pMFT-1, N-terminal His-tag	Ap	This study
His_6-pHop1C	pHop1C, N-terminal His-tag	Ap	This study
His_6-pHop1CΔ328-607	pHop1CΔ328-607, N-terminal His-tag	Ap	This study
His_6-pHop1C-TCC	pHop1C-TCC, N-terminal His-tag	Ap	This study
His_6-pHop1C-TCC-TCC	pHop1C-TCC-TCC, N-terminal His-tag	Ap	This study
His_6-pHop2B	pHop2B, N-terminal His-tag	Ap	This study
His_6-pHop3D	pHop3D, N-terminal His-tag	Ap	This study
pMAL-c2X	Cloning vector, multiple cloning site downstream of a maltose-binding protein (MBP), inducible Tac promoter, strong ribosome binding site	Ap	New England BioLabs; Walker *et al.*, 2010/ S. D. Colloms
pJIP28	T4 gene *39* (codon optimized)	Km	J. Provan/ S. D. Colloms
pMH39	pMAL-c2X, XbaI/HindIII gene fragment which encodes amino acids 473-516 of T4 DNA topoisomerase subunit gp39	Ap	This study
pTrc-GFP	pTrc99A, NcoI/PstI gene fragment which encodes a GFP superfolder with a C-terminal His-tag, multiple cloning site upstream of GFP superfolder	Ap	This study

pMH60	pTrc-GFP, NcoI/KpnI gene fragment which encodes amino acids 2-31 of T4 DNA topoisomerase subunit gp60	Ap	This study
pKT25-zip	T25 fragment of *B. pertussis CyaA* with a C-terminus leucine zipper GCN4 yeast protein (zip), multiple cloning site between T25 and the zip fragment	Km	Karimova *et al.*, 2001/ P. C. M. Fogg
pKT25-gp39$^{473-516}$	pKT25-zip, PstI/XbaI gene fragment which encodes amino acids 473-516 of T4 DNA topoisomerase subunit gp39	Km	This study
pUT18	T18 fragment of *B. pertussis CyaA,* multiple cloning site upstream of T18 fragment	Ap	Karimova *et al.*, 2001/ P. C. M. Fogg
pUT18-gp60$^{2-31}$	pUT18, HindIII/XbaI gene fragment which encodes amino acids 2-31 of T4 DNA topoisomerase subunit gp60	Ap	This study
pUC71K	Cloning vector	Ap, Km	S. D. Colloms
pA	pUC19, HindIII/EcoRI Tn4430 transposon cassette	Ap	This study
p4	pA, PstI/PstI Km fragment subcloned from pUC71K	Ap, Km	This study
pOD2	MiniTn-Km and an expression cassette for resolvase TnpI and a C-terminal His-tagged transposase TnpA	Cm, Km	B. Hallet/ O. Dudekem

2.8 Buffer solutions.

Electrophoresis

50 x TAE: 242 g Tris base, 18.5 g Na_2EDTA, made up to 1 liter in distilled water, adjusted to pH 8.6 with 57 ml of glacial acetic acid. Stored at ambient temperature. The 50 x TAE was diluted to 1 x TAE in distilled water when used for gel electrophoresis of DNA or the 1 x TAE was provided in the laboratories.

DNA gel loading dye: Gel Loading Dye, Purple (6X) (15% Ficoll®-400, 60 mM EDTA, 19.8 mM Tris-HCl, 0.48% SDS, 0.12% Dye 1 (pink/red), 0.006% Dye 2 (blue), pH 8.0 at 25°C) as supplied by New England BioLabs. The 6 x dye was added to a final concentration of 1 x dye to DNA samples run on agarose gels.

Protein gel running buffer: NuPAGE™ MES SDS Running Buffer (20X) (1 M MES (2-(*N*-morpholino)ethanesulfonic acid), 1 M Tris Base, 2% SDS, 20 mM EDTA, pH 7.3) as supplied by Thermofisher with pre-cast NuPAGE™ 4-12% Bis-Tris gels. The 20 x buffer was diluted to 1 x buffer in distilled water when used for polyacrylamide gel electrophoresis of protein in pre-cast Bis-Tris gels. For pre-cast Novex™ 10-20% Tricine gels (Thermofisher), 10 x Tricine SDS running buffer (1 M Tris Base, 1 M Tricine, 1% SDS, in H_2O, pH 8.3) was diluted to 1 x buffer in distilled water and used for electrophoresis.

Protein sample buffer: 4 x Bolt™ LDS sample buffer (424 mM Tris-HCl, 564 mM Tris base, 8% LDS (lithium dodecyl sulfate), 40% glycerol, 2.04 mM EDTA, 0.88 mM SERVA Blue G250, 0.7 mM phenol red, pH 8.5) as supplied by Thermofisher used in combination with samples resolved on NuPAGE™ 4-12% Bis-Tris gel. The 4 x sample buffer was diluted to 2 x sample buffer in distilled water and stored at 4°C. 5% β-mercaptoethanol was added immediately prior to use or as 20% in 2 x sample buffer stored at 4°C. For samples resolved on Novex™ 10-20% Tricine gels, 2 x Tricine SDS sample buffer (450 mM Tris/HCl pH 8.45, 12% glycerol, 4% SDS, 0.0025% Coomassie brilliant blue G, 0.0025% phenol red) was used.

Protein transfer buffer: NuPAGE™ Transfer Buffer (20X) (500 mM bicine, 500 mM Bis-Tris (free base), 20 mM EDTA, pH 7.2) as supplied by Thermofisher. The 20 x buffer was diluted to 1 x buffer in distilled water when used for transfer of proteins from NuPAGE™ 4-12% Bis-Tris gels to solid 0.45 µm nitrocellulose membranes for western blotting during electrophoresis.

Buffers for DNA work

1 x TE buffer: 10 mM Tris/HCl pH 8.0, 1 mM EDTA/NaOH pH 8.0, in distilled water. Stored at ambient temperature.

10 x restriction buffers: as supplied with restriction enzymes (New England BioLabs, Promega).

10 x ligation buffer: as supplied with T4 DNA ligase (New England BioLabs, Thermofisher).

$CaCl_2$ solution: 60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES/NaOH pH 7.0, in distilled water (or just 75 mM $CaCl_2$ and 15% glycerol, in distilled water). Solution was sterilized by filtration through 0.22 µm filter membrane and stored at ambient temperature. Solution was precooled on ice prior to use in competent cells transformation with plasmid DNA.

Buffers for protein work

Protein purification (section 2.24)

1 x TBS: 50 mM Tris/HCl pH 8.0, 150 mM NaCl, in distilled water.

Lysozyme: 100 mg/ml (in distilled water) stored at -20°C.

Binding buffer (HisGraviTrap): 40 mM Tris/HCl pH 8.0, 300 mM NaCl, in distilled water.

Wash buffer (HisGraviTrap): 40 mM Imidazole, in Binding buffer (HisGraviTrap).

Elution buffer (HisGraviTrap): 400 mM Imidazole, in Binding buffer (HisGraviTrap).

Binding/Wash buffer (MBPTrap): 20 mM Tris/HCl pH 8.0, 200 mM NaCl, 1 mM EDTA/NaOH pH 8.0, in distilled water.

10 mM Maltose Elution buffer (MBPTrap): 3.6 g D-(+)-maltose monohydrate (Sigma) dissolved in 1 liter of Binding/Wash buffer (MBPTrap).

Western blotting (section 2.30)

10 x PBS: 1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 17 mM KH_2PO_4, in distilled water. 10 x PBS was diluted to 1 x PBS in distilled water before use.

PBS-T1: 0.1% TWEEN-20, in 1 x PBS

PBS-T3: 0.3% TWEEN-20, in 1 x PBS

Blocking solution: 1 g dried skimmed milk (Premier Food Groups) dissolved in 20 ml of PBS-T1.

Dialysis, size exclusion chromatography and microscale thermophoresis (sections 2.25, 2.26 and 2.27)

Dialysis buffer: 50 mM Tris/HCl pH 8.0, 100 mM NaCl, 0.1 mM DTT, in distilled water.

SEC running buffer (assay buffer for size exclusion chromatography): as Dialysis buffer, but prepared in double distilled water. Vacuum 0.22 µm filtered sterilized to remove gas and particulate matter.

5 x SEC running buffer: as Dialysis buffer, but prepared in double distilled water and 5 times concentrated.

Gel-filtration standard: 100 µl of Bio-Rad Gel-filtration standard (a mixture of 5 proteins: Thyroglobulin (bovine), γ-globulin (bovine), Ovalbumin (chicken), Myoglobin (horse) and Vitamin B12) (Bio-Rad), 100 µl of 5 x SEC running buffer, brought to 1 ml in double distilled water. Particulate matter removed by centrifugation (13,000 rpm, 1 minute) before use.

MST running buffer (assay buffer for microscale thermophoresis): as Dialysis buffer, but prepared in water for HPLC. Vacuum 0.22 µm filter sterilized to remove particulate matter.

Stain and de-stain solutions

Coomassie blue solution: 0.04% Coomassie brilliant blue G 250, 25% methanol, 10% acetic acid glacial, in distilled water.

De-stain solution: 40% methanol, 7% acetic acid glacial, in distilled water. This solution was used with polyacrylamide gels after coomassie blue staining or directly after electrophoresis of radiolabelled proteins.

2.9 Antibiotics. The antibiotic concentrations used throughout for both liquid and plate selections were as follows:

Antibiotic	Stock solution	Selective concentration
Ampicillin (Ap)	100 mg/ml (distilled water)	100 µg/ml
Kanamycin (Km)	50 mg/ml (distilled water)	50 µg/ml

All stock solutions were stored at -20°C.

Antibiotics were added to molten L-Agar precooled to 55°C at ambient temperature.

2.10 Indicator. X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) was used in combination with the host strain BTH101 and the pKT25-zip and pUT18 vectors + derivatives, providing a blue-white screen for the bacterial two-hybrid assay for protein-protein interaction studies. Visual inspection of white colonies indicated an absence of interaction while blue colonies indicated presence of an interaction. X-gal (20 mg/ml, in dimethyl formamide) was stored in the dark at -20°C and added to L-Agar at a final concentration of 40 µg/ml. IPTG (isopropyl-β-D-thiogalactoside) was added to a final concentration of 0.5 mM to medium containing X-gal.

2.11 Growth conditions. Liquid culture for transformation, DNA preparations and expression of recombinant proteins were routinely grown in L-Broth at 37°C with vigorous shaking. L-Agar solid media was used and plates were generally incubated overnight at 37°C (and 30°C for bacterial two-hybrid assay). Antibiotic, IPTG and X-gal supplements were used as required.

Liquid cultures of bacterial strains and transformed bacterial strains were stored in 50% glycerol at -80°C. Colony forming units (CFUs) were counted and individual colonies were used.

2.12 Plasmid DNA isolation using QIAprep Spin Miniprep kit. 3 ml cultures of stationary phase plasmid-containing cells were harvested by centrifugation (8,000 rpm, 3 minutes). The plasmid DNA was then purified using buffers and materials supplied with the kit according to the manufacturer's instructions (Qiagen). 1-10 µl of this DNA was suitable for restriction enzyme digests, sequencing, transformation, PCR or *in vitro* translation reactions. Plasmid DNA in buffer EB was stored at -20°C.

2.13 Transformation with plasmid DNA. Plasmid DNA was introduced into different host strains. The recipient strain was made competent for DNA uptake by calcium chloride chemical transformation method:

Large scale competent cells preparation:

An overnight culture of the recipient strain was diluted 1 in 250 into 50 ml of L-Broth and grown until an optical density of 0.5 was reached. Cells were harvested by centrifugation at 2,500 rpm for 7 minutes at 4°C and the pellet was re-suspended in 10 ml of $CaCl_2$ solution. The cells were pelleted again by centrifugation at 2,500 rpm for 5 minutes at 4°C, the pellet was re-suspended in 10 ml of $CaCl_2$ solution and left on ice for 30 minutes. The cells were harvested by centrifugation at 2,500 rpm for 5 minutes at 4°C, the pellet was re-suspended in 2 ml of $CaCl_2$ solution and 100 µl was aliquoted in eppendorf tubes. The aliquots were left on ice before use or stored at -80°C.

Small scale competent cells preparation:

An overnight culture of the recipient strain was diluted 1 in 50 into 10 ml of L-Broth and grown until an optical density of 0.5 was reached. 1 ml of culture was aliquoted in eppendorf tubes, centrifuged at 8,000 rpm for 3 minutes at 4°C and the cell pellet was re-suspended in 1 ml of $CaCl_2$ solution. Cells were pelleted again by centrifugation at 8,000 rpm for 3 minutes at 4°C, the cell pellet was re-suspended in 100 µl of $CaCl_2$ solution and left on ice for 60 minutes before use or stored at -80°C.

Transformation of competent cells:

100 µl aliquots of competent cells were mixed with DNA (~ 50 ng of plasmid DNA or ~ 2 µl of a ligation/digestion reaction) and, after gentle mixing, were left on ice for 15 minutes. The cells were ‘heat shocked’ at 37°C for 5 minutes and returned on ice for 5 more minutes. 900 µl of L-Broth was added to the transformation mixture and incubated at 37°C for 60 minutes to allow expression of plasmid genes. Cells were pelleted by centrifugation (8,000 rpm, 3 minutes) and the pellet was resuspended in 200 µl of supernatant. 100 µl of the transformation mixture was spread out onto selective plates.

2.14 Restriction digest of DNA. Restriction enzyme digests were carried out in a volume of 20 µl, containing 0.2-1 µg DNA and 1 x restriction buffers made up to the final volume with distilled or nuclease-free water. Restriction enzymes were used in excess at 5-10 units per 1 µg DNA. Reactions were incubated at 37°C for 2 hours. The enzymes were inactivated by heating at 65°C or 80°C for 20 minutes as described by the manufacturer or by addition of loading dye.

2.15 Ligation of DNA fragments. Restriction fragments were ligated in a volume of ~ 10 µl, containing 1 x ligation buffer and 400 units of T4 DNA ligase (New England BioLabs)

or 1 unit of T4 DNA ligase (Thermofisher). A 5:1-100:1 insert to vector ratio of fragments was used (generated using NEBioCalculator software from New England BioLabs). Ligation reactions were carried out at ambient temperature for a minimum of 2 hours and no longer than an overnight. A small volume of ligation mix (~ 2 µl) was used for transformation of competent cells.

2.16 Annealing of DNA oligonucleotides. 2 or 4 complementary DNA oligonucleotides with built-in sticky ends were annealed and used directly in ligation to vector fragments with compatible overhangs. 100 µM stock solutions of each oligonucleotide was made up in TE buffer as recommended by IDT. 10 µl of each oligo were then mixed in an eppendorf tube and brought up to a final volume of 100 µl with TE buffer. The tube was placed in a floatie and floated in an 85°C water bath for 5 minutes. 200 ml of 85°C water was removed to a beaker and the tube was left floating in the 85°C water for 60 minutes at ambient temperature until water temperature dropped to 30°C on a thermometer to allow annealing. Annealed oligonucleotides were diluted 1 in 100 into 100 µl of nuclease-free water and 1.5 µl aliquot was used in ligation. Oligonucleotides that were annealed for plasmid construction are listed in Table 2.3.

2.17 Pfu-blunt-end-fill. By using this technique, blunt ends are generated from sticky ends and then the blunt ends are ligated (to remove unwanted restriction sites from vectors). Reactions were done in a final volume of 10 µl, containing 0.25 µg digested DNA, 2 mM dNTP nucleotide mix, 1 x polymerase buffer and 2 units of 'cloned Pfu' DNA polymerase enzyme as supplied by Thermofisher. Reactions were carried out at 72°C for 30 minutes. The blunt ends were then ligated in a volume of 10 µl, containing 5 µl of above reaction, 1 x ligation buffer and 400 units of T4 DNA ligase (New England BioLabs). Adjusted with distilled water. The reactions were left overnight at ambient temperature to ligate the blunt ends. A small volume of ligation mix (~ 2 µl) was used for transformation of competent cells.

2.18 Amplification of DNA using TaKaRa Ex Taq kit. Linear DNA used in cloning or in *in vitro* translation reactions was amplified using the buffers, nucleotides and DNA polymerase supplied with the kit according to the manufacturer's instructions (TaKaRa). Appropriate templates and primers were used (Table 2.4). Distilled or nuclease-free water was used to adjust reaction volume. Reactions were carried out in a thermocycler with the PCR cycling protocol according to the kit's manufacturer's instructions (except that denaturation was done at 95ºC instead of 98ºC and polymerization at 68ºC instead of 72ºC). DNA was then purified from the reactions using buffers and materials supplied with the

'PureLink Quick PCR purification' kit as described in the manufacturer's instructions (Thermofisher). 1-10 µl of this DNA was suitable for agarose gel analysis, restriction enzyme digests or *in vitro* translation reactions.

2.19 Insertion and deletion mutagenesis of DNA sequences from plasmids using Q5 Site-directed Mutagenesis kit. DNA was mutated using the master mix buffer supplied with the kit according to the manufacturer's instructions (New England BioLabs). Appropriate templates and primers were used (Table 2.5) for plasmid construction. Distilled or nuclease-free water was used to adjust final volumes. Reactions were carried out in a thermocycler with the PCR cycling protocol according to the manufacturer's instructions. Gel loading dye was then added to the reaction and run on an agarose gel to purify amplified DNA. After electrophoresis, the gel was visualized on a UV light box and bands of interest were excised. DNA was then purified using buffers and materials supplied with the 'Monarch Gel Extraction' kit according to the manufacturer's instructions (New England BioLabs). 1 µl of this DNA was then added to a kinase, ligase, DpnI (KLD) reaction. KLD reactions were done in a volume of 10 µl, containing 1 x ligation buffer, 400 units of T4 DNA ligase, 20 units of DpnI and 10 units of T4 polynucleotide kinase as supplied by New England BioLabs. Final volume was adjusted with distilled or nuclease-free water. Reactions were carried out at ambient temperature for 60 minutes. A small volume of this KLD mixture (~ 2 µl) was used for transformation of competent cells.

2.20 Nucleotide substitutions using QuikChange Lightning Site-Directed Mutagenesis kit. DNA was mutated using the buffers, solutions, nucleotides and DNA polymerase supplied with the kit according to the manufacturer's instructions (Agilent Technologies). Appropriate templates and primers were used (Table 2.6). Final volume was adjusted with distilled or nuclease-free water. Reactions were carried out in a thermocycler with the PCR cycling protocol described by the kit's manufacturer's instructions. 20 units of DpnI supplied with the kit was then added to the reaction and it was incubated at 37°C for 2 hours. A small volume of this mixture (~ 2 µl) was used for transformation of competent cells.

2.21 DNA sequencing. DNA sequencing reactions were performed on plasmids by the DNA sequencing services at the University of Dundee. ~ 600 ng of plasmid DNA was prepared in a final volume of 30 µl with nuclease-free water (~ 20 ng/µl plasmid DNA) and shipped to the facility. Oligonucleotide primers for sequencing were provided by the facility or 10 µl of 3 µM primer per sample were sent alongside samples. Sequences were analysed using ApE software and BLAST to confirm successful mutation or ligation.

Table 2.3. List of oligonucleotides that were annealed.

DNA	**Complementary oligonucleotide (5'-3')**	**Top oligonucleotide (5'-3')**
pMH60	CCATTGAAGAGCTCTTAATAGAACGACGAACATTGTTTTGCAGTTTATATTTTTTCATGT and CGACGCTACTCGAGTCAATTTTTACAAATTTCGC	CATGGCGAAATTTGTAAAAATTGACTCGAGTAGCGTCGACATGAAAAAATATAAACTGCA and AAACAATGTTCGTCGTTCTATTAAGAGCTCTTCAATGGGTAC
pUT18-gp60$^{2-31}$	CTAGACCCATTGAAGAGCTCTTAATAGAACGACGAACATTGTTTTGCAGTTTATATTTTT and TCATGTCGACGCTACTCGAGTCAATTTTTACAAATTTGA	AGCTTCAAATTTGTAAAAATTGACTCGAGTAGCGTC GACATGAAAAAATATAAACTGCAA and AACAATGTTCGTCGTTCTATTAAGAGCTCTTCAATGGGT
pHopΔgap / pHopΔgap$_{T7}$	CTAGTACCTAATCCATCGTGATCAGCGTCTGTCATAATAGCGACATTCGCATAGTTCATTGAAGAGCT	CTTCAATGAACTATGCGAATGTCGCTATTATGACAGACGCTGATCACGATGGATTAGGTA

Table 2.4. List of primers used to amplify DNA.

DNA	Reverse primer (5'-3')	Forward primer (5'-3')	Template
pMH39	AAAAAAAGCTTTTATAGATTTTTTCGTAATTCACCTACTG	AAAAATCTAGAGCTTTTGAAGAAAAAGAA GATGG	pJIP28
pTrc-GFP	AAAAACCATGGGCGCGGGTACCCGTAAAGGCGAGGAACTGTTTAC	AAACTGCAGTTAGTGATGGTGATGGTGATGGGATCCCTTATAGAGCTCGTCCATTCCGTG	pFM211
pKT25-gp39$^{473-516}$	AAAAATCTAGATTATAGATTTTTTCGTAATTCACCTACTG	AAAAACTGCAGGAGCTTTTGAAGAAAAAGAAGATGG	pMH39
pHop$_{T7}$-MBP pHop-L47$_{T7}$-MBP pHopΔgap$_{T7}$-MBP pHop-Stop$_{T7}$-MBP	AAAAAAGGATCCAATCCTTCCCTCGATCCCGA	AAAAAACCATGGGCCATCACCATCACCATCACAAAATCGAAGAAGGTAAACTGGTAATCTG	pMAL-c2X
linear templates	GTACCCGGGGATCCTCTAGA	AGCTCTAATACGACTCACTATAGGG	pHop$_{T7}$ pHop-L47$_{T7}$ pHopΔgap$_{T7}$ pHop-Stop$_{T7}$

Table 2.5. List of primers used for insertion and deletion mutagenesis.

	DNA	**Reverse primer (5'-3')**	**Forward primer (5'-3')**	**Template**
insertion	pHop1A/pHop1B/ pHop1C	ATTTCTATA GATAGCCC GAAGGCTA TCCGCAATC AAGCTCGA CC	*<u>AC</u>CTCATA ATTAAGAG ATTATTGGA CCGATCATC GAGCCCGCT	pMFT-1
	pHop2A/pHop2B	ATTTCTATA GCTAGCCCG AAGGCTAG CCGGTCGG GTCAGCGTC	*<u>ACC</u>TCATA ATTAAGAG ATTATTGGC CCTAAGTCG TGGTGGGG GC	pMFT-1
	pHop3D	ATTTCTACA GGTAGCCC GAAGGCTA CCCGTCGCG CTCGACTTC	ACCTCATAA TTAAGAGAT TATTGGGGG CCGGTTCAG GTTCGTC	pMFT-1
	His_6-pMFT-1 His_6-pHop2B His_6-pHop3D His_6-pHop1C-TCC His_6-pHop1C-TCC-TCC His_6-pHop1CΔ328-607 ¥His_6-pHop1C	ATGGTGATG CATATGTAT ATCTCCTTC TTAAAG	CACCATCAC GACACGTA CGCGGGTG CT	pMFT-1 pHop2B pHop3D pHop1C-TCC pHop1C-TCC-TCC pHop1CΔ328-607 N/A
deletion	pHop1CΔ328-607	TCCAATAAT CTCTTAATT ATGAG	TAATAAGG ATCCGAGCT C	pHop1C
	pHop2BΔ538-607	GCCAATAAT CTCTTAATT ATGAG	TAATAAGG ATCCGAGCT C	pHop2B

(*): heterogeneous population of truncated and full-length primer. Nucleotide deletions at 5' end are indicated with an underline bar. **(¥)** Plasmid His_6-pHop1C was constructed by subcloning NcoI-MluI fragment from plasmid pHop1C into plasmid His_6-pMFT-1. Primer sequences were generated by NEBaseChanger™ tool available with mutagenesis kit (New England BioLabs).

Table 2.6. List of primers used for substitution mutagenesis.

DNA	Reverse primer (5'-3')	Forward primer (5'-3')	Template
pMFT-1 TAA	CATACCACTCAGCG GGCTCGATTTATTA TCCGCAATCAAGCT CGACCGGC	GCCGGTCGAGCTTGA TTGCGGATAATAAAT CGAGCCCGCTGAGT GGTATG	pMFT-1
pMFT-1 G327S	CGGGCTCGATGATC GGGGAGCAATCAA GCTCGACC	GGTCGAGCTTGATTG CTCCCCGATCATCGA GCCCG	pMFT-1
pHop1C-TCC	GCGGGCTCGATGAT CGGGGAAATAATCT CTTAATTATGAGGT ATTTCTATAGA	TCTATAGAAATACCT CATAATTAAGAGATT ATTTCCCCGATCATC GAGCCCGC	pHop1C
pHop1C-TCC-TCC	TAGATAGCCCGAAG GCTAGGAGCAATCA AGCTCGACCGG	CCGGTCGAGCTTGAT TGCTCCTAGCCTTCG GGCTATCTA	pHop1C-TCC
pHop-L47 pHop-L47$_{T7}$	GCTGCCACTAGTAC CCAGTCCAATAATC TCTTAATTATGAGG	CCTCATAATTAAGAG ATTATTGGACTGGGT ACTAGTGGCAGC	pHop pHop$_{T7}$
pHop-Stop pHop-Stop$_{T7}$	GCTGCCACTAGTAC CTTATCCAATAATC TCTTAATTATGAGG	CCTCATAATTAAGAG ATTATTGGATAAGGT ACTAGTGGCAGC	pHop pHop$_{T7}$

Primer sequences were generated by Agilent Primer design software available with mutagenesis kit (Agilent Technologies).

2.22 Gel electrophoresis of DNA.

Agarose gels: 0.8-1% agarose gels were used.

Agarose powder was dissolved in 40 ml or 80 ml of 1 x TAE buffer in a microwave and precooled at ambient temperature prior to use. Distilled water was added to account for evaporation. Horizontal gels were used to analyse restriction digests, PCR or DNA preparations. Gels were run for 60 minutes at 6.5 V/cm in gel tanks containing 1 x TAE buffer. Gels were either stained after electrophoresis in 0.5 µg/ml ethidium bromide (Sigma) or prior to casting by adding 0.5 x gel red/green nucleic acid stain (Biotium) to the agarose solution. The DNA was visualized on a UV light box or Bio-Rad Gel Doc UV transilluminator. Gels were photographed using the Bio-Rad Gel Doc camera or using a camera.

2.23 Extraction of DNA fragments from agarose gels for plasmid construction. After electrophoresis, the gel was visualized on a long wavelength (~ 360 nm) UV light box and bands of interest were excised. The DNA was then purified using buffers and materials supplied with the kit according to the manufacturer's instructions ('Monarch Gel Extraction' kit from New England BioLabs or 'QIAquick Gel Extraction' kit from Qiagen).

2.24 Expression and purification of recombinant proteins. Recombinant proteins made in this study (Table 2.7 and Table 2.8) were over-expressed in host strain C41 and purified with a standard protocol as follows. An overnight culture of plasmid-containing C41 cells was diluted 1 in 100 into 100 ml of L-Broth (unless stated otherwise) and grown (about 2 hours) until an optical density of 0.5 was reached (2 x 10^8 cells/ml). IPTG was then added at a final concentration of 1 mM to induce protein expression for 3 hours. Cells were harvested by centrifugation (10,000 rpm, 10 minutes at 4°C with Beckman rotor JA-14) and the pellets were stored at -20°C for an overnight. Cell pellets were thawed on ice and resuspended in 10 ml of 1 x TBS. One tablet of 'cOmplete mini EDTA-free protease inhibitor cocktail' (Sigma) and 200 µl of lysozyme were added to the cell suspension, mixed, and left on ice for 30 minutes. The cells were lysed by passing twice through a French Press cell at a pressure of 8,000 psi. The cell debris was pelleted by centrifugation (17,000 rpm, 30 minutes at 4°C with Beckman rotor JA-17) and the supernatant fraction was retained. The supernatant was further separated by ultra-centrifugation (42,000 rpm, 60 minutes at 4°C with Beckman rotor Ti-50 series). The supernatant fraction was kept and left on ice before use. His-tagged recombinant proteins were isolated by using a HisGraviTrap purification column according to the manufacturer's instructions (GE Healthcare), while MBP-tagged proteins were

isolated by use of a 5 ml MBPTrap HP purification column according to the manufacturer's instructions (GE Healthcare). The columns were washed with 10 ml and 20 ml of Binding buffer (for HisGraviTrap and MBPTrap, respectively), then, the supernatant fractions ('P') were applied to the corresponding column, and the flow-through ('F') was collected. The columns were then washed with 10 ml and 20 ml of Wash buffer (for HisGraviTrap and MBPTrap, respectively) and the washes (W_1, W_2, W_3 etc.) were collected. To elute the column-bound proteins, 3 ml and 6 ml of elution buffer was used (for HisGraviTrap and MBPTrap, respectively) and elutions (E_1, E_2, E_3 etc.) were collected. Elutants were collected as 1 ml fractions consecutively. Gravity flow and syringe force were used (for HisGraviTrap and MBPTrap, respectively). Small volumes were used for protein gel analysis, measurement of protein concentration and fluorescence. Elution samples containing purified protein were suitable for dialysis and subsequent use in size exclusion chromatography / gel-filtration and microscale thermophoresis for protein-protein interaction studies. For re-use of the HisGraviTrap and MBPTrap purification columns, columns were washed and re-generated according to the manufacturer's instruction (GE Healthcare).

Table 2.7. Properties of recombinant proteins made in this study used in Chapter 5.

Plasmid	Recombinant protein	Features	Phenotype	M_w (kDa)	Solubility	A_{280} Molar Extinction Coefficient ($M^{-1}cm^{-1}$)	A_{280} 1 mg/ml Extinction Coefficient	mg/ml	W-Y-C count	Oligomeric state
pMH39	MBP-gp39$^{473-516}$	MBP-tagged T4 gp39 residues 473-516	Light brown	48	soluble	77350	1.6	~5 (100 µM)	10W-15Y-0C	dimer
pMH60	gp60$^{2-31}$-sfGFP-His$_6$	His-tagged GFP superfolder with N-terminal T4 gp60 residues 2-31	Green fluorescent	31	soluble	20400	0.6	~10 (300 µM)	1W-10Y-2C	monomer

*Molecular weight (M_w) in kilo daltons (kDa). Absorbance at 280 nm (A_{280}). Tryptophan (W), Tyrosine (Y) and cysteine (C) amino acid residues W-Y-C count in protein. Protein concentrations (mg/ml) were measured using a Nanodrop (Thermofisher), 2 µl sample and the M_w + Extinction Coefficient (M^{-1} cm^{-1}) setting. Blanks were used as necessary. Data obtained using EMBOSS pepstat software by inputting amino acid sequence of protein.

Table 2.8. Properties of recombinant proteins made in this study used in Chapter 3.

Plasmid	Recombinant protein	Features	Phenotype	M_w (kDa)	Solubility
pHop pHop-L47 pHopΔgap pHop-Stop	Full-length protein	His-tagged gene *60*-iLOV protein	Green fluorescent	19	soluble
$pHop_{T7}$ $pHop\text{-}L47_{T7}$ $pHop\Delta gap_{T7}$ $pHop\text{-}Stop_{T7}$	Truncated protein	His-tagged gene *60* protein	Not fluorescent	6	insoluble/ degradation
$pHop_{T7}$-MBP $pHop\text{-}L47_{T7}$-MBP $pHop\Delta gap_{T7}$-MBP $pHop\text{-}Stop_{T7}$-MBP	Full-length protein	His-tagged MBP-gene *60*-iLOV protein	Green fluorescent	62	soluble
	Truncated protein	His-tagged MBP-gene *60* protein	Not fluorescent	48	soluble

2.25 Dialysis/Buffer exchange. Elution sample containing about 5-10 mg purified protein were injected into a '0.5-3 ml 10,000 MWCO dialysis membrane cassette' (Thermofisher) and left floating in 500 ml of Dialysis buffer with stirring overnight in a cold room (with 2 changes of 500 ml of Dialysis buffer, 2 hours, the following day). Protein in Dialysis buffer (+ traces of imidazole or maltose-EDTA) were then transferred to an eppendorf tube and left on ice (or stored at 4°C) before use in protein gel analysis, size exclusion chromatography or microscale thermophoresis.

2.26 Size exclusion chromatography for protein-protein interaction studies. Size exclusion chromatography was performed on an AKTA system machine (Amersham Bioscience) coupled to a UV detector (absorbance set to wavelengths of 280 nm and 488 nm) and a 30 ml capacity Superose 12 10/300 GL column (GE Healthcare). The flow rate of SEC running buffer was 1 ml/min for 500 µl sample (in combination with a 500 µl loop). The standard curve used for determination of apparent molecular weights was generated by injection of 500 µl of Gel-filtration standard into the column (Graph 5.1). For individual analysis of a protein, dialyzed protein sample was mixed with SEC running buffer (at a 1:1 molar ratio) in a final volume of 500 µl with SEC running buffer and applied to the column. For interaction studies, two proteins were mixed in an eppendorf tube at an approximate 1:1 molar concentration ratio in a final volume of 500 µl, left for 15 minutes at ambient temperature and then returned on ice before injecting it into the column. Elutents were collected in 1 ml fractions. Small volumes were used for protein gel analysis. Chromatograms were opened using pycorn web-tool software which is a script to extract data from UNICORN result (.res) files and plot them (Ahmed, 2017).

2.27 Microscale Thermophoresis for protein-protein interaction studies. Microscale thermophoresis was performed on a Monolith NT.115 machine (NanoTemper Technologies) and using Monolith NT.115 standard capillaries. As the target protein was intrinsically fluorescent, no fluorescent label was necessary to monitor the interaction. The blue-detector was used for detection of GFP fluorescence ($\lambda_{excitation}$= 488 nm, $\lambda_{emission}$= 507 nm) as described in the manufacturer's instructions. The binding assay was performed according to the manufacturer's instruction (except that the assay buffer used was MST running buffer instead of PBS) (NanoTemper Technologies). The dialyzed fluorescent 'target' protein and the 'ligand' protein were prepared at a final concentration of 1 µM in MST running buffer. A 16-steps serial dilution (10^0, 10^{-1}, 10^{-2}, 10^{-3} etc.) of the 'ligand' protein in a final volume of 20 µl with MST running buffer was prepared in PCR tubes. 10 µl of the 1 µM fluorescent 'target' protein was then added to each tube and mixed by pipetting up-and-down (for a final

concentration of 0.5 µM of fluorescent 'target' protein for suitable fluorescence counts > 400 counts). Aggregates were usually removed by centrifugation (13,000 rpm, 5 minutes, 4ºC) prior to loading of the capillaries. Capillaries were filled by dipping into the samples, placed on the holder tray and inserted into the MST machine. Thermophoresis was carried out at a temperature of 25°C (machine set to: cold region start (-1 s) and end (0 s), Hot region start (4 s) and end (5 s)), excitation power of 2-5% and MST power of 40%. Dissociation constants (k_d) were calculated by the integrated Monolith NT.115 software (MO.Affinity Analysis v2.3). The binding assay was repeated 3 more times, except that the 'ligand' protein was prepared at a final concentration of 2 µM, 2 µM and 10 µM in MST running buffer for the 16-steps serial dilution (performed by J. Southall).

2.28 Cell lysates extract. Plasmid-containing C41 cells were grown in the presence of 1 mM of IPTG overnight to express protein. 100 µl of culture was aliquoted in an eppendorf tube and cells were pelleted by centrifugation (8,000 rpm, 3 minutes). Pellets were resuspended in 20 µl of L-Broth. 100 µl of sample buffer containing 5% β-mercaptoethanol was added and the lysate suspension was incubated at 95°C for 10 minutes. A small volume (~ 20 µl) was used for protein gel electrophoresis.

2.29 Electrophoresis of proteins in SDS-polyacrylamide gels. The electrophoresis of protein followed the procedure described by Thermofisher using pre-cast NuPAGE™ 4-12% Bis-Tris gels or Novex™ 10-20% Tricine gels. Gels were placed in vertical gel mini-tank after removing the comb. The running buffer supplied with the gels was poured into gel tanks. Prior to loading, the protein samples were mixed with protein sample buffer and boiled at 95°C for 10 minutes. The gel was run at a constant voltage of 200 V for 30 minutes (for Bis-tris gels) or 125 V for 90 minutes (Tricine gels), leaving the dye front just above the bottom of the gel. Gels containing proteins were stained in coomassie blue solution, used for western blotting or fixed in de-stain solution and dried at 65ºC for 90 minutes.

2.30 Western blotting for detection of His-tagged recombinant proteins. Proteins were transferred from gels to a solid 0.45 µm nitrocellulose membrane (Thermofisher) at a constant voltage of 30 V for 60 minutes in transfer buffer supplied by Thermofisher. The membrane was soaked in blocking solution overnight at 4°C. The membrane was washed three times for 5 minutes in PBS-T1 and incubated with primary antibody for 60 minutes (1 µl monoclonal mouse anti-His tag antibody primary antibody (Ms mAb to 6X His tag® [HIS.H8] antibody (Abcam)) in 5 ml of PBS-T1) at 4°C with rotation. The blot was re-washed with PBS-T1 and incubated with secondary antibody for 60 minutes (2 µl goat anti-mouse IRDye 680RD secondary antibody (LI-COR) in 5 ml of PBS-T1) at 4°C with rotation.

The blot was washed three times for 5 minutes in PBS-T3 then PBS-T1. The protein was visualized using a LI-COR CLx scanner and wavelength channels 700-800 nm. The blot could also be stained in ponceau S solution (0.1% ponceau S dye, 1% acetic acid, in distilled water; present in the laboratory) for 5 minutes and then de-stained in distilled water for 5 minutes at ambient temperature with slow shaking.

2.31 *In vitro* translation. *In vitro* coupled transcription-translation reactions were done using the '*E.coli* T7 S30 extract system for circular DNA' and the '*E. coli* S30 Extract System for Linear Templates' kits as supplied and described in the manufacturer's instructions (Promega) (except that an additional 80 units of T7 RNA polymerase was added (Promega) and reactions were carried out at 37°C for 60 minutes). Pellets were resuspended in 20 µl sample buffer and run on 4-12% Bis-Tris gels. Gels were fixed in de-stain solution for 60 minutes, dried at 65°C for 60 minutes and placed in a light-proof cassette with a Kodak film overnight. Films were developed using a Xomat processor machine. Band intensities (Rf) were measured by 'GelAnalyzer 19.1' software.

2.32 Fluorescence measurements.

iLOV fluorescence. iLOV fluorescence was detected between emission wavelengths of 480-600 nm with an excitation wavelength of 450 nm (as described in Chapman *et al.*, 2008) on a Fluorolog fluorimeter machine (Horiba). 400 µl of purified recombinant proteins in elution sample was transferred to cuvettes supplied with the Fluorolog machine for fluorescence detection. Appropriate blanks were used.

GFP and RFP fluorescence. GFP and RFP fluorescence images were acquired using a Typhoon FLA9500 scanner (GE Healthcare). Settings for GFP were 473 nm laser excitation and the band-pass blue emission filter (BPB1 filter/530DF20) (emission wavelength 506-545 nm). Settings for RFP were 532 nm laser excitation and the long-pass green emission filter (LPG filter/575LP) (emission wavelength > 575 nm). PMT (voltage) was 400 and pixel was 50 µm. 1 ml culture of stationary phase plasmid-containing DS941 or DS953 cells were harvested by centrifugation (8,000 rpm, 3 minutes) and the cell pellet was re-suspended in 1 ml of 1 x PBS. Cells were harvested again by centrifugation (8,000 rpm, 3 minutes) and the cell pellet was re-suspended in 1 ml of 1 x PBS. 200 µl of samples were transferred to individual wells of a flat-bottom 96-well plate for fluorescence imaging.

2.33 gBlocks. The gBlocks used in this study for plasmid construction are listed in Table 2.9.

2.34 *In vivo* recombination assay. *In vivo* recombination assays were carried out as shown in Figure 2.1.

2.35 Bacterial two-hybrid complementation assay. Bacterial two-hybrid complementation assays were carried out as shown in Figure 2.2.

Table 2.9. gBlocks. 20 random nucleotides 'end' sequences on these gBlocks that get cut off when they are cloned are indicated with an underline bar.

DNA	**Nucleotide sequence (5'-3') His_6-T4 gene *60*-iLOV gBlock.**
pHop $pHop_{T7}$	GCTAGTATACGCCTAAAGCTTTCACACAGGAAACAGACCATGGCGCATCA CCATCACCATCACGGATCCAAATTTGTAAAAATTGACTCGAGTAGCGTCGA CATGAAAAAATATAAACTGCAGAACAATGTTCGTCGTTCTATTAAGAGCTC TTCAATGAACTATGCGAATGTCGCTATTATGACAGACGCTGATCACGATGG ATAGCCTTCGGGCTATCTATAGAAATACCTCATAATTAAGAGATTATTGGA TTAGGTACTAGTGGCAGCGGCTCCGGTACCATAGAGAAGAATTTCGTCATC ACTGATCCTAGGCTTCCCGATAATCCCATTATCTTTGCATCAGACGGCTTTC TTGAATTGACAGAGTATTCGCGCGAGGAAATATTGGGGAGAAATGCCCGG TTTCTTCAGGGGCCAGAGACAGATCAAGCGACTGTCCAGAAGATAAGAGA CGCAATTAGAGATCAGAGGGAGACTACTGTGCAGTTGATAAACTACACTA AAAGCGGAAAGAAATTCTGGAACTTACTCCACCTGCAACCTGTGCGTGACC AGAAGGGAGAGCTTCAATACTTCATCGGTGTGCAGCTCGATGGAAGTGAC CATGTATAATAATCTAGACCGAATTCGACGTACGATTC
DNA	**Nucleotide sequence (5'-3') Tn4430gp39_gp60 v5 frame 1 gBlock (Tn4430 transposon)**
pA	AAACATTGACCAGATACAGTAAGCTTGCCGGCGACTCTAGCTCGAGATGC ATCCATGGGGGTACCGCCAGCATTTCGGAAAAAAACCACGCTAAGGGGCC CGCTTTTGAAGAAAAAGAAGATGGCGAGTGGTTTACTTTCGAGCTAAATGG CGATACAATTATCGTAAATGAAAATGATGAAGTACAGATTAATGGTAAAT GGATAACAGTAGGTGAATTACGAAAAAATCTATAACTTAAGAAAATAATA CAACACAATATTAATTGTGTTGTATTAGGTGACGCGTTAAATATAAATCTA GGGGTTTAACGCAACACAATTTATCGATAAATAAATACTTTTAGACGCAAC ACAATTTATAGACGCGGAGGAAATCTAGACTGCAGACCGGTTTTAAGAAG GAGATATACAAATGAAATTTGTAAAAATTGACTCGAGTAGCGTCGACATG AAAAAATATAAACTGCAAAACAATGTTCGTCGTTCTATTAAGAGCTCTTCA ATGGCTAGCGAATTCGTCTTTGACACGTTTACACG

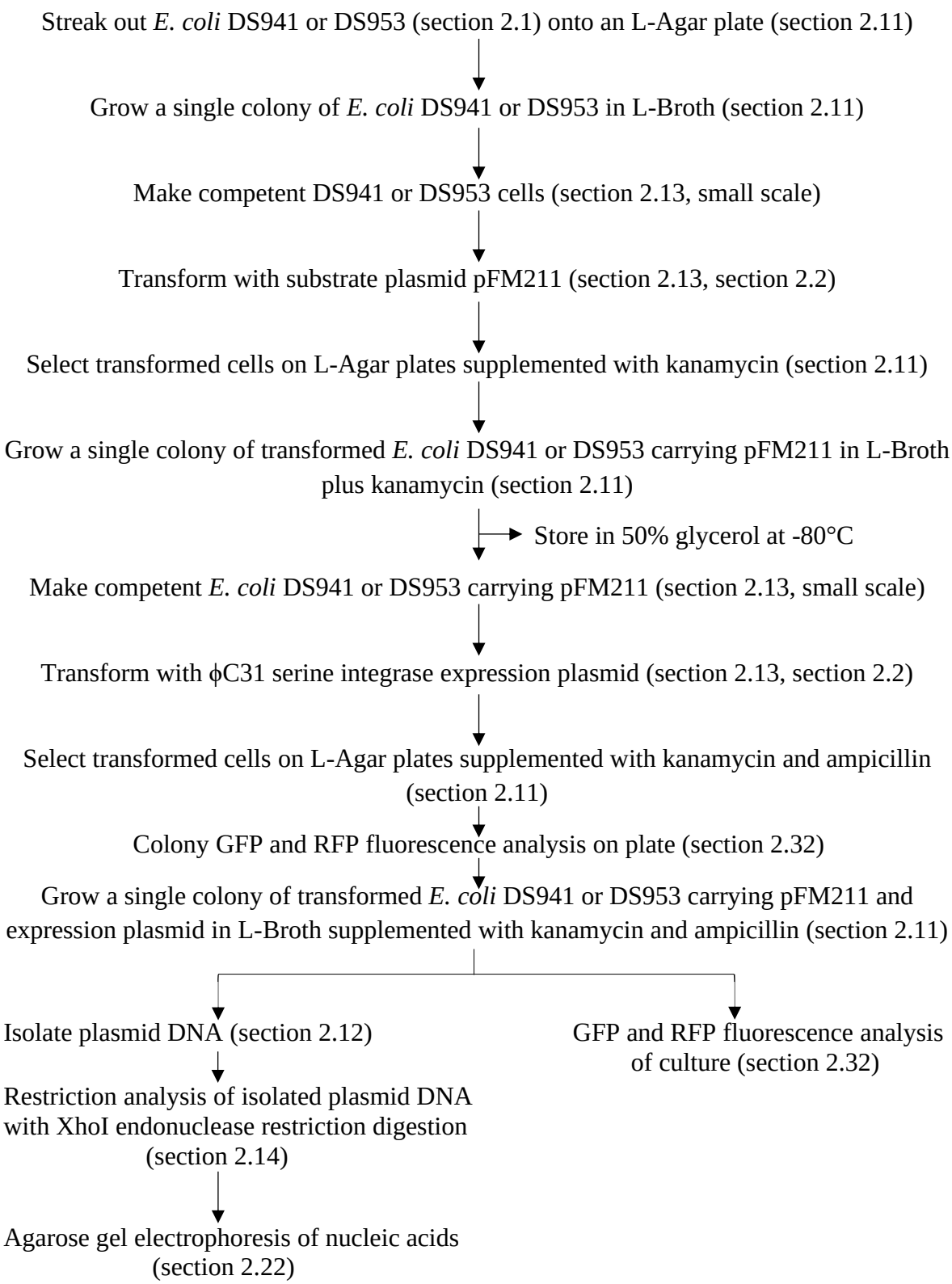


Figure 2.1. Flow diagram summarising the method used for *in vivo* recombination assays.

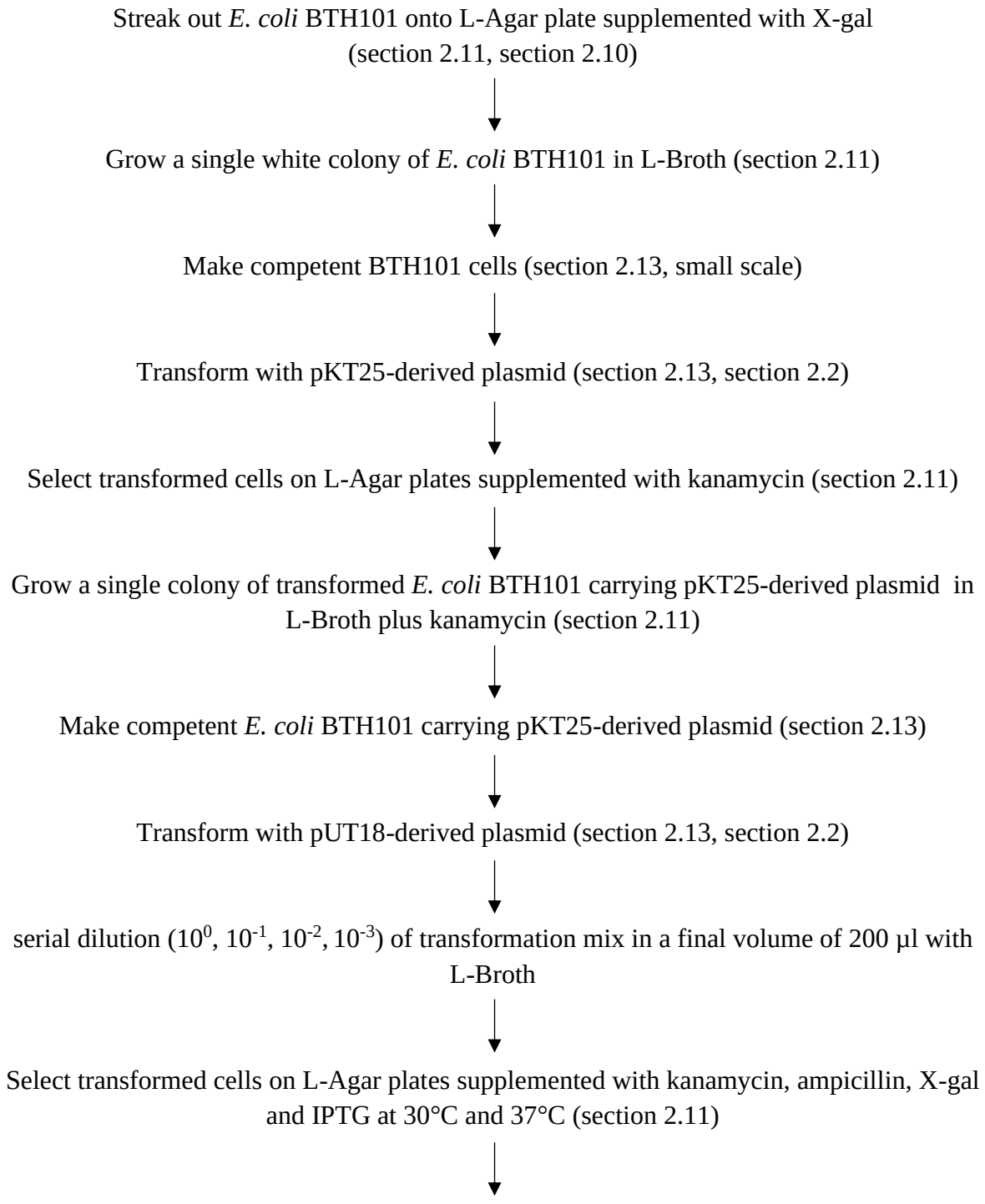


Figure 2.2. Flow diagram summarising the method used for bacterial two-hybrid complementation assays.

CHAPTER THREE

RESULTS AND DISCUSSION:

Efficient translational bypassing promoted by the *cis*-acting signals and nascent peptide of bacteriophage T4 gene *60*

3.0 Overview

Bacteriophage T4 DNA topoisomerase gene *60* contains a 50 nucleotide non-coding region within the coding sequence of its mRNA. Half of the elongating ribosomes bypass this region during translation, resulting in synthesis of a full-length DNA topoisomerase subunit (gp60) from the discontinuous reading frame. Essential features necessary for this translational bypassing are a nascent peptide stalling motif, identical codons bordering the bypassed region and a stop codon contained in a stem-loop structure forming at the beginning of the bypass region. In this work, a His-tagged gene *60-iLOV* reporter system was made to probe these bypassing stimulatory elements. Efficient translational bypassing was observed when a modified segment of gene *60* fused to iLOV at its 3' end and a His-tag at its 5' end was expressed in *E. coli* and in *in vitro* translation systems from *E. coli*. These findings confirmed the previously characterized translational bypassing stimulatory elements of T4 gene *60* and revealed that a suggested 3' feature in gene *60* is likely unnecessary for bypassing.

3.1 Introduction

Translation of the mRNA into protein by the ribosome and associated translational components is invariably performed in all living cells. The cell's highly evolved protein synthesis machinery is programmed to translate each mRNA according to a set of decoding rules that relies on successive interpretation of non-overlapping triplet nucleotide codons (Crick *et al.*, 1961). While most translational events abide to this standard program with high fidelity (Mohler and Ibba, 2017), there are many instances in which canonical translation is altered in response to *cis*-acting signals within the mRNA and nascent polypeptides (Baranov *et al.*, 2015). These events, termed translation reprogramming or recoding, are generally utilized for regulatory purposes and serve to expand the coding capacity of genes (Rodnina *et al.*, 2020).

One example of reprogramming is translational bypassing, wherein the ribosome produces a single polypeptide chain from a discontinuous reading frame. Although translational bypassing was believed to be rare and unique in translation of bacteriophage T4 DNA topoisomerase gene *60* (Huang *et al.*, 1988), this recoding event has also been observed in expression of several yeast mitochondrial genes (Lang *et al.*, 2014; Nosek *et al.*, 2015) and other programmed synthetic gene constructs (Gallant *et al.*, 2003; Gallant *et al.*, 2004; Herr *et al.*, 2004; Wills *et al.*, 2008; Bucklin *et al.*, 2005).

During translation of bacteriophage T4 gene *60* mRNA, which encodes a type II DNA topoisomerase subunit (termed gp60) (Seasholtz and Greenberg, 1983), the elements that promote bypassing include a particular sequence in the nascent peptide, identical 'take-off' and 'landing' codons flanking the bypassed region and a stop codon contained in an mRNA stem loop (termed 'take-off' stem loop) that forms around the take-off codon (Weiss *et al.*, 1990) (Figure 3.1).

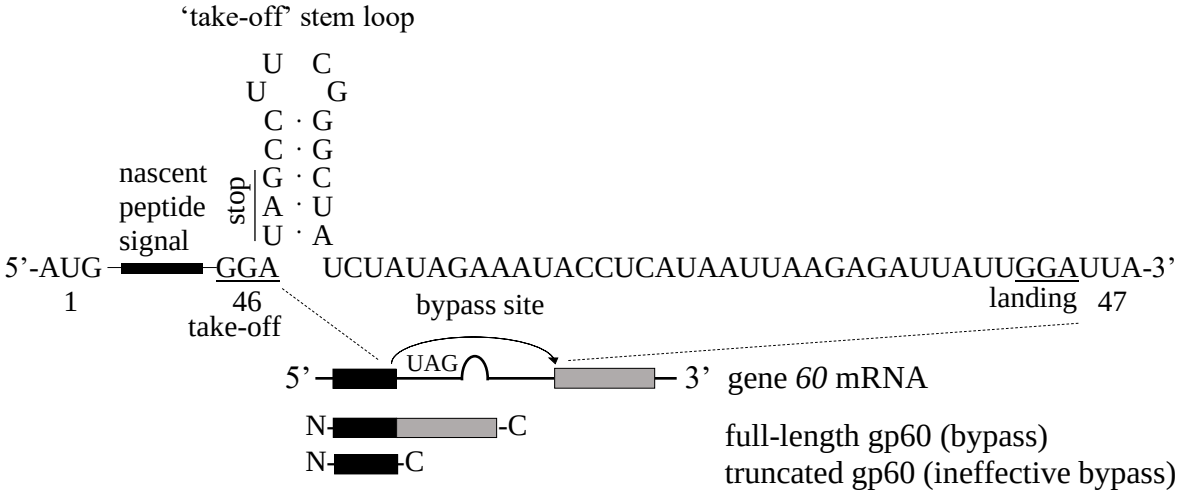


Figure 3.1. Signal elements in gene *60* mRNA that promote translational bypassing. (i) matched 'take-off' and 'landing' codons GGA, (ii) the in-frame UAG stop codon (stop) in the mRNA 'take-off' stem loop and (iii) the nascent peptide signal. The bypassed region starts at the stop codon and ends at the landing site (50 nucleotides long). The bypassing event occurs when tRNAs occupy the 'take-off' site and is stimulated by the 'take-off' stem loop. The matched 'landing' site codon allows the translational apparatus to re-attach or 'land' on the mRNA after 'take-off' and resume translation, and the nascent peptide induces a translational pause prior to bypass for an efficient reprogramming event (Chen *et al.*, 2015; Agirrezabala *et al.*, 2017). Translation of gene *60* mRNA yields two distinct polypeptides at close to a 1:1 ratio (Maldonado and Herr, 1998): (1) full-length gp60 that results from bypass and (2) truncated gp60 that results from ineffective bypass (Herr *et al.*, 2000).

Other evidence also suggests that 5' and 3' RNA stem loop structures flanking the bypass site are necessary for efficient bypassing (Samatova *et al.*, 2014). However, this is debatable, since a study revealed that substitution of the 3' sequence of gene *60* by the coding region of *lacZ* had negligible effect on bypassing efficiency (Weiss *et al.*, 1990).

The exact mechanism by which these translational bypassing stimulatory elements operate are not entirely understood (Agirrezabala *et al.*, 2017), and in some cases, they may be influenced by auxiliary factors such as the ribosomal protein L9 (Herbst *et al.*, 1994), availability of EF-G (Klimova *et al.*, 2019) and ribosomal loading on the mRNA or polysomes (O'Loughlin *et al.*, 2020). Translational bypassing in gene *60* benefits phage T4 since it enables synthesis of the full-length functional T4 topoisomerase subunit gp60 from the discontinuous reading frame. This subunit is necessary for formation of an active topoisomerase complex (Seasholtz and Greenberg, 1983), which is an enzyme essential for viral replication, recombination, and transcription. New methods for probing bypass signals would provide additional routes of investigation for unraveling the mechanism of translational bypassing and would serve as a platform for engineering new tools to reprogram translation for biotechnology and synthetic biology.

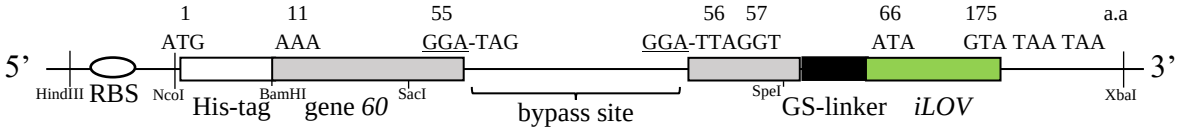
3.2 Results

3.2.1 Design of a His-tagged gene *60-iLOV* reporter system

A His-tagged gene *60-iLOV* reporter system was designed to probe the elements of bacteriophage T4 gene *60* that promote translational bypassing (Figure 3.2).

Briefly, the gene construct contains a modified segment of T4 gene *60* from codons 2 to 46, encoding the first 45 amino acids of gp60 from the lysine to the take-off glycine, and including the nascent peptide, the 50 nucleotide bypass site and the last 2 codons after the landing codon fused downstream to a GS-linker and *iLOV* gene sequence (*iLOV*, which encodes a fluorescent reporter, lacks initiator AUG codon so that its level of expression is dependent on translational bypassing). The His-tag fused upstream to the gene *60* segment allows purification of all resulting protein products by affinity chromatography and serves to exclude sequences that initiate internally at AUG or other start codons. When the elongating ribosome reaches the bypass site, it will either terminate translation at the UAG stop codon that compromises the first 3 nucleotides of the bypass sequence releasing a 55-amino acid truncated non-fluorescent recombinant protein of ~ 6 kDa, or it will engage in productive bypass producing a 175-amino acid full-length fluorescent recombinant protein of ~ 19 kDa. The relative stoichiometry of the protein outputs is dependent on the efficiency of the reprogramming event. By analysis of purified protein products on SDS-PAGE and fluorescence of purified protein products, one can deduce translational bypassing activity and measure its relative efficiency.

A.



B.

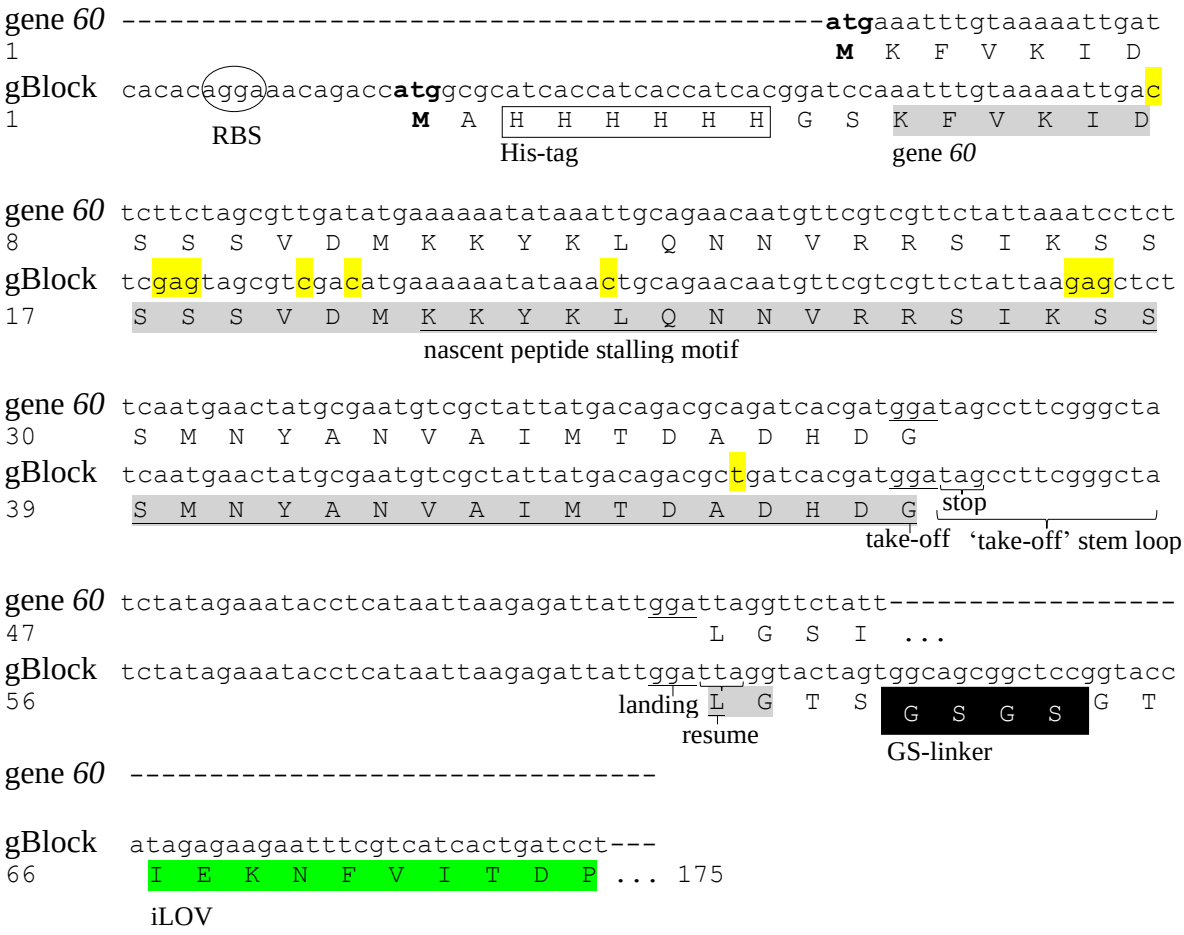


Figure 3.2. His-tagged gene *60-iLOV* reporter system. A. Schematic representation of the reporter system composed of a ribosome binding site (RBS) from pTrc99A, 6 histidine codons (His-tag)**,** a modified segment of T4 gene *60* starting from second codon to the last 2 codons after the landing site (grey) including the 50 nucleotide bypass site, a GS-linker (black) and *iLOV* lacking an initiation codon (green). Take-off and landing codons GGA are underlined. The codon/amino acid (a.a) numbers are shown (not drawn to scale). Unique restriction sites used in this study are shown. **B.** Nucleotide sequence (5'-3') and derived amino acid sequence of the reporter system synthesized as a gblock oligonucleotide by Integrated DNA technologies (IDT) (Chapter 2, Table 2.9). Wild-type bacteriophage T4 gene *60* (accession no. HM_137666.1) is shown for comparison. Nucleotide mismatches are highlighted in yellow (silent mutations that still encode same amino acids as wild-type gp60 that were made to introduce unique restriction sites for future cloning purposes).

3.2.2 Efficient translational bypassing *in vivo*

The gene construct (Figure 3.2) was chemically synthesized as a gblock oligonucleotide and cloned into an expression plasmid under the Trc promoter resulting in plasmid construct termed pHop (Figure 3.3A). This plasmid was introduced into *E. coli* C41, a derivative strain of BL21 (DE3) suitable for production of proteins at high levels from various promoters including Trc promoter (Miroux and Walker, 1996), for expression and purification of the resulting protein products. Analysis of the nickel affinity purified protein products on a protein gel showed a band corresponding in molecular weight to the full-length product at ~ 19 kDa (Lanes 6-7, Figure 3.3B). The identity of the band was further supported by a control construct termed pHopΔgap which had an exact deletion of the bypass site (Figure 3.3A), and whose standard translation generated a single product corresponding in molecular weight to the full-length product at ~ 19 kDa (Lanes 11-12, Figure 3.3B).

As a control for size validation of the truncated product, construct termed pHop-Stop was made in which the leucine TTA resume codon (the codon following the landing codon) was changed to a TAA stop codon by substitution mutagenesis of pHop (Figure 3.3A; Chapter 2-section 2.20). The stop codon would prevent ribosomes from resuming translation after bypass and signal termination of translation resulting in a truncated 6 kDa product. The truncated product at ~ 6 kDa was not detected (Lanes 20-24, Figure 3.3B), indicating its degradation or insolubility. Surprisingly, a less intense band corresponding in molecular weight to full-length product at ~ 19 kDa was observed (Lane 23, Figure 3.3B). This was probably due to a low-level of natural suppression of termination in *E. coli* C41 (Angius *et al.*, 2018) after bypass or aberrant landing at the next in-frame glycine GGU/GGT codon #57 in pHop-Stop (Figure 3.3A). The identity of this product could be verified by future mass spectrometry or protein sequencing experiments to confirm whether it is a product of stop codon read through or aberrant landing.

Another construct, termed pHop-L47, was made in which the leucine TTA resume codon (#56) was changed to a more preferred leucine CTG codon (in an *E. coli* context) by substitution mutagenesis of pHop (Figure 3.3A; Chapter 2-section 2.20), to test the effect of codon bias on ribosomes resuming translation and to serve as a codon optimized derivative of pHop, where both encode identical proteins. Similarly, a band corresponding in molecular weight to the full-length product at ~ 19 kDa was observed (Lanes 18-19, Figure 3.3B).

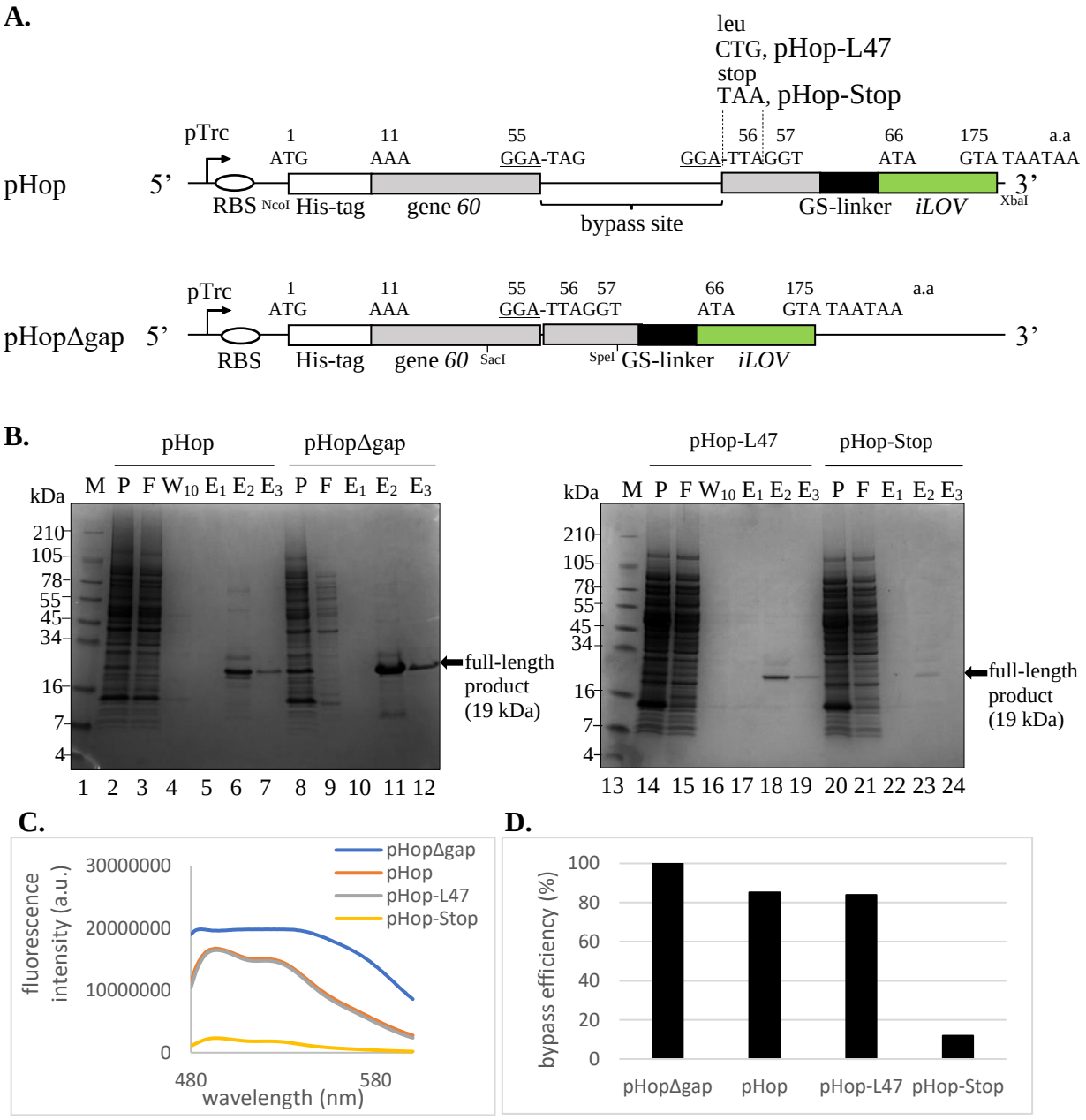


Figure 3.3. Structure and expression of His-tagged gene *60-iLOV* fusions in *E. coli* C41. **A.** Diagram of His-tagged gene *60-iLOV* fusions present in a modified pTrc99A expression vector. Expression signals and coding regions of these constructs are as indicated. **B.** The plasmid constructs (Figure 3.3A) were introduced and expressed in *E. coli* C41 (50 ml culture) induced with 1 mM IPTG for 3 hours, the protein products were purified using nickel affinity chromatography, resolved on a 10-20% Tricine gel and stained with coomassie blue. SeeBlue Plus2 pre-stained protein standard (M), supernatant prior to column passage (P), flow-through (F), final wash (W_{10}), consecutive elution of nickel column-bound protein with 400 mM imidazole (E_1, E_2, E_3). **C.** iLOV fluorescence emission spectra from the purified protein in E_2 (Figure 3.3B). **D.** Estimated bypass efficiency (%) = (fluorescence intensity at 495 nm of pHop, pHop-L47 or pHop-Stop x 100%) ÷ (fluorescence intensity at 495 nm of pHopΔgap) (Figure 3.3C).

The presence of the full-length product in the nickel column purified fractions was further confirmed by fluorescence and detection of the distinctive iLOV fluorescence emission spectra (Figure 3.3C), which is characterized by two peaks at wavelength 495 nm and 520 nm (Chapman *et al.*, 2008).

An accurate bypass efficiency could not be calculated (ratio of full-length to truncated product (Maldonado and Herr, 1998)), since the truncated product at ~ 6 kDa was not detected (Figure 3.3B). However, a bypass efficiency was estimated by comparing the fluorescence emission intensity at wavelength 495 nm of the test samples pHop and pHop-L47 to the fluorescence emission intensity at wavelength 495 nm of the control sample pHopΔgap (Figure 3.3C):

$$Bypass\ efficiency\ (\%) = \frac{Fluorescence\ intensity\ (a.u.)\ at\ 495\ nm\ from\ test\ construct\ \ x\ 100\%}{Fluorescence\ intensity\ (a.u.)\ at\ 495\ nm\ from\ pHop\Delta gap\ (control)}$$

This calculation assumes that recovery of proteins from the nickel column is the same in all experiments. This seems a reasonable assumption (or was the best that could be done in the absence of quantifiable 6 kDa truncated protein). This resulted in an estimated bypass efficiency of 85% in both test constructs pHop and pHop-L47 compared to the control construct pHopΔgap (100%) (Figure 3.3D). Similarly, a bypass + stop codon read through or aberrant landing efficiency of 12% in control construct pHop-Stop was estimated (Figure 3.3D), since most elongating ribosomes would bypass and then terminate translation at the TAA resume stop codon (Figure 3.3A).

The His-tagged gene *60-iLOV* reporter system used is a valid approach to demonstrate that *E. coli* ribosomes engage in effective translational bypassing when encountering the bypass signals of T4 gene *60* and that the recoding process is efficient (85%). Similar bypass efficiencies have been reported previously for other reporter systems *in vivo* (Weiss *et al.*, 1990), which initially identified this minimal segment of gene *60* containing all the elements necessary for effective translational bypassing. Moreover, codon bias of the resume codon following the landing site codon appears to have no effect on ribosomes resuming translation or bypassing since similar bypass efficiencies were obtained from the constructs with TTA or CTG leucine resume codons (pHop and pHop-L47, respectively) (Figure 3.3D).

3.2.3 Efficient translational bypassing *in vivo* compared to *in vitro*

The gene construct (Figure 3.2) was cloned in a different expression vector under the control of the T7 promoter resulting in a plasmid construct termed $pHop_{T7}$ (Figure 3.4A) (subcloning was not possible since the T7 vector (pGEM-4Z) lacks an RBS, however, the gene construct was pre-designed with the RBS sequence from pTrc99A), for coupled transcription-translation *in vitro* in the presence of radioactive ^{35}S-Methionine. Products of *in vitro* transcription and translation were observed (without nickel affinity purification) by SDS-PAGE followed by autoradiography. A band corresponding in molecular weight to the full-length product at ~ 19 kDa was detected (Lane 1, Figure 3.4B). The identity of this band was confirmed by control construct $pHop\Delta gap_{T7}$ which had an exact deletion of the bypass site (Figure 3.4A), and whose standard translation resulted in full-length product at ~ 19 kDa (Lane 3, Figure 3.4B). A second major band corresponding in molecular weight to ~ 17 kDa was also observed (Lane 3, Figure 3.4B), which could have resulted from internal initiation of translation. A similar molecular weight band has been previously observed (Maldonado and Herr, 1998), corresponding to a protein product of internal initiation at methionine #13 in T4 gene *60*.

As a control for the truncated product, construct termed $pHop\text{-}Stop_{T7}$ was made in which the leucine TTA resume codon was changed to a TAA stop codon by substitution mutagenesis of $pHop_{T7}$ (Figure 3.4A; Chapter 2-section 2.20), however, the truncated product at ~ 6 kDa could not be clearly detected (Lane 4, Figure 3.4B). Another construct, termed $pHop\text{-}L47_{T7}$, was made in which the leucine TTA resume codon was changed to a more preferred leucine CTG codon by substitution mutagenesis of $pHop_{T7}$ (Figure 3.4A; Chapter 2-section 2.20), to test the effect of codon bias on ribosomes resuming translation and to serve as a codon optimized derivative of $pHop_{T7}$. Similarly, a band corresponding in molecular weight to the full-length product at ~ 19 kDa was observed (Lane 2, Figure 3.4B). An accurate bypassing efficiency *in vitro* could not be measured since the truncated product was not detected. However, a bypassing efficiency was estimated by using band intensity relative to the control $pHop\Delta gap_{T7}$ (Figure 3.4B). This resulted in a bypassing efficiency of 15% for both test construct $pHop_{T7}$ and $pHop\text{-}L47_{T7}$. Similar bypass efficiencies were reported previously for other translational bypassing assays performed *in vitro* (Samatova *et al.*, 2014). Future work could verify if fluorescence accurately reflects expression of the full-length fusion protein and further confirm the identity of the bands by western blotting using an anti-iLOV antibody (Gawthorne *et al.*, 2012).

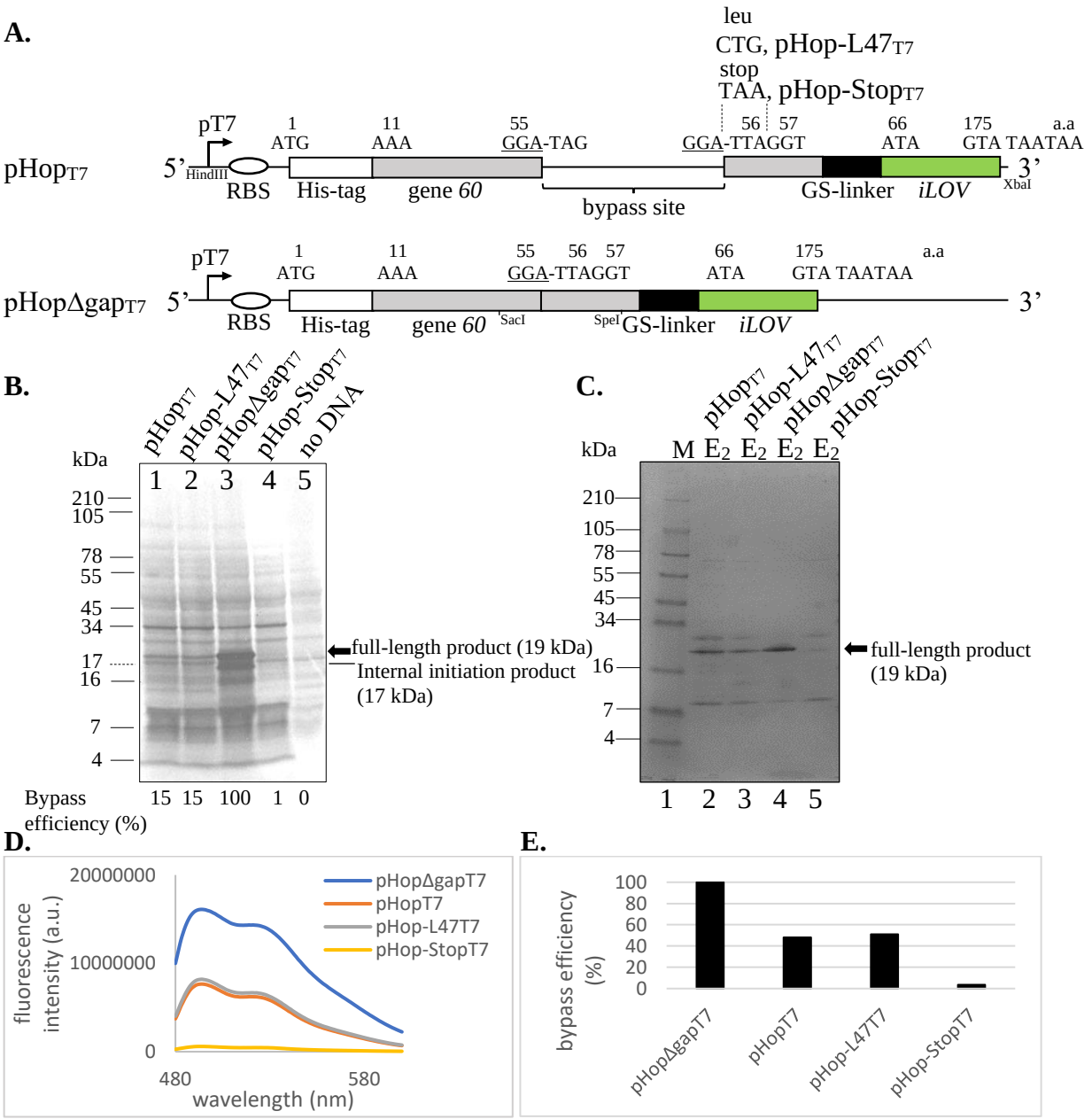


Figure 3.4. Structure and expression of His-tagged gene *60-iLOV* fusions *in vivo* and *in vitro*. **A.** Diagram of His-tagged gene *60-iLOV* fusions present in a pGEM-4Z expression vector. **B.** Linear DNAs were amplified from these constructs (Figure 3.4A; Chapter 2-section 2.18) and used for *in vitro* translation using commercially available kit. Protein products were resolved on a 10-20% Tricine gel, fixed, dried and placed with a film in light-proof cassette for 24 hours before development. **C.** The plasmid constructs (Figure 3.4A) were expressed in *E. coli* C41 (50 ml culture) induced with 1 mM IPTG for 3 hours, the protein products were purified using nickel affinity chromatography, resolved on a 10-20% Tricine gel and stained with coomassie blue. SeeBlue Plus2 pre-stained protein standard (M), second elution of nickel column-bound protein with 400 mM imidazole (E_2). **D.** iLOV fluorescence emission spectra from the purified protein in sample E_2 (Figure 3.4C). **E.** Estimated bypass efficiency (%) (as in Figure 3.3D).

The plasmid constructs (Figure 3.4A) were also introduced into *E. coli* C41, which expresses the T7 RNA polymerase, for protein expression and purification. Analysis of the nickel affinity purified protein products showed a major band corresponding in molecular weight to the full-length product at ~ 19 kDa (Figure 3.4C), similar to previous results (Figure 3.3). The presence of the full-length product in the purified protein samples was further confirmed by fluorescence and detection of the distinctive iLOV fluorescence emission spectra (Figure 3.4D). While an accurate bypassing efficiency *in vivo* could not be measured, it was estimated as previously described, resulting in a bypassing efficiency of ~ 50% for both constructs pHop$_{T7}$ and pHop-L47$_{T7}$, and at less than 10% for pHop-Stop$_{T7}$ (Figure 3.4E).

The reasons behind bypass efficiency being relatively low *in vitro* is unclear. The relative contribution of the bypassing elements may vary *in vivo* and *in vitro*, such as the changes in the components of the translational system *e.g.* the mRNA stability (Maldonado and Herr, 1998) and mRNA folding (Schroeder *et al.*, 2002), that could affect bypassing efficiency.

3.2.4 Bypassing activity is decreased by addition of N-terminal His-tagged MBP

To detect both the full-length and truncated protein products, a His-tagged maltose-binding protein (MBP) coding region was amplified from pMAL-c2X (Chapter 2-section 2.18) and cloned into a modified version of the constructs under the T7 promoter (Figure 3.4A, with one BamHI site removed from the vector backbone to keep unique site for cloning) resulting in plasmid constructs termed pHop$_{T7}$-MBP, pHopΔgap$_{T7}$-MBP, pHop-L47$_{T7}$-MBP and pHop-Stop$_{T7}$-MBP (Figure 3.5A). The His-tagged MBP should increase the solubility of all protein products fused to it (Pryor and Leiting, 1997; Kapust and Waugh, 1999), increase their size by 42 kDa and prevent potential degradation − facilitating purification of the truncated product. The plasmid constructs were expressed in *E. coli* C41 and the protein products were purified by nickel affinity chromatography using the His-tag. Bands corresponding in molecular weight to the full-length product at 61 kDa (19 kDa + 42 kDa) and the truncated product at 48 kDa (6 kDa + 42 kDa) were detected (Figure 3.5B). A bypassing efficiency was calculated by ratio of full-length to truncated product which resulted in bypassing efficiencies less than <1% in pHop$_{T7}$-MBP and pHop-L47$_{T7}$-MBP (Figure 3.5C).

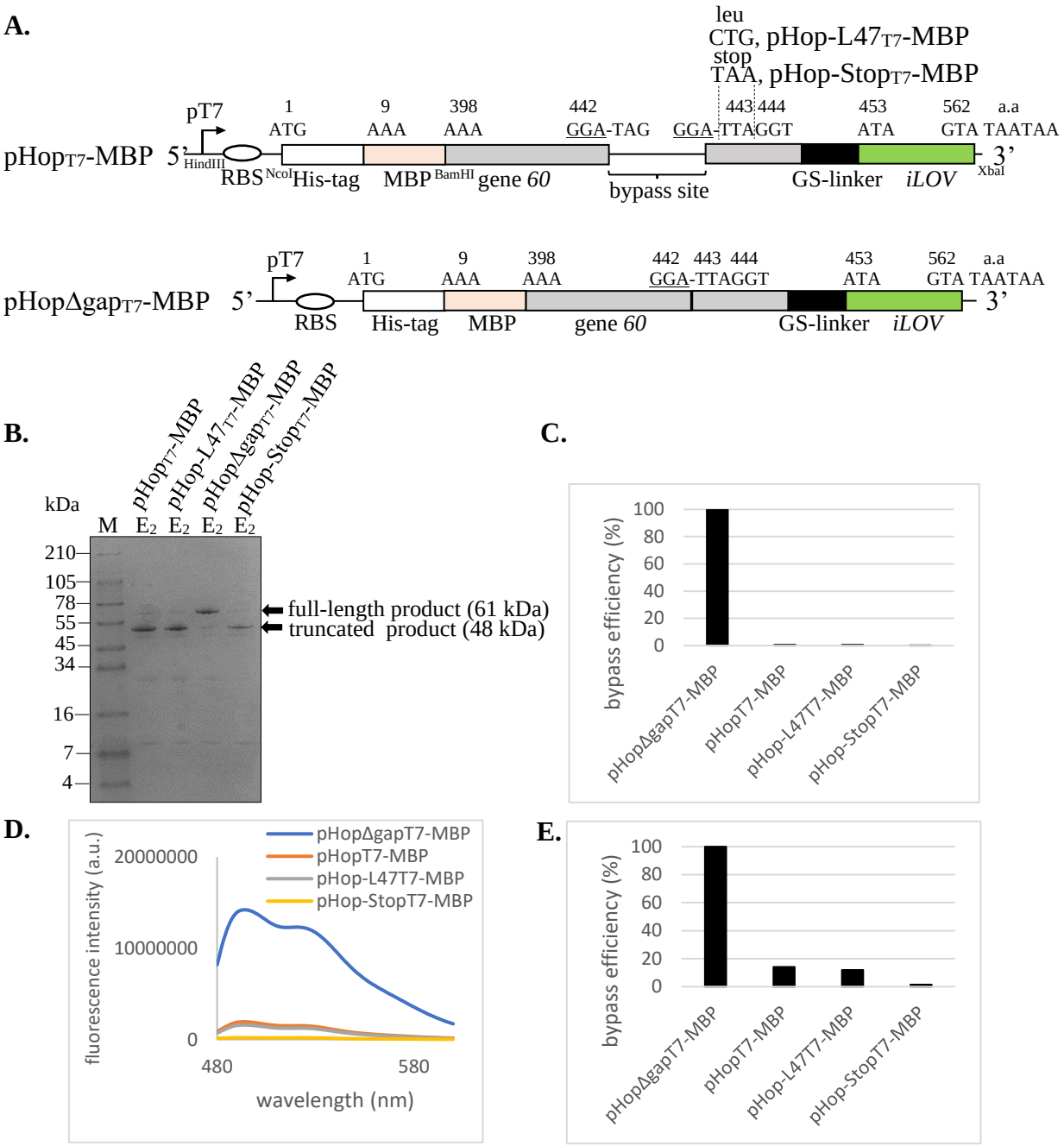


Figure 3.5. Structure and expression of modified translational bypassing reporter systems in *E. coli* C41. A. Diagram of His-tagged MBP-gene *60-iLOV* fusion constructs in modified pGEM-4Z vector. **B.** The plasmid constructs (Figure 3.5A) were expressed in *E. coli* C41 (50 ml culture) induced with 1 mM IPTG for 3 hours, the protein products were purified using nickel affinity chromatography, resolved on a 10-20% Tricine gel and stained with coomassie blue. SeeBlue Plus2 pre-stained protein standard (M), second elution of nickel column-bound protein with 400 mM imidazole (E_2). **C.** Bypass efficiency (%) = (full-length band intensity (Rf)) ÷ (full-length band intensity (Rf) + truncated band intensity (Rf)) (Figure 3.5B); Rf was measured using GelAnalyzer 19.1 software. **D.** iLOV fluorescence emission spectra from the purified protein in sample E_2 (Figure 3.5B). **E.** Estimated bypass efficiency (%) by means of fluorescence as in Figure 3.3D.

The presence of the full-length product in the nickel column purified proteins was further confirmed by detection of the distinctive iLOV fluorescence emission spectra (Figure 3.5D). The estimated bypass efficiency by means of iLOV fluorescence was 15% in both test constructs pHop$_{T7}$-MBP and pHop-L47$_{T7}$-MBP compared to the control construct pHopΔgap$_{T7}$-MBP (100%), and ~ 2% for pHop-Stop$_{T7}$-MBP (Figure 3.5E). Presence of other minor bands were probably contaminants not removed by purification steps (Figure 3.5B).

Fusion of the His-tagged MBP to the constructs appears to decrease bypass efficiency (compare bypass efficiencies: 50% in constructs pHop$_{T7}$, pHop-L47$_{T7}$, pHopΔgap$_{T7}$ and pHop-Stop$_{T7}$ lacking the His-tagged MBP fusion (Figure 3.4E) to 15% in the derived constructs with the MBP fusion (Figure 3.5E)). Although this optimized reporter system allows accurate estimation of bypass efficiencies, the large fusion in turn affected bypassing activity. It has been previously reported that large additions at the N-terminus of a gene *60* construct had a negligible effect on translational bypassing activity and efficiency (Maldonado and Herr, 1998; Wills *et al.*, 2008). It is unclear how a large addition at the N-terminus of these constructs here affects bypass efficiency. Previously, a cross-linking experiment has shown that the N-terminal part of gene *60* nascent peptide present outside the ribosomal exit tunnel could interact with the small ribosomal subunit near the decoding region and that this interaction could contribute to bypassing (Choi *et al.*, 1998). The large His-tagged MBP addition could have affected this interaction, potentially decreasing bypass efficiency.

3.3 Discussion

RNA and nascent peptides play a significant role in reprogramming translation of the genetic information. Naturally occurring elements have provided many examples of RNA structures and sequences, as well as nascent peptide motifs that reprogram the protein synthesis machinery (Atkins and Gesteland, 2010). Development of simple methods that probe and characterize recoding stimulatory elements is necessary for providing mechanistic insights into these recoding events. The translational bypassing reporter system developed in this work is a valid approach to demonstrate that *E. coli* ribosomes engage in efficient translational bypassing when encountering the *cis*-acting RNA signals and the nascent peptide stalling motif from bacteriophage T4 gene *60*. Similar reporter systems of translational bypassing have been previously used to probe those elements whereby the

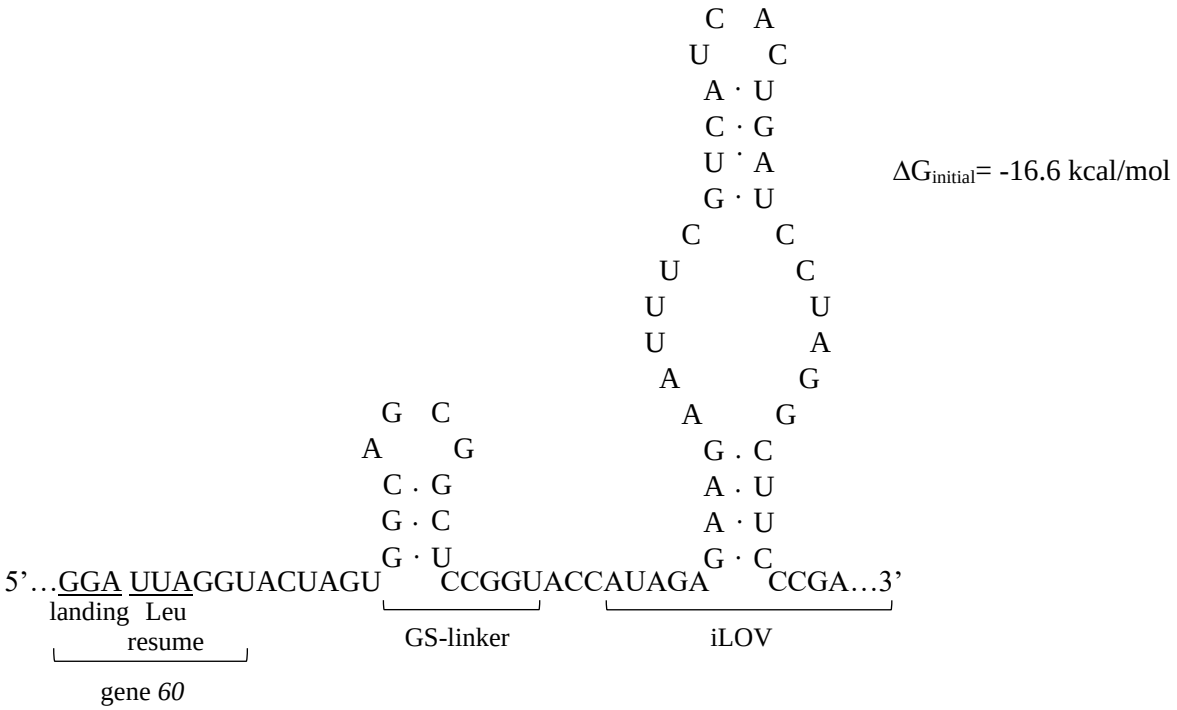
minimal segment of gene *60* was directly fused downstream to the *lacZ* coding region to estimate bypassing through β-galactosidase expression (Weiss *et al.*, 1990). Our system differs in that it allows observation of translational bypassing via fluorescence and allows purification of all His-tagged protein products with their visualization on protein gels, potentially excluding false positive results that could arise from internal initiation of translation.

The control experiments performed validate the hypothesis that the synthesis of the full-length product results from full translation of the mRNA programmed with T4 gene *60* bypass signals by translational bypassing. Importantly, it demonstrates that the system can be used to screen for authentic translational bypassing stimulatory elements. Overall, these results establish the reporter system as an effective tool for the probing of translation reprogramming elements.

The present findings and previous *in vivo* reporter systems (Weiss *et al.*, 1990) showed that translational bypassing was observed when the sequence 2 codons downstream of the landing site codon of bacteriophage T4 gene *60* was replaced by the reporter *iLOV* (in this work) or *LacZ* coding sequence (Weiss *et al.*, 1990). This confirmed that the 3' sequence of gene *60* is not involved in bypassing and only serves as a landing platform for the ribosome. These findings appear to contradict previous suggestions that an RNA feature (a stem loop) in the 3' sequence of gene *60* is necessary for bypassing (Samatova *et al.*, 2014). One possible explanation for these observations is that an RNA structure in the 3' coding region is not sequence-specific and can be provided by different mRNA sequences, although further experiments will be required to confirm this.

Chemical probing of the structure of gene *60* mRNA showed the 3' region to be unstructured (Todd and Walter, 2013), similar to previous computational predictions (Le *et al.*, 1993). The predicted RNA structure of *iLOV* coding sequence in our reporter system shows presence of a weak structural RNA element (Figure 3.6A), compared to the proposed RNA structure downstream of the landing site in wild-type gene *60* (Figure 3.6B). Therefore, it is likely that a 3' RNA feature or weak potential secondary structure elements that may exist is unnecessary for bypassing. Taken together, this data and previous data (Weiss *et al.*, 1990), indicate that the 3' sequence is dispensable in translational bypassing and can be replaced by a variety of sequences without affecting bypass activity.

A.



B.

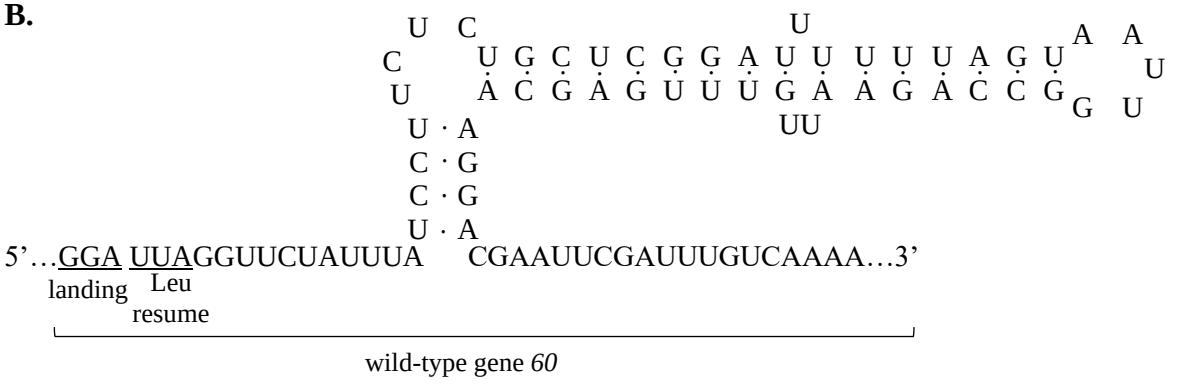


Figure 3.6. Sequence and putative secondary structure of the downstream part of mRNA. A. A potential RNA structure downstream of the landing codon in the His-tagged gene *60-iLOV* fusion predicted by Mfold web server (Zuker, 2003). **B.** Proposed RNA structure downstream of the landing codon in wild-type gene *60* mRNA (Samatova *et al.*, 2014).

Future genetic analysis could be used to probe the effect of mutagenesis of the translational bypassing stimulatory elements of gene *60* in this reporter system on bypassing activity. Future work could also reveal the reason behind the decrease in bypassing efficiency after fusion of the His-tagged MBP to the reporter system (would unfolded MBP also affect bypass efficiency or is there something unique about the folded form of MBP). Beyond probing of the translational bypassing elements, this reporter system potentially provides a useful method for emerging biotechnologies. This work demonstrated that the translational bypassing elements of T4 gene *60* are capable of reprogramming translation in recombinant gene contexts. This has potential applications to the field of reprogramming translation through RNA, which utilizes programmed recoding constructs for the controlled expression of designer proteins. This minimal sequence of gene *60* used in this reporter system could be used to control the relative stoichiometry of two distinct protein outputs by encoding them in a single gene. Systems using frameshift elements for such purposes have been previously described (Anzalone *et al.*, 2016) and could by replaced with bypass elements. Generally, this strategy provides a framework for engineering new devices for biotechnology and synthetic biology (see Chapter 4).

CHAPTER FOUR

RESULTS AND DISCUSSION:

Translational bypassing devices that regulate gene expression in bacteria

4.0 Overview

Translation can be reprogrammed by a wide variety of RNA features and nascent peptides. Translational bypassing is one such reprogramming mechanism that utilizes *cis*-acting signals within the mRNA to redirect a proportion of ribosomes into a new translation reading frame that gives rise to an alternative protein product. This work establishes a strategy for engineering *de novo* translational bypassing devices and their application for regulation of protein synthesis in bacteria. First, translational bypassing devices were designed based on the 50-nucleotide bypass element in bacteriophage T4 gene *60*, commercially chemically synthesized as oligonucleotides, and inserted into the coding region of bacteriophage ϕC31 serine integrase gene *int* by mutagenesis approaches. Expression of the ϕC31 serine integrase was then followed using a well-established *in vivo* recombination assay in *E. coli* and translation systems of *E. coli*. Evidence is presented here that demonstrates that the full-length ϕC31 serine integrase is expressed from the discontinuous translation frames by translational bypassing, and that the engineered translational bypassing devices control the relative stoichiometry of two major distinct protein outputs from one gene. These results establish the translational bypassing devices as tools to regulate gene expression in bacteria and reveal for the first time that the 50-nucleotide bypass element of gene *60* alone can promote a low-level translational bypassing event when present in an alternative genetic context. Overall, this demonstrates the potential for harnessing translational bypassing for biotechnology and expands our synthetic capabilities to manipulate biological systems.

4.1 Introduction

Translational bypassing enables ribosomes to ‘ignore’ a precise mRNA interval of several dozen nucleotides. Well-characterized bypassed sequences are present in bacteriophage T4 gene *60* (Huang *et al.*, 1988) and yeast mitochondrial genes (Lang *et al.*, 2014; Nosek *et al.*, 2015). Although protein translation by the ribosome and its associated translational components occurs with high fidelity (Ramakrishnan, 2002; Ogle and Ramakrishnan, 2005), there are instances when translation is temporarily altered to change the protein output of a gene (Baranov *et al.*, 2015). This reprogramming enables synthesis of proteins containing non-canonical amino acids (such as selenocysteine and pyrrolysine) or expression of multiple distinct proteins from an individual gene or mRNA (Atkins and Gesteland, 2010).

The bypassing mechanism in gene *60* utilizes *cis*-acting RNA elements and a nascent peptide to signal the recoding event at a specific location along the mRNA (Weiss *et al.*, 1990; Samatova *et al.*, 2014; Agirrezabala *et al.*, 2017). This recoding mechanism is similar in yeast mitochondrial genes, although less constrained to the sequence context and requirement of a nascent peptide stalling motif (Nosek *et al.*, 2015). Translational bypassing key signals are composed of two features: (i) matched codons (denoted 'take-off' and 'landing' site codons) present at a defined distance from each other, and (ii) a stop codon contained in a stimulatory RNA stem loop structure (termed 'take-off' stem loop) adjacent to the take-off site that stimulates bypass (Rodnina *et al.*, 2020). When encountering translational bypassing signals in an mRNA, a proportion of elongating ribosomes are programmed to skip many nucleotides, placing the translation apparatus in an alternative reading frame. This, in turn, alters the amino acid composition of the polypeptide that is synthesized downstream of the bypass site. Generally, translational bypassing serves to establish a specific stoichiometry of full-length and truncated proteins, reminiscent of frameshifting (Farabaugh, 1996). Regardless of variation in mRNA transcript levels or translational activity, the stoichiometry of bypass to non-bypass protein products remains constant for the given translational bypassing stimulatory signals (Maldonado and Herr, 1998).

The site-specific recoding and the ability to set a relative abundance of two distinct proteins from one gene, offers an attractive opportunity to engineer genetic devices that regulate protein synthesis. While some forms of translation reprogramming or recoding have been employed for biotechnology, including -1 programmed ribosomal frameshifting (Anzalone *et al.*, 2016; Matsumoto *et al.*, 2018; Dinman, 2019), 2A peptides (Szymczak *et al.*, 2004; Fang *et al.*, 2005; De Felipe *et al.*, 2006) and redefinition of stop codons to specify unnatural amino acids in cells (Liu and Schultz, 2010), other mechanisms of translation reprogramming remain unexplored despite their potential application for biotechnology and synthetic biology. The success of these devices that have been used in medical research and gene regulation, is attributed to the modularity of genetic device engineering, that could be applied beyond previously explored devices.

This work identified translational bypassing as a potential modular mechanism for developing new gene-regulatory tools. Ideally, translational bypassing devices are short in length, have well-defined structures and signal translational bypassing. However, translational bypassing signals are diverse in both sequence and size (Herr *et al.*, 2000), and in some cases, require additional elements for bypassing activity (Herbst *et al.*, 1994;

O'Loughlin *et al.*, 2020; Klimova *et al.*, 2019). These limitations place constraints on the scope of available parts, the flexibility of engineering, and the achievable dynamic range of translational bypassing devices. A translational bypassing system would be useful for biotechnology in that it would enable synthesis of two distinct protein products from one gene at defined stoichiometric ratios.

To address this, a general strategy for engineering *de novo* translational bypassing devices and their application for regulation of protein synthesis in bacteria is established in this work. Engineered devices, derived from the 50-nucleotide bypass element of bacteriophage T4 gene *60*, demonstrate the ability to signal a low-level translational bypassing event when present in genetic contexts unrelated to gene *60 i.e.* in the coding region of bacteriophage ϕC31 serine integrase gene *int*, resulting in full-length (bypass) and truncated (non-bypass) protein products. The advantages of these constitutively active translational bypassing devices over other gene-regulatory devices are discussed, with the potential of creating ligand-responsive translational bypassing switches. Together, this establishes translational bypassing devices as tools for biotechnology and synthetic biology, expanding our capabilities to manipulate biological systems and opening new perspectives for studies of recoding events.

4.2 Results

To investigate whether the 50-nucleotide bypass element of gene *60* alone is sufficient for promoting translational bypassing, which would be ideal for translational bypassing device construction, the bypass element was introduced into the coding region of bacteriophage ϕC31 serine integrase gene *int* (this gene was selected since a well-established *in vivo* recombination assay that can follow ϕC31 serine integrase expression is available (Zhao *et al.*, 2019)).

4.2.1 Selecting locations in ϕC31 serine integrase gene *int* to insert the bypass element

The nucleotide sequence of ϕC31 serine integrase gene *int* was analyzed for homology to T4 gene *60* to determine insertion sites, since an efficient translational bypassing event was previously shown to require a glycine 'take-off' codon (Bucklin *et al.*, 2005). The alignment would also allow analysis of amino acid sequence homology to T4 gp60 to identify whether a nascent peptide stalling motif prior to the bypass site is present, which was previously shown to be involved in an efficient bypassing event (Weiss *et al.*, 1990; Samatova *et al.*, 2014; Chen *et al.*, 2015). Based on the alignment of various locations

of the ϕC31 serine integrase gene *int* to T4 gene *60* (Figure 4.1), three locations to introduce the bypass element after glycine codons G45, G327 and G537 were selected.

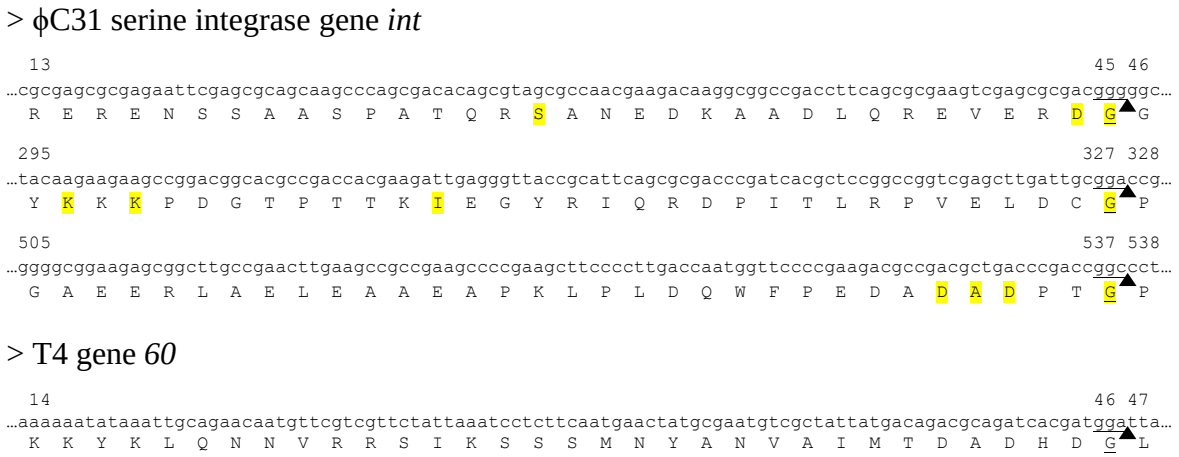


Figure 4.1. Alignment of ϕC31 serine integrase gene *int* to T4 gene *60*. Part of the nucleotide sequence (5'-3') and derived amino acid sequence of different parts of bacteriophage ϕC31 serine integrase gene *int* and bacteriophage T4 gene *60*. Codon/amino acid number location in the coding region is indicated. Amino acid similarities to T4 gp60 are highlighted in yellow. 'Take-off' glycine codon is underlined. Insertion site of the bypass element is shown by a solid triangle. Sequences upstream of the 'take-off' codon are shown for comparison to the nascent peptide stalling motif of T4 gene *60*.

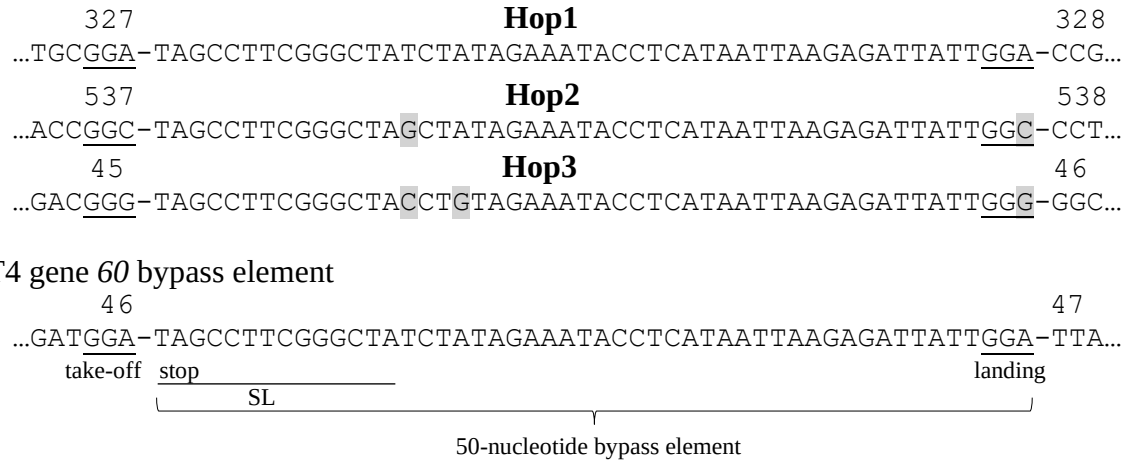
Few amino acid sequence similarities in these regions between ϕC31 serine integrase and T4 gp60 were found (Figure 4.1), suggesting that the amino acid sequences would not function as a stalling peptide sequence. However, they could possibly be changed, if necessary, since several of these regions are not conserved among integrases based on previous alignments (Van Duyne and Rutherford, 2013; Olorunniji *et al.*, 2019).

Since the selected 'take-off' glycine codons G45, G327 and G537 are GGG, GGA and GGC, respectively (Figure 4.1), the bypass element was designed with built-in alteration to the 'landing' glycine codon to match the identity of the take-off glycine codon it is inserted after (where necessary) and was termed Hop1, Hop2 and Hop3, accordingly (Figure 4.2A). Hop1 (introduced after G327 in ϕC31 serine integrase coding region) is identical to the 50-nucleotide bypass element in T4 gene *60* (Figure 4.2A), while Hop2 (introduced after G537) and Hop3 (introduced after G45) contain additional alterations (Figure 4.2A). It is known that matched take-off and landing glycine codons GGA, GGC or GGG have similar bypassing activities (Bucklin *et al.*, 2005), suggesting that these alterations would not affect

the translational bypassing activity of the bypass element. The other changes made in Hop2 and Hop3 were designed to extend the stem loop sequence to a similar length to that found in gene *60* – it extends into the take-off codon (although during translation this structure is unwound by the ribosome) (Figure 4.2B). Gene *60* bypass elements with similar alterations had negligible effect on bypassing activities (Weiss *et al.*, 1990), suggesting that these other changes would not affect the translational bypassing activity of these bypass elements.

A.

> ϕC31 bypass element



> T4 gene *60* bypass element

B.

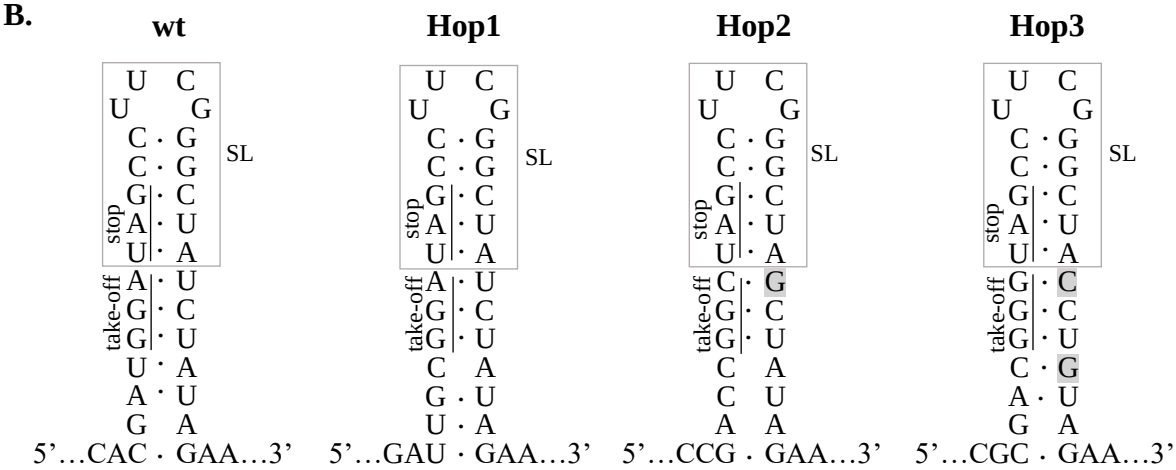


Figure 4.2. Design of bypass elements. A. Nucleotide sequence (5'- 3') of the designed bypass element (or "translational bypassing device") termed Hop1, Hop2 and Hop3. Translational bypassing signals present within the bypass element are indicated (take-off and landing codons are underlined, in-frame TAG stop codon (stop) contained in the take-off stem loop (SL)). Nucleotide mismatches in ϕC31 bypass element compared to T4 gene *60* bypass element are highlighted in grey (mutations that maintain the matching identity of take-off and landing codons and complementarity in the stem-loop as wild-type (wt) gene *60*'s bypass element). Chemically synthesized as oligonucleotide primers by Integrated DNA technologies (Chapter 2, Table 2.5). **B.** Structure of the bypass elements.

The ϕC31 bypass elements Hop1, Hop2 and Hop3 (Figure 4.2) were commercially chemically synthesized as oligonucleotides (Table 2.5) and individually inserted into the coding region of ϕC31 serine integrase gene *int* at their specified locations by insertion mutagenesis of ϕC31 serine integrase expression plasmid pMFT-1 using Q5 DNA polymerase (Chapter 2, section 2.19). The full-length plasmid was amplified with forward and reverse primers that incorporated the insertions, circularized by ligation, and then recovered by transformation of *E. coli*. This resulted in 3 new programmed ϕC31 serine integrase expression plasmid termed pHop1C, pHop2B and pHop3D, respectively (Figure 4.3).

Accidental derivatives with shorter bypass elements were also obtained. The resulting plasmid constructs termed pHop1A and pHop1B respectively contained a 49-nucleotide and 48-nucleotide bypass element after G327 (Figure 4.4), instead of the planned 50-nucleotide bypass element Hop1 (Figure 4.2). The plasmid termed pHop2A contained a 47-nucleotide bypass element after G537 (Figure 4.4), instead of the planned 50-nucleotide bypass element Hop2 (Figure 4.2). The shorter derivatives probably occurred due to incomplete truncated synthesis of the oligonucleotides (Dressman *et al.*, 2006). Gene *60* bypass elements with shorter lengths are involved in efficient translational bypassing (Weiss *et al.*, 1990; Samatova *et al.*, 2014), suggesting that these deletions would not affect the bypassing activity of the unplanned bypass elements.

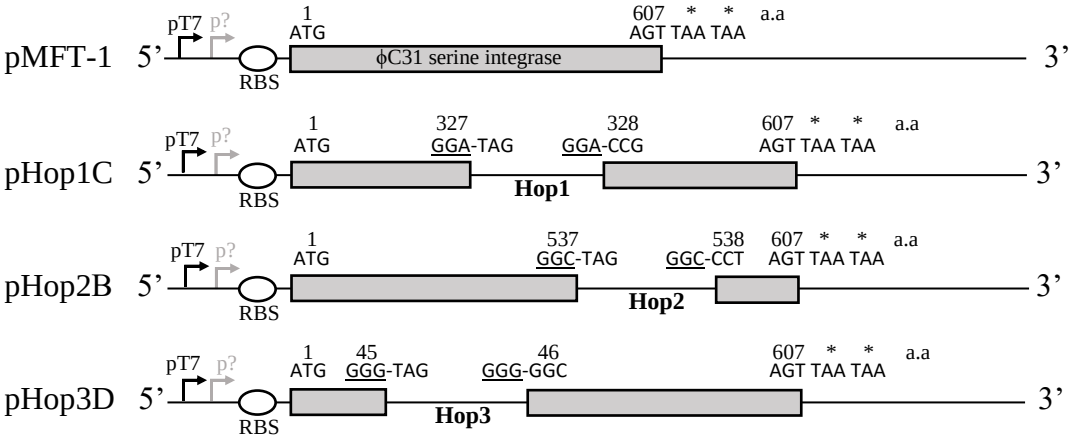


Figure 4.3. Insertion of the bypass element into the coding region of bacteriophage ϕC31 serine integrase gene *int*. Inducible T7 promoter (pT7), constitutive unknown promoter (p?), ribosome binding site (RBS), coding region of wild type ϕC31 serine integrase (gene *int*); continuous or discontinuous reading frame (grey) interrupted by Hop1, Hop2 or Hop3 bypass element (Figure 4.2). Take-off and landing codons are underlined. Codon/amino acid (a.a) number location is indicated. ϕC31 serine integrase expression plasmid (linear form) name is shown. Not drawn to scale.

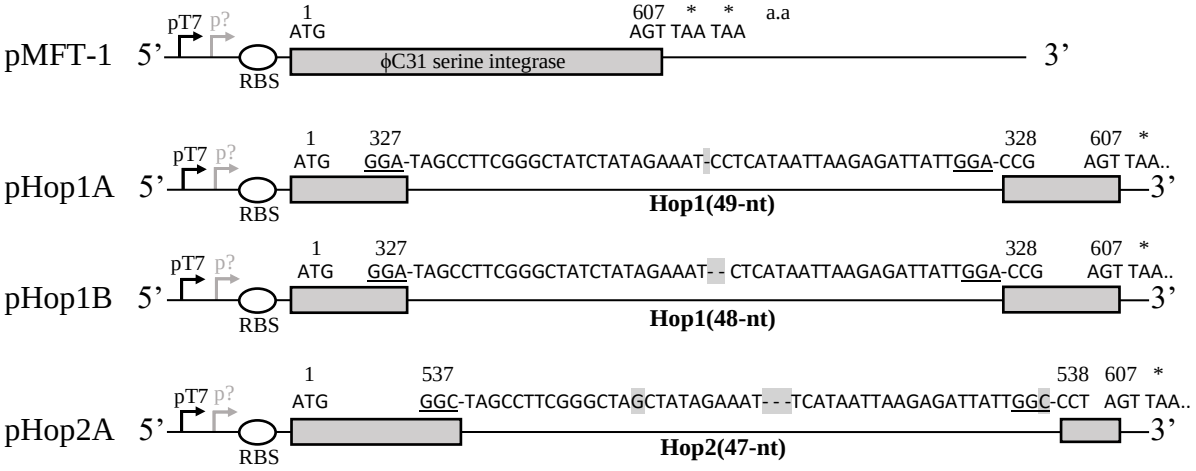


Figure 4.4. Unplanned bypass element insertion into the coding region of bacteriophage ϕC31 serine integrase gene *int*. Constructs with the shorter bypass element Hop1(49-nt), Hop1(48-nt) or Hop2(47-nt) derivatives of Hop1 or Hop2 (Figure 4.2) are shown. Relevant genetic information is as indicated (Figure legend as in Figure 4.3). Not drawn to scale.

4.2.2 An *in vivo* screening for translational bypassing

It was reasoned that the bypass element would promote translational bypassing to allow synthesis of the full-length ϕC31 serine integrase from the discontinuous reading frame. The in-frame TAG stop codon (stop) at the start of each bypass element would only allow the synthesis of a truncated ϕC31 serine integrase (Figure 4.3 and Figure 4.4), while the combination of the two reading frames encode the full-length ϕC31 serine integrase.

Bacteriophage ϕC31 serine integrase is an enzyme that catalyzes site-specific DNA recombination reactions. The protein binds specifically to pairs of recombination sites (termed *attP* and *attB*) and brings them together by protein-protein and protein-DNA interactions. The enzyme then catalyzes recombination by strand cleavage, exchange by subunit rotation and re-ligation (Van Duyne and Rutherford, 2013). Because of their specificity, efficiency, and highly unidirectional recombination (Stark, 2017), serine integrases have been used to make inversion switches (Figure 4.5). In a typical module, two recombination sites flank a promoter sequence. Expression of the serine integrase promotes inversion of the orientation of the promoter sequence, switching between expression of two genes that are divergently transcribed from the module (Figure 4.5) (Zhao *et al.*, 2019).

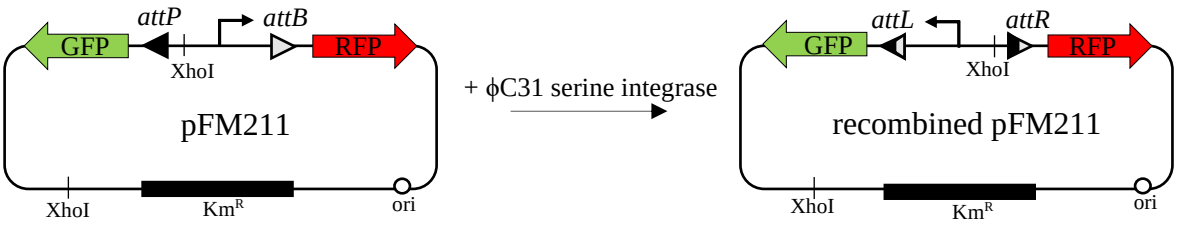


Figure 4.5. Inversion switch by expression of ϕC31 serine integrase. Substrate plasmid pFM211 carries ϕC31 recombination sites *attP* and *attB* flanking a constitutively active promoter (arrow) and an XhoI restriction site. ϕC31 serine integrase catalyzed recombination between *attP* and *attB* sites results in inversion in the direction of the promoter sequence and the position of the XhoI site, changing the gene expression status and restriction pattern of the module.

E. coli carrying the substrate plasmid pFM211 and a ϕC31 serine integrase expression plasmid (compatible) is analyzed by GFP and RFP fluorescence, as well as XhoI restriction analysis of plasmid DNA recovered from it, to deduce the recombination state of pFM211 (Figure 4.5) and in turn the recombination activity of the expressed ϕC31 serine

integrase (Zhao *et al.*, 2019). This well-established *in vivo* recombination assay was used to follow expression of ϕC31 serine integrase from the programmed ϕC31 serine integrase expression plasmid and screen for translational bypassing activity (Figure 4.3 and Figure 4.4). Ineffective bypass would result in a truncated ϕC31 serine integrase that would not recombine substrate plasmid pFM211 leading to RFP fluorescence and XhoI restriction patterns termed "PB" (1.3 kbp and 4.2 kbp), whereas bypassing would result in a full-length ϕC31 serine integrase that would recombine pFM211 leading to GFP fluorescence and XhoI restriction patterns termed "LR" (1.7 kbp and 3.8 kbp).

E. coli strain DS941 already carrying the substrate plasmid pFM211 (Figure 4.5) was transformed with the ϕC31 serine integrase expression plasmid (Figure 4.3 and Figure 4.4) for expression of the ϕC31 serine integrase, and recombination activity was assessed by fluorescence analysis of the cells (Figure 4.6). In the absence of ϕC31 serine integrase, pFM211 was not recombined leading to RFP expression (Lane 1, Figure 4.6A), whereas in the presence of ϕC31 serine integrase produced from pMFT-1 (Figure 4.3), pFM211 was recombined leading to GFP expression (Lane 2, Figure 4.6A). These two controls show that GFP expression occurs only when the ϕC31 serine integrase is expressed to recombine pFM211 (note that binding of integrase to *attL* leads to reduced transcription of the GFP gene and therefore relatively low green fluorescence in the presence of pMFT-1). The ϕC31 serine integrase expressed from the programmed ϕC31 serine integrase expression plasmids, except pHop3D, recombined pFM211 leading to GFP expression (Lanes 3-8, Figure 4.6A). These observations are consistent across the colony population (Figure 4.6A). These results indicate that translational bypassing occurred, except for pHop3D, producing ϕC31 serine integrase that recombined pFM211 leading to GFP expression (Figure 4.6B).

Plasmid DNA was recovered from one corresponding colony (Figure 4.6), cut with XhoI and visualized on agarose gel to further assess recombination activity (Figure 4.7A). In the absence of ϕC31 serine integrase, pFM211 was still not recombined leading to RFP expression and "PB" band patterns (1.3 kbp and 4.2 kbp) (Lanes 2-3, Figure 4.7A), whereas in the presence of ϕC31 serine integrase produced from pMFT-1, pFM211 was recombined leading to GFP expression and "LR" band patterns (1.7 kbp and 3.8 kbp) (Lane 4, Figure 4.7A). These controls show that GFP expression and "LR" band patterns occur only when the ϕC31 serine integrase is expressed to recombine pFM211. The ϕC31 serine integrase expressed from the programmed ϕC31 serine integrase plasmids, except pHop3D, recombined pFM211 leading to GFP expression and "LR" band patterns (Lanes 5-10, Figure 4.7A).

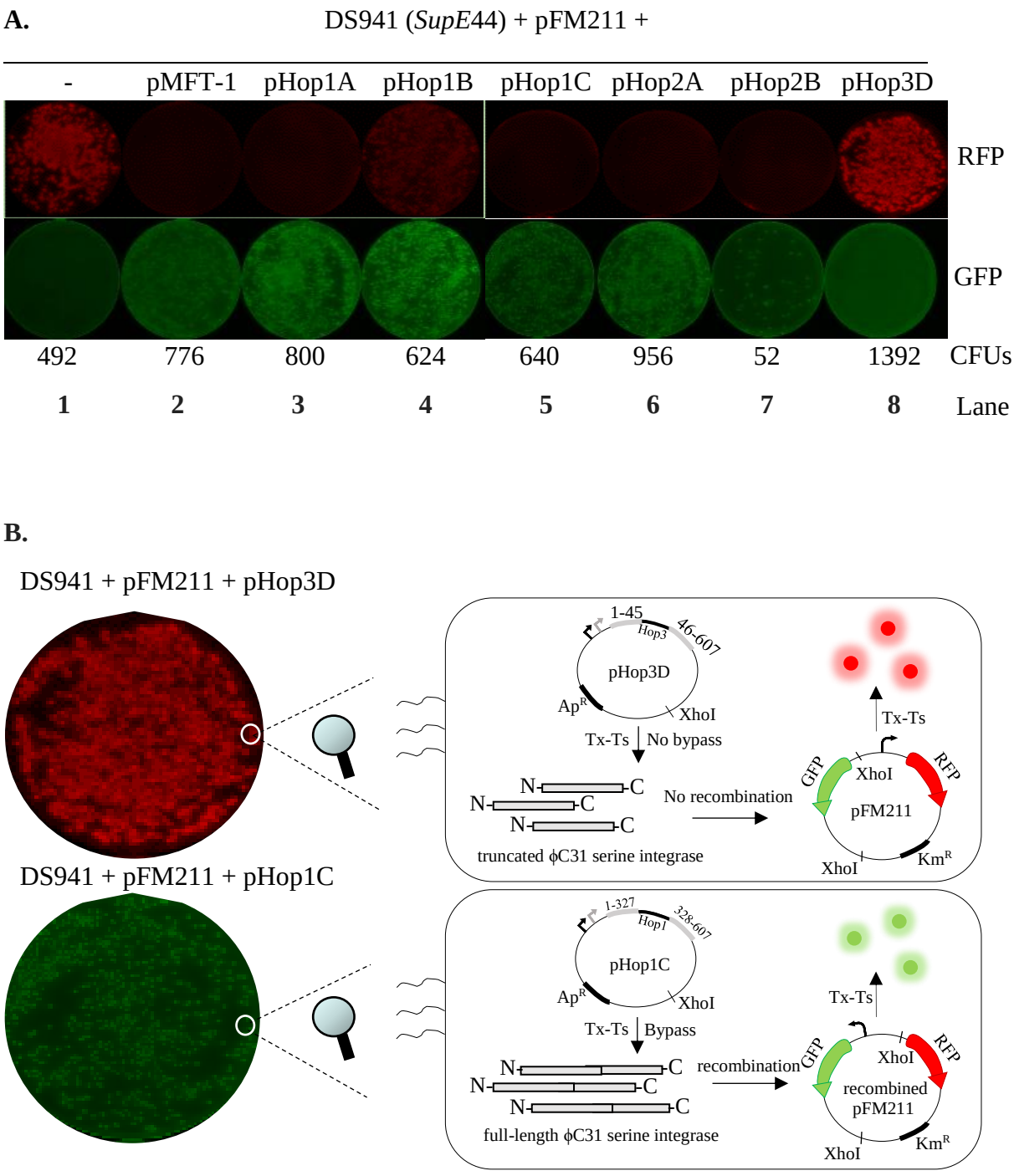


Figure 4.6. *In vivo* recombination activity of ϕC31 serine integrase produced from the ϕC31 serine integrase expression plasmids. A. *E. coli* DS941 already carrying substrate plasmid pFM211 (Figure 4.5) was transformed with the ϕC31 serine integrase expression plasmid (Figure 4.3 and Figure 4.4) and plated onto appropriate selective L-Agar plates grown for 18 hours at 37°C. GFP and RFP fluorescence of the colonies was visualized using Typhoon FLA9500 scanner. Scans of 9 cm L-Agar plates containing 50-1000 transformant colonies are shown. **B.** Schematic illustration of the cause behind the green and red fluorescent colonies. Tx-Ts: Transcription-Translation. CFUs: colony forming units.

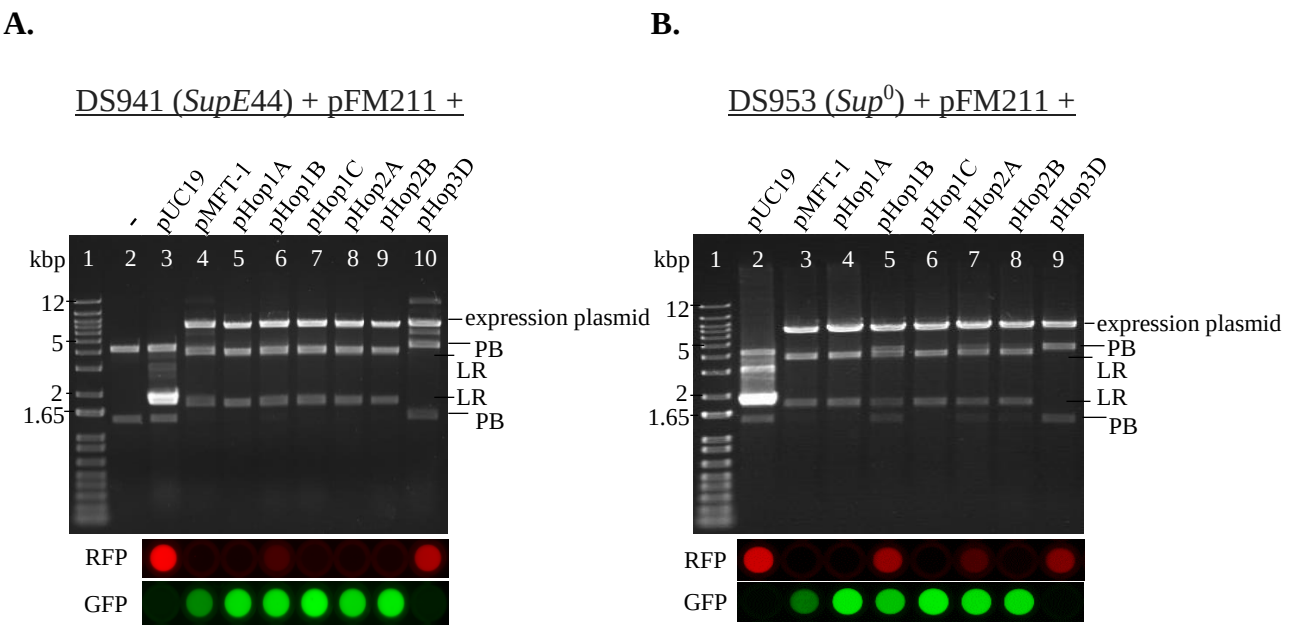


Figure 4.7. *In vivo* recombination activity of ϕC31 serine integrase produced from the ϕC31 serine integrase expression plasmids. A-B. *E. coli* DS941 or DS953 already carrying substrate plasmid pFM211 (Figure 4.5) was transformed with the ϕC31 serine integrase expression plasmid (Figure 4.3 and Figure 4.4) or pUC19. Plasmid DNA was isolated, cut with XhoI and visualized on a 1% agarose gel. Expression plasmid is linearized by XhoI whereas pUC19 remains uncut and circular. Band patterns 'PB' represent non-recombined pFM211 and 'LR' represent recombined pFM211. GFP and RFP fluorescence of cell culture was visualized using Typhoon FLA9500 scanner.

The *in vivo* recombination assay was carried out in *E. coli* DS941 which contains the *supE*44 mutation that could allow stop codon read through, where all three stop codons are predicted to be suppressed by *SupE*44 mutation (Singaravelan *et al.*, 2010). Even if stop codon read through occurred, the wrong translation frame would be used (the bypass element is 47-50 nucleotides long, except 48 nucleotides long in pHop1B (Figure 4.4)), resulting in an out-of-frame ϕC31 serine integrase. Therefore, the recombination assay was repeated in *E. coli* DS953, a strain that lacks this suppressor mutation, to remove possible stop codon read through products. The results were similar, except for incomplete recombination that was observed with pHop2A and pHop2B (Lanes 7-8, Figure 4.7B) and a repeatable incomplete recombination was observed with pHop1B (compare Lane 6, Figure 4.7A to Lane 5, Figure 4.7B). Incomplete recombination might be due to some substrate pFM211 escaping recombination, reflecting low levels of ϕC31 serine integrase being produced. These results indicate that translational bypassing occurred, except for pHop3D, producing

observable amounts of ϕC31 serine integrase that recombined pFM211 leading to GFP expression and "LR" band patterns.

Absence of recombination from pHop3D indicates that bypass is very inefficient for this construct. However, the cause for poor bypass is unknown. One possible explanation would be aberrant landing at other suitable 'landing' sites resulting in out-of-frame integrases (the resume codon following the landing site GGG codon in this construct is a glycine GGC codon (Figure 4.3) − there are alternative overlapping out of frame GGG landing codons). An alternative explanation could be that the bypass element is inhibiting gene expression from this construct. It has been previously reported that translation can be inhibited by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences (Borujeni *et al.*, 2017; Borujeni and Salis, 2016; Borujeni *et al.*, 2013), suggesting that the bypass element near the initiation site (after G45) might affect translation initiation.

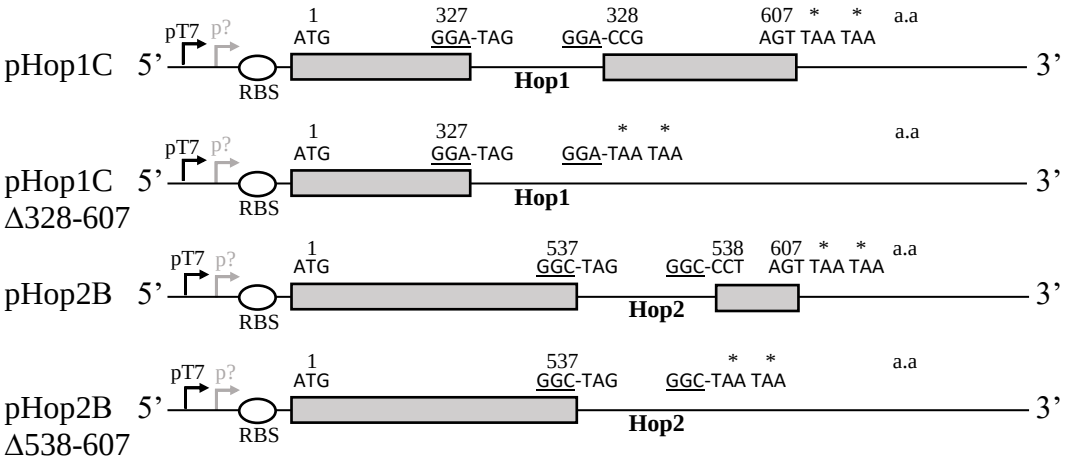
4.2.3 Full-length ϕC31 serine integrase is required for recombination activity

The results shown in Figure 4.6 and Figure 4.7 appear to show that a full-length functional ϕC31 serine integrase was produced from the programmed ϕC31 serine integrase expression plasmids, except pHop3D. An alternative explanation might be that a truncated ϕC31 serine integrase produced from termination at the TAG stop codon at the beginning of the bypass element could be functional, and this would lead to recombination activity in the absence of a full-length product. Therefore, the entire nucleotide sequence encoding amino acids 328-607 and 538-607 after the bypass element were deleted in plasmids pHop1C and pHop2B, respectively (Figure 4.3) by deletion mutagenesis (Chapter 2, section 2.19, Table 2.5) resulting in plasmids termed pHop1CΔ328-607 and pHop2BΔ538-607, respectively (Figure 4.8A).

E. coli strain DS953 already carrying the substrate plasmid pFM211 (Figure 4.5) was transformed with the truncated ϕC31 serine integrase expression plasmid (Figure 4.8A) and recombination activity was assessed by restriction analysis of plasmid DNA recovered from the cells (Figure 4.8B). In the presence of the truncated ϕC31 serine integrase, pFM211 was not recombined leading to "PB" band patterns (Lanes 4-5, Figure 4.8B). This indicated that truncated ϕC31 serine integrases are inactive and lose their recombination activity. These results are consistent with previous reports that C-terminal truncated ϕC31 serine integrases are inactive and lose their recombination activity (Olorunniji *et al.*, 2019; Weinberg *et al.*, 2019; Fogg *et al.*, 2018), as truncated integrases lack domains necessary for recombination activity (Van Duyne and Rutherford, 2013). It is therefore likely that the recombination

activity observed with pHop1A, pHop1B, pHop1C, pHop2A and pHop2B plasmids was from the full-length ϕC31 serine integrase and not a truncated product.

A.



B. DS953 (*Sup*^{0}) + pFM211 +

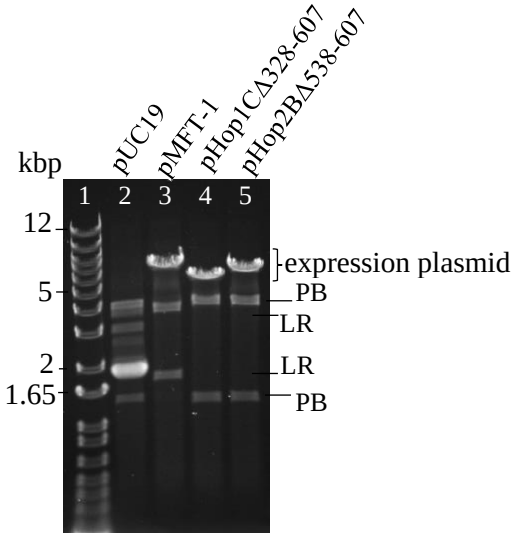


Figure 4.8. *In vivo* recombination inactivity of truncated ϕC31 serine integrase. A. Truncated derivatives of ϕC31 serine integrase expression plasmids pHop1C and pHop2B lacking the second reading frame of the integrase after the bypass element are shown (Figure legend as in Figure 4.3). **B.** *E. coli* DS953 already carrying substrate plasmid pFM211 (Figure 4.5) was transformed with the ϕC31 serine integrase expression plasmid (Figure 4.8A), plasmid DNA was isolated, cut with XhoI and visualized on a 1% agarose gel.

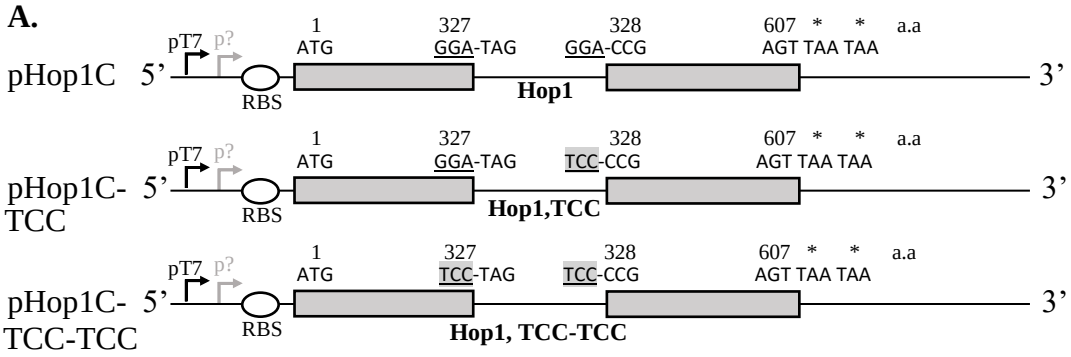
4.2.4 Translational bypassing-mediated synthesis of ϕC31 serine integrase requires matched codons flanking either end of the bypass element

The results suggest the presence of the full-length functional ϕC31 serine integrase, which could only have been the product of translational bypassing. A proportion of elongating ribosomes did not terminate at the in-frame UAG stop codon but skipped the bypass element and continued translation downstream to synthesize the full-length enzyme, apart from pHop3D. A clear requirement for translational bypassing is matched take-off and landing codons (Weiss *et al.*, 1990). To test this, the take-off and landing codons were altered so they were either matched or unmatched. TCC (Ser) codons were chosen based on previous reports that this pair confers an efficient translational bypassing event (Bucklin *et al.*, 2005; Wills *et al.*, 2008). To unmatch the take-off and landing GGA codons in pHop1C (Figure 4.3), the landing GGA codon was changed to a TCC codon by substitution mutagenesis of pHop1C (Chapter 2, section 2.20), resulting in plasmid construct termed pHop1C-TCC (Figure 4.9A). The integrase (truncated) produced from this expression plasmid did not recombine pFM211 and "PB" band patterns were obtained from restriction analysis of plasmid DNA recovered from the cells (Lane 6, Figure 4.9B). This indicated that translational bypassing was abolished when the take-off and landing codons were unmatched.

To re-match the take-off and landing codons in pHop1C-TCC (Figure 4.9A), the take-off GGA codon was changed to a TCC codon by substitution mutagenesis of pHop1C-TCC (Chapter 2, section 2.20), resulting in plasmid construct termed pHop1C-TCC-TCC (Figure 4.9A). This plasmid did give partial recombination (Lane 7, Figure 4.9B), demonstrating that the matched TCC take-off and landing codons allowed production of some full length integrase. This indicated that translational bypassing was partially restored when the take-off and landing codons were re-matched. The ϕC31 serine integrase produced from pHop1C-TCC-TCC (Figure 4.9A) contains mutation G327S (a result of take-off codon alteration). TCC (Ser) codons were initially also chosen because the serine mutation would be tolerated in the serine integrase and it would not have a deleterious effect on recombination activity as the change is in a non-conserved region based on previous alignments (Van Duyne and Rutherford, 2013). As a control, mutation G327S was introduced into ϕC31 serine integrase expression plasmid pMFT-1 (Figure 4.3) by substitution mutagenesis (Chapter 2, section 2.20) resulting in plasmid construct termed pMFT-1 G327S. The ϕC31 serine integrase mutant produced from this plasmid recombined

pFM211 and "LR" band patterns were obtained (Lane 4, Figure 4.9B), similar to the wild-type ϕC31 serine integrase produced from pMFT-1 (Lane 3, Figure 4.9B).

A.



B.

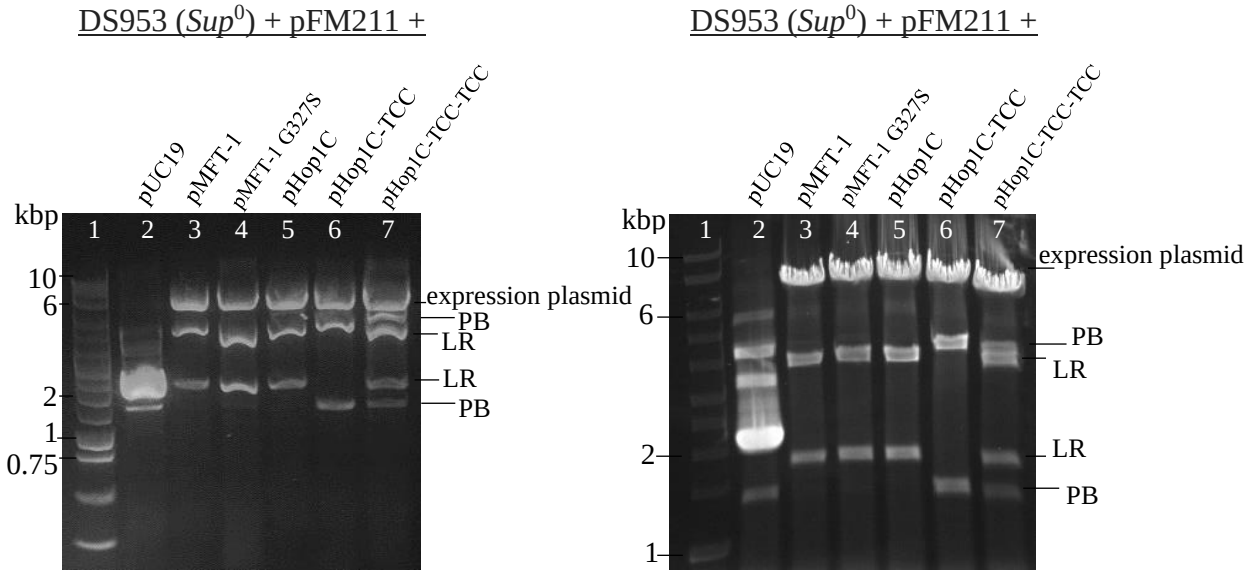


Figure 4.9. ***In vivo*** **recombination activity of ϕC31 serine integrase synthesized from the programmed ϕC31 serine integrase expression plasmids with matched or unmatched codons. A.** Whether the take-off and landing codons (underlined) are matched or unmatched in the ϕC31 serine integrase expression plasmid is indicated (highlighted in grey). Relevant genetic information is shown (Figure legend as in Figure 4.3). **B.** *E. coli* DS953 already carrying substrate plasmid pFM211 (Figure 4.5) was transformed with the ϕC31 serine integrase expression plasmid (Figure 4.9A and Figure 4.3) or pUC19. Plasmid DNA was isolated, cut with XhoI and visualized on a 1% agarose gel. Assay was repeated twice on different occasions. DNA electrophoresis was slightly disrupted due to voltage errors or agarose gel non-homogenity distribution.

Incomplete recombination with pHop1C-TCC-TCC (Lane 7, Figure 4.9B) was possibly due to bypass being less effective with TCC than with GGA as take-off and landing codons, or aberrant landing within the bypass element at other serine codons leading to ineffective production of full-length functional ϕC31 serine integrase. Future experiments could include varying the take-off and landing codon to see if a pattern can be detected.

Overall, the results show that the bypass element Hop1 (identical to T4 gene *60* 50-nucleotide bypass element), Hop2, and the unplanned shorter derivatives (Hop1(49-nt), Hop1(48-nt) and Hop2(47-nt)) present in the coding region of bacteriophage ϕC31 serine integrase gene *int* promoted translational bypassing, resulting in synthesis of full-length functional ϕC31 serine integrase that was observed through its recombination activity by the *in vivo* recombination assay. An alternative explanation would be that recombination was not from translational bypass, but from expression of two separate peptides (from translational initiation after the bypass sequence) and that the two separate peptides came together to form a functional protein. However, this is unlikely due to the failure of the mismatched landing and take-off codons (TCC) to give recombination and the restoration of recombination by re-matching these codons.

Although the recombination assay allows deduction of translational bypassing, accurate estimation of bypassing efficiency is not possible, since the effect of the truncated integrase (ineffective bypass) produced during the same time as the full-length integrase is not visible (the assay is a very sensitive method where even a small number of full-length ϕC31 serine integrase molecules can promote almost complete recombination effect).

4.2.5 Estimation of bypassing efficiency

To measure bypassing efficiency (ratio of full-length to truncated integrase), the ϕC31 serine integrase expression plasmid (Figure 4.3, Figure 4.8A and Figure 4.9A) was directly expressed in an *in vitro* translation system from *E. coli* in the presence of ^{35}S-Methionine. Products of *in vitro* coupled transcription and translation were observed by SDS-PAGE followed by autoradiography (Figure 4.10A).

For the wild-type and mutant ϕC31 serine integrase produced from pMFT-1 and pMFT-1 G327S, respectively, a predominant band corresponding in molecular weight to the full-length ϕC31 serine integrase at ~ 67 kDa was observed (Lane 1 and Lane 7, Figure 4.10A). The lower bands were possibly a result of internal translation initiation and/or incomplete translation products from the plasmids. The majority of ϕC31 serine integrase product from all other plasmids corresponded in molecular weight to truncated ϕC31 serine

integrase at ~ 38 kDa (Lanes 2, 8 and 9, Figure 4.10A) and ~ 59 kDa (Lane 3, Figure 4.10A). The identity of these bands were further confirmed by truncated control constructs pHop1CΔ328-607 and pMFT-1 TAA (contains two stop codons TAA after G327 in ϕC31 serine integrase gene *int* made by substitution mutagenesis of pMFT-1 (Chapter 2, section 2.20); serves as an additional truncated integrase control) that generated a band corresponding in molecular weight to truncated integrase at ~ 38 kDa (Lanes 5 and 6, Figure 4.10A and Lane 3, Figure 4.10B) and pHop2BΔ538-607 which resulted in a truncated integrase at ~ 59 kDa (Lane 5, Figure 4.10B). The truncated product from pHop3D at ~ 5 kDa could not be clearly detected (Lane 4, Figure 4.10A). The band corresponding in molecular weight to full-length ϕC31 serine integrase at ~ 67 kDa was not clearly detected (Lanes 2, 3 and 9, Figure 4.10A). This indicated that the majority of translational product from the programmed ϕC31 serine integrase expression plasmids was truncated ϕC31 serine integrase.

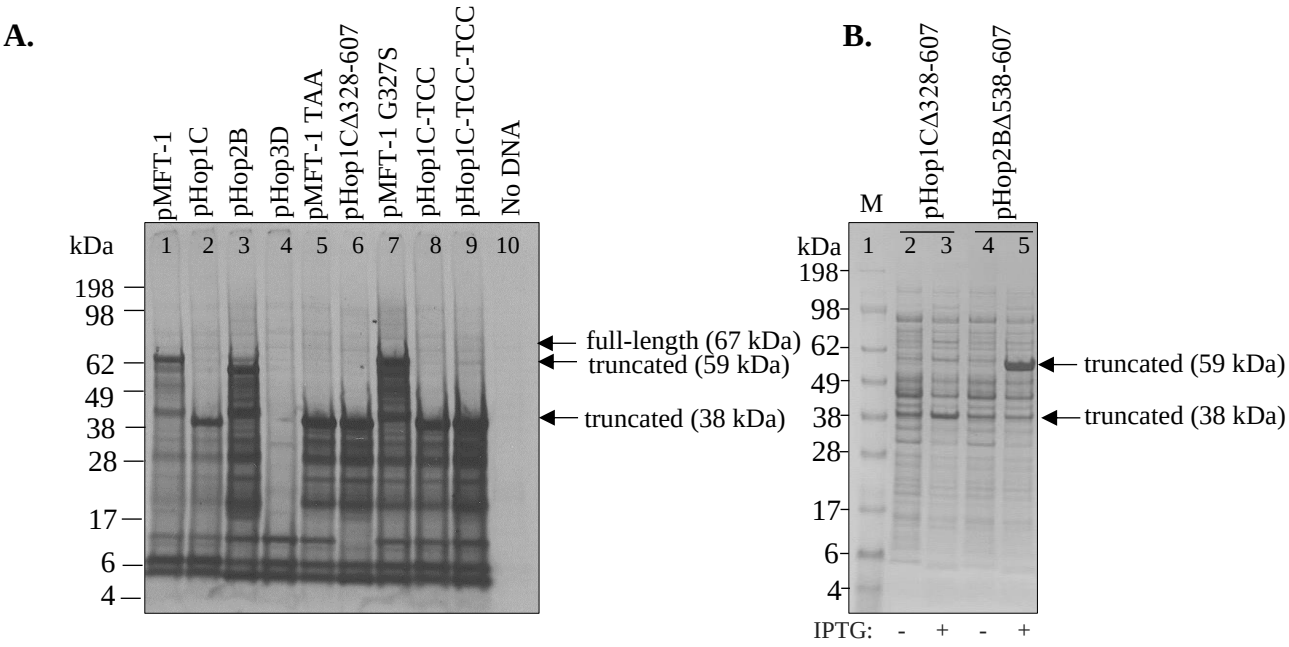


Figure 4.10. Expression of ϕC31 serine integrase from the ϕC31 serine integrase expression plasmids. A. ϕC31 serine integrase expression plasmid (Figure 4.3, Figure 4.8A and Figure 4.9A) was directly used for *in vitro* translation using commercially available kits. Protein products were resolved on a 4-12% Bis-Tris gel, fixed, dried and placed with a film in a light-proof cassette for 24 hours before being developed using an Xomat processor. **B.** ϕC31 serine integrase expression plasmid (Figure 4.8A) was introduced in *E. coli* C41, cultures were grown in the presence (+) or absence (-) of 1 mM IPTG for gene expression overnight and whole cell extracts were analyzed on a 4-12% Bis-Tris gel stained with coomassie blue solution. Major protein bands indicated with an arrow. The content of each lane is shown. Protein marker (M): SeeBlue Plus2 pre-stained protein standard.

However, the undetectable full-length ϕC31 serine integrase product that was not observed on the protein gel (Figure 4.10A) was detected through its recombination activity by the *in vivo* recombination assay, indicating that a minority of translational product were the full-length ϕC31 serine integrase.

To better visualize the protein products (Figure 4.10), an N-terminal His-tag was inserted in ϕC31 serine integrase expression plasmid pMFT-1, pHop1CΔ328-607, pHop1C-TCC, pHop1C-TCC-TCC, pHop2B and pHop3D by insertion mutagenesis (Figure 4.3, Figure 4.8A and Figure 4.9A; Chapter 2, section 2.19) resulting in plasmid constructs termed His_6-pMFT-1, His_6-pHop1CΔ328-607, His_6-pHop1C-TCC, His_6-pHop1C-TCC-TCC, His_6-pHop2B and His_6-pHop3D, respectively (partially sequenced). The His-tag was chosen based on previous reports that an N-terminal His-tagged ϕC31 serine integrase exhibits similar recombination activity as an un-tagged integrase (McEwan *et al.*, 2011), indicating that an N-terminal His-tag would not affect the assays. The His-tag would allow visualization of only His-tagged protein products by western blotting.

The His-tagged ϕC31 serine integrase expression plasmids were introduced into *E. coli* C41 for expression and whole cell extracts were analyzed by western blotting (Figure 4.11A). For the His-tagged ϕC31 serine integrase produced from His_6-pMFT-1, a predominant band corresponding in molecular weight to the full-length integrase at ~ 67 kDa was observed (Lane 2, Figure 4.11A). The lower bands were degradation or incomplete translational products whereas the higher band was a strong dimer (Lane 2, Figure 4.11A). The majority of His-tagged ϕC31 serine integrase produced from all other expression plasmids, except His_6-pHop3D and His_6-pHop1C-TCC, corresponded in molecular weight to truncated integrase at ~ 38 kDa (Lane 3, Lane 4 and Lane 6, Figure 4.11A) and ~ 59 kDa (Lane 7, Figure 4.11A). This indicated that most elongating ribosomes were synthesizing a truncated His-tagged ϕC31 serine integrase.

The truncated product from His_6-pHop3D at ~ 5 kDa was not detected (Lane 8, Figure 4.11A). The truncated product from His_6-pHop1C-TCC at ~ 40 kDa (Lane 5, Figure 4.11A) was probably a stop codon read through product in *E. coli* C41, that has been reported to naturally suppress stop codons at low frequencies (Angius *et al.*, 2018) or a colony with a mutant version of the plasmid. When this experiment was repeated using another colony, the expected truncated ~ 38 kDa product was observed (Lane 5, Figure 4.11B), supporting the idea that a mutant version of the plasmid was responsible for the larger protein product seen in Lane 5 of Figure 4.11A.

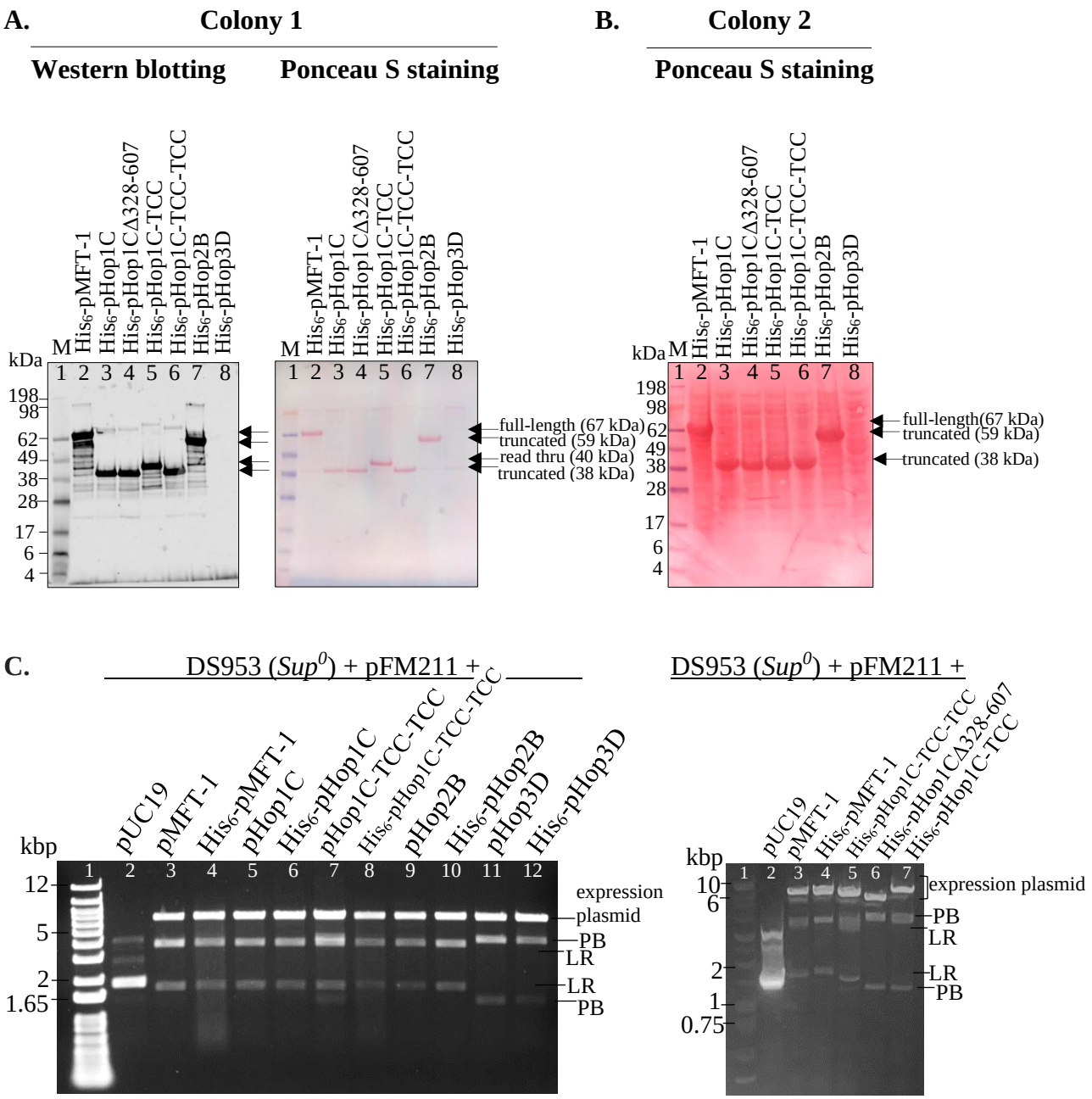


Figure 4.11. *In vivo* expression and recombination activity of N-terminal His-tagged ϕC31 serine integrases. A-B. *E. coli* C41 was transformed with the corresponding plasmid and grown on selective L-Agar plates for 18 hours at 37°C. Cultures were made from a transformant colony, cells were lysed, and the extracts were resolved on a 4-12% Bis-Tris gel, transferred to a nitrocellulose membrane and His-tagged proteins were then detected with an anti-His tag antibody. Blots were also stained afterwards in Ponceau S. Major protein bands indicated with an arrow. The content of each lane is shown. SeeBlue Plus2 pre-stained standard (M). **C.** *E. coli* DS953 already carrying substrate plasmid pFM211 (Figure 4.5) was transformed with the His-tagged ϕC31 serine integrase expression plasmid or pUC19. Plasmid DNA was isolated from one transformant colony (left) or 10 pooled colonies (right gel), cut with XhoI and visualized on a 1% agarose gel.

The higher molecular weight bands were dimers of truncated or stop codon read through product, and the lower bands degradation or incomplete translation products (Figure 4.11A). If these are dimers, indicates that β-mercaptoethanol (5%) used was not enough for denaturation. This assertion could be tested by increasing β-mercaptoethanol concentration or purifying the His-tagged products with affinity chromatography followed by size exclusion chromatography for molecular weight determination.

Although the full-length functional His-tagged ϕC31 serine integrase product was not detected on the blots (Figure 4.11A and Figure 4.11B), it was detected through its recombination activity by the recombination assay (Figure 4.11C), indicating that a minority of ribosomes engaged in effective bypassing.

However, an accurate conclusion can not be made for this experiment. Although the expression and recombination results obtained from the programmed His-tagged and un-tagged ϕC31 serine integrase expression plasmids were similar, His-tagged constructs were partially sequenced and will likley require complete sequencing of the ϕC31 serine integrase coding region.

4.3 Discussion

The results demonstrate that the engineered bypass elements (or "translational bypassing devices") including Hop1 which is identical to the 50-nucleotide bypass element of bacteriophage T4 gene *60*, Hop2, and the unplanned shorter derivatives present in the coding region of bacteriophage ϕC31 serine integrase gene *int* promoted a low-level translational bypassing event in *E. coli*. The bypass elements contained the minimal requirements for signaling translational bypassing which included matched take-off and landing codons and an in-frame stop codon contained in a stimulatory take-off stem loop. The low-level translational bypassing event could be explained by the absence of other *cis*-acting signals such as a nascent peptide stalling motif and probably other RNA features shown to be required for a highly efficient translational bypassing event in a gene *60* context (Weiss *et al.*, 1990; Samatova *et al.*, 2014; Chen *et al.*, 2015). Interestingly, this autonomous function of gene *60* bypass element (or 'Hop1' in this work) agrees with its origin as a foreign genetic element inserted into gene *60* by a mobile genetic element (Bonocora *et al.*, 2011).

The ability of these engineered bypass elements (short non-coding functional sequences) to regulate gene expression in bacteria and control the relative abundance of truncated and full-length protein products from one gene is important for emerging biotechnologies. These first-generation constitutively active translational bypassing devices could be used to set the relative abundance of two distinct proteins by encoding them in a single gene. This control offers precision and consistency that can be exploited for various biotechnological purposes, especially those that depend on the balance of two functionally interacting proteins. Furthermore, these devices may be applicable to eukaryotic systems, such as mitochondria of yeast cells, where translational bypassing naturally occurs (Lang *et al.*, 2014; Nosek *et al.*, 2015). While these devices should have utility in higher eukaryotes, it is likely that they will need to be fashioned based on the properties of the biological setting of interest.

The strategy established in this work could be adapted for the discovery of novel translational bypassing stimulatory sequences, such as screening for a larger library of functional translational bypassing devices by random mutagenesis techniques (Cirino *et al.*, 2003; Tobias, 2003), enabling a more diverse regulation of the protein outputs stoichiometry. Moreover, this may be applicable to the construction of inducible translational bypassing devices with switchable functions, that provide distinct advantages over other reprogramming gene-control strategies. Whereas frameshifting devices (Anzalone *et al.*, 2016; Matsumoto *et al.*, 2018; Dinman, 2019) are coding segments, translational bypassing devices are skipped rather than translated, and thus, are not part of the protein outputs. This enables their use in many genes of interest without concern for their effect on protein function. Ligand-responsive translational bypassing devices could be achieved by means of anti-sense RNA (asRNA) approaches (Yang *et al.*, 2019), that could be adapted for the purpose of turning translational bypassing ‘on’ or ‘off’ in response to a ‘trigger’ compound (Figure 4.12), reminiscent of riboswitches (Ausländer and Fussenegger, 2014; Green *et al.*, 2014). While this approach has not been tested, it merits further investigation.

With some ingenuity and creativity, it should be possible to conceive and construct various configurations of translational bypassing devices that can be used to perform complex logic operations and other regulatory functions in the cell (Figure 4.12). Overall, the tools and methodology developed here will empower efforts to develop customized genetic programs and expand our capabilities to manipulate biological systems.

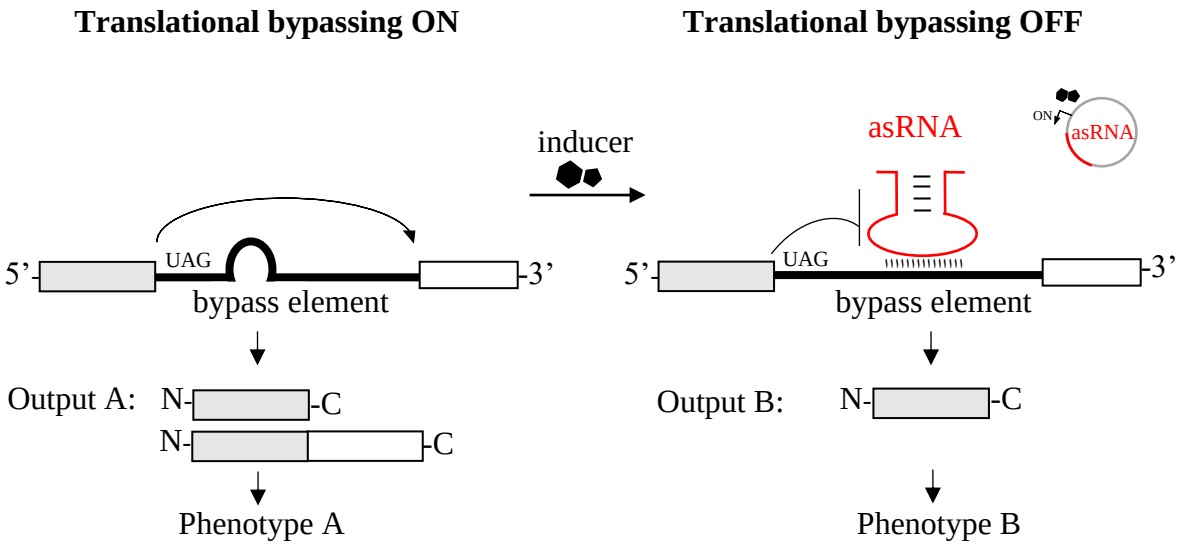


Figure 4.12. Ligand-responsive translational bypassing switches. Through base pairing between the designed anti-sense RNA (asRNA) and target mRNA, the resulting double-strand RNA can inhibit translational bypassing by steric hindrance. The plasmid for asRNA expression can be easily constructed and conditionally controlled by various promoters in response to an inducer. As a result, translational bypassing devices can be controlled. This form of regulation can be used for achieving diverse phenotypic performances in genetically modified organisms.

CHAPTER FIVE

RESULTS AND DISCUSSION:

Protein interaction domains from *mobA* insertion element in bacteriophage T4 DNA topoisomerase subunits gp39 and gp60

5.0 Overview

Protein-protein interactions are fundamental for regulation of protein activity and cellular processes, and numerous methods have emerged to predict and detect them. Previous computational data have identified a potential interaction interface between bacteriophage T4 DNA topoisomerase subunits gp39 (amino acids 473-516) and gp60 (amino acids 2-31), that may play an important role in assembly of an active topoisomerase. However, there is no direct evidence confirming this interaction interface. This work used various protein-protein interaction assays to demonstrate that these regions interact. When these peptides were fused to recombinant protein fragments, strong functional protein complexes formed that were detected by size exclusion chromatography, microscale thermophoresis and the bacterial two-hybrid system. These findings have important implications for studies of topoisomerases and development of new tools that control protein function.

5.1 Introduction

Bacteriophage T4 encodes a type II topoisomerase with properties more similar to those of eukaryotic enzymes than to those of the bacterial DNA gyrase (Kreuzer and Jongeneel, 1983). Because of this similarity, T4 topoisomerase has become an important target model for anti-microbial and anti-cancer drugs (Huff *et al.*, 1989; Kreuzer, 1998). The enzyme is a complex formed by the products of three T4 genes; gene *52* which encodes the 442-amino acid subunit (gp52), gene *39* which encodes the 516-amino acid subunit (gp39) and gene *60* which encodes the full-length 160-amino acid subunit (gp60) plus a shorter form (Huang, 1986; Huang *et al.*, 1988; Seasholtz and Greenberg, 1983). Unlike T4, other related phages have only two subunits for their type II topoisomerases (Figure 5.1). T4 gp52 is equivalent to the gyrA subunit of DNA gyrase while gp39 + gp60, or the longer gp39 of T4-related phages, is most closely related to the gyrB subunit of DNA gyrase (Figure 5.1). There is no structure currently available for the T4 topoisomerase and the interaction interfaces between its subunits are unidentified.

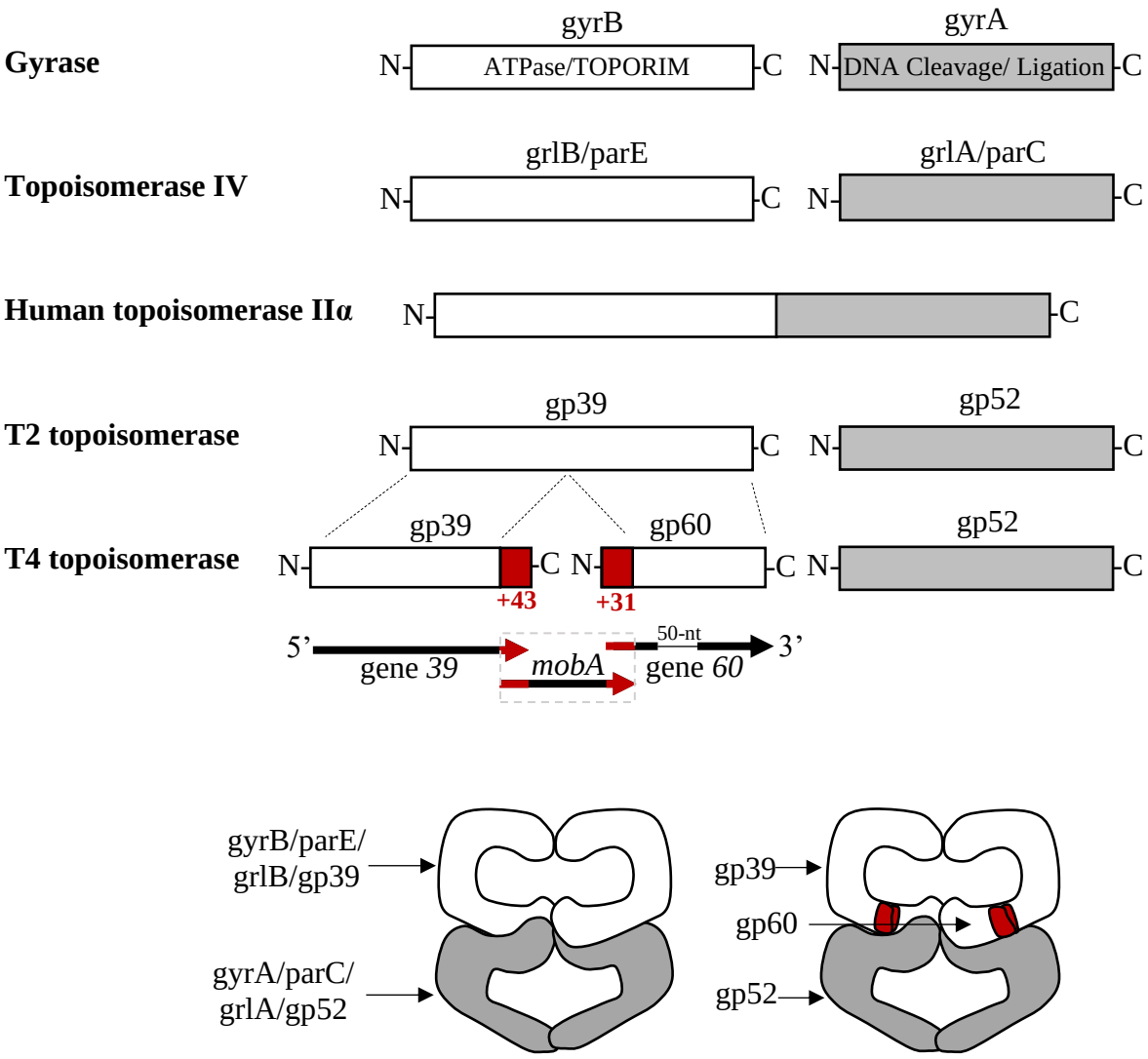


Figure 5.1. Domain structure alignment of type II topoisomerases. Bacteriophage T4 topoisomerase is homologous to T2 topoisomerase. However, during the evolution of T4, the gp39 subunit was split into two independent polypeptides, gp39 and gp60. This fragmentation was the consequence of insertion of a foreign genetic mobile element (denoted *mobA*) into T4 topoisomerase gene *39*, splitting it into genes *39* and *60* (Bonocora *et al.*, 2011). *mobA* insertion contributed to additional sequences in the form of amino acid extensions at the C- and N-terminal ends of the split subunits (in red) [as well as termination and initiation signals], that are absent in gp39 homologues, and likely provide an interaction platform to bring the independently expressed proteins together and restore activity. At the same time, the 50-nucleotide bypass element (50-nt) was also inserted into the newly formed T4 gene *60*. A schematic representation of type II topoisomerase structure is shown.

Previous studies suggested that the C-terminus of T4 gp39 (amino acids 473 to 516) and the N-terminus of T4 gp60 (amino acids 2 to 31); amino acids that are present in the T4 protein and not topoisomerases with a single gp39 polypeptide, would interact to hold the two subunits together (Crona *et al.*, 2011; Bonocora *et al.*, 2011; Wills *et al.*, 2008; Todd and Walter, 2013). However, there is no confirmation that the regions interact. In this study, the interaction between these two regions is demonstrated by various protein-protein interaction methods. Fusion of these amino acid extensions to recombinant protein fragments results in formation of strong functional protein complexes that are detected by size exclusion chromatography, microscale thermophoresis and the bacterial two-hybrid system.

Interestingly, these interacting C- and N-terminal regions are of foreign origin that originated from a mobile genetic element insertion (*mobA*) into the T4 progenitor topoisomerase subunit gene *39* (Figure 5.1). Instead, these residues are part of the N-terminal region of gp60 that was added by *mobA* (Figure 1.3.1). In bacteriophage genomes, various genes, including topoisomerase genes, are commonly targeted by mobile genetic elements − free standing and intron-encoded homing endonucleases and inteins (Petrov *et al.*, 2010; Petrov *et al.*, 2010; Swithers *et al.*, 2009; Sandegren *et al.*, 2005; Baranov *et al.*, 2006; Dassa *et al.*, 2009). The data presented here is consistent with these *mobA*-derived regions functioning as protein interaction domains that are likely necessary to restore topoisomerase activity. This may represent a solution distinct from RNA and protein splicing to restore function from a fragmented coding region and may be a general mechanism to neutralize fragmentation of essential genes by mobile genetic elements (Figure 5.2).

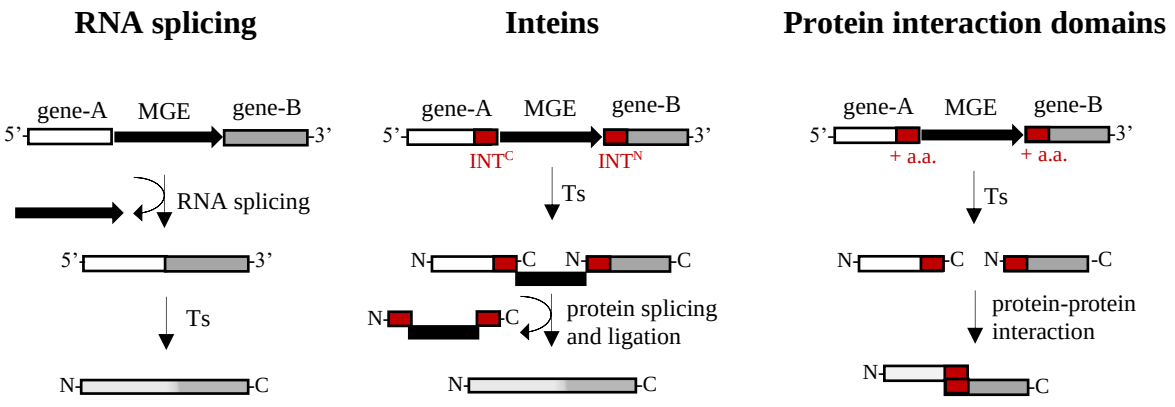


Figure 5.2. Mobile genetic element mechanisms used to restore function of genes they interrupt. MGE: mobile genetic element. Ts: Translation. INT: protein splicing element. a.a.: amino acids extensions.

5.2 Results

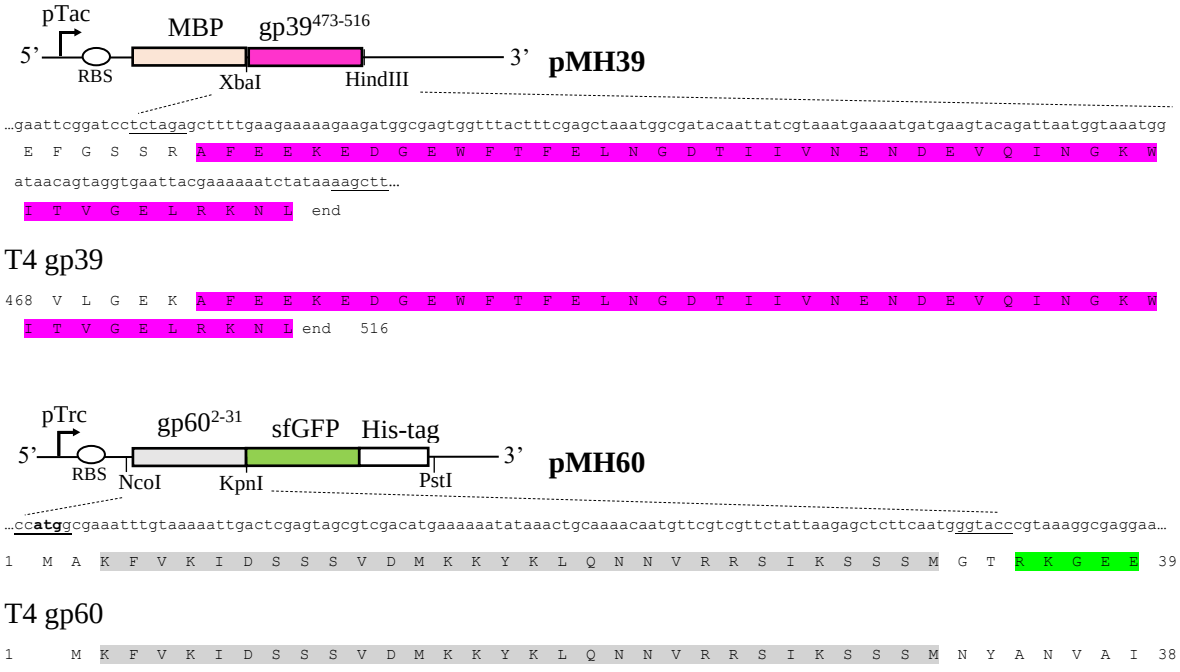
The interaction of amino acids 473 to 516 at the C-terminus of T4 topoisomerase subunit gp39 (termed here gp39$^{473-516}$) with amino acids 2 to 31 at the N-terminus of T4 topoisomerase subunit gp60 (termed here gp60$^{2-31}$; N-terminal initiator methionine excluded for cloning purposes) was investigated to confirm previous computational data which stated that the C- and N-terminal regions of T4 gp39 and T4 gp60 would interact to hold the two subunits together.

5.2.1 Recombinant proteins

It was reasoned that two proteins fused to interacting peptides would associate and form a protein complex that could be observed by size exclusion chromatography. For this purpose, two recombinant proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ were made that respectively contained gp39$^{473-516}$ fused to the C-terminus of a maltose-binding protein (MBP) and gp60$^{2-31}$ fused to the N-terminus of a GFP superfolder (sfGFP) with a C-terminal His-tag (encoded by pMH39 and pMH60, respectively) (Figure 5.3A). The fusion partners were chosen due to their ease of purification and GFP superfolder so that it could be tracked by fluorescence or by absorbance at wavelength of 488 nm (Pédelacq *et al.*, 2006). Fluorescence and absorbance are different − though in this case related (MBP lacks an absorbance peak at 488 nm, whereas GFP superfolder (sfGFP) shows an absorbance peak at 488 nm that can be used to track it).

The proteins were expressed in *E. coli* strain C41 carrying the expression plasmid pMH39 and pMH60 (Figure 5.3A), the MBP-tagged protein was purified using sugar affinity chromatography while the His-tagged protein was purified using nickel affinity chromatography (Chapter 2, section 2.24) , dialyzed (Chapter 2, section 2.25), and analyzed on SDS-PAGE prior to size exclusion chromatography (Figure 5.3B). MBP-gp39$^{473-516}$ showed a band corresponding in molecular weight to ~ 48 kDa (Lane 3, Figure 5.3B), while gp60$^{2-31}$-sfGFP-His$_6$ showed a major band corresponding in molecular weight to ~ 31 kDa (Lane 2, Figure 5.3B). These two samples were then used in size exclusion chromatography for analysis of interaction.

A.



B.

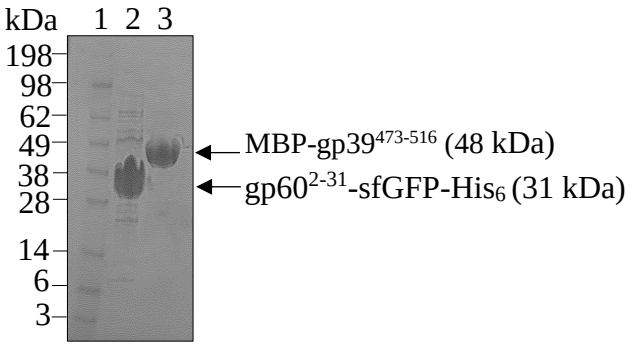


Figure 5.3. Expression of recombinant proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$. A. Part of nucleotide sequence (5'-3') and derived amino acid sequence of expression plasmids pMH39 and pMH60 for MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ proteins, respectively. Amino acid sequences of T4 topoisomerase subunits gp39 and gp60 are shown for comparison. Unique restriction sites used for cloning are underlined (Chapter 2, Table 2.2). **B.** Proteins were expressed in *E. coli* C41 carrying pMH39 and pMH60 (Figure 5.3A), the MBP-tagged protein was purified using MBPTrap column, while the His-tagged protein was purified using HisGraviTrap column. The column-bound proteins were eluted and elution samples (E_2 for gp60$^{2-31}$-sfGFP-His$_6$ (~ 10 mg/ml) and E_4+E_5 for MBP-gp39$^{473-516}$ (~ 5 mg/ml)) were dialyzed. A small volume of dialyzed purified protein sample was analyzed on a 4-12% Bis-Tris gel stained with coomassie blue prior to its use in size exclusion chromatography. Lane 1: SeeBlue Plus2 pre-stained protein standard. Lane 2: gp60$^{2-31}$-sfGFP-His$_6$. Lane 3: MBP-gp39$^{473-516}$.

5.2.2 Interaction of gp39$^{473-516}$ with gp60$^{2-31}$ *in vitro*

It was reasoned that the two proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ would form a protein complex that could be detected by size exclusion chromatography. In size exclusion chromatography, the elution of a protein or of a protein complex relates with its molecular weight. This provides a simple method for evaluating interactions between proteins. A solution containing a protein and a 'ligand' protein at same concentrations is applied in a small volume to the column and the material is resolved in the usual way. The elution positions of the protein and ligand in the mixture are compared with those of the protein and ligand when each is chromatographed individually on the same column. If a complex has formed between the protein and ligand, the complex will elute earlier than either protein alone (monitored by protein absorbance at 280 nm (and 488 nm in this study)). Moreover, the molecular weights can be measured from the elution profiles by comparing to a standard molecular weight curve generated under the same conditions.

Individual analysis of recombinant protein MBP-gp39$^{473-516}$ by size exclusion chromatography (Chapter 2, section 2.26), showed that MBP-gp39$^{473-516}$ behaved as a dimer in solution (Figure 5.4A). MBP-gp39$^{473-516}$ showed to have a molecular weight of 48 kDa (Lane 3, Figure 5.3B), however, when chromatographed it had a molecular weight of 100 kDa (Figure 5.4A). The elution volume of this protein was ~ 12 ml (Figure 5.4A), and by inputting this elution profile on the standard molecular weight curve (Graph 5.1B), a 100 kDa molecular weight is measured. This indicated that MBP-gp39$^{473-516}$ eluted as a dimer (2 molecules of MBP-gp39$^{473-516}$, 48 kDa + 48 kDa). MBP has been previously reported to exist as a monomer and a dimer under various conditions (Richarme, 1982). Previous data also show that when MBP is fused to other peptides it tends to form dimers (Momin *et al.*, 2019), where both MBP and the peptides fused to it provoke dimerization.

Individual analysis of recombinant protein gp60$^{2-31}$-sfGFP-His$_6$ by size exclusion chromatography (Chapter 2, section 2.26), showed it was a monomer (Figure 5.4B). gp60$^{2-31}$-sfGFP-His$_6$ had a molecular weight of 31 kDa (Lane 2, Figure 5.3B), that was consistent with the molecular weight obtained when chromatographed (Figure 5.4B). The elution volume of this protein was ~ 13.5 ml (Figure 5.4B), and by inputting this elution profile on the standard molecular weight curve a ~ 37 kDa molecular weight is measured (Graph 5.1B). sfGFP has been previously reported to exist as a monomer (Pédelacq *et al.*, 2006; Costantini *et al.*, 2012), consistent with our data. Moreover, the detected absorbance peak at 488 nm (Figure 5.4B) is a characteristic of sfGFP (Pédelacq *et al.*, 2006) used to track it.

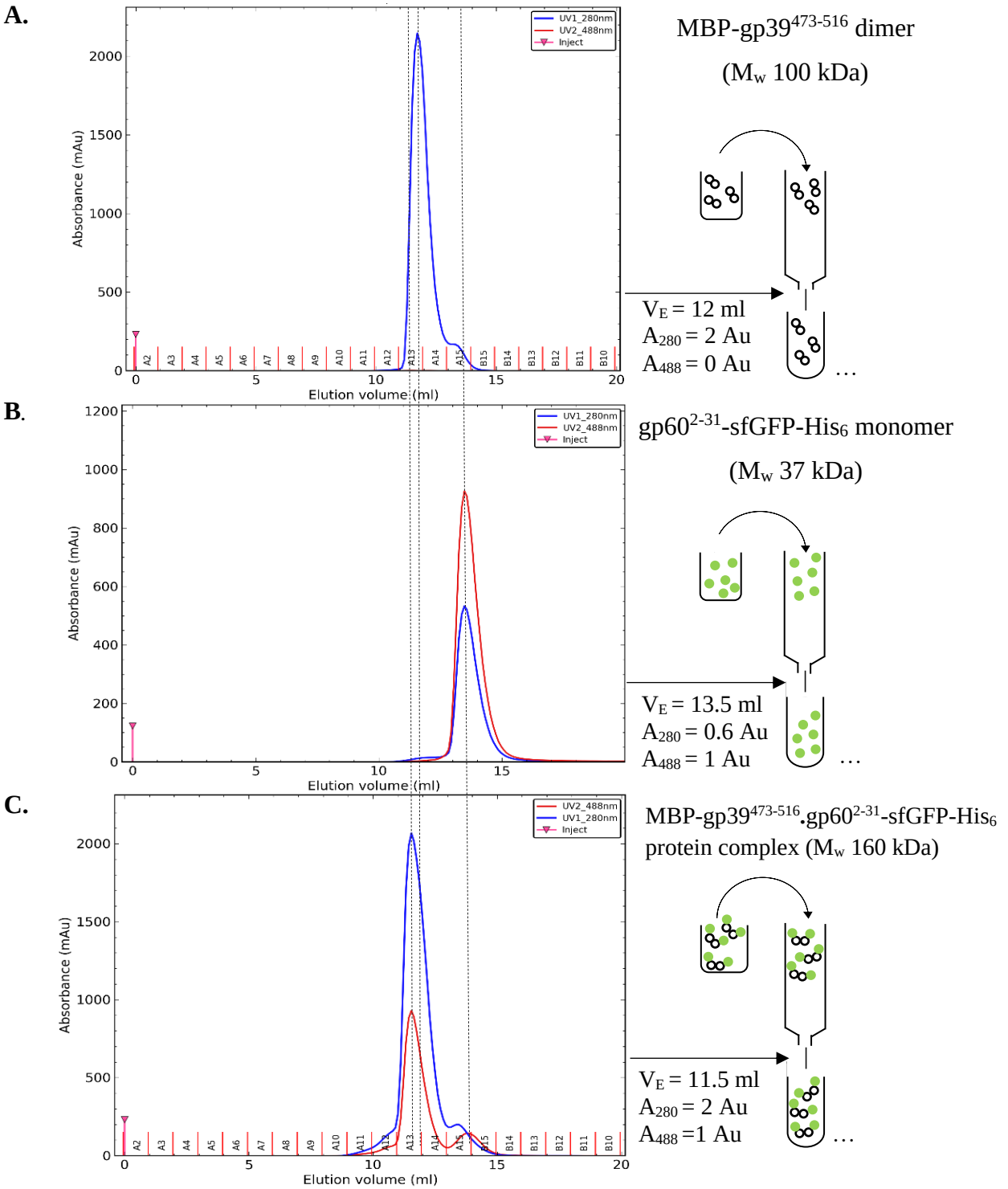
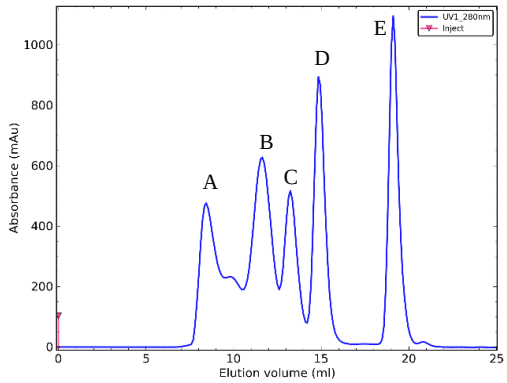


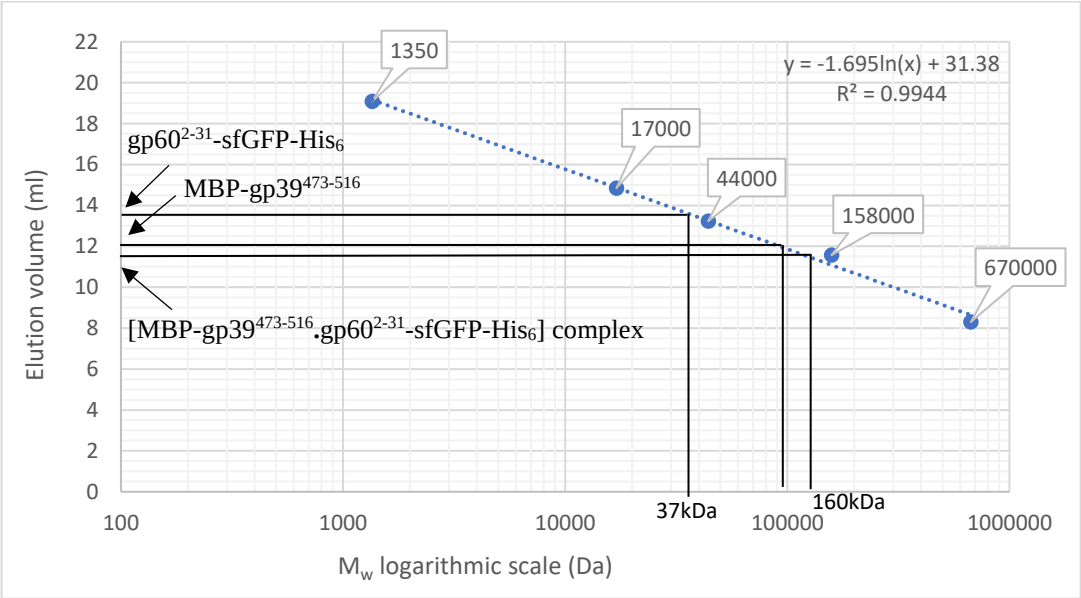
Figure 5.4. Size exclusion chromatography profiles. Protein samples (Figure 5.3B) were prepared at equimolar concentrations in SEC running buffer (**A-B**) or mixed (**C**), applied to the column and resolved by AKTA system machine (Chapter 2, section 2.26). **A.** Chromatogram of MBP-gp39$^{473-516}$. **B.** Chromatogram of gp60$^{2-31}$-sfGFP-His$_6$. **C.** Chromatogram of MBP-gp39$^{473-516}$.gp60$^{2-31}$-sfGFP-His$_6$ complex. Absorbance (mAu) at 280 nm (blue) (A_{280}) and 488 nm (red) (A_{488}) against elution volume (ml) is plotted. Time of injection of samples (pink line) and fractionation A2-B10 are shown. Apparent molecular weight (M_w) in kilo daltons (kDa). V_E: elution volume (dashed line).

A.



	Gel-filtration standard (Bio-Rad)	Molecular weight (Da)	Elution volume (ml)
A	Bovine thyroglobulin	670,000	8.3
B	Bovine γ-globulin	158,000	11.57
C	Chicken ovalbumin	44,000	13.23
D	Equine myoglobin	17,000	14.84
E	Vitamin B-12	1,350	19.08

B.



Graph 5.1. Generation of standard molecular weight curve. Gel-filtration standard sample (a mixture of 5 proteins) (Bio-Rad) was prepared in SEC running buffer, applied to the column, and resolved by AKTA system machine. **A.** Chromatogram of Gel-filtration standard. Absorbance (mAu) at wavelength 280 nm (blue) against elution volume (ml) is plotted. Time of injection of sample (pink line). Elution volumes and molecular weights of protein samples chromatographed are shown in the table. **B.** Standard molecular weight curve was generated by plotting elution volume (ml) against log-Molecular weight (Da) of Gel-filtration standard (Graph 5.1A).

The absorbance peak at 488 nm is double than that at 280 nm in the gp60$^{2-31}$-sfGFP-His$_6$ sample (Figure 5.4B). Since the molar extinction coefficient of this protein at absorbance 488 nm (A_{488}) is double than that at 280 nm (A_{280}) (Pédelacq *et al.*, 2006), it results in a greater absorbance at 488 nm than 280 nm according to Beer's Law:

$$Absorbance\ (a.u.) = concentration\ (M) x\ Molar\ extinction\ coefficient\ (M^{-1}\ cm^{-1}) x\ L\ ;\ where\ L = 1\ cm$$

Furthermore, the absorbance peak at 280 nm was ~ 2 (a.u.) for MBP-gp39$^{473-516}$ sample (Figure 5.4A) and ~ 0.6 (a.u.) for gp60$^{2-31}$-sfGFP-His$_6$ (Figure 5.4B). The A_{280} extinction coefficient at 1 mg/ml is 1.5 for MBP-gp39$^{473-516}$ and 0.6 for gp60$^{2-31}$-sfGFP-His$_6$ (Chapter 2, Table 2.7), showing that the same molar concentration of each protein was used in the assays.

For the interaction assay using size exclusion chromatography, an equimolar concentration of recombinant proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ were mixed and chromatographed (Chapter 2, section 2.26). The mixture of the two proteins was shown to be a stable [MBP-gp39$^{473-516}$**.**gp60$^{2-31}$-sfGFP-His$_6$] protein complex (Figure 5.4C). The elution volume of this protein complex was ~ 11.5 ml (Figure 5.4C), and by inputting this elution profile on the standard molecular weight curve (Graph 5.1B), a ~ 160 kDa molecular weight is measured. This protein complex formed by association of MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ was probably a heterotetramer in the form of an MBP-gp39$^{473-516}$ dimer (~ 100 kDa) with 2 molecules of gp60$^{2-31}$-sfGFP-His$_6$ (~ 37 kDa + 37 kDa). The presence of the absorbance peak at 488 nm, a characteristic of sfGFP, confirms that gp60$^{2-31}$-sfGFP-His$_6$ is co-eluting with MBP-gp39$^{473-516}$ that contributes to a greater absorbance peak at 280 nm (Figure 5.4C).

These observations were further validated by analysis of the elutent samples (Figure 5.4) on an SDS-PAGE (Figure 5.5), which showed the presence of both proteins co-eluting from the mixture sample. The bands corresponding in molecular weights to MBP-gp39$^{473-516}$ at 48 kDa and gp60$^{2-31}$-sfGFP-His$_6$ at 31 kDa were detected in the elution sample (Lanes 8-9, Figure 5.5). The other bands were probably degradation, aggregation and some unbound protein indicated by the small negligible absorbance peaks at various elution profiles (Figure 5.4 and Figure 5.5).

Overall, these results indicate that MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ formed a protein complex, likely mediated by their gp39$^{473-516}$ and gp60$^{2-31}$ moieties. Both MBP and sfGFP have been previously used to monitor protein-protein interactions between peptides fused to them *in vitro* (Pinto *et al.*, 2009; Pinto and Baiker, 2012).

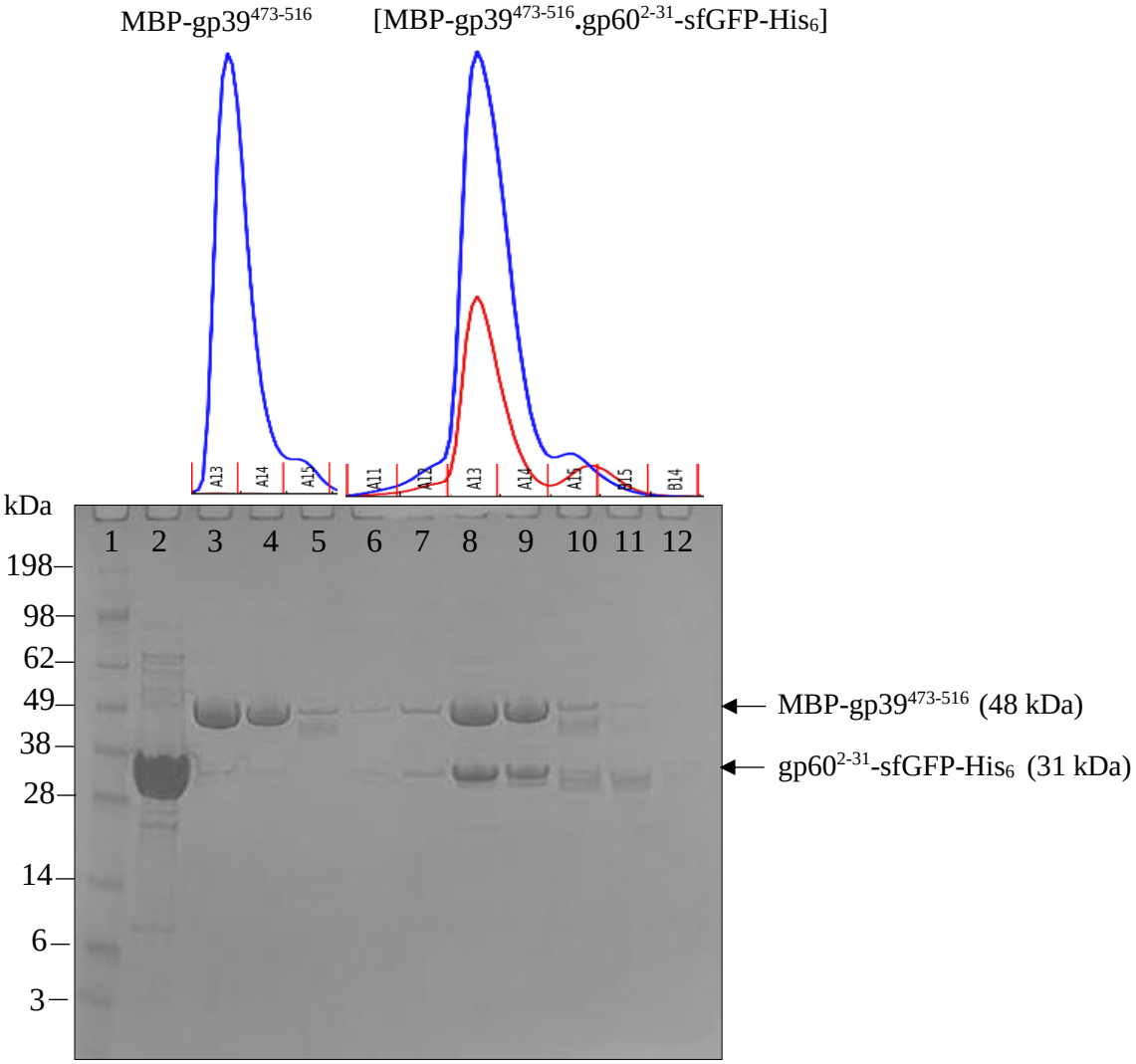


Figure 5.5. Fractionation samples of MBP-gp39$^{473-516}$ and MBP-gp39$^{473-516}$.gp60$^{2-31}$-sfGFP-His$_6$ protein complex. A small volume of the fractionation samples obtained from size exclusion chromatography (Figure 5.4) was analyzed on a 4-12% Bis-Tris gel stained with coomassie blue. Lane 1: SeeBlue Plus2 Pre-stained protein standard. Lane 2: gp60$^{2-31}$-sfGFP-His$_6$ (Figure 5.3B). Lanes 3-5: MBP-gp39$^{473-516}$ fractionation samples A13, A14 and A15, respectively (Figure 5.4A). Lanes 6-12: MBP-gp39$^{473-516}$.gp60$^{2-31}$-sfGFP-His$_6$ protein complex fractionation samples A11, A12, A13, A14, A15, B15 and B14, respectively (Figure 5.4C).

To determine the strength of interaction between MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ *in vitro*, the dissociation constant (k_d) of the protein complex was measured using MicroScale Thermophoresis (MST) (Monolith NT.115 instrument). The technique is based on the directed movement of molecules in temperature gradients, which is referred to as thermophoresis, and depends on molecule size, charge, and hydration shell (Mueller *et al.*, 2017). Since at least one of these parameters is typically affected upon binding of a ligand (resulting in distinct thermophoretic movements of the unbound and bound states that can be detected by optics) (Figure 5.6), this method can be used to analyze any kind of biomolecular interaction.

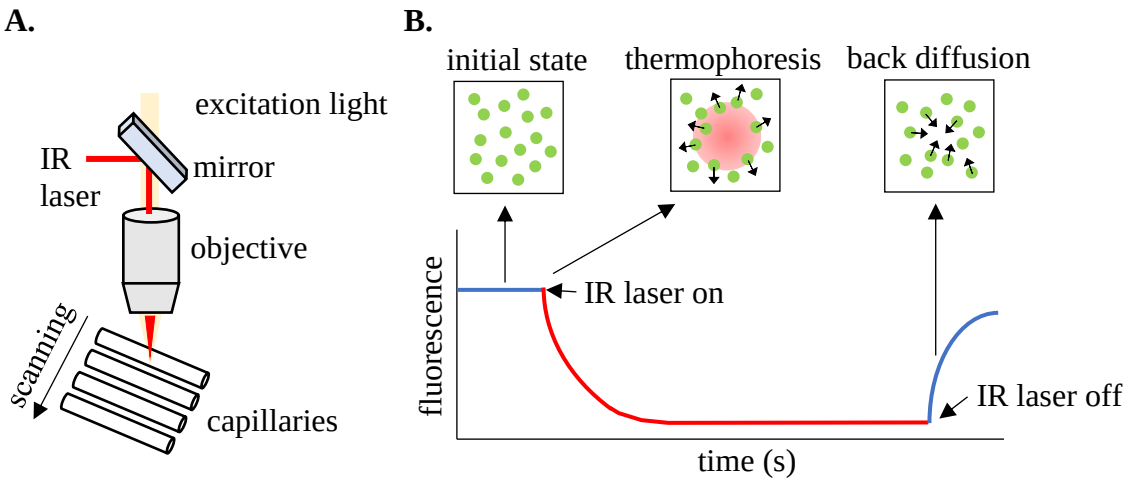


Figure 5.6. Setup of an MST experiment. A. The emitted fluorescence signal of the optically visible molecule (intrinsically fluorescent or with an attached fluorophore) is detected by the optics focus in the center of capillaries. An infrared (IR) laser is used to establish a temperature gradient in the observation window of the optical system. The thermophoretic movement of the molecules in solution can be monitored by changes in fluorescence signal in the detection window of the optical system. **B.** MST trace in a capillary – movement profile of molecules in a temperature gradient. After an initial cold phase (laser off), the laser is turned on to establish a temperature gradient. The increase in temperature causes the molecules to diffuse away from the heated spot in the observation optical window decreasing the fluorescence signal yield upon heat induction. After the laser is turned back off, molecules diffuse back and the fluorescent signal yield increases. In a typical MST experiment, 16 capillaries containing the same concentration of an optically visible 'target' molecule and a decreasing concentration of an unlabeled 'ligand' molecule (to establish different ratios of the binding partners) are used in a single run to obtain binding parameters such as the binding affinity.

Again, the proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ were expressed in *E. coli* C41 carrying pMH39 and pMH60, respectively (Figure 5.3A), purified by affinity chromatography (Chapter 2, section 2.24), dialyzed (Chapter 2, section 2.25), and analyzed on an SDS-PAGE prior to microscale thermophoresis (MST). Bands corresponding in molecular weights to 48 kDa for MBP-gp39$^{473-516}$ and 31 kDa for gp60$^{2-31}$-sfGFP-His$_6$ were observed (Lanes 2-3, Figure 5.7), as previously observed (Lanes 2-3, Figure 5.3B). These two samples were then used in MST to determine the dissociation constant (k_d) of the protein complex.

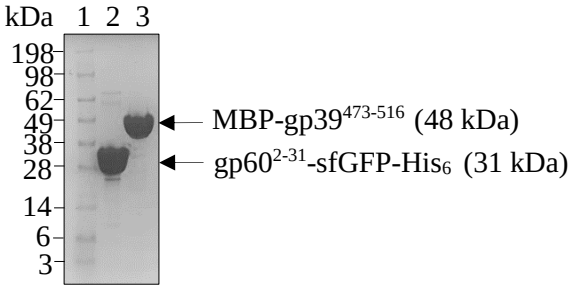


Figure 5.7. Expression of recombinant proteins MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$. The proteins were expressed in *E. coli* C41 carrying pMH39 and pMH60 (Figure 5.3A), the MBP-tagged protein was purified using MBPTrap column, while the His-tagged protein was purified using HisGraviTrap column. The column-bound proteins were eluted and elution samples (E_2 for gp60$^{2-31}$-sfGFP-His$_6$ (~ 10 mg/ml or 300 µM) and E_4+E_5 for MBP-gp39$^{473-516}$ (~ 5 mg/ml or 150 µM)) were dialyzed. A small volume of dialyzed sample was analyzed on a 4-12% Bis-Tris gel stained with coomassie blue solution prior to its use in microscale thermophoresis. Lane 1: SeeBlue Plus2 pre-stained protein standard. Lane 2: gp60$^{2-31}$-sfGFP-His$_6$. Lane 3: MBP-gp39$^{473-516}$.

As one of the proteins, gp60$^{2-31}$-sfGFP-His$_6$, was intrinsically fluorescent (fluorescent 'target' molecule), protein labelling was not necessary to monitor the interaction. A fluorescence pre-test scan was performed to determine the concentration of gp60$^{2-31}$-sfGFP-His$_6$ required for a fluorescence signal above 400 fluorescence counts (detectable and acceptable). A final concentration of 0.5 µM gave suitable counts and was used throughout the MST experiment. A serial dilution of the 'ligand' protein MBP-gp39$^{473-516}$ was prepared and mixed with a constant concentration of the fluorescent 'target' protein

gp60$^{2-31}$-sfGFP-His$_6$ to establish different ratios of the binding partners (Chapter 2, section 2.27).

The samples were loaded into capillaries and analyzed in the MST instrument by subsequently scanning each capillary. Prior to the binding measurement, a capillary scan was performed to determine molecule adsorption to the capillary walls, aggregation, or pipetting errors. The fluorescence obtained was approximately uniform across the 16 capillaries (Figure 5.8A), indicating negligible signs of aggregation, pipetting errors or molecule adsorption to the capillary.

In the initial time of the experiment (up to 0 second time point) cold phase, homogeneity between samples is present (Figure 5.8B). After the solution in the capillaries is locally heated up by the IR laser, the raise in temperature causes a steep drop in the fluorescence signal (Figure 5.8B). The fluorescence signal decreases faster as the ligand concentration decreases (Figure 5.8B), since mass diffusion dictates the kinetics of depletion (*i.e.* the fluorescent molecules that formed a complex will diffuse away slower than the unbound molecules, so the fluorescence signal is detected longer in the observation optical window of the MST instrument (Figure 5.6B)). These well-defined MST traces (Figure 5.8B), indicate an interaction between gp60$^{2-31}$-sfGFP-His$_6$ and MBP-gp39$^{473-516}$.

According to the Monolith NT.115 instrument manufacturer's (NanoTemper technologies), a local temperature difference ΔT leads to a local change in molecule concentration quantified by the Soret coefficient S_T: $c_{hot}/c_{cold} = \exp(-S_T\Delta T) \approx 1-(S_T\Delta T)$. The normalized fluorescence (F_{norm}) measures this concentration ratio, plus a temperature dependence of the dye. In the linear approximation, $F_{norm} = F_{hot}/F_{cold} = 1-(ST - dF/dT)\ \Delta T$. Within the serial dilution experiment F_{norm} changes according to:

$$F_{norm} = (1-x)\ F(A)_{norm} + x\ F(AT)_{norm} \qquad (1)$$

Where $F(A)_{norm}$ is the contribution of the unbound fluorescent molecule A, $F(AT)_{norm}$ the contribution of the complex of the fluorescent molecule A and its interacting ligand T, and x the fraction of fluorescent molecules that formed the complex.

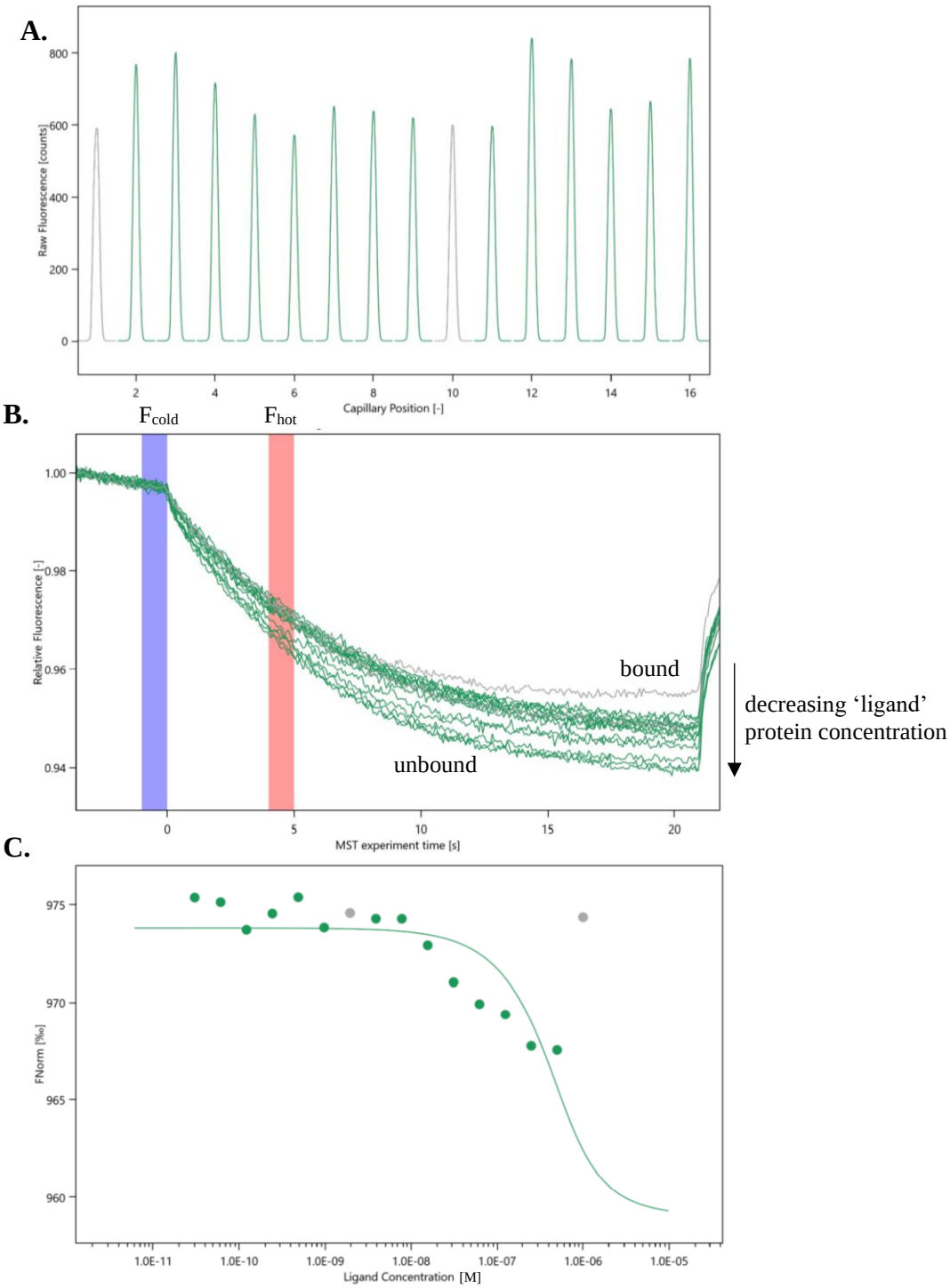


Figure 5.8. MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ binding parameters determined with microscale thermophoresis. 16 capillaries containing a fixed concentration of fluorescent 'target' protein gp60$^{2-31}$-sfGFP-His$_6$ (0.5 µM) and a decreasing concentration of the 'ligand' protein MBP-gp39$^{473-516}$ (16-step serial dilution of 1 µM) were analyzed in a single run (Monolith NT.115). **A.** Capillary scan. Raw fluorescence (counts) against capillary position (1-16). **B.** MST Trace. Relative fluorescence (a.u.) against MST experiment time (s). **C.** K_d-Fit curve. Normalized fluorescence of the MST traces (‰) (Figure 5.8B) against ligand concentration (M). Screenshots of analysis report obtained by the MO.Affinity Analysis v2.3 software.

By increasing the concentration of the unlabeled ligand and mixing it with a constant concentration of fluorescent molecule, the fraction of complexes increases until all fluorescent molecules formed complexes with the ligand. Thus, the fraction of bound molecules x can be derived from the measured change in normalized fluorescence F_{norm}.

Plotting the normalized fluorescence F_{norm} from the MST traces (Figure 5.8B) against the concentration of the 'ligand' protein MBP-gp39$^{473-516}$ results in a binding curve from which the dissociation constant (k_d) can be derived (Figure 5.8C). The dissociation constant k_d is obtained by fitting the binding curve with the quadratic solution for the fraction of fluorescent molecules that formed the complex, calculated from the law of mass action:

$$A + T \rightleftharpoons [A \,.\, T] \text{ complex}$$

$$k_d = [A] * [T] / [AT] \quad ; \text{at 50\% [T] bound the } k_d = [A]$$

where [A] is the concentration of free fluorescent molecule, [T] the concentration of free ligand and [AT] is the concentration of complexes of [A] and [T].

The free concentrations of A and T are $[A]=[A_0]-[AT]$ and $[T]=[T_0]-[AT]$, respectively, where $[A_0]$ is the known concentration of the fluorescent molecule and $[T_0]$ is the known concentration of added ligand. This leads to a quadratic fitting function for [AT]:

$$[AT]=1/2*(([A_0]+[T_0]+ k_d)-(([A_0]+[T_0]+ k_d)^2 - 4*[A_0]*[T_0]))^{1/2}) \qquad (2)$$

The concentration of fluorescent molecule $[A_0]$ is kept constant during the experiments and the concentration of ligand $[T_0]$ is varied by serial dilution. The signal obtained in the measurement (equation 1) directly corresponds to the fraction of fluorescent molecules that formed the complex $x = [AT]/[A_0]$ which can be fitted with the derived formula to obtain k_d. This resulted in a dissociation constant of $k_d = 1.8248x10^{-7} \pm 5.6534x10^{-6}$ (M), measured by the MST software.

The MST experiment was repeated (but with varying the 'ligand' protein MBP-gp39$^{473-516}$ concentration) three times on different occasions (Figure 5.9). The capillary scans showed an approximately uniform fluorescence across the 16 capillaries in each independent assay (Figure 5.9A), indicating negligible signs of aggregation, pipetting errors or molecule adsorption to the capillary. Moreover, the non-uniformity of fluorescence among the independent assays (Figure 5.9A) was probably a result of protein degradation.

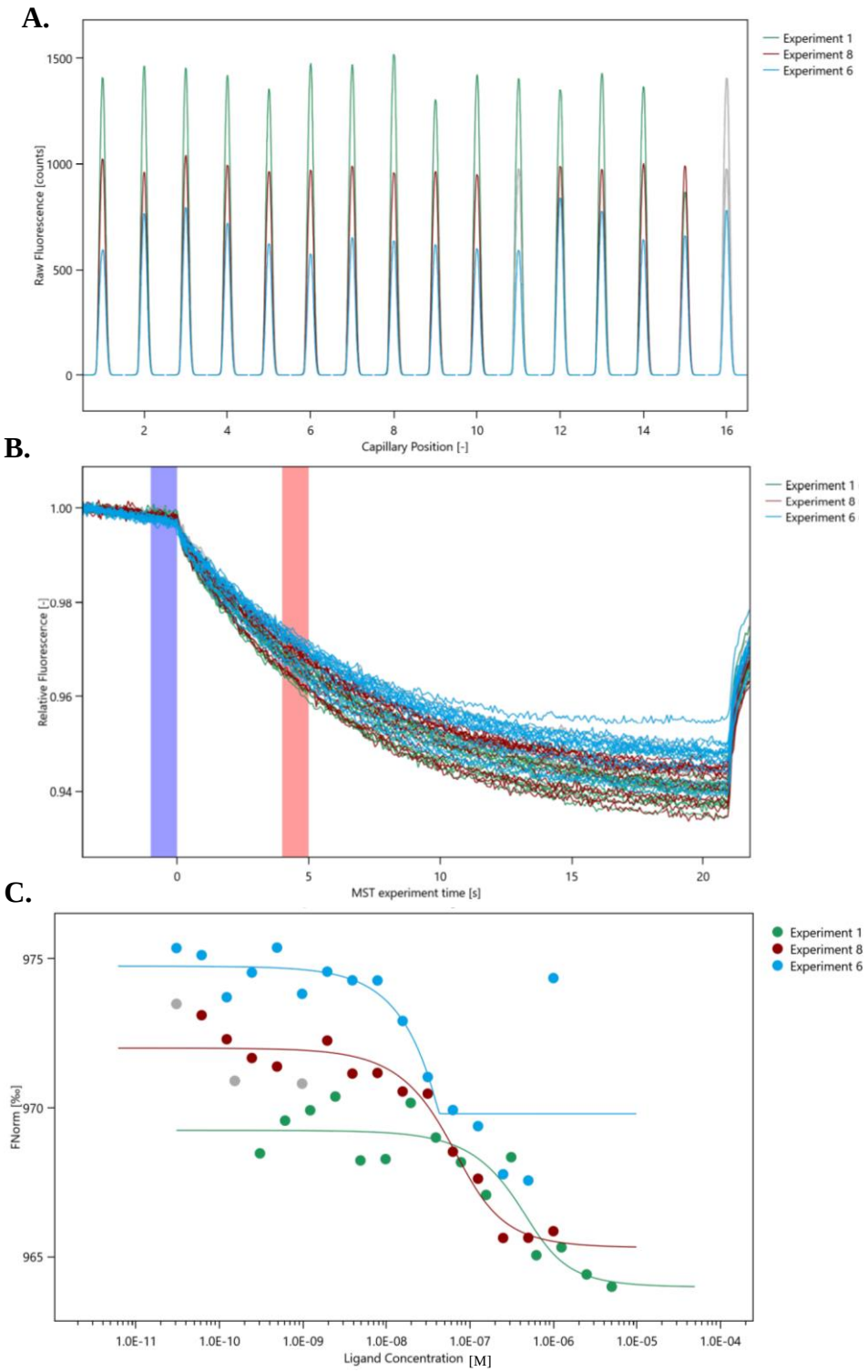


Figure 5.9. MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$ binding parameters determined with microscale thermophoresis. 16 capillaries containing a fixed concentration of fluorescent 'target' protein gp60$^{2-31}$-sfGFP-His$_6$ (0.5 µM) and a decreasing concentration of the 'ligand' protein MBP-gp39$^{473-516}$ were analyzed in a single run (Monolith NT.115). The experiment was repeated three times on different occasions, but with varying the 'ligand' protein MBP-gp39$^{473-516}$ concentration (Experiment 1: 16-steps serial dilution of 10 µM MBP-gp39$^{473-516}$; Experiment 8 and 6: 16-steps serial dilution of 2 µM MBP-gp39$^{473-516}$). **A.** Capillary scans. **B.** MST Traces. **C.** K_d-Fit curves. Figure legend as in Figure 5.8.

The MST traces of gp60$^{2-31}$-sfGFP-His$_6$ binding to MBP-gp39$^{473-516}$ provided well-defined curves (Figure 5.9B), indicating an interaction between the two proteins. The quantitative information on binding affinity was extracted from the K_d-Fit curves (Figure 5.9C), resulting in dissociation constant of $k_d = 1.6301 \times 10^{-7} \pm 1.8403 \times 10^{-7}$ (M) (experiment 1), $k_d = 3.0303 \times 10^{-8} \pm 3.3098 \times 10^{-8}$ (M) (experiment 8), $k_d = 1.3848 \times 10^{-12}$ (M) (experiment 6).

The average dissociation constant from the 4 independent binding assays was calculated (Table 5.1). As the k_d is smaller than the concentration of the fluorescent 'target' protein gp60$^{2-31}$-sfGFP-His$_6$ used in the assays, indicates that most molecules are in complex form than in free form (reflecting a strong interaction between MBP-gp39$^{473-516}$ and gp60$^{2-31}$-sfGFP-His$_6$).

Experiment	$k_d \pm k_d$ confidence (M)	Reference
	$1.8248 \times 10^{-7} \pm 5.6534 \times 10^{-6}$	Figure 5.8C
1	$1.6301 \times 10^{-7} \pm 1.8403 \times 10^{-7}$	Figure 5.9C
8	$3.0303 \times 10^{-8} \pm 3.3098 \times 10^{-8}$	Figure 5.9C
6	$1.3848 \times 10^{-12} \pm$ N/A	Figure 5.9C
Average k_d	$9.39 \times 10^{-8} \pm 9.21 \times 10^{-8}$	

Table 5.1. Dissociation constant (k_d) of [gp60$^{2-31}$-sfGFP-His$_6$. MBP-gp39$^{473-516}$] protein complex as determined by MST. Dissociation constants were measured by the integrated MO.Affinity Analysis v2.3 software (Monolith NT.115).

However, lack of curvature (Figure 5.8C and Figure 5.9C) makes accurate fitting of the data difficult or impossible. Thus, an accurate k_d could not be measured. The experiment could be repeated with protein labelling for a more sensitive fluorescence signal detection or by increasing the number of capillary samples to obtain more data points for the k_d-Fit curves.

5.2.3 Interaction of gp39$^{473-516}$ with gp60$^{2-31}$ *in vivo*

Two-hybrid systems are *in vivo* genetic tests for detection of protein-protein interactions. Numerous two-hybrid systems exist that are suitable for screening of functional interactions between two proteins in yeast or bacteria (Mehla *et al.*, 2017). Because gp39$^{473-516}$ and gp60$^{2-31}$ are derived from phage-encoded bacterial proteins, a bacterial two-hybrid system was used to examine their interaction, as the screening conditions are more biologically similar to the proteins' natural environment. The bacterial two-hybrid system is based on the principle of restoring protein activity upon non-covalent reconstitution of split protein fragments. One fragment of a modular protein is fused to a polypeptide X, and the other fragment is fused to a polypeptide Y. If the polypeptides X and Y interact upon co-expression, the modular protein is reconstituted, regains its activity, and its activity is detected through a reporter gene.

The most common used bacterial two-hybrid system is the Bacterial Adenylate Cyclase-based Two-Hybrid System (BACTH) which is based on the reconstitution of adenylate cyclase activity in *E. coli* (Karimova *et al.*, 1998) (Figure 5.10). It exploits the fact that the catalytic domain of adenylate cyclase from *Bordetella pertussis* (Ladant and Ullmann, 1999) consists of two complementary fragments, termed T25 and T18, that are not active when physically separated. When these two fragments are fused to interacting polypeptides, X and Y, heterodimerization of these hybrid proteins results in functional complementation between T25 and T18 fragments, leading to cyclic AMP (cAMP) synthesis. cAMP produced associates with CAP, a catabolite activator protein. The [CAP**.**cAMP] complex is a regulator of gene expression in *E. coli* (Ullman and Danchin, 1980). It turns on the expression of several genes, including genes of the *lac* operons involved in lactose catabolism. Therefore, bacteria become able to hydrolyze lactose or lactose-analog X-gal and can be easily distinguished on indicator media (hydrolysis of X-gal results in an intensely blue-colored product that turns cells blue).

This approach is performed in *cya*-deficient *E. coli* strains (such as BTH101), that have non-functional adenylate cyclases incapable of cAMP synthesis, and therefore, are catabolically repressed (Karimova *et al.*, 2001). This strain expresses β-galactosidase only when the catalytic domain fragments capable of complementing are provided exogenously. Two compatible expression plasmids pKT25-zip and pUT18 that respectively encode the T25 and T18 fragments have been previously constructed and are available for fusions with polypeptides of interest to detect interactions in *cya- E. coli* (Karimova *et al.*, 2001).

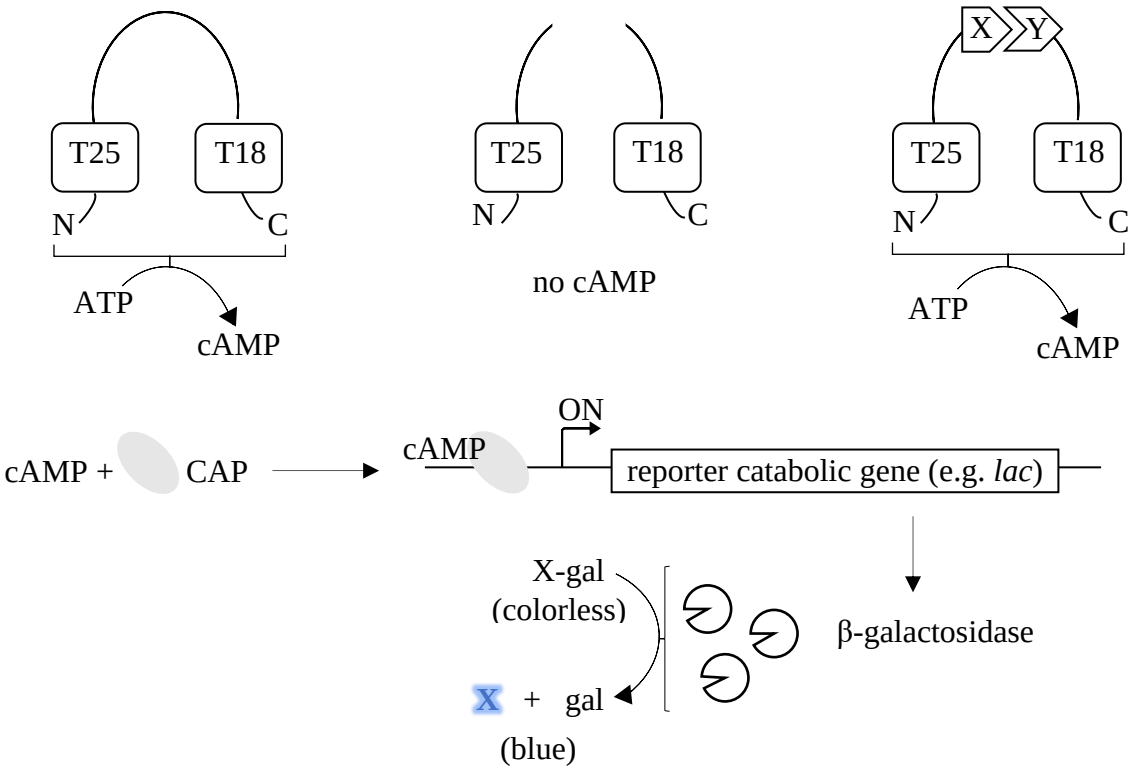
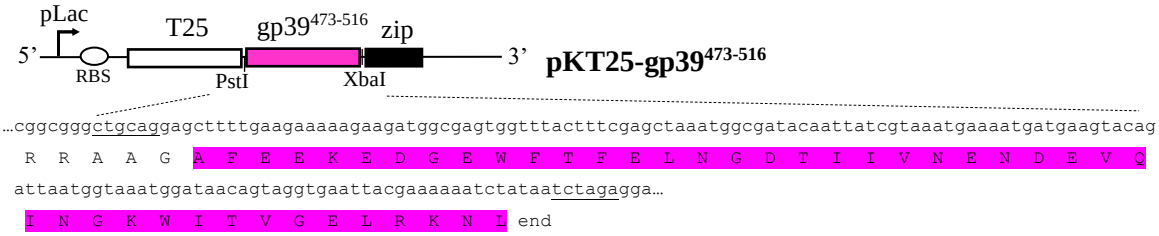


Figure 5.10. Principle of the Bacterial Adenylate Cyclase-based Two-Hybrid System. The T25 and T18 fragments of the catalytic domain of *B. pertussis* adenylate cyclase are shown in different configurations. The intact catalytic domain catalyzes cAMP synthesis. The separated T25 and T18 fragments are unable to interact and no cAMP synthesis occurs. The interacting polypeptides X and Y bring together T25 and T18 and restore cAMP synthesis. cAMP synthesis leads to characteristic phenotypes on indicator plates.

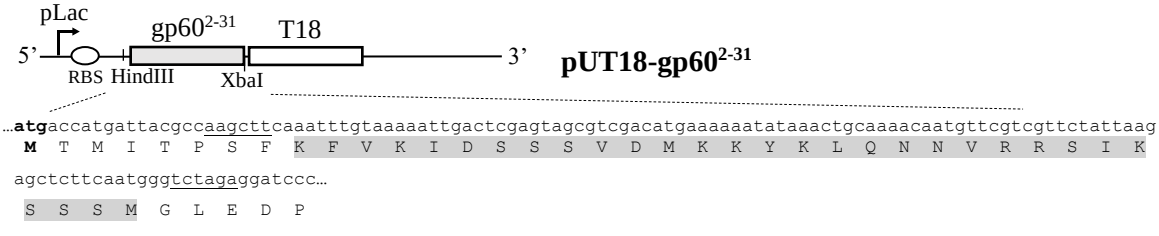
To examine the interaction of gp39$^{473\text{-}516}$ with gp60$^{2\text{-}31}$ *in vivo*, the bacterial two hybrid system described above was performed in *cya- E. coli* strain BTH101. Peptide gp39$^{473\text{-}516}$ was fused in-frame to the C-terminus of the T25 fragment and gp60$^{2\text{-}31}$ was fused in-frame to the N-terminus of the T18 fragment (encoded by plasmids pKT25-gp39$^{473\text{-}516}$ and pUT18-gp60$^{2\text{-}31}$, respectively) (Figure 5.11A). *E. coli* BTH101 was transformed with pKT25-gp39$^{473\text{-}516}$ and pUT18-gp60$^{2\text{-}31}$ (Figure 5.11A) and complementation was analyzed by visual inspection of colony color (Figure 5.11B). When *E. coli* BTH101 carried pKT25-gp39$^{473\text{-}516}$ and pUT18-gp60$^{2\text{-}31}$ it yielded blue colonies (Figure 5.11B), indicating a functional association between the expressed hybrid proteins which was mediated by the gp39$^{473\text{-}516}$ and gp60$^{2\text{-}31}$ moieties. This was further confirmed by the control experiments showing that complementation does not occur in absence of either one or both of gp39$^{473\text{-}516}$ and gp60$^{2\text{-}31}$ moieties (Figure 5.11B).

A.



T4 gp39

```
468 V L G  E  K  A F E E K E D G E W F T F E L N G D T I I V N E N D E V Q
    I N G K W I T V G E L R K N L end   516
```



T4 gp60

```
1                   M  K F V K I D S S S V D M K K Y K L Q N N V R R S I K
  S S S M N Y A N V A I 38
```

B.

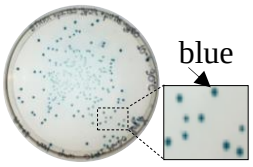
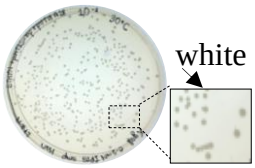
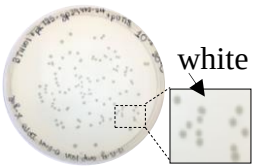
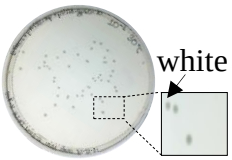
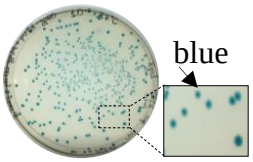
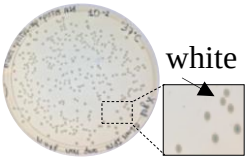
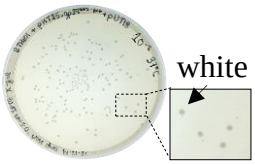
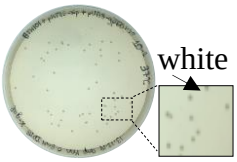
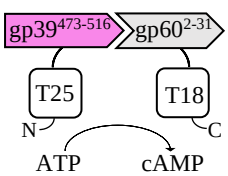
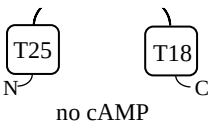
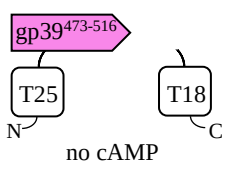
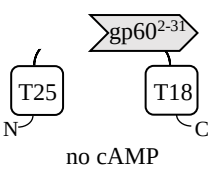
	***E. coli* BTH101 +**			
	pKT25-gp39$^{473-516}$ + pUT18-gp60$^{2-31}$	pKT25-zip + pUT18	pKT25-gp39$^{473-516}$ + pUT18	pKT25-zip + pUT18-gp60$^{2-31}$
30°C				
37°C				
				

Figure 5.11. Analysis of complementation in *E. coli* strain BTH101. A. Part of nucleotide sequence (5'-3') and derived amino acid sequence of bacterial two-hybrid system plasmids pKT25-gp39$^{473-516}$ and pUT18-gp60$^{2-31}$ encoding T25-gp39$^{473-516}$ and gp60$^{2-31}$-T18 hybrid proteins, respectively. Amino acid sequences of T4 topoisomerase subunits gp39 and gp60 are shown for comparison**. B.** Colony color phenotype of BTH101 carrying relevant expression plasmids on L-Agar plates supplied with X-gal (40 µg/ml), IPTG (0.5 mM) and antibiotics as necessary; grown at either 30°C or 37°C for 24 hours. A 10^{-2} dilution of the initial transformation mix was plated to prevent cell crowding. Schematic illustration of the expressed hybrid proteins from the corresponding plasmid constructs (bottom).

5.2.4 Application of gp39$^{473-516}$ and gp60$^{2-31}$ interaction

The interaction of gp39$^{473-516}$ with gp60$^{2-31}$ has important implications for development of synthetic tools that can be used to control protein activity by protein interaction domains. Various approaches have been developed to control protein activity by reconstitution of split protein fragments using inteins (Olorunniji *et al.*, 2019) and inducible dimerization domains (Weinberg *et al.*, 2019). A protein of interest is split into two fragments and each one is fused to an intein or inducible dimerization domain. Split protein fragments tagged with inducible dimerization domains or inteins can be rendered non-functional until an inducer molecule is administered, enabling spatiotemporal control of gene expression beyond those at the DNA and RNA levels (Di Ventura and Mootz, 2018). However, identification of functional split sites in proteins is challenging, because of problems that rely on rational inference of split sites from secondary structure alignment of homologous proteins, and beacuse active proteins can lose function by splitting and tagging them.

To overcome this challenge, a new method for screening functional split sites in proteins was developed. The method relies on the random insertion of an engineered transposon Tn*4430* and subsequent insertion of polypeptides gp39$^{473-516}$ and gp60$^{2-31}$ within the target gene (Figure 5.12). Transposable elements have been previously used to deliver sequences into target genes (Hallet *et al.*, 1997; Ho *et al.*, 2021; Gregory *et al.*, 2010), and can be adapted to identify functional split sites in proteins.

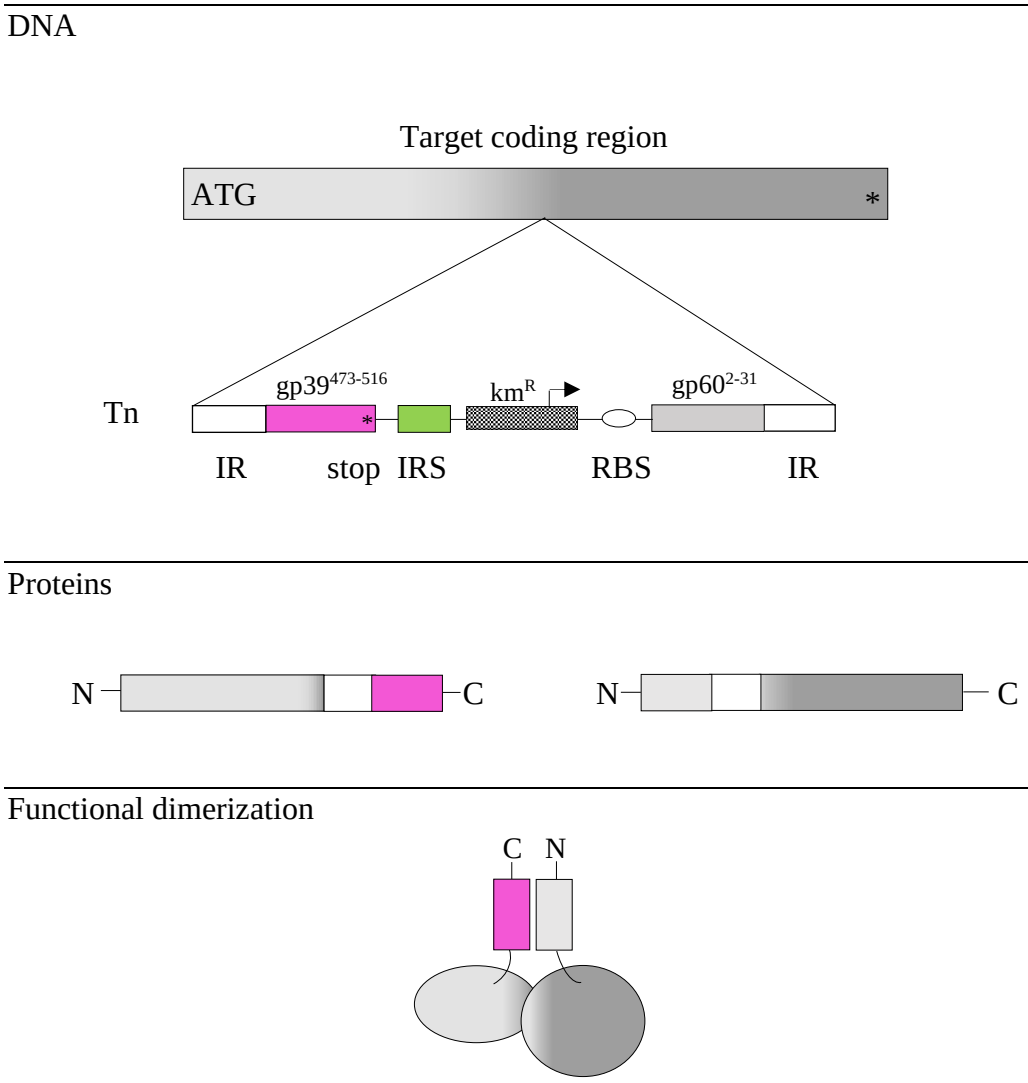


Figure 5.12. Transposon tagging scheme. The customized transposon Tn*4430* cassette (Tn) composed of: (i) transposase TnpA terminal inverted repeat (IR) sites and TnpI resolvase internal recombination site (IRS), (ii) interaction domains derived from *mobA* ($gp39^{473-516}$: amino acids 473-516 of T4 gp39, $gp60^{2-31}$: amino acids 2-31 of T4 gp60), (iii) kanamycin resistance gene (km^R) to select for insertion and (iv) initiation signals (a promoter (arrow) and a ribosome binding site (RBS)) for translation downstream**,** allows random insertions of the interaction domains into a target coding region without disrupting reading frames. Libraries can be screened to identify insertions that maintain target protein function, allowing localization of functional split sites.

The transposon cassette (Tn) (Figure 5.12) was designed and commercially chemically synthesized as a gBlock oligonucleotide (Figure 5.13).

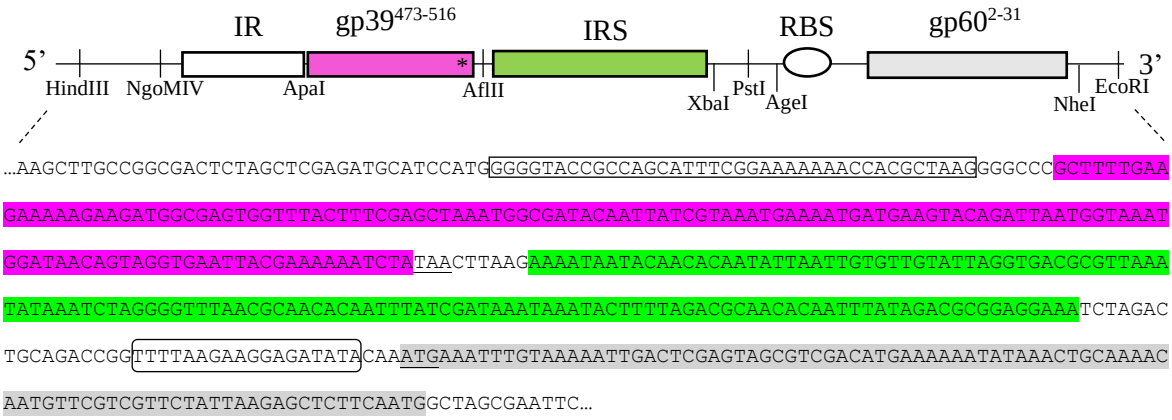


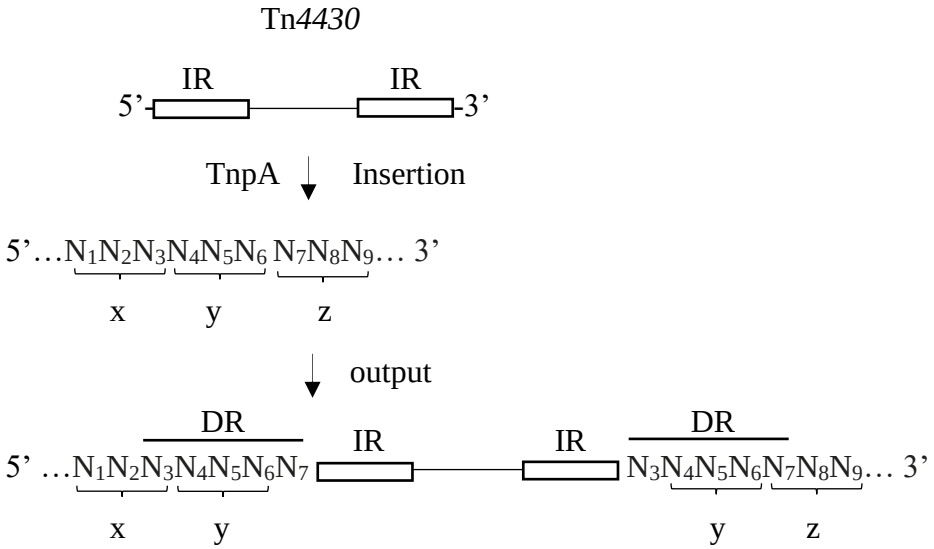
Figure 5.13. Transposon cassette. Structure and nucleotide sequence (5'-3') of a customized transposon Tn*4430* cassette consisting of an inverted repeat (IR) (white box), internal recombination site (IRS) (green), coding fragments for amino acids 473-516 of T4 topoisomerase subunit gp39 ($gp39^{473-516}$) (purple) and amino acids 2-31 of T4 gp60 ($gp60^{2-31}$) (grey), and a ribosome binding site (RBS) (circle). Unique restriction sites are shown. Synthesized as a gBlock oligonucleotide by IDT (Chapter 2, Table 2.9).

The transposon was designed with one terminal inverted repeat sequence (IR) instead of two, since presence of two terminal IRs would affect gBlock synthesis. Therefore, unique restriction sites NgoMIV and NheI were introduced that would allow future cloning of the transposon into a plasmid vector pOD2 (a gift from Prof. B. Hallet and O. Dudekem) that already contains the second terminal IR in the appropriate frame. A unique restriction site PstI was added that would allow subcloning of the km^R fragment (containing a promoter sequence (personal communication from S. D Colloms)) from plasmid pUC71K into the transposon in the appropriate frame to complete it. Other unique restriction sites were added so that the building blocks of the transposon could be easily exchanged, for example, with inducible dimerization domains for delivery into target genes, enabling ligand-responsive re-constitution of the split proteins.

Transposon Tn*4430*, a Tn*3*-related transposon from *Bacillus thuringiensis*, transposes efficiently in *E. coli* and duplicates 5 bp of host sequences at the insertion point (Mahillon and Lereclus, 1988; Hallet *et al.*, 1997) (Figure 5.14A). Therefore, the transposon was designed so that its insertion would not disrupt reading frames − only insertions after

the first nucleotide within a codon would maintain the proper translation frame (Figure 5.14B)**.**

A.



B.

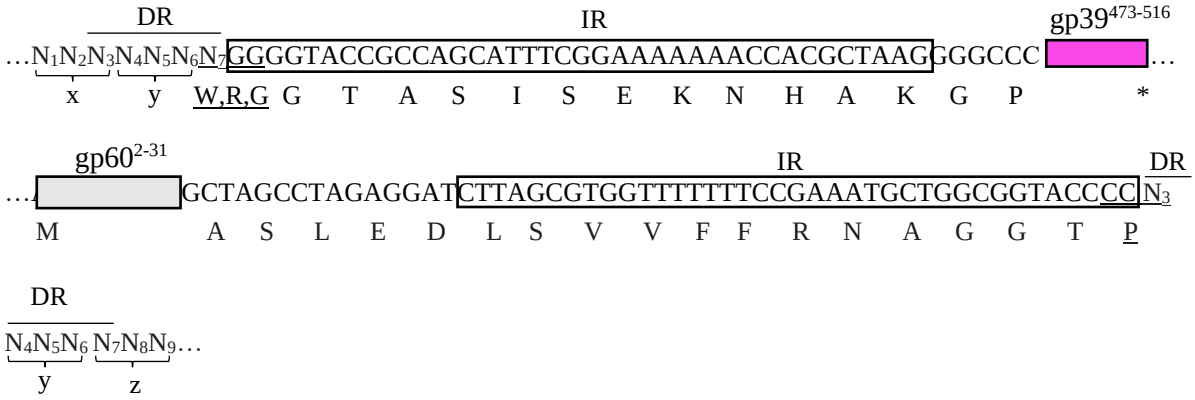


Figure 5.14. Principle of transposon insertion method. A. Transposon Tn*4430* with the terminal inverted repeat (IR) sequences represented by a white box. The string of nucleotides denoted N_1N_2 etc. represents the sequence of the target coding region. Insertion site of the transposon within a codon z catalyzed by transposase TnpA is shown. Horizontal line indicates the 5 nucleotides duplicated during Tn*4430* insertion (direct repeats (DR)). This position contains a duplication in the upstream codon y**,** and codons whose identity depends on the nucleotides present in the IR site downstream N_7 or upstream N_3. **B.** Only insertions at this specific position of the engineered transposon Tn*4430* (Figure 5.12 and Figure 5.13-after completion) would maintain the proper translation reading frame of the split fragments. Insertion in other positions would take translation out-of-frame or introduce stop codons.

The transposon cassette (Figure 5.13) was first cloned into pUC19 plasmid vector at the HindIII-EcoRI sites to allow cloning of a km^R fragment within the cassette, and the resulting plasmid construct was termed pA (Figure 5.15A). The km^R fragment was then subcloned from plasmid pUC71K into pA at the PstI site, and the resulting plasmid construct was termed p4 (Figure 5.15B).

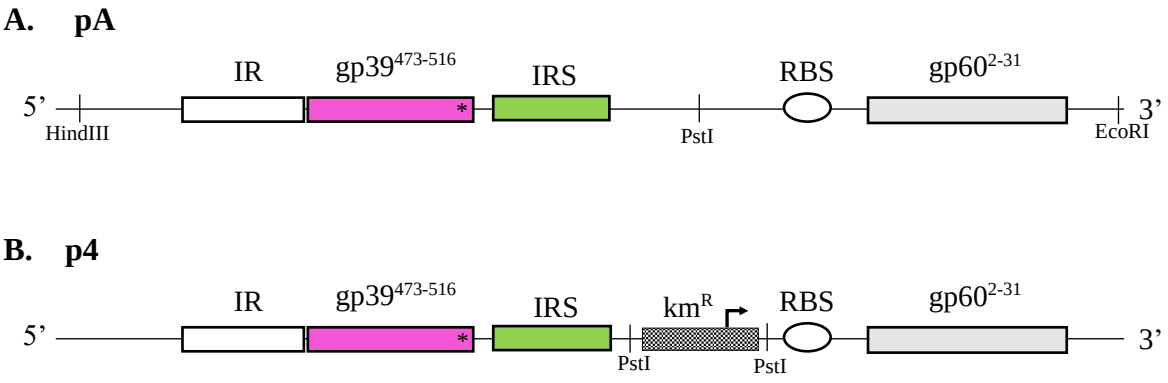


Figure 5.15. Transposon cassette construction. Relevant genetic information is shown (Figure legend as in Figure 5.13). The km^r fragment contains an internal promoter (personal communication from S. D. Colloms) for translation downstream.

Construction of the complete system *i.e* subcloning of the cassette from p4 (Figure 5.15B) into plasmid pOD2 that introduces the second terminal IR site to complete it, as well as testing the method was not possible due to time constraints, and would merit further investigation due to its important application to biotechnology and synthetic biology.

5.3 Discussion

The results demonstrate that amino acids 473-516 at the C-terminus of T4 topoisomerase subunit gp39 (referred to as $gp39^{473-516}$) interact with amino acids 2-31 at the N-terminus of T4 topoisomerase subunit gp60 (referred to as $gp60^{2-31}$). The protein-protein interaction assays used which include size exclusion chromatography, microscale thermophoresis and the bacterial two-hybrid system, are well-established methods and have been previously used to detect and analyze protein-protein interactions (Phizicky and Fields,

1995; Mueller *et al.*, 2017; Pinto *et al.*, 2009; Pinto and Baiker, 2012; Karimova *et al.*, 1998; Karimova *et al.*, 2001; Khaleel *et al.*, 2011; Fogg *et al.*, 2018). Future genetic and structural studies would be required to identify which specific amino acid residues are involved in the interaction interface of $gp39^{473-516}$ with $gp60^{2-31}$. The bacterial two-hybrid system would be an ideal genetic system to screen mutants quickly and easily in these interacting peptides of gp39 and gp60 to find out which amino acids are important for this interaction. Furthermore, future experiments using the bacterial two-hybrid system could be conducted to determine if regions outside of gp60(2-31) and gp39(473-516) might be involved in the interaction.

Protein interaction interfaces are thought to be distinguishable from the rest of the protein surface by their greater degree of residue conservation (Caffrey *et al.*, 2004; Mintseris and Weng, 2005). Multiple sequence alignments of T4 topoisomerase fragments $gp39^{473-516}$ and $gp60^{2-31}$ to other phages' fragmented topoisomerases show a high degree of conservation in numerous amino acid residues (Figure 5.16), that might suggest the main contacts involved in the interaction between $gp39^{473-516}$ and $gp60^{2-31}$.

Interestingly, these interacting C- and N-terminal regions of T4 topoisomerase gp39 and gp60 ($gp39^{473-516}$ and $gp60^{2-31}$, respectively) originated from a mobile genetic element insertion (*mobA*) into the T4 progenitor topoisomerase subunit gene *39*, and are absent in gp39 homologues (that encode the subunit as a single polypeptide) (Figure 5.1). The *mobA* insertion split the T4 progenitor gene *39* into two independent genes, gene *39* and gene *60*, contributing to these additional amino acid extensions, as well as the 50-nucleotide bypass element (Bonocora *et al.*, 2011), and the re-constituted topoisomerase complex retains function (Seasholtz and Greenberg, 1983). A similar case of gene fragmentation by a mobile genetic element *mobE* was demonstrated in phage Aeh1 ribonucleotide reductase gene *nrdA* (Friedrich *et al.*, 2007). The insertion created a unique genes-in-pieces arrangement, where *nrdA* was split into two genes, *nrdA-a* and *nrdA-b*, and the reconstituted complex retained function. The *mobE* insertion also contributed to additional amino acid sequences at the C- and N- terminal regions of the split subunits, which were shown to be necessary for re-assembly of the fragmented ribonucleotide reductase subunits (by protein-protein interaction of the regions) and enzymatic activity (Crona *et al.*, 2011). Therefore, it is likely that the unique C- and N- terminal regions of T4 topoisomerase gp39 and gp60 contributed by the *mobA* insertion are essential for re-assembly of an active topoisomerase.

It appears that insertion of these mobile genetic elements are phenotypically neutral in that their presence as an intervening sequence does not disrupt the function of the surrounding gene.

Aside from revealing a possible interaction site between T4 topoisomerase subunits gp39 and gp60, the interaction of $gp39^{473-516}$ with $gp60^{2-31}$ has important application to biotechnology and synthetic biology. These peptides, combined with a mobile element, could be used to screen for functional split sites in proteins, overcoming split site identification problems and providing a direct and easy way to isolate a large library of constructs instead of having to individually construct split fragments into vectors (Figure 5.12). Although design and construction of such a system started, construction and testing of the system was not finished due to time constraints and would merit further investigation. Other applications of these types of peptides would include stabilizing protein-protein complexes for cryo-EM studies.

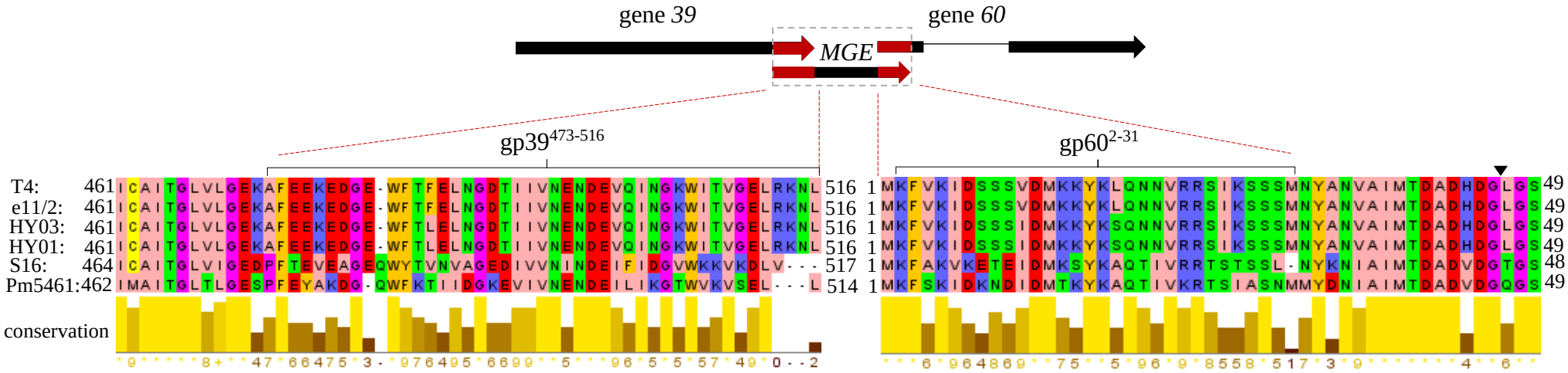


Figure 5.16. Phage genes for the type II DNA topoisomerase. Screenshot of the Jalview 2.11.1.4 Desktop showing an alignment of amino acid sequences surrounding sites of insertion of a mobile genetic element (*MGE*) in phage T4 and other phages' topoisomerase gene *39* and *60*. Zappo colors in Jalview: Aliphatic/hydrophobic residues ILVAM (pink), Aromatic FWY (orange), Positive KRH (blue), Negative DE (red), Hydrophilic STNQ (green), conformationally special PG (purple) and Cysteine C (yellow). The bypass element in gene *60* that stimulates translational bypassing is shown by a solid triangle. Accession number: T4 (HM_137666.1), e11/2 (NC_024125), HY03 (NC_031047), HY01 (NC_027349), S16 (NC_020416) and Pm5461 (NC_028762).

CHAPTER SIX

CONCLUSION AND REMARKS

The first objective of this work was to probe the elements of bacteriophage T4 DNA topoisomerase gene *60* that promote translational bypassing. Probing of these elements using a His-tagged gene *60-iLOV* reporter system developed in this work that contained a segment of gene *60* previously shown to contain all elements necessary for efficient bypassing (Weiss *et al.*, 1990) fused downstream to a fluorescent iLOV reporter and upstream to a purification His-tag (see Figure 3.2), and expressed in *E. coli* and in *in vitro* translation systems from *E. coli*, led to the identification and verification of the set of gene *60* elements necessary for efficient translational bypassing. These signals included: (i) the matched GGA codons (termed 'take-off' and 'landing' codons) flanking either end of the bypassed region that allow the translation apparatus to re-attach or 'land' on the mRNA after 'take-off' and resume translation, (ii) the in-frame TAG stop codon contained in the mRNA stem loop (termed 'take-off' stem loop) that stimulates bypass and (iii) the nascent gp60 peptide stalling signal. These results were consistent with previously characterized gene *60* bypassing elements (Weiss *et al.*, 1990). The translation systems and genetic assays employed were feasible approaches to track translation of the His-tagged gene *60-iLOV* fusion by *E. coli* ribosomes and visualize the synthesized proteins, to deduce efficient ribosomal bypass over the non-coding region of T4 gene *60*.

This work then aimed to use the gene *60* bypassing elements to create translational bypassing devices that work in other genetic contexts to regulate gene expression and control the relative stoichiometry of two distinct protein outputs from one gene for biotechnological purposes. The successful achievement of this goal relied on de-novo-designed translational bypassing devices based on the 50-nucleotide bypass element of T4 gene *60* (see Figure 4.2 and Figure 4.4) and examination of their translational bypassing activity within bacteriophage ϕC31 serine integrase gene *int* context. One such device termed Hop1, which is identical to the 50-nucleotide bypass element in T4 gene *60* (see Figure 4.2), promoted a low-level translational bypassing event in this genetic context unrelated to T4 gene *60*. This data suggests that the 50-nucleotide bypass element in gene *60* alone is enough to induce translational bypassing (contains the minimum set of elements necessary for bypassing that include matched codons bordering the bypass element and a stop codon contained in the take-off stem loop). Moreover, this low-level translational bypassing event is attributed to the absence of other features such as a nascent peptide stalling motif and RNA elements flanking the bypassed site that promote a highly efficient bypassing event (Weiss *et al.*, 1990; Samatova *et al.*, 2014; Agirrezabala *et al.*, 2017).

The proposed mechanism of translational bypassing promoted by these functional devices follows the principles of translational bypassing in gene *60* (Figure 6.1). The mRNA stem loop (take-off stem loop) re-folds in the *E. coli* ribosomal A-site preventing termination at the in-frame UAG stop codon (stop) and inducing take-off by destabilizing the peptidyl-tRNA anticodon codon interaction at the take-off codon in the P-site. The translational complex then bypasses the region without decoding it and re-joins the mRNA at the matched landing codon where the peptidyl-tRNA anticodon codon interaction is stabilized and translation continues. This in turn, enables ribosomes to synthesize full length functional protein from the discontinuous frame. The remaining ribosomes fail to bypass and produce a truncated protein form in the first reading frame.

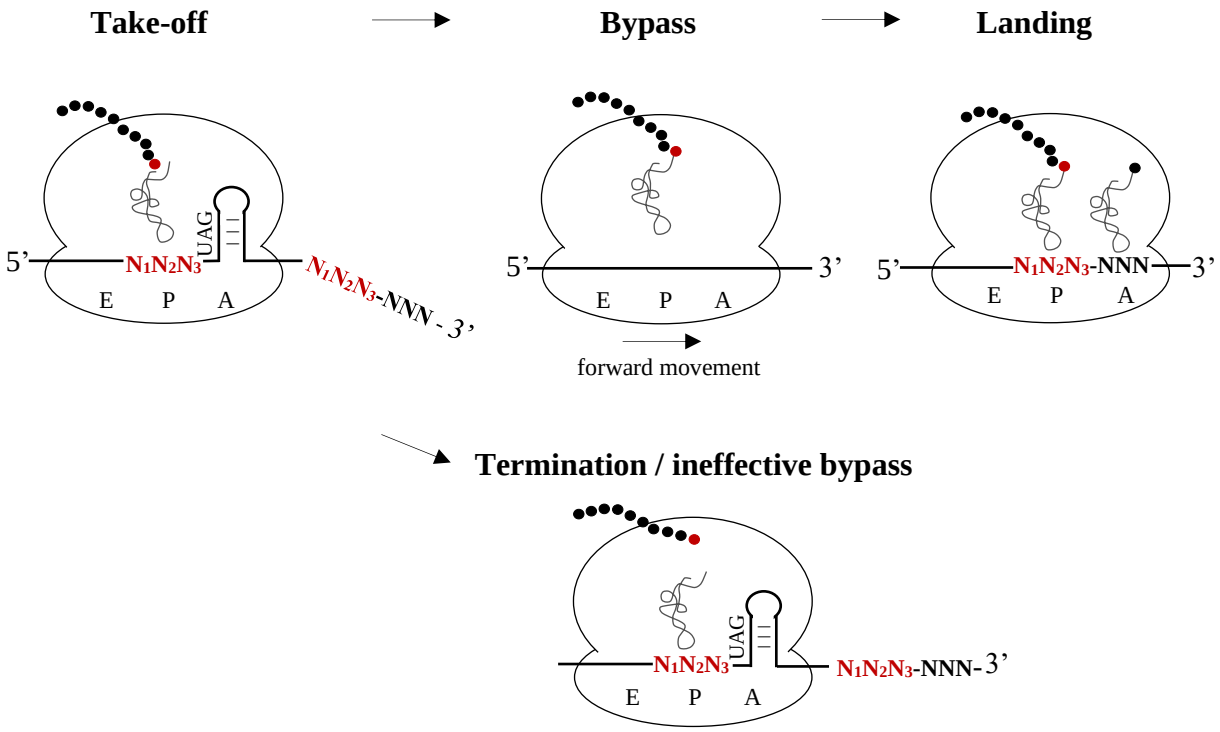


Figure 6.1. A model for translational bypassing. Peptidyl-tRNA detaches from the 'take-off' codon, bypasses the translational bypassing device (bypass element, non-coding region or interruption) and re-pairs at the matched 'landing' codon where elongation resumes at the next in-frame codon. The matched take-off and landing codons bordering the bypass region are shown as $N_1N_2N_3$ (red). A proportion of ribosomes successfully bypass to resume translation downstream and synthesize a full length protein product, while the rest terminate translation in the first reading frame resulting in truncated protein.

Future genetic and structural studies on these recombinant genes and translation could provide insight on the nascent peptide folding and mRNA folding in the ribosome. Hopefully, the development of a method that allows real-time visualization of translation will also enhance our understanding of the mechanism of translational bypassing.

The ability of these devices (Figure 4.2 and Figure 4.4) to regulate gene expression and its application to synthetic biology is intriguing and deserves further study. The devices could be optimized by introduction of a nascent peptide 'stalling' sequence upstream of the bypass site but would be part of the protein outputs; a translational pause would increase the efficiency of the bypassing event and set a different ratio of the protein outputs. Other elements could also be altered such as the matched codons flanking the interruption or interruption size yielding a wide range of efficiencies that set different ratios of the protein outputs.

Lastly, this work investigated a previously predicted interaction interface between T4 DNA topoisomerase subunits gp39 and gp60 (amino acids 473-516 of T4 gp39 with amino acids 2-31 of T4 gp60). The protein-protein interaction assays (size exclusion chromatography, microscale thermophoresis and bacterial two-hybrid assay) that were used demonstrated that amino acids 473-516 of T4 gp39 interact with amino acids 2-31 of T4 gp60, confirming the presence of the predicted interaction between these regions. The assays performed were feasible approaches to detect this interaction. The interaction between these short peptide sequences was then employed in the development of a Tn4430-transposon based screening approach for identification of functional split sites in proteins (see Figure 5.12). This transposon allows random insertion of these T4 peptides into target proteins without disrupting reading frames, in turn, libraries can be screened for insertions that maintain target protein function by dimerization allowing identification of functional split sites. This transposon system has not yet been tested due to time constraints and merits further investigation due to its importance to synthetic biology. This transposon system would overcome split sites identification problems that rely on secondary structure alignment of homologous protein and that functional proteins can be inactivated by splitting them. The system could be tested with site-specific DNA recombinases enabling their spatiotemporal control which is useful in gene therapy, biological counting devices and assembly of DNA fragments.

REFERENCES

Agirrezabala, X., Samatova, E., Klimova, M., Zamora, M., Gil-Carton, D., Rodnina, M. V. and Valle, M. (2017). Ribosome rearrangements at the onset of translational bypassing. *Science Advances*, ***3***, e1700147.

Ahmed, Y. L. (2017). Pycorn 0.19. Retrieved from https://pypi.org/project/pycorn/

Amann, E., Ochs, B. and Abel, K-J. (1988). Tightly regulated trc promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, ***69***, 301-315.

Angius, F., Ilioaia, O., Amrani, A., Suisse, A., Rosset, L., Legrand, A., Abou-Hamdan, A., Uzan, M., Zito, F. and Miroux, B. (2018). A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. *Scientific Reports*, ***8***, 8572.

Anzalone, A. V., Lin, A. J., Zairis, S., Rabadan, R. and Cornish, V. W. (2016). Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. *Nature Methods*, ***13***(5), 453-458.

Atkins, J. F. and Gesteland, R. F. (2010). *Recoding: Expansion of decoding rules enriches gene expression*. New York Dordrecht Heidelberg London, Springer. Nucleic acid and molecular biology, **24.**

Ausländer, S. and Fussenegger, M. (2014). Toehold gene switches make big footprints. *Nature*, ***516,*** 333-334.

Ausländer, S., Stücheli, P., Rehm, C., Ausländer, D., Hartig, J. S. and Fussenegger, M. (2014). A general design strategy for protein-responsive riboswitches in mammalian cells. *Nature Methods*, ***11***, 1154-1160.

Bachmann, B. J. (1972). Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriology Reviews*, ***36***(4), 525-557.

Baranov, P. V., Atkins, J. F. and Yordanova, M. M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nature Reviews Genetics*, ***16***, 517-529.

Baranov, P. V., Gesteland, R. F. and Atkins, J. F. (2002). Recoding: translational bifurcations in gene expression. *Gene*, ***286***, 187-201.

Baranov, P. V., Gurvich, O. L., Fayet, O., Prère, M. F., Miller, W. A., Gesteland, R. F., Atkins, J. F. and Giddings, M. C. (2001). RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *NAR*, ***29***, 264-267.

Baranov, P. V., Fayet, O., Hendrix, R. W. and Atkins, J. F. (2006). Recoding in bacteriophages and bacterial IS elements. *Trends in Genetics*, ***22***(3), 174-181.

Baron, C. and Böck, A. (1991). The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *The Journal of Biological Chemistry*, ***266***, 20375-20379.

Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P. and Peltz, S. W. (2004). The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *Journal of Virology*, ***78***(4), 2082-2087.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, ***277***, 1453-1474.

Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991). Selenocysteine: the 21st amino acid. *Molecular Microbiology*, ***5***, 515-520.

Bonocora, R. P., Zeng, Q., Abel, E. V. and Shub, A. D. (2011). A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. *PNAS*, ***108***, 16351-16356.

Borujeni, A. E. and Salis, H. M. (2016). Translation Initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society*, ***138***, 7016-7023.

Borujeni, A. E., Channarasappa, A. S. and Salis, H. M. (2013). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *NAR*, ***42***(4), 2646-2659.

Borujeni, A. E., Cetnar, D., Farasat, I., Smith, A., Lundgren, N. and Salis, H. M. (2017). Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *NAR*, ***45*** (9), 5437-5448.

Brierley, I. (1995). Ribosomal frameshifting viral RNAs. *Journal of General Virology*, ***76***(8), 1885-1892.

Bucklin, D. J., Wills, N. M., Gesteland, R. F. and Atkins, J. F. (2005). P-site pairing subtleties revealed by the effects of different tRNAs on programmed translational bypassing where anticodon re-pairing to mRNA is separated from dissociation. *Journal of Molecular Biology*, ***345***, 39-49.

Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J. and Huang, E. S. (2004). Are protein-protein interfaces more conserved in sequence than the rest of the protein surface?. *Protein Science*, ***13***, 190-202.

Chapman, S., Faulkner, C., Kaiserli, E., Garcia-Mata, C., Savenkov, E. I., Roberts, A. G., Oparka, K. J. and Christie, J. M. (2008). The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *PNAS*, ***105***(50), 20038-20043.

Chatterjee, A., Guo, J., Lee, H. S. and Schultz, P. G. (2013). A genetically encoded fluorescent probe in mammalian cells. *Journal of the American Chemical Society*, ***135***, 12540-12543.

Chen, J., Coakley, A., O'Connor, M., Petrov, A., O'Leary, S. E., Atkins, J. F. and Puglisi, J. D. (2015). Coupling of mRNA structure rearrangement to ribosome movement during bypassing of non-coding regions. *Cell*, ***163***, 1267-1280.

Choi, K. M., Atkins, J. F., Gesteland, R. F. and Brimacombe, R. (1998). Flexibility of the nascent polypeptide chain within the ribosome contacts from the peptide N-terminus to a specific region of the 30S subunit. *European Journal of Biochemistry*, ***255***, 409-413.

Cirino, P. C., Mayer, K. M. and Umeno, D. (2003). Generating mutant libraries using error-prone PCR. *Methods in Molecular Biology*, ***231***, 3-9.

Costantini, L. M., Fossati, M., Francolini, M. and Snapp, E. L. (2012). Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic*, ***13***(5), 643-649.

Craigen, W. J., Cook, R. G., Tate, W. P. and Caskey, C. T. (1985). Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *PNAS*, ***82***, 3616-3620.

Craigen, W. J. and Caskey, C. T. (1986). Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature*, ***322***, 273-275.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, ***227***, 561-563.

Crick, F. H., Barnett, L., Brenner, S. and Watts-Tobin, R. (1961). General nature of the genetic code for proteins. *Nature*, ***192***, 1227-1232.

Crick, F. H. C. (1966). Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, ***19***(2), 548-555.

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*, ***38,*** 367-379.

Crona, M., Moffatt, C., Friedrich, N. C., Hofer, A., Sjoberg, B. and Edgell D. R. (2011). Assembly of a fragmented ribonucleotide reductase by protein interaction domains derived from a mobile genetic element. *NAR*, ***39***(4), 1381-1389.

Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O. and Pietrokovski, S. (2009). Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *NAR*, ***37***(8), 2560-2573.

De Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C. and Ryan, M. D. (2006). E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends in Biotechnology*, ***24***(2), 68-75.

de la Torre, D. and Chin, J. W. (2021). Reprogramming the genetic code. *Nature Reviews Genetics*, ***22,*** 169-184.

Dinman, J. D. (2019). Translational recoding signals: Expanding the synthetic biology toolbox. *Journal of Biological Chemistry*, ***294***(19), 7537-7545.

Di Ventura, B. and Mootz, H. D. (2018).Switchable inteins for conditional protein splicing. *Journal of Biological Chemistry*, ***400***, 467-475.

Doronina, V. A., Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D. and Brown, J. D. (2008). Site-specific release of nascent chains from ribosomes at a sense codon. *Molecular and Cellular Biology*, ***28***(13), 4227-4239.

Dressman, H. K., Barley-Maloney, L., Rowlette, L. L., Agris, P. F. and Garcia-Blanco, M. A. (2006). Assessing incomplete deprotection of microarray oligonucleotides in situ. *NAR*, ***34***(19), e131.

Fang, J., Qian, J. J., Yi, S., Harding, T. C., Tu, G. H., VanRoey, M. and Jooss, K. (2005). Stable antibody expression at therapeutic levels using the 2A peptide. *Nature Biotechnology*, ***23,*** 584-590.

Farabaugh, P. J. (1996). Programmed translational frameshifting. The *Annual Review of Genetics*, ***30***, 507-528.

Farabaugh, P. J., Kramer, E., Vallabhaneni, H. and Raman, A. (2006). Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. *Journal of Molecular Evolution*, ***63***, 545-561.

Fischer, N., Neumann, P., Bock, L. V., Maracci, C., Wang, Z., Paleskava, A., Konevega A. L., Schroder, G. F., Grubmuller, H., Ficner, R., Rodnina, M. V. and Stark, H. (2016). The pathway to GTPase activation of elongation factor SelB on the ribosome. *Nature*, ***540***, 80-85.

Fogg, P. C. M., Younger, E., Fernando, B. D., Khaleel, T., Stark, M. W. and Smith, M. C. M. (2018). Recombination directionality factor gp3 binds ϕC31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *NAR*, ***46****(3)*, 1308-1320.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K. and Agrawal, R. K. (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature*, ***376***, 441-444.

Friedrich, N. C., Torrents, E., Gibb, E. A., Sahlin, M., Sjöberg, B. M. and Edgell, D. R. (2007). Insertion of a homing endonuclease creates a genes-in-pieces ribonucleotide reductase that retains function. *PNAS*, ***104***(15), 6176-6181.

Gallant, J., Bonthuis, P. and Lindsley, D. (2003). Evidence that the bypassing ribosome travels through the coding gap. *PNAS*, ***100***(23), 13430-13435.

Gallant, J., Bonthuis, P., Lindsley, D., Cabellon, J., Gill, G., Heaton, K., Kelley-Clarke, B., MacDonald, L., Mercer, S., Vu, H. and Worsley, A. (2004). On the role of the starved codon and the takeoff site in ribosome bypassing in *Escherichia coli*. *Journal of Molecular Biology*, ***342***(3), 713-724.

Garofalo, R., Wohlgemuth, I., Pearson, M., Lenz, C., Urlaub, H. and Rodnina, M. V. (2019). Broad range of missense error frequencies in cellular proteins. *NAR*, ***47***(6), 2932-2945.

Gawthorne, J. A., Reddick, L. E., Akpunarlieva, S. N., Beckham, K. S. H., Christie, J. M., Alto, N. M., Gabrielsen, M. and Roe, A. J. (2012). Express Your LOV: An Engineered Flavoprotein as a Reporter for Protein Expression and Purification. *PLoS ONE*, ***7***(12), e52962.

Giedroc, D. P. and Cornish, P. V. (2009). Frameshifting RNA pseudoknots: structure and mechanism. *Virus Research*, ***139***, 193-208.

Green, A. A., Silver, P. A., Collins, J. J. and Yin, P. (2014). Toehold switches: de-novo-designed regulators of gene expression. *Cell*, ***159***(4), 925-939.

Gregory, J. A., Becker, E. C., Jung, J., Tuwatananurak, I. and Pogliano, K. (2010). Transposon assisted gene insertion technology (TAGIT): A tool for generating fluorescent fusion proteins. *PLoS ONE*, ***5***(1), e8731.

Hallet, B., Sherratt, D. J. and Hayes, F. (1997). Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *NAR*, ***25***(9), 1866-1867.

Herbst, K., Nichols, L., Gesteland, R. and Weiss, R. (1994). A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene *60* from Bacteriophage T4. *PNAS*, ***91***(26), 12525-12529.

Herr A. J., Atkins, J. F. and Gesteland, R. F. (1999). Mutations which alter the elbow region of $tRNA2\ Gly$ reduce T4 gene 60 translational bypassing efficiency. *The EMBO Journal*, ***18***(10), 2886-2896.

Herr, A. J., Gesteland, R. F. and Atkins, J. F. (2000). One protein from two open reading frames: mechanism of a 50 nt translational bypass. *The EMBO Journal*, ***19***, 2671-2680.

Herr, A. J., Atkins, J. F. and Gesteland, R. F. (2000). Coupling of open reading frames by translational bypassing. *Annual Review of Biochemistry*, ***69***(1), 343-372.

Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. and Atkins, J. F. (2004). Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *Journal of Biological Chemistry*, ***279***(12), 11081-11087.

Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. and Atkins, J. F. (2001). Drop-off during ribosome hopping. *Journal of Molecular Biology*, ***311***, 445-452.

Herr, A. J., Nelson, C. C., Wills, N. M., Gesteland, R. F. and Atkins, J. F. (2001). Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene *60* bypassing signals during ribosome slippage on mRNA. *Journal of Molecular Biology*, ***309***, 1029-1048.

Ho, T. Y. H., Shao, A., Lu, Z., Savilahti, H., Menolascina, F., Wang, L., Dalchau, N. and Wang, B. (2021). A systematic approach to inserting split inteins for Boolean logic gate engineering and basal activity reduction. *Nature Communications*, ***12***, 1-12.

Huang, W., Ao, S., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R. and Fang, M. (1988). A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science*, ***239***, 1005-1012.

Huang, W. M. (1986). Nucleotide sequence of a type II DNA topoisomerase gene. Bacteriophage T4 gene *39*. *NAR*, ***14***(19), 7751-7765.

Huff, A. C., Leatherwood, J. K. and Kreuzer, K. N. (1989). Bacteriophage T4 DNA topoisomerase is the target of antitumor agent 4'-(9-acridinylamino) methanesulfon-m-anisidide (m-AMSA) in T4-infected *Escherichia coli*. *PNAS*, ***86***(4), 1307-1311.

Hüttenhofer, A., Heider, J. and Böck, A. (1996). Interaction of the *Escherichia coli fdhF* mRNA hairpin promoting selenocysteine incorporation with the ribosome. *NAR*, ***24***(20), 3903-3910.

Ibba, M. and Söll, D. (2000). Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, ***69***, 617-650.

Ito, K. and Chiba, S. (2013). Arrest peptides: *cis*-acting modulators of translation. *Annual Review of Biochemistry*, ***82***, 171-202.

Itoh, Y., Bröcker, M. J., Sekine, S., Hammond, G., Suetsugu, S., Söll, D. and Yokoyama, S. (2013). Decameric SelA•tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation. *Science*, ***340***(6128), 75-78.

Jacks, T. and Varmus, H. E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science*, ***230***, 1237-1242.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. and Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, ***331***, 280-283.

Jakubowski, H. and Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiology and Molecular Biology Reviews*, ***56***, 412-429.

Jenner, L. B., Demeshkina, N., Yusupova, G. and Yusupov, M. (2010). Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nature Structural and Molecular Biology*, ***17***, 555-560.

Jørgensen, F. and Kurland, C. G. (1990). Processivity errors of gene expression in *Escherichia coli*. *Journal of Molecular Biology*, ***215***, 511-521.

Kapust, R. B. and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science*, ***8***(8), 1668-1674.

Karimova, G., Ullmann, A., Pidoux, P. and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *PNAS*, ***95***, 5752-5756.

Karimova, G., Ullmann, A. and Ladant, D. (2001). Protein-protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *Journal of Molecular Microbiology and Biotechnology*, ***3***(1), 73-82.

Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. and Smith, M. C. M. (2011). A phage protein that binds φC31 integrase to switch its directionality. *Molecular Microbiology*, ***80***(6), 1450-1463.

Klimova, M., Senyushkina, T., Samatova, E., Peng, B. Z., Pearson, M., Peske, F. and Rodnina, M. V. (2019). EF-G–induced ribosome sliding along the noncoding mRNA. *Science Advances*, ***5***, eaaw9049.

Knight, R. D., Freeland, S. J. and Landweber, L. F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, ***2***, 49-58.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, ***361***, 13-37.

Kreuzer, K. N. and Jongeneel, C. V. (1983). *Escherichia coli* phage T4 topoisomerase. *Methods in Enzymology*, ***100***, 144-160.

Kreuzer, K. (1998). Bacteriophage T4, a model system for understanding the mechanism of type II topoisomerase inhibitors. *Biochimica et Biophysica Acta*, ***1400***, 339-347.

Kromayer, M., Wilting, R., Tormay, P. and Böck, A. (1996). Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. *Journal of Molecular Biology*, ***262***, 413-420.

Ladant, D. and Ullmann, A. (1999). *Bordetella pertussis* adenylate cyclase: a toxin with multiple talents. *Trends in Microbiology*, ***7***, 172-176.

Lang, B. F., Jakubkova, M., Hegedusova, E., Daoud, R., Forget, L., Brejova, B. and Nosek, J. (2014). Massive programmed translational jumping in mitochondria. *PNAS*, ***111***(16), 5926-5931.

Le, S., Chen, J. and Maizel Jr., J.V. (1993). Identification of unusual RNA folding patterns encoded by bacteriophage T4 gene *60*. *Gene*, ***124***, 21-28.

Liu, C. C. and Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annual Review of Biochemistry*, ***79***, 413-444.

Mahillon, J. and Lereclus, D. (1988). Structural and functional analysis of Tn*4430*: identification of an integrase-like protein involved in the co-integrate-resolution process. *The EMBO Journal*, ***7***(5), 1515-1526.

Maldonado, R. and Herr, A. J. (1998). Efficiency of T4 gene 60 translational bypassing. *Journal of Bacteriology*, ***180***, 1822-1830.

Matsumoto, S., Caliskan, N., Rodnina, M. V., Murata, A. and Nakatani, K. (2018). Small synthetic molecule-stabilized RNA pseudoknot as an activator for -1 ribosomal frameshifting. *NAR*, ***46***(16), 8079-8089.

McEwan, A. R., Raab, A., Kelly, S. M., Feldmann, J. and Smith, M. C. M. (2011). Zinc is essential for high-affinity DNA binding and recombinase activity of ϕC31 integrase. *NAR*, ***39***(14), 6137-6147.

Mehla, J., Caufield, J. H., Sakhawalkar, N. and Uetz, P. (2017). A comparison of two-hybrid approaches for detecting protein-protein interactions. *Methods in Enzymology*, ***586***, 333-358.

Miller, J. H. and Albertini, A. M. (1983). Effects of surrounding sequence on the suppression of nonsense codons. *Journal of Molecular Biology*, ***164***, 59-71.

Milligan, R. A. and Unwin, P. N. (1986). Location of exit channel for nascent protein in 80S ribosome. *Nature*, ***319***, 693-695.

Mintseris, J. and Weng, Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. *PNAS*, ***102***(31), 10930-10935.

Miroux, B. and Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, ***260***(3), 289-298.

Mohler, K. and Ibba, M. (2017). Translational fidelity and mistranslation in the cellular response to stress. *Nature Microbiology*, ***2,*** 17117.

Momin, A. A., Hameed, U. and Arold, S. T. (2019). Passenger sequences can promote interlaced dimers in a common variant of the maltose-binding protein. *Scientific Reports*, ***9***(1), 20396.

Mueller, A. M., Breitsprecher, D., Duhr, S., Baaske, P., Schubert, T. and Längst, G. (2017). Microscale thermophoresis: a rapid and precise method to quantify protein–nucleic acid interactions in solution. *Methods in Molecular Biology*, ***1654***, 151-164.

Nakamura, Y. and Ito, K. (2011). tRNA mimicry in translation termination and beyond. *WIREs RNA*, ***2***, 647-668.

Nakatogawa, H. and Ito, K. (2001). Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Molecular Cell*, ***7***, 185-192.

Nakatogawa, H. and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell*, ***108***, 629-636.

Norrander, J., Kempe, T. and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*, ***26***(1), 101-106.

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. and Schultz, P. G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, ***244***, 182-188.

Nosek, J., Tomaska, L., Burger, G. and Lang, B. F. (2015). Programmed translational bypassing elements in mitochondria: structure, mobility, and evolutionary origin. *Trends in Genetics*, ***31***(4), 187-194.

Ogle, J. M. and Ramakrishnan, V. (2005). Structural insights into translational fidelity. *Annual Review of Biochemistry*, ***74***, 129-177.

Oliver, D., Norman, J. and Sarker, S. (1998). Regulation of *Escherichia coli secA* by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. *Journal of Bacteriology*, ***180***, 5240-5242.

Olorunniji, F. J., Lawson-Williams, M., McPherson, A. L., Paget, J. E., Stark, W. M. and Rosser, S. J. (2019). Control of ϕC31 integrase-mediated site-specific recombination by protein trans-splicing. *NAR*, ***47***(21), 11452-11460.

O'Loughlin, S., Capece, M. C., Klimova, M., Wills, N. M., Coakley, A., Samatova, E., O'Connor, P. B. F., Loughran, G., Weissman, J. S., Baranov, P. V., Rodnina, M. V., Puglisi, J. D. and Atkins, J. F. (2020). Polysomes bypass a 50-nucleotide coding gap less efficiently than monosomes due to attenuation of a 5' mRNA stem-loop and enhanced drop-off. *Journal of Molecular Biology*, ***432***(16), 4369-4387.

OpenWetWare. (2014). Escherichia coli/Codon usage. Retrieved from https://openwetware.org/mediawiki/index.php?title=Escherichia_coli/Codon_usage&oldid=820473.

Pape, T., Wintermeyer, W. and Rodnina, M. V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *The EMBO Journal*, ***17***, 7490-7497.

Park, S. and Lee, J. W. (2021). Detection of coronaviruses using RNA toehold switch sensors. *International Journal of Molecular Sciences*, ***22***(4), 1772.

Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, ***24***, 79-88.

Peske, F., Rodnina, M. V. and Wintermeyer, W. (2005). Sequence of steps in ribosome recycling as defined by kinetic analysis. *Molecular Cell*, ***18***(4), 403-412.

Petrov, V. M., Ratnayaka, S. and Karam, J. D. (2010). Genetic insertions and diversification of the PolB-type DNA polymerase (gp43) of T4-related phages. *Journal of Molecular Biology*, ***395***, 457-474.

Petrov, V. M., Ratnayaka, S., Nolan, J. M., Miller, E. S. and Karam, J. D. (2010). Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology Journal*, ***7***, 292.

Phizicky, E. M. and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiology and Molecular Biology Reviews*, ***59***(1), 94-123.

Pinto, M. G. and Baiker, A. (2012). LuMPIS: luciferase-based MBP-pull-down protein interaction screening system. *Methods in Molecular Biology*, ***815***, 265-275.

Pinto, V., Villegas, M. G., Peter, J. M., Haase, J., Haas, R., Lotz, J., Muntau, A. S., Baiker, A. C. and Haas, J. (2009). LuMPIS-a modified luminescence-based mammalian interactome mapping pull-down assay for the investigation of protein-protein interactions encoded by GC-low ORFs. *Proteomics*, ***9***, 5303-5308.

Pryor, K. D. and Leiting, B. (1997). High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expression and Purification*, ***10***, 309-319.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell*, ***108***, 557-572.

Richarme, G. (1982). Associative properties of the *Escherichia coli* galactose binding protein and maltose binding protein. *Biochimica et Biophysica Acta*, ***105***, 476-481.

Rodnina, M. V., Korniy, N., Klimova, M., Karki, P., Peng, B., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E. and Peske, F. (2020). Translational recoding: canonical translation mechanisms reinterpreted. *NAR*, ***48***(3), 1056-1067.

Rodnina, M. V. and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annual Review of Biochemistry*, ***70***, 415-435.

Rodnina, M. V., Savelsbergh, A., Katunin, V. I. and Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature*, ***385***(6611), 37-41.

Rudinger, J., Hillenbrandt, R., Sprinzl, M. and Giegé, R. (1996). Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *The EMBO Journal*, ***15***, 650-657.

Samatova, E., Konevega, A. L., Wills, N. M., Atkins, J. F. and Rodnina, M. V. (2014). High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide. *Nature Communications*, ***5***.

Sandegren, L., Nord, D. and Sjöberg, B. M. (2005). SegH and Hef: two novel homing endonucleases whose genes replace the *mobC* and *mobE* genes in several T4-related phages. *NAR*, ***33***(19), 6203-6213.

Schroeder, R., Grossberger, R., Pichler, A. and Waldsich, C. (2002). RNA folding *in vivo*. *Current Opinion in Structural Biology*, ***12***(3), 296-300.

Scolnick, E., Tompkins, R., Caskey, T. and Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. *PNAS*, ***61***(2), 768-774.

Seasholtz, A. F and Greenberg, G. R. (1983). Identification of bacteriophage T4 gene 60 product and a role for this protein in DNA Topoisomerase*. *Journal of Biological Chemistry*, ***258***(2), 1221-1226.

Seidelt, B., Innis, C. A., Wilson, D. N., Gartmann, M., Armache, J. P., Villa, E., Trabuco, L. G., Becker, T., Mielke, T., Schulten, K., Steitz, T. A. and Beckmann, R. (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, ***326***(5958), 1412-1415.

Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D. and Brown, J. D. (2012). 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *NAR*, ***40***(7), 3143-3151.

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B. P. and Yusupov, M. (2009). A structural view of translation initiation in bacteria. *Cellular and Molecular Life Sciences*, ***66***, 423-436.

Singaravelan, B., Roshini, B. R. and Munavar, M. H. (2010). Evidence that the *supE44* mutation of *Escherichia coli* is an amber suppressor allele of *glnX* and that it also suppresses ochre and opal nonsense mutations. *Journal of Bacteriology*, ***192***(22), 6039-6044.

Srinivasan, G., James, C. M. and Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, ***296***, 1459-1462.

Stark, W. M. (2017). Making serine integrases work for us. *Current Opinion in Microbiology*, ***38***, 130-136.

Summers, D. K. and Sherratt, D. J. (1984). Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell*, ***36***(4), 1097-1103.

Swithers, K. S., Senejani, A. G., Fournier, G. P. and Gogarten, J. P. (2009). Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evolutionary Biology*, ***9***, 303.

Szymczak, A., Workman, C., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F. and Vignali, D. A. A. (2004). Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnology*, ***22***, 589-594.

Théobald-Dietrich, A., Giegé, R. and Rudinger-Thirion, J. (2005). Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins. *Biochimie*, ***87***, 813-817.

Tobias A. V. (2003) Preparing Libraries in *Escherichia coli. Methods in Molecular Biology*, ***231***, 11-16.

Todd, G. C. and Walter, N. G. (2013). Secondary structure of bacteriophage T4 gene *60* mRNA: Implications for translational bypassing. *RNA*, ***19***, 685-700.

Topp, S. and Gallivan, J. P. (2007). Guiding bacteria with small molecules and RNA. *Journal of the American Chemical Society*, ***129***(21), 6807-6811.

Tourigny, D. S., Fernández, I. S., Kelley, A. C. and Ramakrishnan, V. (2013). Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science*, ***340***, 1235490.

Ullman, A. and Danchin, A. (1980). Role of cyclic AMP in regulatory mechanisms in bacteria. *Trends in Biochemical Sciences*, ***5***(4), 95-96.

Urbonavicius, J., Qian, Q., Durand, J. M., Hagervall, T. G. and Björk, G. R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *The EMBO Journal*, ***20***, 4863-4873.

Van Duyne, G. D. and Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. *Critical Reviews in Biochemistry and Molecular Biology* , ***48***(5), 476-491.

Voss, N. R., Gerstein, M., Steitz, T. A. and Moore, P. B. (2006). The geometry of the ribosomal polypeptide exit tunnel. *Journal of Molecular Biology*, ***360***(4), 893-906.

Walker, I. H., Hsieh, P. C. and Riggs, P. D. (2010). Mutations in maltose-binding protein that alter affinity and solubility properties. *Applied Microbiology and Biotechnology*, ***88***(1), 187-197.

Wang, L., Xie, J. and Schultz, P. G. (2006). Expanding the genetic code. *Annual Review of Biophysics and Biomolecular Structure*, ***35***, 225-249.

Wang, J., Xie, J. and Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society*, ***128***, 8738-8739.

Weinberg, B. H., Cho, J. H., Agarwal, Y., Pham, N., Caraballo, L. D., Walkosz, M., Ortega, C., Trexler, M., Tague, N., Law, B., Benman, W., Letendre, J., Beal, J. and Wong, W. W. (2019). High-performance chemical- and light-inducible recombinases in mammalian cells and mice. *Nature Communications*, ***10***(1), 4845.

Weiss, R. B., Huang, W. M. and Dunn, D. M. (1990). A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene *60*. *Cell*, ***62***, 117-126.

Wills, N. M., O'Connor, M., Nelson, C. C., Rettberg, C. C., Huang, W. M., Gesteland, R. F. and Atkins, J. F. (2008). Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *The EMBO Journal*, ***27***, 2533-2544.

Yang, Y., Wang, J., Zhang, R. and Yan, Y. (2019). Antisense RNA elements for downregulating expression. *Methods in Molecular Biology*, ***1927***, 23-35.

Yonath, A., Leonard, K. R. and Wittmann, H. G. (1987). A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science*, ***236***(4803), 813-816.

Youngman, E. M., McDonald, M. E. and Green, R. (2008). Peptide release on the ribosome: mechanism and implications for translational control. *Annual Review of Microbiology*, ***62***, 353-373.

Zaher, H. S. and Green, R. (2009). Fidelity at the molecular level: lessons from protein synthesis. *Cell*, ***136***(4), 746-762.

Zhang, Y., Baranov, P. V., Atkins, J. F. and Gladyshev, V. N. (2005). Pyrrolysine and selenocysteine use dissimilar decoding strategies. *Journal of Biological Chemistry*, ***280***, 20740-20751.

Zhao, J., Pokhilko, A., Ebenhoh, O., Rosser, S. J. and Colloms, S. D. (2019). A single-input binary counting module based on serine integrase site-specific recombination. *NAR*, ***47***, 4896-4909.

Zinoni, F., Heider, J. and Böck, A. (1990). Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *PNAS*, ***87***, 4660-4664.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *NAR*, ***31***(13), 3406-3415.

END